

**UNIVERSIDADE FEDERAL DE ITAJUBÁ**  
**PROGRAMA DE PÓS-GRADUAÇÃO**  
**EM CIÊNCIA E TECNOLOGIA DA COMPUTAÇÃO**

Extração de Conhecimento com Hipercaixas Difusas

Lênio Oliveira Prado Júnior

**Itajubá, Agosto de 2013**

**UNIVERSIDADE FEDERAL DE ITAJUBÁ**  
**PROGRAMA DE PÓS-GRADUAÇÃO**  
**EM CIÊNCIA E TECNOLOGIA DA COMPUTAÇÃO**

Lênio Oliveira Prado Júnior

Extração de Conhecimento com Hipercaixas Difusas

Dissertação submetida ao programa de Pós-Graduação em Ciência e Tecnologia da Computação como parte dos requisitos para obtenção do título de Mestre em Ciência e Tecnologia da Computação.

Área de Concentração: Matemática da Computação

Orientador: Prof. Dr. Carlos Henrique Valério de Moraes

**Agosto de 2013**

**Itajubá**

**UNIVERSIDADE FEDERAL DE ITAJUBÁ**  
**PROGRAMA DE PÓS-GRADUAÇÃO**  
**EM CIÊNCIA E TECNOLOGIA DA COMPUTAÇÃO**

Lênio Oliveira Prado Júnior

Extração de Conhecimento com Hipercaixas Difusas

Dissertação aprovada por banca examinadora em 19 de Agosto de 2013, conferindo ao autor o título de **Mestre em Ciência e Tecnologia da Computação**.

**Banca Examinadora:**

Prof. Dr. Carlos Henrique Valério de Moraes (Orientador)

Prof. Dr. Ahmed Ali Abdalla Esmín (UFLA-Lavras)

Prof. Dr. Enzo Seraphim (Unifei)

**Itajubá**

**2013**

# Resumo

Este trabalho apresenta uma metodologia para extração de regras difusas a partir da análise de dados classificados. Dos dados são obtidas regras difusas, que permitem a classificação de novos elementos com determinado grau de aceitação em cada um dos grupos existentes. A metodologia é baseada em técnicas de detecção de colisão de objetos utilizadas em jogos e simulação 3D, com adição de lógica difusa, criando assim hipercaixas difusas. A técnica é supervisionada, trabalhando com classes de dados previamente classificadas. Possui boa capacidade de generalização, sendo capaz de alcançar altos níveis de tolerância para diversos cenários de dados de entrada. O aumento no número de casos analisados não eleva o custo computacional proporcionalmente, devido ao fato da metodologia desenvolvida utilizar apenas as regiões limítrofes dos grupos.

# Abstract

This paper presents a methodology for extracting fuzzy rules based on the analysis of previously classified data. Fuzzy rules are obtained from the data. This rules allows the classification of new elements with a certain degree of relevancy in each of the existing groups. The methodology is based on object collision detection techniques used in games and 3D simulation, with fuzzy logic added for creation of fuzzy N-dimensional bounding box. The technique is supervised, working with previously classified data classes. It has good generalization capability, being able to achieve high levels of tolerance for different input data sets. The increase in the number of cases analyzed does not increase the computational cost proportionally, due to the fact that the methodology developed uses only the groups borders regions.

# Agradecimentos

A minha família, que me apoiou incondicionalmente em todos os momentos e me deu força para continuar.

Ao meu orientador, Carlos Henrique Valério de Moraes, pela orientação, dedicação, e pela presteza com que sempre me ajudou.

A todos os meus amigos, que contribuíram para que o objetivo fosse alcançado e me ajudaram direta ou indiretamente.

Ao meu amigo, Emerson Assis de Carvalho, que enfrentou junto o desafio do mestrado, colaborando sempre nas etapas decisivas da trajetória.

Ao meu professor e mestre Marcos Alberto de Carvalho, que me incentivou a realizar o mestrado. Agradeço-o de coração por todo o conhecimento transmitido e pela confiança depositada.

Dedido este trabalho à minha falecida irmã, Maria Alice Dias Prado, desejando a Deus que, de onde quer que ela esteja, comemore junto comigo mais essa grande conquista. Obrigado por ter me proporcionado tantas alegrias.

# Sumário

**Lista de Figuras**

**Lista de Tabelas**

**Lista de Algoritmos** p. 13

**1 Introdução** p. 14

1.1 Contexto . . . . . p. 14

1.2 Motivação . . . . . p. 17

1.2.1 Conjuntos Aproximados (Rough Sets): . . . . . p. 18

1.2.2 Redes Neurais Artificiais (Artificial Neural Networks): . . . . . p. 19

1.3 Objetivos . . . . . p. 20

1.3.1 Objetivo Geral . . . . . p. 20

1.3.2 Objetivos Específicos . . . . . p. 21

1.4 Organização do trabalho . . . . . p. 22

**2 Revisão Bibliográfica** p. 23

2.1 Descoberta de conhecimento em bases de dados . . . . . p. 23

2.1.1 Árvores de decisão . . . . . p. 24

2.1.2 Regras de Associação . . . . . p. 25

2.1.3	Agrupamento de dados . . . . .	p. 26
2.1.4	Conjuntos Aproximados (Rought Sets) . . . . .	p. 26
2.1.5	Redes Neurais . . . . .	p. 27
2.2	Trabalhos Correlacionados . . . . .	p. 28
<b>3</b>	<b>Metodologia</b>	<b>p. 31</b>
3.1	Colisão de Objetos . . . . .	p. 31
3.1.1	Volume em Esfera . . . . .	p. 33
3.1.2	Volume de Caixas Alinhadas ao Eixo - AABB . . . . .	p. 34
3.1.3	Volume de Caixas Orientadas - OBB . . . . .	p. 36
3.1.4	Teorema do Eixo de Separação - SAT . . . . .	p. 37
3.1.5	Considerações . . . . .	p. 37
3.1.6	Justificativa - Escolha do Volume Envolvente . . . . .	p. 39
3.2	Conjuntos Difusos . . . . .	p. 41
3.3	Teoria da Possibilidade . . . . .	p. 42
3.4	Funções de Pertinência . . . . .	p. 45
3.4.1	Função de pertinência Triangular . . . . .	p. 46
3.4.2	Função de pertinência Trapezoidal . . . . .	p. 47
3.4.3	Função de pertinência de Gauss ou Gaussiana . . . . .	p. 49
3.4.4	Considerações . . . . .	p. 49
3.5	Média Aritmética e Desvio Padrão . . . . .	p. 50
3.5.1	Utilização do Desvio Padrão . . . . .	p. 54



<b>4</b>	<b>Desenvolvimento</b>	p. 55
4.1	Técnica Proposta . . . . .	p. 55
4.2	Criação da hipercaixa envolvente . . . . .	p. 56
4.3	Testes de Colisão . . . . .	p. 61
4.4	Obtenção das Áreas de Incerteza . . . . .	p. 66
4.5	Classificador Difuso . . . . .	p. 69
4.6	Sistema de inferência difuso . . . . .	p. 71
4.7	Geração das Regras de Classificação . . . . .	p. 73
4.8	Ferramentas de desenvolvimento . . . . .	p. 76
<b>5</b>	<b>Testes e Resultados</b>	p. 77
5.1	Bases de Dados utilizadas . . . . .	p. 77
5.2	Metodologia de Testes . . . . .	p. 82
5.3	Medidas de Avaliação . . . . .	p. 85
5.4	Testes . . . . .	p. 86
<b>6</b>	<b>Conclusão</b>	p. 90
<b>7</b>	<b>Trabalhos Futuros</b>	p. 92
	<b>Referências</b>	p. 93
	<b>Apêndice A - Gráficos dos resultados obtidos</b>	p. 97
	<b>Apêndice B - Exemplo de Regras Difusas para Base Wisconsin</b>	p. 102
	Funções Difusas . . . . .	p. 102

Regras Difusas . . . . . p. 113

# Lista de Figuras

1	Passos para transformação de dados brutos em conhecimento . . .	p. 24
2	Árvore de decisão: Maioridade e menoridade . . . . .	p. 25
3	Colisão entre os objetos A e B . . . . .	p. 31
4	Detecção de não colisão utilizando esferas . . . . .	p. 33
5	Falsa colisão entre objetos . . . . .	p. 34
6	Volume envolvente AABB . . . . .	p. 35
7	Detecção de não colisão utilizando AABB . . . . .	p. 35
8	Detecção de não colisão utilizando OBB . . . . .	p. 36
9	Linha e Eixo de Separação entre dois objetos. . . . .	p. 38
10	AABB - Eixo de coordenadas do sistema e OBB - Eixo de coordenadas orientado . . . . .	p. 40
11	Representação da função de pertinência triangular . . . . .	p. 46
12	Representação da função de pertinência trapezoidal . . . . .	p. 48
13	Representação da função de pertinência de Gauss . . . . .	p. 49
14	Distribuição normal evidenciando a regra '68-95-99,7' . . . . .	p. 53
15	Representação gráfica para as classes Círculo e Triângulo . . . . .	p. 57
16	Representação quadridimensional dos dados - Hipercaixa e Hiperplanos . . . . .	p. 58

17	Classes Círculo e Triângulo envolvidas pelo volume obtido via AABB . . . . .	p. 59
18	Representação da pertinência entre hipercaixas envolventes . . . . .	p. 59
19	Passos para verificar a colisão . . . . .	p. 63
20	Classes Círculo e Triângulo estão colidindo nos eixos X1 e X2 . . . . .	p. 65
21	Hipercaixa da classe Triângulo é dividida com base no ponto médio . . . . .	p. 65
22	Nova divisão da hipercaixa da classe Triângulo. Não há mais colisões . . . . .	p. 66
23	Hipercaixa sigma envolvendo as classes . . . . .	p. 67
24	Listagem de hipercaixas que devem ser envolvidas pelas hipercaixas sigma . . . . .	p. 68
25	Cálculo dos valores de entrada para o classificador difuso . . . . .	p. 70
26	Valores representados pela função trapezoidal . . . . .	p. 72
27	Valores da função trapezoidal obtidos pelo classificador difuso . . . . .	p. 72
28	Passos para a criação dos conjuntos difusos . . . . .	p. 73
29	Algoritmo de geração das regras do sistema de inferência difuso . . . . .	p. 74
30	Exemplo de regra gerada pelo sistema de inferência difuso . . . . .	p. 75
31	Passos para a criação do sistema de inferência difuso . . . . .	p. 75
32	Diagrama UML do projeto . . . . .	p. 76
33	Divisão da base para treinamento e testes . . . . .	p. 83
34	Variação do percentual treinamento e testes . . . . .	p. 84
35	Revocação e Precisão . . . . .	p. 86
36	Resultados obtidos para a base de dados Iris . . . . .	p. 87
37	Comparativo utilizando a Precisão . . . . .	p. 88

38	Comparativo utilizando a Revocação . . . . .	p. 89
39	Comparativo utilizando a Medida F . . . . .	p. 89
40	Base Pima . . . . .	p. 98
41	Base Thyroid . . . . .	p. 99
42	Base Vehicle . . . . .	p. 99
43	Base Wine . . . . .	p. 100
44	Base Wisconsin . . . . .	p. 101

# Lista de Tabelas

1	Problema de discretização - Conjuntos Aproximados . . . . .	p. 19
2	Valores para os grupos de dados A e B . . . . .	p. 50
3	Cálculo do desvio e quadrado dos desvios para o Grupo A . . . . .	p. 51
4	Cálculo do desvio e quadrado dos desvios para o Grupo B . . . . .	p. 51
5	Cálculo da variância e do desvio padrão para os Grupos A e B . . . . .	p. 52
6	Tabela contendo os itens de dados das classes Círculo e Triângulo . . . . .	p. 56
7	Valores das classes Círculo e Triângulo . . . . .	p. 61
8	Resultado dos cálculos utilizando-se a Equação 3.2 . . . . .	p. 62
9	Bases de dados utilizadas para testes . . . . .	p. 77
10	Resultados obtidos com a técnica proposta . . . . .	p. 88
11	Funções difusas da base Wine . . . . .	p. 103
12	Regras difusas da base Wine . . . . .	p. 114

## Lista de Algoritmos

- 1 Verifica colisão entre hipercaixas . . . . . p.62
- 2 Define o percentual de contato . . . . . p.63
- 3 Algoritmo utilizado para cálculo do três sigma . . . . . p.67
- 4 Passos executados em cada variação de modelo . . . . . p.85

# 1 Introdução

## 1.1 Contexto

O acúmulo de informações é uma preocupação crescente pois, segundo Piatetsky-Shapiro e Frawley (1991), estima-se que a quantidade de informação no mundo dobra a cada 20 meses. Um dos motivos para isso é a automação de atividades comerciais, que produz um fluxo sempre crescente de dados, uma vez que diversos tipos de transações, como venda de produtos, uso do cartão de crédito, entre outros, precisam ser armazenadas. Para realizar o armazenamento de todo esse volume de informações é necessário a utilização de um mecanismo que facilite o acesso aos dados, como o uso de banco de dados.

A partir da década de 1970, com o surgimento dos primeiros bancos de dados, o problema de armazenamento de informações começou a ser solucionado. Banco de dados é uma coleção lógica de dados, mantida em um ou mais arquivos, organizados para facilitar o armazenamento eficiente, modificação e recuperação de informações relacionadas (SELINGER et al., 1979).

Apesar de todo esse volume de dados armazenados, existem informações implícitas embutidas nesse amontoado de dados que ainda não foram devidamente analisadas. Tais dados podem revelar padrões, áreas de interesse, períodos favoráveis para determinados tipos de operações e inúmeras outras informações de extrema utilidade.

A maioria dos gerenciadores de bancos de dados possuem métodos para recu-



peração de informações, como por exemplo, o total de vendas do ano, custo médio de um produto e listagem dos itens mais consumidos durante o ano. Este tipo de resultado está disponível instantaneamente ao usuário.

As funções normalmente disponibilizadas pelos gerenciadores de banco de dados, como consultas, junções, agrupamentos e contagens, não são suficientes para o problema de análise de dados, apesar de possuírem fundamentação matemática sólida, conforme detalhado em Codd (1972). Entre os problemas na utilização das técnicas disponibilizadas por tais sistemas, há o fato dos resultados desta abordagem serem complexos e difíceis de interpretar.

Os dados, analisados e apresentados de forma inteligente, são um valioso recurso para ser usado como vantagem competitiva (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). Neste cenário, surge a necessidade da descoberta de conhecimento implícito no volume de dados, e para tal, são necessárias técnicas avançadas para análise inteligente e entendimento dos dados, o que representa um grande desafio científico e prático à comunidade da ciência da computação (HAN; FU, 1995).

As técnicas de análise inteligente devem identificar padrões de interesse. O resultado deve descrever de maneira compreensível e concisa o conhecimento adquirido.

De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), a descoberta de conhecimento em bases de dados é a extração não trivial de informação implícita, previamente desconhecida, e potencialmente útil. A extração não trivial refere-se ao emprego de métodos que vão além da computação estatística.

Um sistema de descoberta deve, além de ser capaz de decidir quais cálculos realizar e quais resultados são suficientemente interessantes para constituir conhecimento no contexto corrente, possuir certo grau de autonomia no processamento de dados e avaliação dos resultados (PIATETSKY-SHAPIRO; FRAWLEY, 1991).

A descoberta de conhecimento nos dados *Knowledge Discovery in Databases* (KDD) é o processo que mapeia dados não analisados, que normalmente são

volumosos para se entender e digerir facilmente, em uma forma mais compacta. Trata-se do desenvolvimento de métodos e técnicas para dar sentido aos dados (FAYYAD; PIATETSKY-SHAPIO; SMYTH, 1996).

O atual problema enfrentado pelos pesquisadores da área de descoberta de conhecimento nos dados não se refere ao armazenamento de dados coletados pelos mais diversos meios geradores de informações, mas sim ao consumo eficiente desse recurso a fim de gerar conhecimento útil que possa ser utilizado para prover melhorias nos cenários onde os próprios dados foram coletados.

Com a finalidade de extrair conhecimento útil, diversas abordagens visam examinar os dados em sua forma bruta. Algumas abordagens realizam a análise de informações buscando como resultado a formação de grupos de interesses, enquanto outras realizam a análise em grupos de interesse previamente formados, buscando como resultado informações que permitam classificar novos elementos dentre os grupos existentes.

Cada uma das técnicas mencionadas possuem pontos que devem ser observados. A análise de dados brutos é um processo lento e dispendioso, e está intimamente relacionado à um profissional com familiaridade com os dados analisados.

Abordagens que buscam realizar o agrupamento de informações muitas vezes necessitam de definição prévia de alguns parâmetros, o que é feito de forma subjetiva, pois não foram encontradas bibliografias que definam valores adequados, como por exemplo, o número de grupos resultantes desejado.

As técnicas de análises de grupos previamente formados leva em consideração que os grupos tenham sido corretamente divididos, e que os dados não possuem informações incorretas, o que nem sempre é verdade.

No contexto desse trabalho, os dados foram analisados através de grupos previamente classificados.

Existem diversos cenários onde a informação é tratada em formas de grupos, como por exemplo, grupos de pessoas que possuem determinada doença, ou grupo

de vinhos, entre outros. O agrupamento é realizado por alguém que possui grande conhecimento na área analisada, e normalmente, como no caso de doenças, são necessários vários anos ou mesmo décadas para que seja possível o correto delineamento das fronteiras que dividem tais grupos.

Os fatores que fizeram com que os elementos pertencessem a determinado grupo foram devidamente analisados por especialistas, garantindo a correta distinção entre os grupos formados. Para quaisquer novos elementos que devam ser classificados, é necessária a intervenção de alguém especializado, a fim de diagnosticar a presença ou ausência de atributos que classifiquem o novo elemento em algum dos grupos existentes.

Cada agrupamento de dados contém informações implícitas que foram determinantes para que o especialista realizasse corretamente a classificação dos elementos.

Métodos que consigam extrair tais informações permitiriam a elaboração de uma ferramenta automatizada que realizaria a classificação de novos elementos dentre os grupos formados, sem o auxílio de um especialista.

Os métodos para realizar tal extração de regras são o foco deste trabalho, e uma ferramenta que utiliza esses métodos é proposta com a finalidade de testar as soluções apresentadas.

Este trabalho é uma adaptação da pesquisa realizada no trabalho de (ALMEIDA, 2010) que utilizava as hipercaixas envolventes para treinamento de redes neurais. Neste trabalho a técnica foi aprimorada a fim de realizar a extração de regras utilizando hipercaixas difusas.

## 1.2 Motivação

O crescente aumento no volume de dados e a necessidade de aproveitamento deste massivo amontoado torna necessário o desenvolvimento de técnicas que permitam automatizar de certa forma parte do processo de análise de informações.

Em um cenário ideal, uma ferramenta deve ser capaz de separar áreas de interesse e obter informações relevantes desses dados.

A abordagem existente para análise baseia-se fundamentalmente em um ou mais analistas, intimamente familiarizados com os dados e os processos, servindo como uma ponte entre dados, usuários e produtos, sendo esta forma de análise lenta, cara e totalmente subjetiva, tornando-se praticamente inviável.

A elaboração de técnicas que auxiliem a obtenção de informações úteis de bases de dados vêm de encontro com a necessidade vigente, uma vez que diminui a necessidade de um analista especializado no processo de análise, e utiliza o poder de processamento do computador, em substituição da análise manual. A metodologia proposta visa atacar parte do problema mencionado.

Comumente, quando os dados são coletados e armazenados, não é feito nenhum tipo de classificação que torne possível separar tais dados por interesse, pois as informações não estão relacionadas entre si. As informações estão em seu estado bruto, em forma de tabelas ou alguma forma de armazenamento similar.

Antes de realizar qualquer tipo de análise sobre esses dados, é necessário lidar com alguns problemas, tais como a falta de informação de determinados atributos, informação supérflua que não altera o resultado com relação à classificação dos objetos e o alto volume de informações acumulados, dificultando a classificação destes dados para a tomada de decisões.

Algumas técnicas existentes visam tratar os problemas mencionados, dentre elas estão:

### **1.2.1 Conjuntos Aproximados (Rough Sets):**

Visam a extração de regras a partir da redução da base de conhecimento a um conjunto mínimo de atributos, eliminando a informação supérflua e encontrando padrões escondidos nos dados (PAWLAK; SKOWRON, 2007).

Esta abordagem sofre do problema na discretização dos valores, onde um determinado conjunto de valores contínuos ora leva a uma decisão ora leva a outra. Na tabela 1 é possível visualizar valores iguais, mas que representam decisões distintas. Embora as sequências 3 e 5 possuam os mesmos valores de atributos, levam a uma decisão distinta.

<b>Sequência</b>	<b>Cor</b>	<b>Tamanho</b>	<b>Material</b>	<b>Decisão</b>
1	Azul	Grande	Liso	Positiva
2	Vermelho	Médio	Indefinido	Neutra
<b>3</b>	<b>Verde</b>	<b>Pequeno</b>	<b>Áspero</b>	<b>Negativa</b>
4	Azul	Grande	Liso	Positiva
<b>5</b>	<b>Verde</b>	<b>Pequeno</b>	<b>Áspero</b>	<b>Neutra</b>

Tabela 1: Problema de discretização - Conjuntos Aproximados

### 1.2.2 Redes Neurais Artificiais (Artificial Neural Networks):

As *Redes Neurais Artificiais* (RNA) processam informações de maneira similar ao cérebro humano. A rede é composta por elementos de processamento interconectados, os neurônios, trabalhando paralelamente a fim de solucionar um problema específico. A rede aprende através de exemplos (WASSERMAN, 1993).

As redes neurais possuem a desvantagem de não demonstrar a lógica interna de seu funcionamento, e por isso são analisadas como caixas pretas, ou seja, apesar de capturar relações implícitas entre os dados de entrada e os dados de saída, tais redes não fornecem as informações obtidas durante o treinamento da rede, o que não torna claro como a rede chegou ao resultado obtido.

As abordagens citadas não deixam claras as características que representam apropriadamente um conjunto de dados.

A plataforma desenvolvida visa detectar as similaridades nos dados que fizeram com que determinados elementos fossem alocados em determinado grupo, de tal maneira que seja possível realizar a separação de novos elementos, ainda não classificados, dentre os grupos formados, levando em consideração os mesmos fatores

que levaram a formação do grupo com os elementos previamente existentes.

A proposta é prover meios para evidenciar os fatores que fizeram com que os dados fossem separados dentre os grupos formados, de tal forma que seja possível a obtenção de regras de classificação que permitam a correta alocação de novos elementos dentro do grupo que mais se assemelha às suas particularidades.

A ideia é extrair a relação que existe entre os dados de um determinado grupo através de regras, que permitam uma correta classificação de novos elementos utilizando inteligência computacional.

## 1.3 Objetivos

### 1.3.1 Objetivo Geral

O objetivo desse projeto é desenvolver uma nova técnica de classificação supervisionada para análise de grupos de informações de modo a identificar em seus atributos as características implícitas que permitiram agrupá-los corretamente dentre os grupos existentes, e com o uso de tais características, extrair regras difusas que permitam a classificação de novos elementos, ainda não classificados, dentre os grupos existentes.

Em contraste com as técnicas existentes mencionadas, a metodologia apresentada é capaz de lidar com possíveis erros no conjunto de dados, bem como é capaz de processar conjuntos de dados cujos valores de decisão variam para um mesmo conjunto de entrada, conhecido como problema de discretização, apresentado no tópico 1.2.1.

Os métodos utilizados para atingir o objetivo do projeto utilizam técnicas amplamente conhecidas e matematicamente comprovadas, e com isso, expõe de forma clara todos os passos necessários para se alcançar o resultado, em contraposição às redes neurais, conhecidas como caixas pretas, que ocultam a lógica utilizada para se alcançar o resultado.

Para realizar a análise foram utilizadas técnicas comumente aplicadas em detecção de colisão em jogos e ambientes 3D, através de projeção utilizando representações gráficas.

Para aumentar o grau de confiabilidade em relação à representatividade de cada elemento dentre os conjuntos formados, foi utilizada a lógica difusa, a fim de permitir uma melhor classificação de cada elemento e possibilitar a visualização de áreas de incerteza, nos casos onde não foi possível capturar corretamente as propriedades delimitadoras dos atributos analisados.

### 1.3.2 Objetivos Específicos

Abaixo são apresentados outros objetivos explorados neste trabalho:

- Realizar uma revisão bibliográfica sobre técnicas de colisão de objetos e lógica difusa, além de avaliar o estado da arte na área de extração e descoberta de conhecimento;
- Desenvolver uma metodologia híbrida de classificação baseada em *Inteligência Artificial* (IA) que permita distinguir os elementos entre os grupos definidos;
- Desenvolver um aplicativo computacional utilizando a metodologia proposta de forma a servir como plataforma de testes;
- Utilizar as diversas bases de testes de domínio público com o algoritmo desenvolvido, a fim de verificar sua eficácia;
- Analisar e validar os resultados desta metodologia buscando a melhor solução para o problema;
- Publicar os resultados obtidos com a técnica desenvolvida.

## 1.4 Organização do trabalho

O trabalho está estruturado da seguinte forma:

No capítulo 2 é realizada uma revisão bibliográfica sobre descoberta de conhecimento em dados.

O capítulo 3 apresenta a fundamentação teórica sobre colisão de objetos, conjuntos difusos e desvio padrão.

O capítulo 4 descreve o desenvolvimento da técnica proposta utilizando os conceitos definidos anteriormente.

No capítulo 5 são apresentados os testes e resultados obtidos com a técnica proposta. A metodologia utilizada nos testes bem como as bases de dados utilizadas são descritas neste capítulo.

O capítulo 6 trata da conclusão do trabalho e o capítulo 7 apresenta sugestões de melhorias e trabalhos futuros.



## 2 Revisão Bibliográfica

### 2.1 Descoberta de conhecimento em bases de dados

Devido a grande quantidade de informações coletadas e armazenadas, há a necessidade de uma nova geração de teorias computacionais e ferramentas para auxiliar a extração de informações. O conjunto dessas ferramentas e teorias é o assunto conhecido como descoberta de conhecimento em bases de dados, ou *Knowledge Discovery in Databases* (KDD) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Data mining é um dos processos envolvidos na descoberta de conhecimento em bases de dados. O processo consiste no mapeamento de dados de baixo nível, que normalmente são volumosos para serem entendidos e digeridos, em outras formas que devem ser mais compactas, como por exemplo, um relatório sintetizado.

A Figura 1 demonstra os passos a serem realizadas até que o conhecimento seja extraído dos dados. Algumas das tarefas que comumente devem ser executados sob os dados são:

- Seleção dos dados que serão analisados;
- Pré-processamento para eliminar possíveis erros nos dados;
- Transformação nos dados para torná-los analisáveis pelo computador;
- O processo de mineração de dados;

- Interpretação e avaliação do resultado obtido;
- Extração do conhecimento obtido.

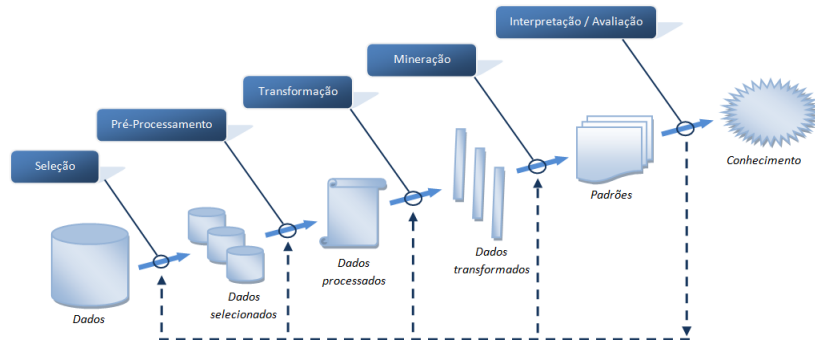


Figura 1: Passos para transformação de dados brutos em conhecimento

Existe na literatura uma grande variedade de métodos de mineração de dados que visam atender a um objetivo único: dar sentido aos dados. Dentre eles estão:

### 2.1.1 Árvores de decisão

Constituem uma poderosa técnica amplamente utilizada em problemas de classificação. A técnica é amplamente utilizada devido ao fato do conhecimento adquirido ser representado a partir de regras, que podem ser expressas em linguagem natural, o que facilita o entendimento por parte de humanos.

O método utilizado pela técnica consiste na subdivisão de uma grandeza em unidades menores, criando nesse processo uma árvore de caminhos e escolhas (QUINLAN, 1986).

Como exemplo de utilização do método, iremos utilizar um contexto de dados onde a grandeza com maior significado é o atributo *idade*.

A primeira divisão dessa grandeza ( $A$ ) separa os elementos dentre duas classes, e o critério de seleção é a idade em que a pessoa alcança a maioridade (18 anos). No próximo passo, a grandeza é subdividida em 2 novas classes, com os valores para

menoridade (incapazes, inimputáveis) e para a maioria (capaz, imputáveis). A Figura 2 exibe árvore de decisão resultante.

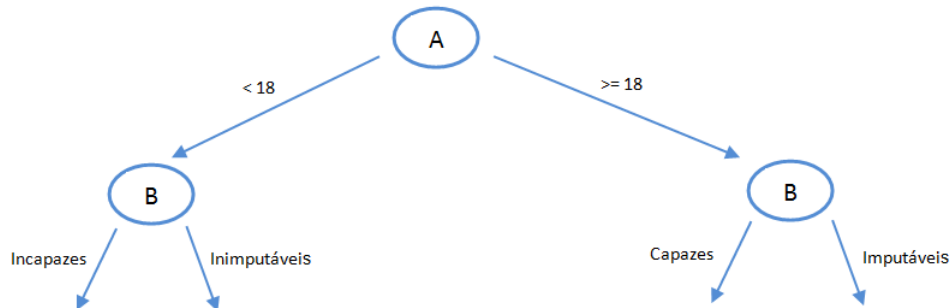


Figura 2: Árvore de decisão: Maioridade e menoridade

O processo se repete, formando assim uma árvore de decisão. Em cada extremidade da árvore encontra-se uma resposta delineando a decisão a ser tomada.

Existem diversos algoritmos baseados na técnica de árvores de decisão, sendo os mais conhecidos: ID3, C4.5 (sucessor do ID3) e CART (HAN; KAMBER, 2000).

### 2.1.2 Regras de Associação

As regras de associação têm como premissa básica encontrar elementos que implicam na presença de outros elementos em uma mesma transação, ou seja, encontrar relacionamentos ou padrões frequentes entre conjuntos de dados (AGRAWAL; IMIELINSKI; SWAMI, 1993).

O conceito de transação indica quais itens foram pesquisados em uma determinada operação de consulta. Comumente, as regras de associação representam padrões existentes em transações armazenadas em grandes bases de dados (AGRAWAL; SRIKANT, 1994).

Com o uso de regras de associação em uma estratégia de mineração, uma possível regra gerada a partir de uma base de dados com registros de itens adquiridos por clientes poderia ser:

*Regra: {cinto, sapato} ⇒ {bolsa}*

Indicando que, com um determinado grau de certeza, um cliente que adquire um cinto e um sapato também adquira uma bolsa.

### 2.1.3 Agrupamento de dados

O agrupamento ou clusterização de dados é um processo que visa segmentar populações heterogêneas em grupos ou segmentos homogêneos. Trata-se da identificação e formação dos grupos de objetos dentro de classes de objetos similares.

Para a formação dos grupos, os objetos são clusterizados baseados no princípio da maximização da similaridade intra classe e minimização da similaridade entre classes.

A ideia deste princípio é a de que os grupos de objetos são formados de tal maneira que objetos dentro do cluster possuem alta similaridade em comparação a outros objetos do grupo, porém são muitos dissimilares em relação a objetos de outros grupos.

O conceito de distância dos dados em relação ao centro de grupo é utilizado para medir a similaridade entre os objetos (EBERLY, 2001).

### 2.1.4 Conjuntos Aproximados (Rought Sets)

A Teoria dos Conjuntos Aproximados, proposta por Zdzislaw Pawlak em 1982, é utilizada para o tratamento de incertezas e imprecisões, e é considerada como uma extensão da Teoria Clássica dos Conjuntos.

A técnica possui propriedades que permitem eliminar variáveis ou atributos irrelevantes através do processo de redução do sistema de informação. Para realizar tal redução, baseia-se no conceito de redutos, que são subconjuntos de propriedades capazes de representar o conhecimento da mesma forma como quando utilizando todo o conjunto de atributos.

O processo de eliminação de propriedades irrelevantes é uma das características dessa teoria.

Outra habilidade da técnica se refere à classificação de objetos. De acordo com suas propriedades, os objetos de dados são agrupados em classes. Os objetos que estão em um mesmo agrupamento são indiscerníveis entre si.

Na maioria das vezes, os dados que estão sendo analisados possuem informações imprecisas ou incompletas, e a teoria dos conjuntos aproximados é capaz de administrar tais imprecisões. Objetos que não podem ser especificados através dos dados disponíveis são classificados por esta teoria através de dois conceitos: as aproximações inferior e superior.

A utilidade desta teoria pode ser comprovada pelo crescente número de aplicações disponíveis que utilizam a técnica, tanto no campo de mineração de dados, agrupamento de dados (clusterização) quanto no processamento de sinais de imagem e som.

Esta técnica é fundamentada em teorias matemáticas, e tal característica possibilita a descoberta de padrões ocultos na base de processamento (PAWLAK, 1982), (PAWLAK, 1991).

### 2.1.5 Redes Neurais

Como uma subárvore ligada à Inteligência Artificial, as *Redes Neurais Artificiais* (RNA) utilizam uma estrutura básica em seu funcionamento, o neurônio, para representar o conhecimento obtido. Tal estrutura busca simular o funcionamento do neurônio do cérebro, e prover dessa maneira a inteligência artificial (LIPPMANN, 1987).

Redes Neurais são tipicamente organizadas em camadas, constituídas de um número de nós interconectados, que possuem uma função de ativação. Padrões são apresentados à rede através das camadas de entrada, que se comunica com uma

ou mais camadas ocultas, onde o processamento é efetuado através de um sistema de conexões ponderadas. As camadas ocultas são ligadas à camada de saída, onde a resposta é apresentada.

Segundo Caudill (1987), as Redes Neurais Artificiais são um sistema de computação constituído por um número de elementos de processamento altamente interconectados, que processam a informação dinamicamente em resposta aos fatores externos.

Apesar do advento das redes neurais datar da década de 40, devido a limitação em sua capacidade de aprendizagem automatizada, a área de pesquisa ficou estagnada até a década de 80. Em 1986, Rumelhart, Hinton e Williams (1986) apresentou a comunidade científica seu trabalho, que deu origem ao algoritmo de retropropagação do erro (Backpropagation) que permitiu a retomada dos estudos relativos às redes neurais.

Associado individualmente a cada um dos neurônios ocultos, e a cada neurônio de saída da rede neural, está um conjunto de pesos (valores ponderados) e um valor simples chamado *bias*. Os pesos e *bias* determinam os valores de saída para um conjunto de valores de entrada.

O algoritmo de retropropagação do erro visa solucionar o desafio de encontrar o conjunto de valores para pesos e *bias* que possibilitem gerar o resultado que melhor representam os dados existentes.

## 2.2 Trabalhos Correlacionados

Na literatura existem inúmeras abordagens relacionadas à extração de conhecimento. Esta seção visa analisar o estado da arte de algoritmos de classificação, listando os principais pontos de cada uma das abordagens adotadas.

No trabalho de (PARPINELLI; LOPES; FREITAS, 2002), um algoritmo para mineração de dados chamado Ant-Miner foi proposto. O objetivo do algoritmo é

realizar a extração de regras dos dados, e foi inspirado tanto pela pesquisa sobre o comportamento de colônias de formigas reais quanto por conceitos de mineração de dados. A comparação de performance entre o algoritmo proposto e o algoritmo CN2, um conhecido algoritmo de mineração de dados para classificação, evidenciou que o Ant-Miner é competitivo com o CN2 em relação à precisão da previsão e que a lista descoberta pelo Ant-Miner é consideravelmente menor do que a lista descoberta pelo CN2.

Em (LIANG; LEE; LEE, 2011) os autores propõem uma extensão do Ant-Miner, com a finalidade de incorporar o conceito de processamento paralelo e agrupamento. A intercomunicação via feromônio entre formigas é uma parte crítica no mecanismo de busca de otimização baseado em colônia de formiga. A extensão proposta visa modificar o projeto de algoritmos a fim de incorporar processamento paralelo. A fim de resolver o problema de classificação de regras, foi projetado um algoritmo com o conceito de mecanismo de escolha de regra multi nível a fim de obter maior precisão da regra induzida.

O algoritmo *Colônia de Formigas* (Ant colony optimization - ACO), foi alterado no trabalho de (OTERO; FREITAS; JOHNSON, 2013) a fim de solucionar o problema da interação de regras, onde o resultado de uma regra afeta as regras que podem ser detectadas subsequentemente uma vez que o espaço de busca é modificado devido à remoção dos exemplos abrangidos pelas regras anteriores. O trabalho apresenta uma nova estratégia, a fim de solucionar o problema mencionado, onde a ordem das regras é implicitamente codificada como valores de feromônio, e a busca é guiada pela qualidade de uma lista de candidatos de regras. De acordo com os testes realizados, a precisão da previsão obtida pelo novo algoritmo de classificação ACO implementando a estratégia de cobertura seqüencial proposta é significativamente mais elevada do que a precisão preditiva dos algoritmos de classificação de indução de regras atuais.

Com a finalidade de explorar um método de classificação mais eficiente, (FENG, 2012) apresenta um algoritmo de classificação supervisionado baseado em imuni-

dade artificial. Sistemas de imunidade artificial são uma classe de sistemas inteligentes baseados em princípios e processos do sistema imunológico dos vertebrados. Estes algoritmos exploram as características de aprendizado e memória do sistema imunológico a fim de resolver problemas. O algoritmo apresentado possui seleção não linear e de clone, regulação e memória imunológicas e outras características do sistema imunológico biológico. Tais características tornam o algoritmo uma nova solução para problemas de classificação supervisionado. Nos trabalhos de (LIANG et al., 2010) e (SOTIROPOULOS; TSIHRINTZIS, 2012) são realizados estudos sob algoritmos de classificação baseados em imunidade artificial.

Um dos problemas mais importantes em qualquer minerador de regras de associação é a interpretação e avaliação das regras obtidas. A maioria dos algoritmos empregam o framework de suporte-confiança para avaliar as regras de associação e classificação. Estudos recentes mostraram que estas medidas não são suficientes para filtrar regras de associação que não são interessantes, mesmo regras fortes podem ser desinteressantes e enganosas. No artigo de (TABOADA et al., 2009), medidas de interesse adicionais baseadas em significância estatística e análise correlacional é proposta. O algoritmo utiliza lógica difusa e redes de programação genética para descobrir regras de associação difusas compreensíveis úteis na classificação. Os resultados do trabalho mostraram as vantagens e efetividade do modelo proposto.

As técnicas mencionadas utilizam metodologias diversificadas a fim de aumentar a performance e a qualidade dos resultados na tarefa de classificação de dados. As abordagens citadas variam significativamente em relação à técnica proposta devido ao fato deste trabalho utilizar técnicas de colisão de jogos em conjunto com lógica difusa a fim de analisar os dados sob uma perspectiva de hipercaixas envolventes, usando técnicas amplamente difundidas na detecção de colisão de objetos para extrair regras difusas de classificação.



## 3 Metodologia

### 3.1 Colisão de Objetos

A detecção de colisão entre objetos é um problema fundamental em inúmeras áreas de estudo. Diversas aplicações são encontradas na robótica, computação gráfica e jogos de computadores 3D (WANG; WANG; LI, 2009).

O objetivo da detecção de colisão é denunciar automaticamente um ou mais contatos geométricos que estão prestes a ocorrer ou que realmente ocorreu entre os objetos.

Gottschalk, Lin e Manocha (1996) evidencia que são necessários métodos que detectem colisões entre objetos de uma maneira eficiente. A detecção de pontos de colisão entre objetos é uma tarefa que demanda grande esforço computacional, e com isso o procedimento pode não ser executado dentro de limites de tempo aceitáveis.

Segundo Eberly (2001), a colisão entre dois objetos  $A$  e  $B$  ocorre quando o objeto  $A$  interfere na geometria do objeto  $B$ , conforme Figura 3.

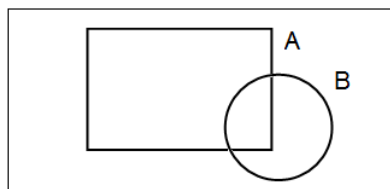


Figura 3: Colisão entre os objetos  $A$  e  $B$

A colisão entre dois ou mais objetos não altera as dimensões e tamanhos das entidades envolvidas. A colisão não causa deformações, apenas afetam a localização ou movimento dos objetos envolvidos.

A detecção de colisão entre objetos se faz necessária para prover um modelo computacional aprimorado, simulando o ambiente real em um ambiente computacional, evitando que objetos se sobreponham, permitindo a simulação de atividades e projetos dentro do espaço virtual da mesma forma como seria em um ambiente real (JIMÉNEZ; THOMAS; TORRAS, 2001).

Tanto em simulações quanto em jogos, o ambiente computacional criado pode possuir centenas ou milhares de objetos, e por vezes, tais objetos possuem, além de suas características básicas, como tamanho e altura, ações de movimento, como rotação, translação, etc.

A detecção de colisão em tais ambientes pode torna-se um gargalo computacional, o que pode comprometer a eficácia de utilização da ferramenta desenvolvida, tornando-a inviável.

Ferramentas e métodos que diminuam o esforço computacional dedicado à detecção de colisões e, conseqüentemente, melhore o tempo de resposta da tarefa de detecção de colisão entre objetos faz-se necessários.

A fim de reduzir esse esforço, são utilizados volumes envolventes, que são estruturas simples que contém objetos complexos, que visam reduzir o número de testes necessários para realizar a tarefa de detecção de colisão (GOTTSCHALK, 2000).

A razão pela qual os volumes envolventes são mais rápidos deve-se ao fato que efetuar o cálculo entre dois volumes simples exige menos esforço computacional do que calcular a colisão entre geometrias complexas.

Entre os volumes envolventes existentes podemos citar:

- Esfera;
- Caixas Alinhadas aos Eixos e;

- Caixas Orientadas.

O método utilizado por cada um desses volumes, bem como os prós e contras de cada modelo é detalhado a seguir:

### 3.1.1 Volume em Esfera

O volume em esfera é o mais simples dos tipos de volumes envolventes. Conforme pode ser observado na Figura 4, os objetos são envolvidos por uma esfera, que é representada pelo centro e raio.

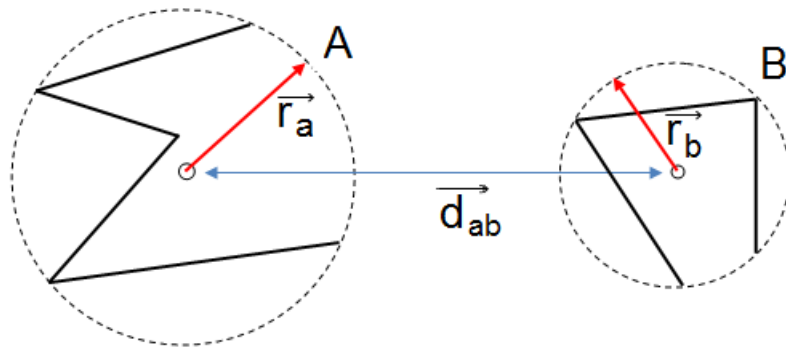


Figura 4: Detecção de não colisão utilizando esferas

Duas esferas estão colidindo quando a distância entre os centros  $\|\vec{d}_{ab}\|$  for menor do que a soma dos raios  $(\vec{r}_a + \vec{r}_b)$ , conforme Equação 3.1.

$$\left\{ \begin{array}{l} \|\vec{r}_a\| + \|\vec{r}_b\| > \|\vec{d}_{ab}\| \rightarrow \text{colide} \\ \|\vec{r}_a\| + \|\vec{r}_b\| = \|\vec{d}_{ab}\| \rightarrow \text{tangencia} \\ \|\vec{r}_a\| + \|\vec{r}_b\| < \|\vec{d}_{ab}\| \rightarrow \text{nao colide} \end{array} \right. \quad (3.1)$$

Equação 3.1: Teste de colisão para o volume em esfera

Na Equação 3.1,  $r$  é o raio da esfera e  $d$  é a distância entre os centros das esferas.

O uso do volume envolvente em esfera possui a vantagem de possuir testes de colisão rápidos. A desvantagem é que o volume tende a acusar falsas colisões com alta frequência.

Uma falsa colisão ocorre quando é detectado contato entre as estruturas que envolvem dois ou mais objetos, porém os objetos em si não estão realmente em contato, o que pode ser observado pela figura 5.

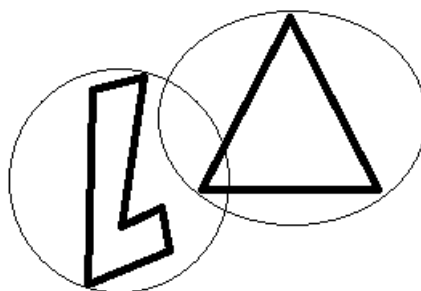


Figura 5: Falsa colisão entre objetos

### 3.1.2 Volume de Caixas Alinhadas ao Eixo - AABB

No volume envolvente de caixas alinhadas aos eixos - *Axis Aligned Bounding Box* (AABB), os objetos em análise são envolvidos por uma caixa, onde cada face é paralela aos eixos do sistema de coordenadas, como ilustra a Figura 6.

Volumes AABB se ajustam melhor ao objeto envolvido, comparado ao volume de esferas. A colisão em volumes AABB ocorre quando a distância entre os centros das caixas é menor do que a soma das distâncias do centro até a lateral de cada caixa, para todas as direções.

Como pode ser visualizado pela Figura 7, se houver ao menos um eixo na qual as projeções não se sobreponham, então não há colisão.

O teste de colisão para o volume AABB é mostrado na Equação 3.2. O cálculo

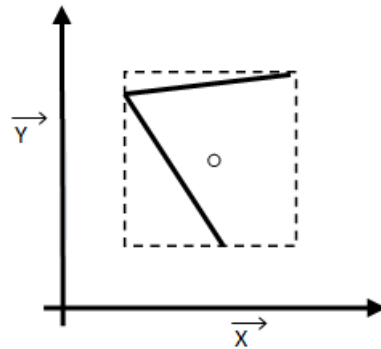


Figura 6: Volume envolvente AABB

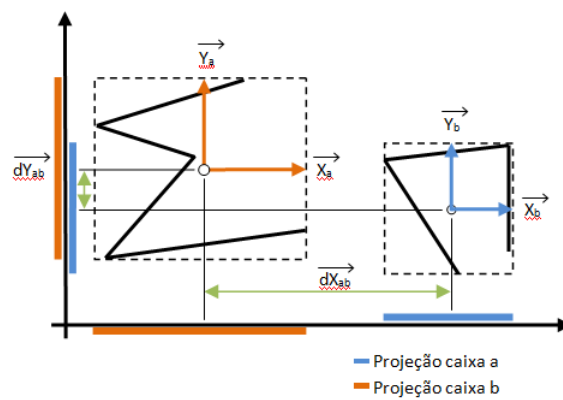


Figura 7: Detecção de não colisão utilizando AABB

é similar ao efetuado para detecção de colisão em volumes em esfera. A colisão não ocorre quando há pelo menos um eixo na qual as projeções não se sobreponham.

$$\begin{cases} \vec{m}_a + \vec{m}_b > \vec{d}_{m_{ab}} \rightarrow \text{colide} \\ \vec{m}_a + \vec{m}_b \leq \vec{d}_{m_{ab}} \rightarrow \text{nao colide} \end{cases} \quad (3.2)$$

Equação 3.2: Teste de colisão para o volume AABB

Na Equação 3.2,  $m$  é o eixo na qual o teste está sendo efetuado,  $m_a$  é a distância entre o centro e a lateral da caixa  $a$  no eixo  $m$  e  $d_{m_{ab}}$  é a distância entre os centros da caixa  $a$  e  $b$ , no eixo  $m$ .

### 3.1.3 Volume de Caixas Orientadas - OBB

O volume de Caixas Orientadas - *Oriented Bounding Box* (OBB) são caixas similares às AABBs, com a diferença de que a orientação nas OBBs se dá de acordo com a orientação do objeto envolvido, enquanto que nas AABBs a orientação se dá em relação aos eixos de coordenadas do sistema.

A Figura 8 ilustra dois objetos envolvidos com o volume de caixas orientadas que não estão colidindo.

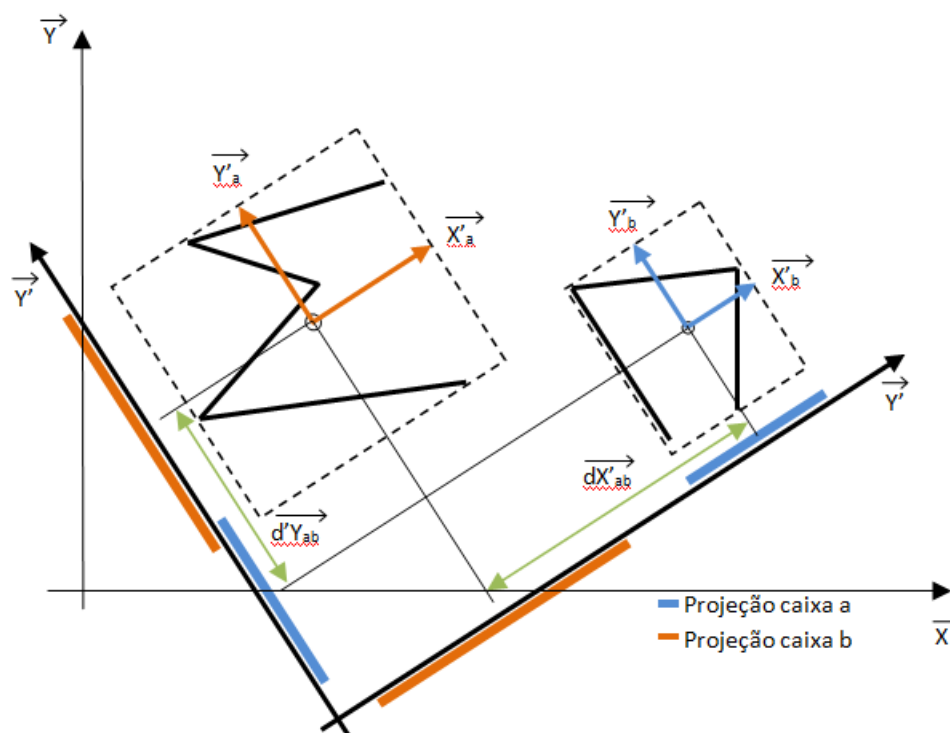


Figura 8: Detecção de não colisão utilizando OBB

As OBBs aumentam a precisão na representação do modelo, ao custo de uma maior utilização de recursos para armazenar informações referentes à representação do objeto.

Os volumes envolventes OBBs possuem testes de colisão mais lentos, porém necessitam de menos testes para se verificar a colisão, enquanto que volumes simples,

como esferas e AABBs, possuem testes de colisão mais rápidos, porém necessitam de mais testes para se verificar a colisão (GOTTSCHALK, 2000).

Outra vantagem das OBBs em relação às AABBs é que a medida que o objeto é rotacionado, as OBBs não necessitam que o volume envolvente seja reconstruído, ao passo que nas AABBs a caixa deve ser reconstruída a cada rotação do objeto envolvido.

### 3.1.4 Teorema do Eixo de Separação - SAT

Apresentado em Gottschalk (1996) o Teorema do Eixo de Separação - SAT (Separating Axis Theorem), determina um conjunto de condições para que dois objetos não estejam se interceptando. O teste verifica se há um plano de separação entre os objetos, realizando a projeção das caixas em cada um dos eixos no espaço.

A projeção das caixas produz um intervalo no eixo, e caso exista algum eixo na qual os intervalos não se sobreponham, esse eixo é chamado de *eixo de separação*.

Caso as sombras se sobreponham, mais testes podem ser necessários para verificar se é possível separar as caixas (GOTTSCHALK; LIN; MANOCHA, 1996).

Os centros das caixas são projetados nos eixos e o raio de cada projeção resultante é calculado. Caso a distância entre os centros das caixas projetadas no eixo for maior do que a soma dos raios, então os objetos estão separados no espaço.

Conforme pode ser observado na Figura 9, após a projeção dos objetos envolvidos em um determinado eixo, há uma linha de separação entre eles. O eixo onde os objetos foram projetados é o Eixo de Separação.

### 3.1.5 Considerações

Segundo Jiménez, Thomas e Torras (2001), o custo computacional dos algoritmos de detecção de colisão dependem não apenas da complexidade do teste básico de interferência, mas também no número de vezes que o teste é aplicado.

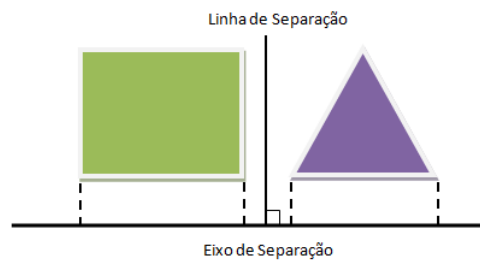


Figura 9: Linha e Eixo de Separação entre dois objetos.

A decisão referente a qual volume envolvente utilizar depende da necessidade da aplicação a ser desenvolvida. Caso a aplicação necessite de alta qualidade na detecção de colisões, é necessário um volume envolvente que se ajuste mais firmemente ao objeto. Para o desenvolvimento de jogos e simulações de resposta rápida, uma contagem simples de superfícies de volumes é mais adequada.

Volumes envolventes simples realizam testes de interseção mais rápidos, enquanto volumes que se ajustam melhor à geometria envolvida oferece precisão a troco de testes de interseção mais lentos.

A detecção de colisão admite várias formulações do problema, dependendo do tipo de saída desejada e as restrições impostas para as entradas. O teste básico de colisão informa apenas se há ou não colisão, não fornecendo maiores detalhes, como o intervalo de tempo até que haja colisão entre objetos que se movimentam.

A principal etapa dos algoritmos de detecção de colisão corresponde a subdivisão do espaço ocupado pelos objetos, a fim de determinar a colisão ou não colisão entre eles (GOTTSCHALK; LIN; MANOCHA, 1996).

O método consiste em dividir o maior eixo do volume envoltório com seu eixo cartesiano correspondente. A divisão de uma geometria a fim de tentar solucionar colisões entre objetos é o único evento que faz com que os testes de colisão tenham que ser aplicados novamente.



### 3.1.6 Justificativa - Escolha do Volume Envolvente

A seguir serão detalhados os motivos que justificam o uso do volume envolvente AABB (Axis Aligned Bounding Box) no contexto deste trabalho em contraposição aos outros volumes mencionados:

- Volume em Esfera:

A geração de regras de identificação de grupos para o volume em esfera torna-se muito complexo quando se tratando de ambientes multidimensionais. Apesar de possuir testes de colisão rápidos e que demandam pouco esforço computacional, o volume em esfera acusa falsas colisões com muita frequência.

- Volume de Caixas Alinhadas ao Eixo:

Os testes de colisão para o volume de caixas alinhadas ao eixo são rápidos e demandam pouco esforço computacional, além de não acusar falsas colisões com tanta frequência, como no volume em esfera.

O motivo de sua maior eficácia em relação às falsas colisões é o fato de que a caixa envolvente gerada com este volume se adapta melhor ao objeto envolvido.

A desvantagem deste volume é que a caixa deve ser recriada toda vez que o objeto é rotacionado. A maior vantagem desse volume no contexto utilizado, e que levou à decisão de utilizá-lo para implementação da técnica, é que as caixas criadas são alinhadas com o eixo de coordenadas do sistema, o que não exige nenhum esforço computacional no momento da geração das regras de identificação, ao contrário do volume orientado.

- Volume de Caixas Orientadas:

Dos volumes envolventes mencionados, o que possui testes de colisão mais lentos e o que exige maior esforço computacional é o volume de caixas orientadas (GOTTSCHALK; LIN; MANOCHA, 1996).

A vantagem desse volume é devido ao fato de que caso o objeto seja rotacionado, a caixa não precisa ser reconstruída, pois as caixas geradas são orientadas pelo eixo do objeto envolvido e não pelo eixo de coordenadas do sistema. Tal vantagem em relação aos outros volumes, no contexto desse trabalho, se tornou o motivo que impossibilitou sua utilização.

Como cada caixa gerada possui seus próprios eixos de orientação, relativos ao objeto envolvido, não é possível fazer a comparação entre os valores de dados cuja caixa está envolvendo com os valores de dados das outras caixas geradas, logo, para possibilitar a comparação, é necessário realizar a transformação dos eixos da caixa, do eixo de coordenadas relativo ao objeto envolvido para o eixo do sistema de coordenadas.

Na Figura 10 é possível visualizar que o eixo referente ao volume orientado é diferente do eixo de coordenadas do sistema, utilizado pelo volume de caixas orientadas.

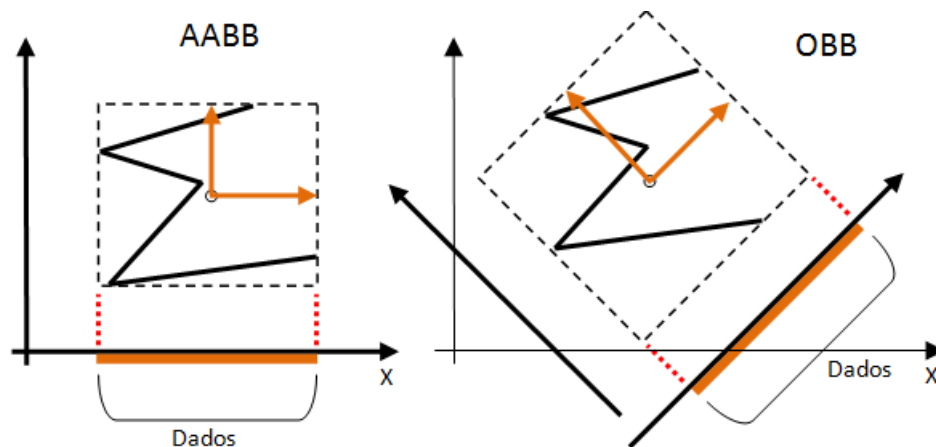


Figura 10: AABB - Eixo de coordenadas do sistema e OBB - Eixo de coordenadas orientado

## 3.2 Conjuntos Difusos

Conjuntos difusos, também chamados de Fuzzy Sets, são classes de objetos que possuem um grau contínuo de pertinência. Tal conjunto é caracterizado por uma função de pertinência, que atribui para cada objeto do conjunto um grau de pertinência variando de zero a um.

A noção de união, interseção, complemento, etc., presente na teoria dos conjuntos, são estendidos para os conjuntos difusos.

Conjuntos na forma matemática usual são ineficientes quando se tratando de classes de dados imprecisos ou expressos em linguagem formal, como por exemplo, "*conjunto de pessoas ricas*" ou "*conjunto de mulheres bonitas*".

O que torna essas classes diferentes das classes que são bem definidas no sentido convencional da matemática é a imprecisão de seus limites. Para classes que possuem fronteiras limítrofes incertas, um objeto pode possuir um grau de pertinência nessa classe variando de totalmente pertinente e não pertinente (ZADEH, 1978).

Algumas palavras definem o contexto de incerteza, como por exemplo, "João é *alto*", "*x é aproximadamente* igual a 20". Nos exemplos mostrados, as palavras escritas em itálico são rótulos para conjuntos difusos.

A relevância dos limites incertos existentes em tais conjuntos se deve ao fato de possibilitar a resolução de problemas que seriam praticamente impossíveis ou inviáveis devido à sua complexidade.

Sistemas de controle de tráfego são um exemplo de problema cuja resolução torna-se possível utilizando conjuntos difusos.

Na equação 3.3 é mostrada a definição de conjunto difuso:

Seja  $X = x$  um espaço de pontos (objetos), onde  $x$  é um elemento qualquer de  $X$ . Um conjunto difuso  $A$  em  $X$  é um conjunto de pares ordenados:

$$A = \{[x, \mu_A(x)]\} \quad x \in X \quad (3.3)$$

Equação 3.3: Definição de conjunto difuso

onde  $A(x)$  é denominado grau de pertinência de  $x$  em  $A$ . Com isso, se  $\mu_A(x)$  obter valores em um espaço  $M$  - denominado espaço de pertinência - então  $A$  é essencialmente uma função de  $X$  para  $M$ . A função  $\mu_A : X \rightarrow M$ , que define  $A$ , é chamada de função de pertinência de  $A$ .

Quando  $M$  contém apenas dois pontos 0 e 1,  $A$  é não difuso e sua função de pertinência reduz-se à característica convencional de função de um conjunto não difuso.

Tal conjunto  $A$  em  $X$  é uma classe sem limites bem definidos, ou seja, uma classe onde um ponto (objeto)  $x$  pode ter um grau de pertinência intermediário entre totalmente pertinente e não pertinente.

Um conjunto difuso pode ser precisamente definido associando-se para cada  $x$  um grau de pertinência em  $A$ . Deduz-se que  $M$  é um intervalo  $[0, 1]$ , onde 0 representa não pertinente e 1 representa totalmente pertinente no conjunto difuso.

### 3.3 Teoria da Possibilidade

Quando a preocupação principal acerca dos dados gira em torno do significado da informação, ao invés do peso de seus valores, o ambiente apropriado para análise de informações é possibilístico, ao invés de probabilístico, o que implica a utilização da teoria da possibilidade, ao invés da teoria da probabilidade (ZADEH, 1978).

A teoria da possibilidade, apresentada por L.A. Zadeh em 1978, é uma teoria matemática para lidar com certos tipos de incertezas, de forma alternativa à teoria da probabilidade. A teoria da possibilidade é fundamentada em conceitos matemáticas dos conjuntos difusos.

A importância da teoria da possibilidade vem do fato que muitas das informações nas quais os humanos se baseiam na tomada de decisões são possibilísticas, ao invés de probabilísticas.

Devido à natureza tipicamente incerta das linguagens naturais, de origem possibilista, é possível construir uma linguagem universal onde a tradução de proposições expressadas em linguagem natural é dada através de um sistema capaz de computar a distribuição da possibilidade de um conjunto de relações difusas de um banco de dados, expressando assim informações de forma sintetizada.

Abaixo é detalhado o conceito de distribuição da possibilidade:

Seja  $X$  uma variável que pode assumir valores do universo  $U$ , com um elemento qualquer de  $U$  denotado por  $u$ :

$$X = u \tag{3.4}$$

onde  $X$  recebe o valor de  $u$ ,  $u \in U$ .

Seja  $F$  um subconjunto difuso de  $U$ , caracterizado pela função de pertinência  $\mu_F$ .

$F$  é uma restrição difusa (*fuzzy restriction*) de  $X$  se  $F$  atuar como um elástico restringindo os valores que podem ser atribuídos a  $X$ , de forma que a atribuição do valor  $u$  para  $X$  tem a seguinte forma:

$$X = u : \mu_F(u) \tag{3.5}$$

onde  $\mu_F(u)$  é interpretado como o grau na qual a restrição representada por  $F$  é satisfeita quando  $u$  é atribuído em  $X$ . Equivalentemente,  $1 - \mu_F(u)$  é o grau na qual a restrição em questão precisa ser apertada a fim de permitir a atribuição de  $u$  para  $X$ .

Seja  $R(X)$  defino como a restrição difusa associada com  $X$ . Assim, para ex-

pressar que  $F$  desempenha o papel restritivo difuso na relação com  $X$ , segue:

$$R(X) = F \quad (3.6)$$

Uma relação dessa forma é chamada *equação de atribuição relacional* porque representa a atribuição de um conjunto difuso para a restrição associada com  $X$ .

Considere a proposição na forma  $p \triangleq X \text{ é } F$ , onde  $X$  é o nome de um objeto, uma variável ou proposição, e  $F$  é o nome do subconjunto difuso  $U$ , como em "*João é inteligente*".

A tradução de tal proposição pode ser expressa como:

$$R(A(X)) = F \quad (3.7)$$

onde  $A(X)$  é um atributo implícito de  $X$  que recebe valores em  $U$ , e (3.7) significa que a proposição  $p \triangleq X \text{ é } F$  tem o efeito de atribuir  $F$  à restrição difusa nos valores de  $A(X)$ .

Como exemplo, seja a proposição: "*João é novo*"

Seja idade um subconjunto difuso de  $U = [0, 100]$ , caracterizado pela função de pertinência:

$$\mu_{novo}(u) = 1 - S(u; 20, 30, 40) \quad (3.8)$$

onde  $u$  é a idade numérica e a função  $S$  é definida por:

$$\begin{aligned} S(u; \alpha, \beta, \gamma) &= 0 && \text{para } u \leq \alpha \\ &= 2\left(\frac{u-\alpha}{\gamma-\alpha}\right)^2 && \text{para } \alpha \leq u \leq \beta \\ &= 1 - 2\left(\frac{u-\gamma}{\gamma-\alpha}\right)^2 && \text{para } \beta \leq u \leq \gamma \\ &= 1 && \text{para } u \geq \gamma \end{aligned} \quad (3.9)$$

na qual o parâmetro  $\beta \triangleq (\alpha + \gamma)/2$  é o ponto de cruzamento, ou seja,  $S(\beta; \alpha, \beta, \gamma) = 0.5$ . Neste caso, o atributo implícito  $A(X)$  é idade (João) e a tradução de “*João é novo*” assume a forma:

$$\text{João é novo} \rightarrow R(\text{Idade}(\text{João})) = \text{novo}. \quad (3.10)$$

Para relacionar o conceito de uma restrição difusa para a distribuição de possibilidade, interpreta-se o membro do lado direito da seguinte maneira:

Seja idade = 28, cujo grau de pertinência é definido como aproximadamente 0.7. Interpreta-se 0.7 como o grau de compatibilidade de 28 com o conceito chamado novo.

Assim obtemos que a proposição “*João é novo*” converte o significado 0.7 do grau de compatibilidade de 28 com novo para o grau de possibilidade que João possui 28 é dada a proposição “*João é novo*”. A compatibilidade do valor de  $u$  com novo é convertido na possibilidade de que o valor de  $u$  obtenha “*João é novo*”.

### 3.4 Funções de Pertinência

Conforme mencionado no tópico anterior, as funções de pertinência definem o grau de verdade de um ou mais elementos dentro de um conjunto. Trata-se de uma função de avaliação que avalia elementos com determinados níveis de pertinência em cada um dos conjuntos existentes.

As funções de pertinência permitem representar um conjunto difuso graficamente. O eixo X representa o universo enquanto o eixo Y representa o grau de pertinência no intervalo  $[0,1]$ . A função de pertinência a ser utilizada depende da representatividade desejada e do número de informações disponíveis para análise da pertinência.

Dentre as funções de pertinência existentes, foram analisadas as seguintes:

- Função de pertinência Triangular;
- Função de pertinência Trapezoidal;
- Função de pertinência de Gauss ou Gaussiana.

### 3.4.1 Função de pertinência Triangular

A função de pertinência triangular possui a representação gráfica mostrada na Figura 11, e pode ser simétrica ou assimétrica:

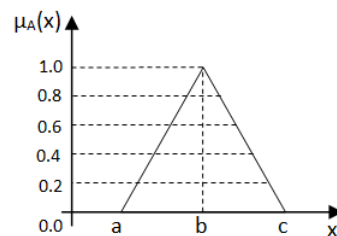


Figura 11: Representação da função de pertinência triangular

A função triangular representa o seguinte cenário:

- Valores menores que  $b$  possuem grau de pertinência variando linearmente entre a faixa:  $0 \leq e < 1.0$ ;
- Valores iguais a  $b$  possuem grau de pertinência igual a 1.0;
- Valores maiores que  $b$  possuem grau de pertinência variando linearmente entre a faixa:  $1.0 > e \geq 0$ .

Um cenário onde a utilização de tal função de pertinência se aplica é mostrado abaixo:

Para um conjunto universo  $[0,100]$ , onde se deseja obter o ponto médio exato do conjunto universo, obtém-se:



- Para valores de  $X$  entre  $[0,49]$ , a pertinência é menor do que 1, com valores lineares;
- Para  $X$  com o valor igual a 50, o grau de pertinência é igual a 1;
- Para valores de  $X$  entre  $[51,100]$ , a pertinência é menor do que 1, com valores lineares.

A função é descrita por três parâmetros  $a$ ,  $b$  e  $c$ , dadas pela expressão 3.11:

$$f(x, a, b, c) = \max \left\{ \min \left( \frac{x-a}{b-a}, \frac{c-x}{c-b} \right), 0 \right\} \quad (3.11)$$

Equação 3.11: Função de pertinência triangular

Onde os parâmetros  $a$  e  $c$  estão localizados nas bases do triângulo, e o parâmetro  $b$  está localizado no topo, conforme mostrado na figura 11.

### 3.4.2 Função de pertinência Trapezoidal

A função trapezoidal deve ser utilizada quando o seguinte cenário é apresentado:

- Valores menores que  $b$  possuem grau de pertinência variando linearmente entre a faixa:  $0 \leq e < 1.0$ ;
- Valores no intervalo entre  $b$  e  $c$  possuem grau de pertinência igual a 1.0;
- Valores maiores que  $c$  possuem grau de pertinência variando linearmente entre a faixa:  $1.0 > e \geq 0$ .

Um cenário onde a utilização de tal função de pertinência se aplica é mostrado abaixo:

Para um conjunto universo  $[0,100]$ , onde se deseja encontrar o período onde um ser humano está na fase adulta, compreendida entre 19 e 59 anos obtém-se:

- Para valores entre  $[0,18]$  a pertinência é menor do que 1, com valores lineares;
- Para valores entre  $[19,59]$ , a pertinência é igual a 1;
- Para valores entre  $[60,100]$ , a pertinência é menor do que 1, com valores lineares.

A função de pertinência trapezoidal possui a representação gráfica mostrada na Figura 12, e pode ser simétrica ou assimétrica:

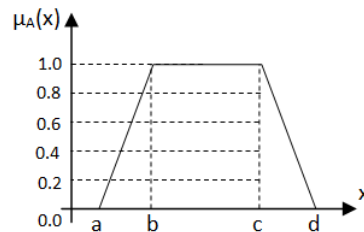


Figura 12: Representação da função de pertinência trapezoidal

A função é descrita por quatro parâmetros  $a$ ,  $b$ ,  $c$  e  $d$ , dadas pela expressão 3.12:

$$f(x, a, b, c, d) = \max \left\{ \min \left( \frac{x-a}{b-a}, 1, \frac{d-x}{d-c} \right), 0 \right\} \quad (3.12)$$

Equação 3.12: Função de pertinência trapezoidal

Onde os parâmetros  $a$  e  $d$  estão localizados nas bases do triângulo, e os parâmetros  $b$  e  $c$  estão localizados no topo, conforme mostrado na Figura 12.

### 3.4.3 Função de pertinência de Gauss ou Gaussiana

Uma utilização da função de Gauss está relacionada com o cálculo de probabilidade de ocorrência de um dado evento dentro de uma distribuição qualquer.

A função de pertinência de Gauss possui a representação gráfica mostrada na Figura 13:

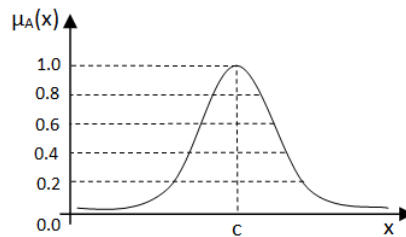


Figura 13: Representação da função de pertinência de Gauss

A função é descrita pelos parâmetros  $c$  e  $\sigma$  e é dada pela expressão 3.13:

$$f(x; \sigma, c) = e^{-\frac{(x-c)^2}{2\sigma^2}} \quad (3.13)$$

Equação 3.13: Função de pertinência trapezoidal

Onde o parâmetro  $c$  indica a distância da origem e  $\sigma$  indica a largura da curva.

### 3.4.4 Considerações

No contexto deste trabalho foi utilizada a função de pertinência trapezoidal. Para cada eixo existente no volume envolvente criado sobre os dados, os valores mínimo e máximo referente ao volume criado representa as variáveis  $b$  e  $c$  mostradas na Figura 12, enquanto valores que estão fora dos intervalos que representam a caixa, chamados de região difusa, representam as variáveis  $a$  e  $d$ .

Maiores detalhes da implementação serão vistos no capítulo que trata do desenvolvimento.

### 3.5 Média Aritmética e Desvio Padrão

A média aritmética é considerada uma medida de tendência central, onde o ponto central de um conjunto de elementos é obtido.

O desvio padrão é a medida mais comum da dispersão estatística, e é representado pelo símbolo sigma ( $\sigma$ ).

Enquanto a média aritmética é uma das formas de se obter o valor intermediário entre os valores de um conjunto, o desvio padrão visa demonstrar o quanto de variação ou dispersão existe entre os elementos do conjunto.

Na Tabela 2 são mostrados os valores obtidos pelo cálculo da média para os elementos dos grupos *A* e *B*:

Grupo	Valor 1	Valor 2	Valor 3	Valor 4	Média
A	28	29	30	31	29,5
B	22	25	36	35	30,5

Tabela 2: Valores para os grupos de dados A e B

O valor da coluna *Média* é obtido através da média aritmética entre os valores das colunas *Valor 1* a *Valor 4*. Como pode ser observado na coluna *Média*, apesar de ambos os grupos possuírem média igual a 29, os valores no grupo *A* variam pouco, enquanto no grupo *B* os valores estão muito dispersos, com alta variação.

Em alguns casos, quando os valores dos elementos de um conjunto possuem um alto grau de dispersão, a média aritmética não consegue capturar de forma adequada o centro de massa do conjunto. O problema é que a média aritmética procura encontrar o ponto intermediário entre os valores dos dados, o que nem sempre é o esperado.

Para este tipo de cenário, onde os dados tendem a estar espalhados dentro do conjunto, utiliza-se outra medida, que evidencia a diferença que existe entre a media e os valores do conjunto, conhecida como variância, ou dispersão. A Equação 3.14 exhibe a fórmula para o cálculo da variância:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (3.14)$$

Equação 3.14: Equação para o cálculo da variância

Nas Tabelas 3 e 4 são calculados os desvios e o quadrado dos desvios para os grupos A e B, utilizando os valores da média mostrados na Tabela 2:

<b>Grupo A</b>	<b>Média</b>	<b>Desvio</b>	<b>Quadrado dos Desvios</b>
28	29,5	-1,5	2,25
29	29,5	-0,5	0,25
30	29,5	0,5	0,25
31	29,5	1,5	2,25

Tabela 3: Cálculo do desvio e quadrado dos desvios para o Grupo A

<b>Grupo B</b>	<b>Média</b>	<b>Desvio</b>	<b>Quadrado dos Desvios</b>
22	29,5	-7,5	56,25
25	29,5	-4,5	20,25
36	29,5	6,5	42,25
35	29,5	5,5	30,25

Tabela 4: Cálculo do desvio e quadrado dos desvios para o Grupo B

Depois de obtidos os valores dos desvios e quadrado dos desvios, é possível calcular a variância e o desvio padrão, utilizando a Equação 3.15:

$$s = \sqrt{\frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (3.15)$$

Equação 3.15: Equação para o cálculo do desvio padrão

Na Tabela 5 estão representados os valores necessários para obter o valor do desvio padrão para os grupos A e B.

Na Coluna 1 é exibida a soma dos quadrados dos desvios, calculados nos passos anteriores e exibidos na Tabela 3 e 4. O valor da Coluna 1 é então dividido pelo

valor da Coluna 2, que representa o número de elementos do grupo, obtendo-se então a dispersão, mostrada na Coluna 3. Na Coluna 4 é exibido o valor do Desvio padrão, que é a raiz quadrada da variância.

Grupo	$\sum(\text{desvios})^2$	Nº elementos	Variância	Desvio Padrão
A	5,00	4	1,2500	1,1180
B	149,00	4	37,2500	6,1033

Tabela 5: Cálculo da variância e do desvio padrão para os Grupos A e B

Nota-se que o desvio padrão para o segundo grupo é maior do que para o primeiro grupo, o que não foi possível visualizar apenas com o cálculo da média aritmética.

É visível que os elementos do grupo B estão mais espalhados (dispersos) do que os elementos do primeiro grupo.

Para realizar o cálculo da variância, os valores obtidos são elevados ao quadrado, e de tal forma não podem ser utilizados para expressar os valores na mesma unidade das variáveis em estudo. Por outro lado, o desvio padrão expressa os valores de dispersão na mesma unidade dos dados em estudo, e para tal, em seu cálculo, o valor da raiz quadrada é obtido.

A utilização do desvio padrão tem por objetivo delimitar de forma mais precisa qual o ponto dentro de um grupo que concentra o maior número de elementos. Os valores dos dados que estão situados próximos a região evidenciada pelo desvio padrão são os que melhor representam os elementos do conjunto, e conseqüentemente são utilizados no contexto deste trabalho a fim de garantir a eficácia dos resultados.

Conforme mostrado anteriormente, o desvio padrão demonstra qual é o grau de dispersão existente entre os dados analisados. Uma baixa dispersão indica que os dados tendem a estar próximos da média, enquanto dados com alta dispersão tendem a estar espalhados.

A fim de obter uma região ainda mais abrangente, que defina melhor a região

onde provavelmente a maioria dos elementos daquele conjunto possa estar, utilizamos a regra conhecida como a regra dos *68-95-99,7*, que representa a propriedade de uma distribuição normal, que define o seguinte:

- 68% dos valores encontram-se a uma distância da média inferior a um desvio padrão;
- 95% dos valores encontram-se a uma distância da média inferior a duas vezes o desvio padrão;
- 99,7% dos valores encontram-se a uma distância da média inferior a três vezes o desvio padrão.

A representação da distribuição dos elementos conforme a regra acima é visualizada na Figura 14, extraída de (WIKIPÉDIA, 2013):

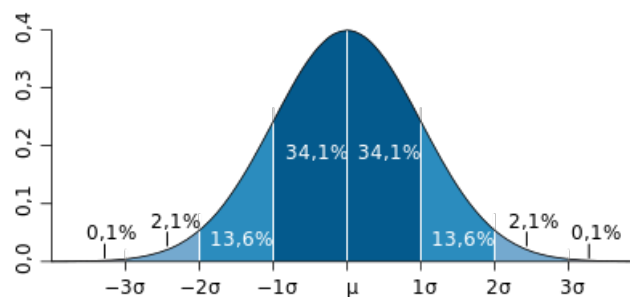


Figura 14: Distribuição normal evidenciando a regra '68-95-99,7'

De acordo com a regra, a área que está no raio de  $1\sigma$  abrange cerca de 68% dos valores de dados; A área no raio de  $2\sigma$  abrange cerca de 95% dos dados e a área no raio de  $3\sigma$  abrange cerca de 99,7% dos dados.

Com a utilização dos valores do desvio padrão  $3\sigma$ , é possível delimitar uma região na qual a quase totalidade dos elementos cujos valores sejam aproximados, ou seja, tendenciam a pertencer a um mesmo conjunto, estarão inclusos.

Levando em consideração que elementos de um mesmo grupo possuem características semelhantes, ou seja, a faixa de valores dos elementos de dados são

próximas, embora possam estar dispersos, é possível utilizar o cálculo do desvio padrão para abranger uma área maior, e de maior significância, permitindo que novos elementos, ainda não classificados, que estejam a uma determinada distância, porém fora da região pertencente ao conjunto sejam corretamente classificados como membros daquele grupo, com algum grau de pertinência.

### 3.5.1 Utilização do Desvio Padrão

No contexto deste trabalho, a utilização do desvio padrão possibilitou a criação de uma região de possibilidade de elementos da classe, onde os elementos dispostos nessa região são classificados como pertencentes à classe com um certo grau de pertinência.

Inicialmente, os elementos de dados previamente agrupados são envolvidos por um volume envolvente, de modo que os dados estão inteiramente contidos dentro desse volume, logo sua pertinência é 1.0 ou 100%.

Após a criação deste volume envolvente, com base nos valores dos dados analisados, os valores do ponto médio e desvio padrão ( $\sigma$ ) são calculados, e utilizando-se esses valores, é criada uma nova caixa, representando a região de incerteza. Maiores detalhes serão vistos no desenvolvimento.



## 4 Desenvolvimento

### 4.1 Técnica Proposta

Para alcançar o objetivo deste trabalho, foram utilizadas técnicas comumente aplicadas para a detecção de colisão de objetos em jogos.

O algoritmo para realizar a detecção de colisão foi dividido em partes, a fim de obter um baixo acoplamento entre as entidades envolvidas e facilitar o detalhamento das tarefas realizadas por cada parte.

As partes do algoritmo de detecção são:

- Criação da hipercaixa envolvente;
- Testes de colisão;
- Divisão das hipercaixas;
- Obtenção das áreas de incerteza;
- Seleção de elementos para testes;
- Criação do classificador difuso;
- Geração das regras classificatórias difusas.

Para tornar mais clara a explicação dos algoritmos, nos itens a seguir serão utilizadas duas classes de dados artificiais para exemplificar o funcionamento: uma

classe representada por Círculos, com 6 pontos, e uma classe representada por Triângulos, com 12 pontos. Os dados mostrados na Tabela 6 foram gerados de forma aleatória, e são utilizados apenas como exemplo para os algoritmos utilizados.

Itens de Dados		
Classe	Plano X1	Plano X2
A (Círculo)	55,23	23,37
	33,01	26,11
	34,22	38,41
	49,21	44,87
	69,77	28,09
	55,06	28,62
B (Triângulo)	97,23	94,11
	4,13	5,89
	99,96	50,7
	92,13	46,72
	91,72	5,16
	6,00	87,39
	40,58	52,47
	22,08	16,38
	23,72	30,16
	30,79	66,52
	48,40	7,38
	46,74	57,94

Tabela 6: Tabela contendo os itens de dados das classes Círculo e Triângulo

Na Figura 15 é mostrada a representação gráfica para as classes Círculo e Triângulo.

## 4.2 Criação da hipercaixa envolvente

No contexto deste trabalho volumes envolventes de caixas alinhadas ao eixo foram utilizados para representar agrupamentos de dados, e possibilitar a geração de regras que permitam classificar novos elementos, ainda não agrupados, dentre os conjuntos existentes, com determinado grau de pertinência em cada um dos grupos.

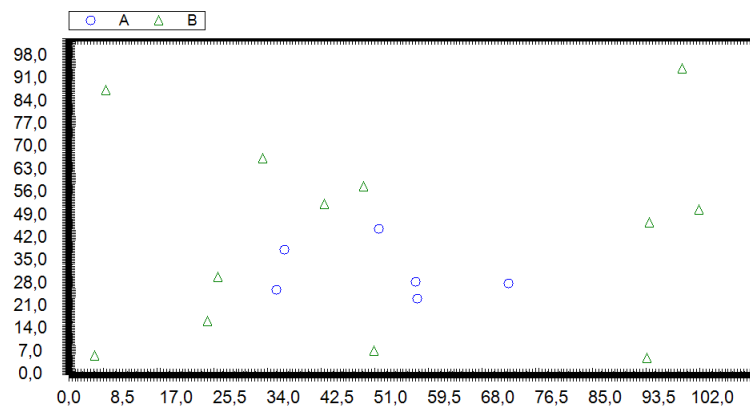


Figura 15: Representação gráfica para as classes Círculo e Triângulo

Para a criação da hipercaixa, é necessário encontrar os valores nos dados que permitam a sua representação através da estrutura envolvente. Cada atributo existente nos dados representa um plano (também conhecido como dimensão), e cada plano é representado pelos limites mínimo e máximo, que determinam a região limítrofe onde todos os dados daquele plano estão incluídos.

Um plano é representado por um eixo do sistema de coordenadas. Quando dois ou mais planos são projetados no eixo de coordenadas, uma representação N-dimensional dos dados é obtida, com isso temos uma hipercaixa formada por vários hiperplanos. Na Figura 16 é possível visualizar uma representação quadridimensional dos dados - a *hipercaixa* e seus *hiperplanos*:

Em cada eixo do sistema de coordenadas são criadas as laterais que representam os valores mínimo e máximo do conjunto de dados (bordas do volume envolvente). As laterais são as fronteiras que delimitam a estrutura envolvente, de forma que, ao final do processo, todos os dados estejam envolvidos.

A Equação 4.1 é utilizada para obtenção dos valores necessários para a criação da hipercaixa:

$$\begin{cases} \textit{Mínimo} & : \min (\textit{dados da dimensão}) \\ \textit{Máximo} & : \max (\textit{dados da dimensão}) \end{cases} \quad (4.1)$$

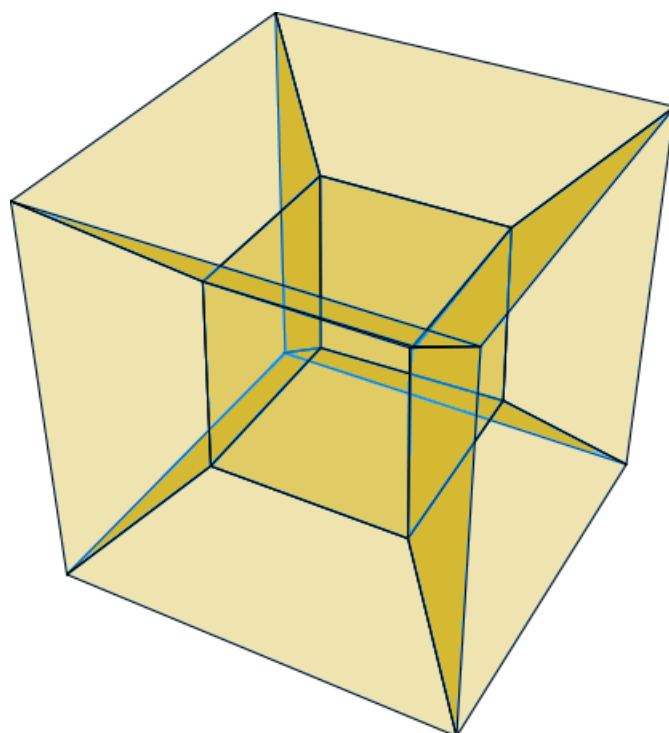


Figura 16: Representação quadridimensional dos dados - Hipercaixa e Hiperplanos

Equação 4.1: Obtenção dos valores para a criação da hipercaixa

De posse dos valores mínimo e máximo, é possível envolver os dados com a criação de bordas limítrofes da hipercaixa. Na Figura 17 é possível visualizar as classes Círculo e Triângulo envolvidas com o volume AABB. Como as classes possuem apenas dois atributos, sua representação é obtida projetando os dois planos no eixo de coordenadas (representação bidimensional):

Na Figura 18 é exibida a representação visual para o conceito de pertinência. As hipercaixas exibidas na figura representam três conjuntos distintos, “Verde”, “Azul” e “Vermelho”. Cada hipercaixa possui duas regiões, uma região de cor opaca (sem transparência), e uma região onde o nível de opacidade começa a diminuir até se tornar totalmente transparente. O intervalo opaco é visualizado na parte central da hipercaixa, enquanto o intervalo ao redor da parte central representa a região onde a opacidade tende a diminuir.

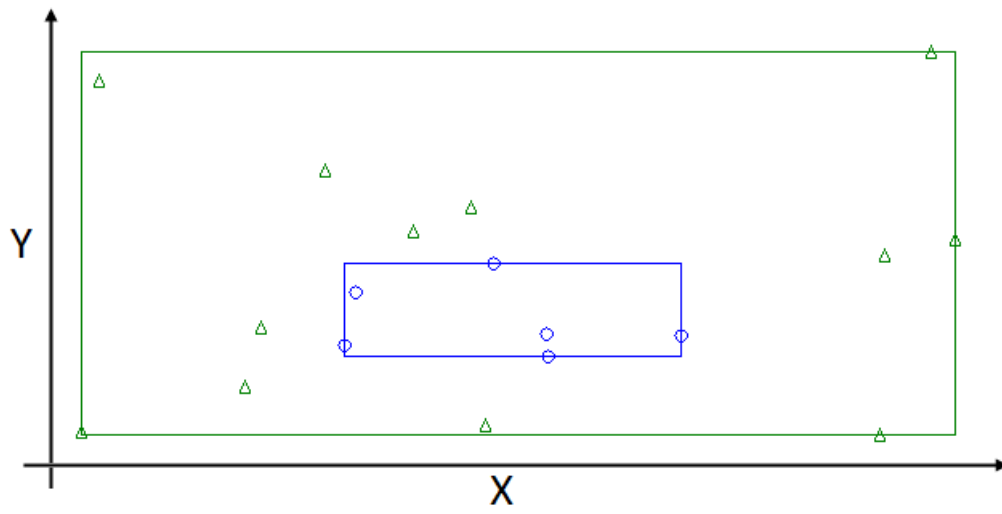


Figura 17: Classes Círculo e Triângulo envolvidas pelo volume obtido via AABB



Figura 18: Representação da pertinência entre hipercaixas envolventes

O intervalo representado pela cor opaca possui pertinência igual a 100%, ou seja, elementos que estejam localizados dentro desta área possuem total representatividade dentro do conjunto representado pela hipercaixa. No intervalo onde a cor começa a ganhar transparência, a representatividade dentro do conjunto não é mais 100%, e sim um percentual que tende a 0% a medida que o nível de opacidade diminui, ou seja, elementos que estejam nessa região possuem algum nível de representatividade dentro do conjunto representado pela hipercaixa mas também possui certo nível de representatividade em um ou mais conjuntos existentes no sistema.

O intervalo degradê, onde é possível visualizar mais de uma cor se misturando, representa a situação onde a pertinência varia entre os conjuntos, ou seja, elementos que estejam nessa região possuem representatividade em um ou mais conjunto do sistema.

A Equação 4.2 é utilizada para obter os valores dos dados da hipercaixa e serão utilizados nos testes de colisão que serão mostrados adiante:

$$\begin{cases} \textit{Centro} & : (\max(\textit{valor}) + \min(\textit{valor})) \div 2 \\ \textit{Raio} & : (\max(\textit{valor}) - \min(\textit{valor})) \div 2 \\ \textit{Médio} & : \textit{média valores da dimensão} \end{cases} \quad (4.2)$$

Equação 4.2: Obtenção dos valores para o teste de colisão da hipercaixa

Um ponto importante em relação à performance é que o custo computacional para a criação da hipercaixa envolvente não aumenta significativamente com o aumento de número de elementos existentes pois o algoritmo utiliza apenas as regiões limítrofes das classes na criação da hipercaixa.

Após a obtenção dos valores, que serão utilizados nos próximos passos, a primeira etapa do algoritmo está concluída.

### 4.3 Testes de Colisão

A fim de realizar os testes de colisão entre cada hipercaixa, o teorema de separação dos eixos (GOTTSCHALK, 2000) é utilizado. Conforme o teorema, se houver pelo menos um eixo na qual as projeções das caixas não se sobreponham, então há um eixo de separação entre elas, e logo as hipercaixas não estão colidindo no espaço. De tal modo, não é necessário testar todos os eixos para se comprovar a não colisão entre as hipercaixas.

Na Tabela 7 são mostrados os valores calculados com base nos dados da Tabela 6 para as classes Círculo e Triângulo. Com os valores de dados da tabela, as classes estão colidindo, o que será evidenciado adiante. Os valores serão utilizados em várias etapas do algoritmo.

Valor	Classe Círculo		Classe Triângulo	
	Eixo X1	Eixo X2	Eixo X1	Eixo X2
Mínimo	33,01	23,37	4,13	5,16
Máximo	69,77	44,87	99,96	94,11
Centro	51,39	34,12	52,05	49,64
Raio	25,7	17,06	26,02	24,82
Médio	49,58	31,75	50,37	43,49

Tabela 7: Valores das classes Círculo e Triângulo

O teste de colisão é realizado entre pares de hipercaixas. O mesmo eixo das duas hipercaixas é selecionado, e então o teste de colisão é efetuado.

O cálculo é realizado com base na Equação 3.2, a fim de verificar se as projeções das hipercaixas estão se sobrepondo e é exibido no Algoritmo 1. O resultado dos cálculos da Equação 3.2 pode ser visualizado na Tabela 8:

---

**Algoritmo 1:** Verifica colisão entre hipercaixas
 

---

```

Data: Hipercaixa cx1, cx2
begin
  double[] grauColisao = new double[numero_dimensoes]
  double Dist_Centros, Soma_Raios = 0
  for i ← 0 to numero_dimensoes do
    Dist_Centros = Mod(cx1.Centro.Plano[i] - cx2.Centro.Plano[i])
    Soma_Raios = (cx1.Raio.Plano[i] + cx2.Raio.Plano[i])
    if Dist_Centros >= Soma_Raios then
      //Nao_ha_colisao
      retorna falso
    else
      grauColisao = Soma_Raios - Dist_Centros
    end if
  end for
  DefinirValoresDeContato(cx1, grauColisao)
  DefinirValoresDeContato(cx2, grauColisao)
  //Ha_colisao
  retorna verdadeiro
end

```

---

Classe	Cálculo	X1	X2
Círculo	Raio	25,7	17,06
Círculo	Centro	51,39	34,12
Triângulo	Raio	26,02	24,82
Triângulo	Centro	52,05	49,64
Soma dos Raios		51,72	41,88
Distância entre os Centros		-0,65	-15,52

Tabela 8: Resultado dos cálculos utilizando-se a Equação 3.2

Como a diferença entre os Centros é menor do que a soma dos Raios em todos os eixos (X1 e X2), há colisão entre as hipercaixas.

Na Figura 19 é possível visualizar os passos necessários para se verificar a colisão entre duas hipercaixas:

Caso as hipercaixas estejam colidindo, então o algoritmo calcula o percentual de contato existente entre as hipercaixas. O valor obtido por tal cálculo será utilizado posteriormente para selecionar qual hipercaixa deverá ser particionada a fim de tentar solucionar a colisão. O cálculo do percentual de colisão utiliza o Algoritmo 2:



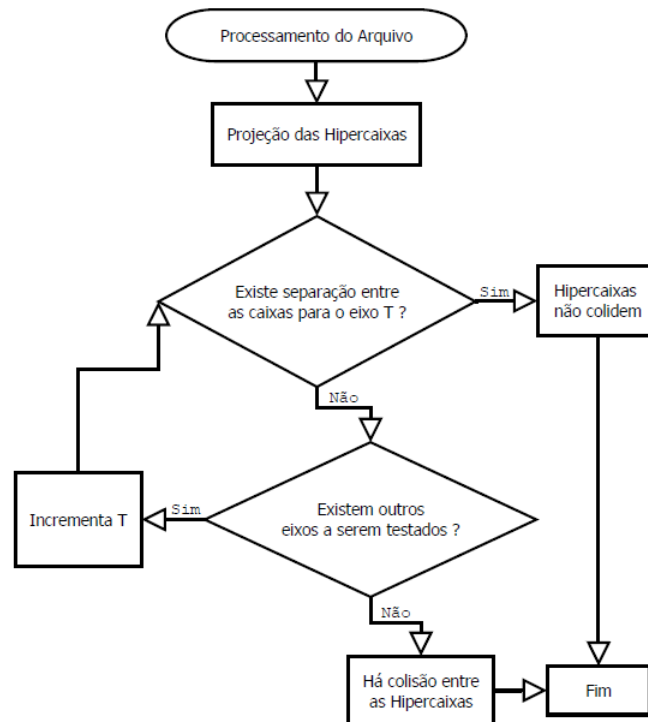


Figura 19: Passos para verificar a colisão

---

### Algoritmo 2: Define o percentual de contato

---

```

Data: Hipercaixa cx, double[] grauColisao
begin
  double[] perc_contato = new double[numero_dimensoes]
  for i ← 0 to numero_dimensoes do
    double max_min = (cx.Maximo.Plano[i] - cx2.Minimo.Plano[i])
    if grauColisao[i] >= max_min then
      | perc_contato[i] = 1 - 0.0000000001
    else
      | perc_contato[i] = grauColisao[i] / max_min
    end if
  end for
  if cx.MinContato == nulo then
    | cx.MinContato = new HiperPontos(perc_contato, classe)
  end if
  if cx.MaxContato == nulo then
    | cx.MaxContato = new HiperPontos(perc_contato, classe)
  end if
  for i ← 0 to numero_dimensoes do
    double min = Min(cx.MinContato.Plano[i], perc_contato[i])
    double max = Min(cx.MaxContato.Plano[i], perc_contato[i])

    cx.MinContato.Plano[i] = min
    cx.MaxContato.Plano[i] = max
  end for
end

```

---

Após a detecção de colisão, são necessários métodos para solucionar o problema, de modo que as projeções das hipercaixas não se sobreponham e seja possível distinguir corretamente as classes.

O método proposto realiza a divisão de uma das caixas envolvidas na colisão, e então realiza novamente o teste de colisão entre as demais para verificar se a colisão foi solucionada. A teste de colisão é aplicado novamente após qualquer divisão realizada nas hipercaixas, pois a divisão pode solucionar a colisão existente entre a hipercaixa dividida e outras envolvidas nas colisões. O passo se repete até não haver mais colisões ou até que não seja mais possível realizar divisões.

Aqui são implementados dois métodos de divisão da hipercaixa. Um dos métodos utiliza a hipercaixa com o maior percentual de contato, enquanto o outro método realiza a divisão na hipercaixa com menor percentual de contato.

O que foi observado nos testes realizados é que o resultado obtido com cada um desses métodos depende dos dados analisados. Em alguns cenários, a divisão pela hipercaixa que possui maior percentual de colisão mostrou-se mais efetiva, enquanto que em outros, a divisão da caixa com menor percentual conseguiu solucionar as colisões de forma mais efetiva.

Por falta de referências bibliográficas do tema, várias alternativas foram tentadas através de experimentações até a obtenção da melhor técnica.

No momento da divisão da hipercaixa utiliza-se o ponto médio, que divide a quantidade de pontos ao meio, e não o centro, que é exatamente a metade da hipercaixa. Isso se deve ao fato de que com a utilização do ponto médio, os pontos mais afastados consequentemente serão separados dos pontos mais aglomerados. Com isso a divisão tende a resolver as colisões mais rapidamente. O fato da divisão das hipercaixas deixá-las desbalanceadas em relação ao número de pontos não causa nenhum impacto no processo.

As Figuras 20, 21 e 22 mostram as divisões efetuadas nas hipercaixas representando as classes Círculo e Triângulo até que todas as colisões estejam resolvidas:

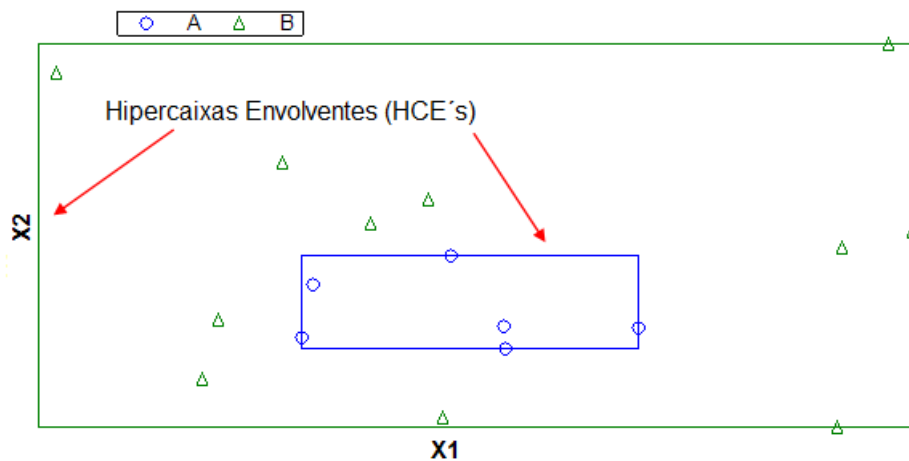


Figura 20: Classes Círculo e Triângulo estão colidindo nos eixos X1 e X2

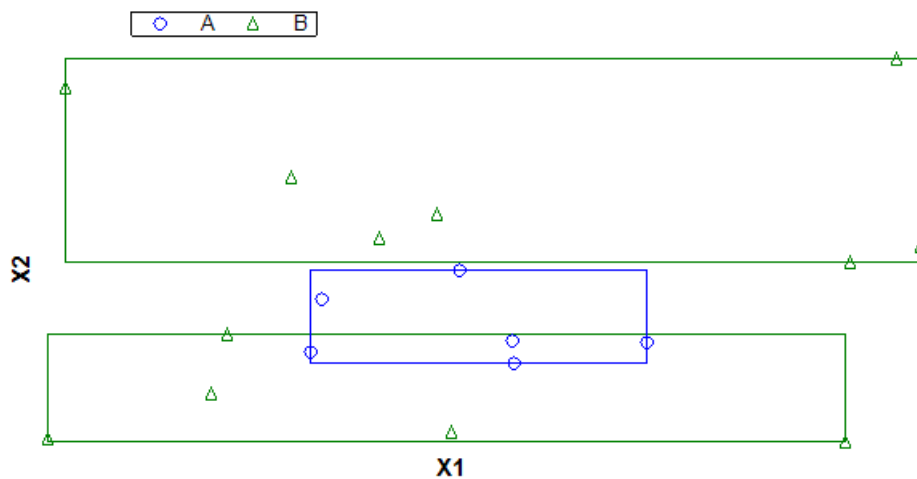


Figura 21: Hipercaixa da classe Triângulo é dividida com base no ponto médio

Após o processo de divisões, que visa solucionar as colisões entre as hipercaixas, a segunda etapa do algoritmo está concluída. Cada classe de dados está representada não mais por uma única hipercaixa, mas por N hipercaixas, resultantes das divisões efetuadas na hipercaixa original, com a finalidade de solucionar as colisões existentes. A partir deste ponto, para efetuar a análise de uma classe, todas as hipercaixas resultantes da divisão da hipercaixa original devem ser analisadas.

A Hipercaixa Envolvente (HCE) é a representação equivalente ao intervalo de valores existente nos dados.

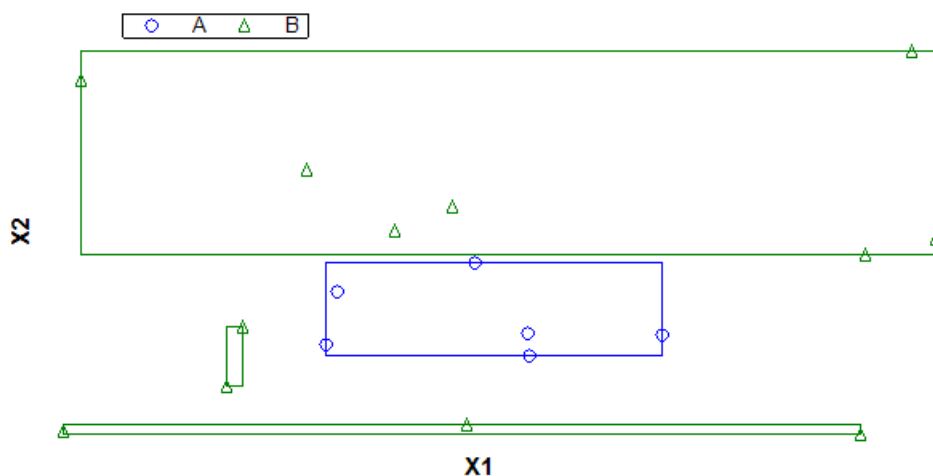


Figura 22: Nova divisão da hipercaixa da classe Triângulo. Não há mais colisões

## 4.4 Obtenção das Áreas de Incerteza

Quando as colisões entre as hipercaixas existentes são solucionadas, os dados podem ser analisados de forma adequada, uma vez que não existem mais regiões no espaço que representam mais de uma classe (colisões).

Como o objetivo é desenvolver uma ferramenta que seja capaz de realizar a classificação de novos elementos de dados não classificados, que podem não estar dentro da HCE, surge a necessidade de realizar a expansão da área de busca da classe.

Com a finalidade de expandir a HCE (ou área de representatividade) de uma classe, mantendo a coerência de pertinência com o grupo da classe, utilizamos o desvio padrão (representado pelo símbolo sigma  $\sigma$ ) para capturar a região que pode abranger novos elementos da mesma classe. O  $6\sigma$  (seis sigma) garante que 99,7% dos elementos estão a uma distância da média inferior a três vezes o desvio padrão. Se utilizarmos essa informação para criar uma região maior de abrangência, pode-se elevar a taxa de acerto para classificação de novos elementos fora dos limites da HCE.

O cálculo do três sigma é mostrado no Algoritmo 3:

---

**Algoritmo 3:** Algoritmo utilizado para cálculo do três sigma
 

---

```

Data: int dimensao, HiperPontos[] pontos, centro
begin
  int num_sigma = 3
  int denominador = Total_Elementos(pontos) - 1
  if denominador == 0 then
    | denominador = 1
  end if
  double total = 0
  foreach ponto in pontos do
    | dif = ponto.Plano[dimensao] - centro.Plano[dimensao]
    | total = Elevar_ao_Quadrado(dif)
  end foreach
  double resultado = total/denominador
  retorna Raiz_Quadrada(resultado) * num_sigma
end

```

---

Utilizando-se o ponto médio, calculado e exibido na Tabela 7, como ponto de início, será criada uma nova estrutura envolvente. O processo consiste em criar uma nova hipercaixa utilizando o valor referente a três vezes o valor do desvio padrão (sigma  $\sigma$ ), o seis sigma. Tal estrutura será chamada Hipercaixa Sigma (HCS). Na Figura 23 é exibida a HCS para as classes Círculo e Triângulo:

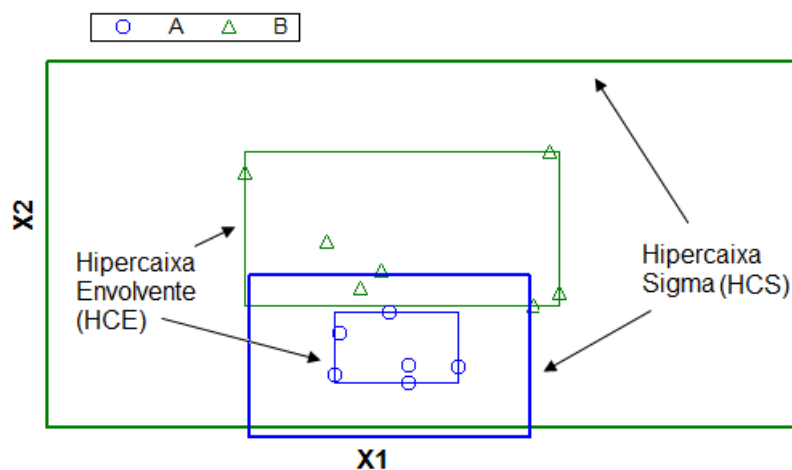


Figura 23: Hipercaixa sigma envolvendo as classes

A utilização do ponto médio como ponto de referência para a criação da hipercaixa sigma se deve ao fato que o valor de três sigma tende a abranger apenas os pontos que estão mais concentrados na hipercaixa, deixando os pontos que estão mais dispersos, ou seja, menos representativos para o grupo em questão, fora da área envolvida pela hipercaixa sigma.

Para a criação da hipercaixa sigma são utilizados os valores *sigma mínimo* e *sigma máximo* do hiperplano, conforme Equação 4.3:

$$\begin{cases} \text{Sigma Mínimo} & : \text{Centro do Plano} - (\text{Desvio Padrão} * \text{Três Sigma}) \\ \text{Sigma Máximo} & : \text{Centro do Plano} + (\text{Desvio Padrão} * \text{Três Sigma}) \end{cases} \quad (4.3)$$

Equação 4.3: Cálculo dos valores sigma mínimo e máximo

Como a hipercaixa sigma é construída com base no ponto médio, a nova estrutura pode não envolver completamente a hipercaixa envolvente que engloba os dados. A fim de exemplificar tal comportamento, a Figura 24 mostra duas classes de dados, representadas pela cor verde e pela cor azul, onde a hipercaixa verde é completamente envolvida pela hipercaixa sigma enquanto que hipercaixa azul possui uma área não envolvida.

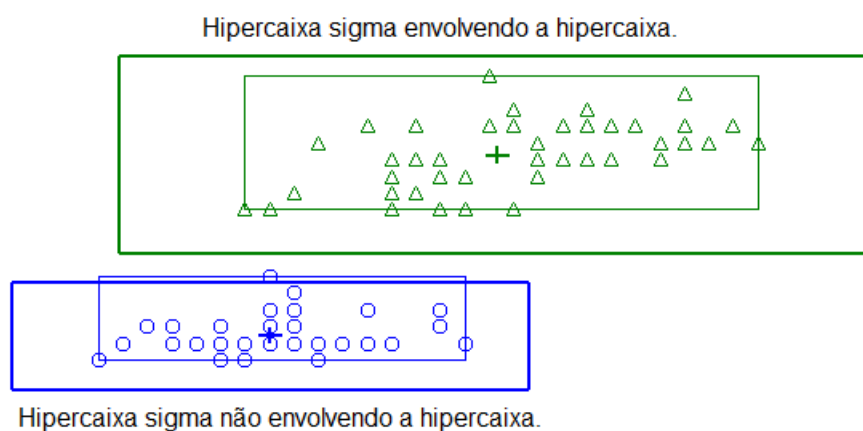


Figura 24: Listagem de hipercaixas que devem ser envolvidas pelas hipercaixas sigma

Novos elementos cujos valores estejam dentro da hipercaixa sigma possuirão alguma representatividade em relação à classe relacionada. Mais adiante será detalhado o processo responsável por atribuir tal valor de representatividade aos elementos.

O fato de haver uma região não envolvida pela HCS impossibilita o cálculo dos valores de representatividade mencionados anteriormente. A fim de resolver tal problema, cada hipercaixa não envolvida completamente pela hipercaixa sigma é analisada e um hiperplano é selecionado para que seja feita uma divisão.

A análise para seleção de qual plano da hipercaixa será utilizado para a divisão leva em consideração o eixo que possui o menor grau de envolvimento pela hipercaixa sigma, ou seja, quanto mais a hipercaixa envolvente naquele plano estiver fora da hipercaixa sigma, maior a probabilidade desse plano ser escolhido para a divisão.

Com tal método de divisão, os pontos mais afastados serão conseqüentemente separados em uma nova hipercaixa. É importante ressaltar que a divisão não é feita na hipercaixa sigma e sim na hipercaixa envolvente.

Após a divisão, serão geradas duas novas hipercaixas, e os cálculos dos valores mostrados na Tabela 7 devem ser reprocessados, agora com bases nos novos valores gerados pela divisão das hipercaixas.

O fato dos pontos mais afastados serem deslocados para uma nova hipercaixa envolvente faz com a nova hipercaixa sigma calculada esteja mais próxima aos pontos que agora estão concentrados, e com isso, as chances de que a hipercaixa envolvente esteja totalmente envolvida pela hipercaixa sigma aumenta consideravelmente.

A divisão das hipercaixas envolventes é realizada até que todas estejam completamente envolvidas pelas hipercaixas sigma.

## 4.5 Classificador Difuso

A próxima etapa consiste na adição de lógica difusa ao processamento a fim de permitir a extração de regras que consigam abstrair o conceito de pertinência para os dados analisados.

Para que seja possível a utilização do classificador difuso, é necessário utilizar uma função de pertinência que defina o grau de verdade de um ou mais elementos dentro de um conjunto. A função avalia os elementos com determinados níveis de pertinência em cada um dos conjuntos existentes. A função de pertinência que será utilizada no método proposto é a função trapezoidal.

O classificador difuso desenvolvido recebe como parâmetro de entrada cada uma das hipercaixas existentes. Cada hipercaixa possui quatro valores de dados que serão utilizados no cálculo: os valores mínimo e máximo da hipercaixa envolvente e os valores mínimo e máximo da hipercaixa sigma.

Na Figura 25 são exibidos os passos para se calcular os valores de cada hipercaixa utilizados no classificador difuso:

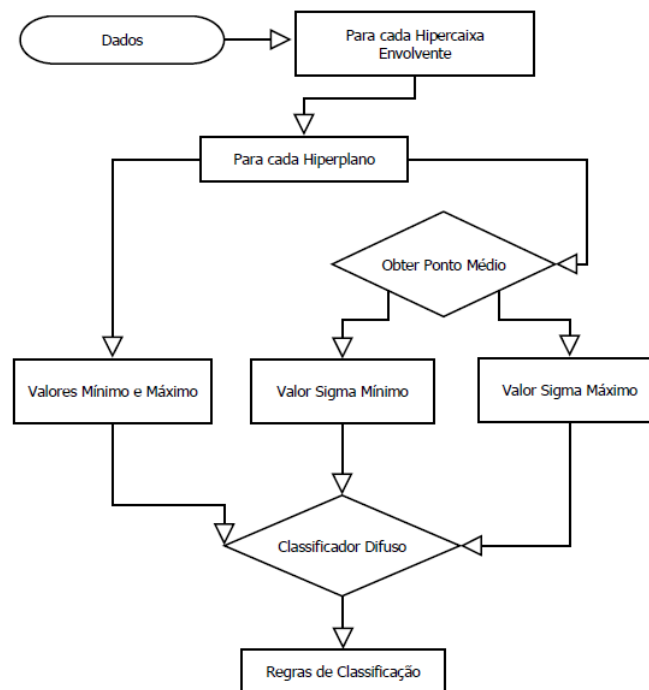


Figura 25: Cálculo dos valores de entrada para o classificador difuso

Após a geração das hipercaixas envolventes e da resolução das colisões, cada um dos hiperplanos resultará em quatro valores de entrada para o classificador difuso.



Para cada hiperplano da hipercaixa são obtidos os valores abaixo:

- Mínimo e máximo conforme Equação 4.1;
- Sigma mínimo e sigma máximo, conforme Equação 4.3.

De posse dos valores obtidos, o classificador difuso pode realizar o processamento das informações.

O classificador difuso desenvolvido é criado e alimentado com as informações das hipercaixas existentes. Após este procedimento, o módulo está preparado para realizar o processamento das informações e então realizar a geração das regras de classificação difusa.

## 4.6 Sistema de inferência difuso

O *Sistema de Inferência Difuso* (SID) realiza o processamento das informações referentes as hipercaixas envolventes criadas sobre os dados. Os valores de entrada recebidos pelo módulo correspondem às fronteiras das hipercaixas envolventes e hipercaixas sigma.

Para efetuar a geração das regras é utilizada a função de pertinência trapezoidal. Cada hiperplano possui quatro valores representativos, conforme mostrado na Figura 26:

A Figura 27 mostra como os valores da função de pertinência trapezoidal são obtidos pelo classificador difuso.

O SID é composto por um conjunto de variáveis linguísticas, que compõem o banco de dados e executam a computação difusa.

Variáveis linguísticas são utilizadas para representação das classes existentes nos dados. Cada variável linguística é representada por conjuntos difusos que são formados pelas HCEs e HCSs.

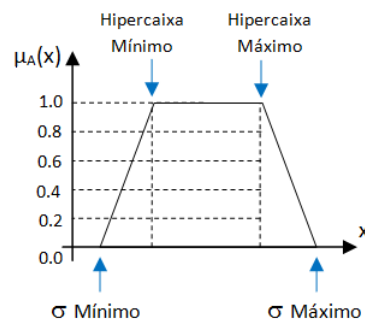


Figura 26: Valores representados pela função trapezoidal

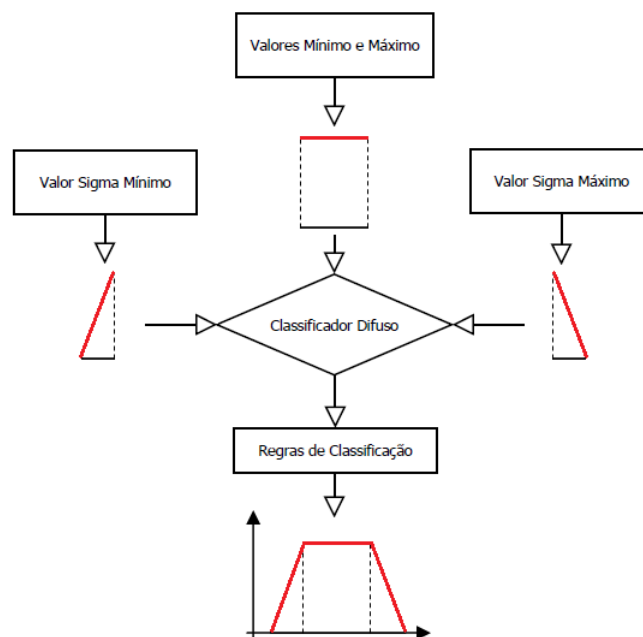


Figura 27: Valores da função trapezoidal obtidos pelo classificador difuso

Por exemplo, se existirem as classes "Motocicleta", "Carros" e "Caminhões", serão necessárias três variáveis linguísticas para representá-los.

Após a definição de classes da variável linguística são criados os conjuntos difusos (HCE e HCS). Cada item no conjunto difuso possui uma função trapezoidal com seus valores de pertinência.

Para cada hiperplano (eixo) existente no agrupamento de dados, deve-se obter valores limítrofes respectivos em cada uma das hipercaixas existentes. Se houve-

rem vinte hipercaixas resultantes após as colisões terem sido resolvidas, os valores limítrofes do hiperplano em cada uma das vinte hipercaixas deve ser computado. A Figura 28 torna mais claro os passos executados para a criação dos conjuntos difusos:

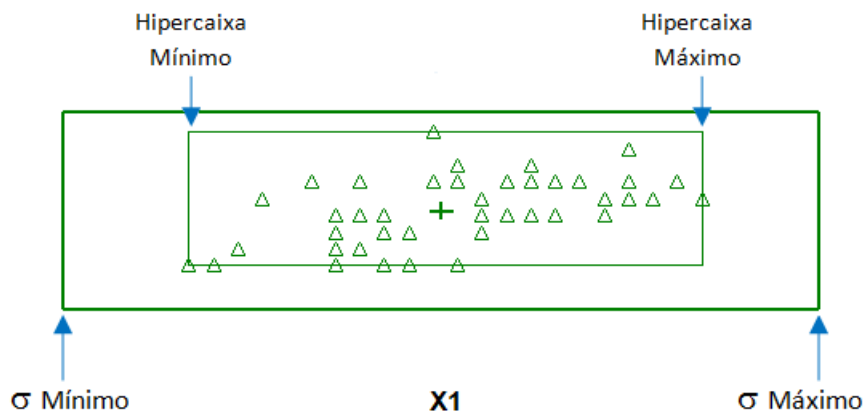


Figura 28: Passos para a criação dos conjuntos difusos

Da mesma maneira que foram criadas hipercaixas para representar os hiperplanos existentes, também é necessário criar variáveis linguísticas que representem cada um dos agrupamentos de dados existentes.

Neste ponto, o SID está preenchido com todas as informações necessárias para que seja possível a extração das regras de classificação.

## 4.7 Geração das Regras de Classificação

Para cada uma das hipercaixas existentes, será gerada um regra no SID. As regras difusas geradas possuem uma estrutura condicional, no formato "*Se Verdade* → *Então*". A cada regra é atribuído um identificador que representa o nome da regra. Este valor não é fundamental, porém é exigido na criação da regra.

No momento em que algum novo ponto não classificado é submetido à avaliação, cada um dos conjuntos contidos na regra são substituídos por um valor

percentual de pertinência. Cada um dos grupos/classes existentes é atribuído para o novo ponto em análise. Cada hipercaixa é uma regra difusa.

O valor situado antes da cláusula "IS" representa um determinado hiperplano da caixa existente no agrupamento de dados. O valor após a cláusula "IS" é substituído pelos valores do hiperplano no objeto que está sendo analisado.

Na Figura 29 é mostrado o algoritmo de geração das regras do SID.

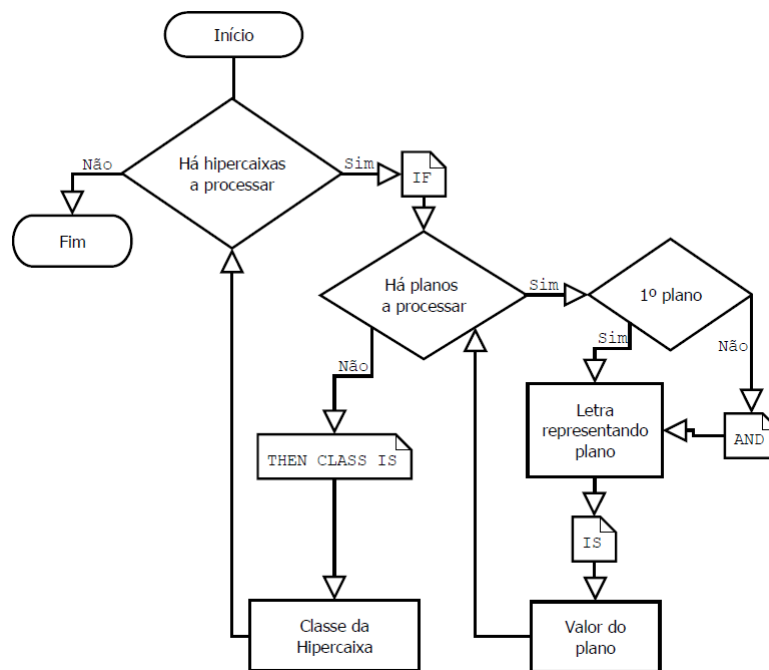


Figura 29: Algoritmo de geração das regras do sistema de inferência difuso

A regra é gerada de forma que cada um dos hiperplanos seja considerado, formando assim a hipercaixa. Na Figura 30 é mostrado um exemplo de regra gerada. A regra de exemplo mostrada indica que se (IF) os valores dos planos AA, BB, CC e DD estiverem (IS) simultaneamente (AND) dentro dos limites de valores definidos pelos conjuntos AA\_1, BB\_1, CC\_1 e DD\_2, então (THEN) o elemento avaliado pela regra possui certo grau de representatividade dentro da classe (ClassName). A avaliação dos elementos utilizados para testes é feita sob todas as regras geradas pelo SID. No B são exibidas as regras geradas para todas



## 4.8 Ferramentas de desenvolvimento

Para o desenvolvimento do aplicativo resultante deste trabalho foram utilizadas as seguintes ferramentas:

- Microsoft Visual Studio utilizando a linguagem C# (C-Sharp).
- AForge.NET, um framework em C# desenvolvido para programadores e pesquisadores no campo de Visão Computacional e Inteligência Artificial, disponível em ["http://www.aforgenet.com/"](http://www.aforgenet.com/).

Na Figura 32 é exibido o Diagrama UML evidenciando a estruturação do projeto e o relacionamento entre as entidades.

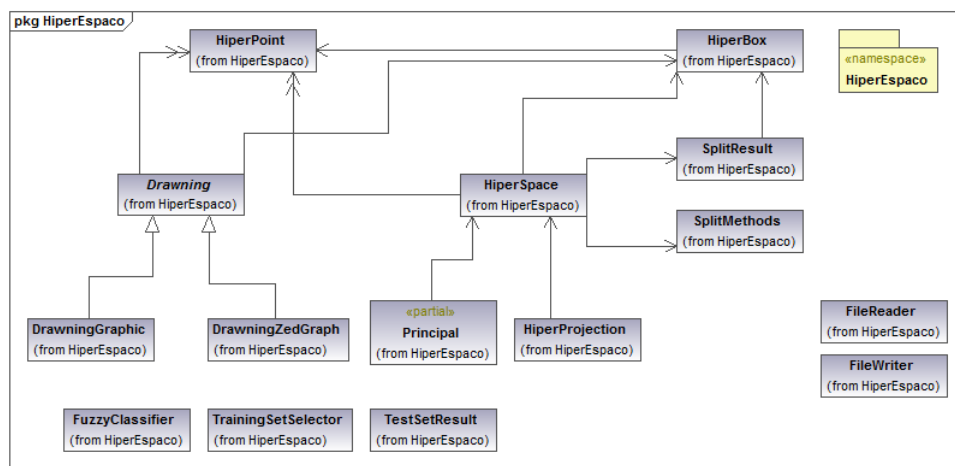


Figura 32: Diagrama UML do projeto

## 5 Testes e Resultados

### 5.1 Bases de Dados utilizadas

Para efetuar a validação do método proposto, foram utilizados bancos de dados mundialmente conhecidos e utilizados. Na realização da avaliação foram utilizados sete bases de dados oriundas do repositório da UCI (<http://archive.ics.uci.edu/ml/>). Na Tabela 9 são exibidas as bases de dados utilizadas nos testes, bem como informações detalhadas a respeito de cada uma delas:

Nome	Número de Classes	Número de Registros	Atributos
Iris	3	150	4
Wine	3	178	13
Vehicle	4	846	18
Pima	2	768	8
Thyroid	3	215	5
Wisconsin	2	699	9

Tabela 9: Bases de dados utilizadas para testes

Abaixo serão descritas as bases de dados, indicando qual o conteúdo e qual a finalidade de cada uma das bases utilizadas.

#### **Iris**

A base de dados Iris contém a classificação de espécies de flores dentre três grupos: Iris-Setosa, Iris-Virgínica e Iris-Versicolor. A base contém 50 elementos de cada espécie, e cada elemento possui quatro atributos: comprimento e largura da pétala e comprimento e largura da sépala. Esta base de dados tornou-se um teste

de caso típico para várias técnicas de classificação em aprendizado de máquina.

O primeiro artigo publicado a utilizar a base de dados Iris foi "The use of multiple measurements in taxonomic problems", por Fisher, R.A no Annals of Eugenics em 1936 (FISHER, 1936).

A base artigo publicado por Fisher é um clássico da área e é frequentemente referenciado. A base Iris representa um problema de fácil resolução, uma vez que a classe Iris Setosa é linearmente separável das demais e, embora as outras duas classes não sejam linearmente separáveis umas das outras, podem ser facilmente separadas com a utilização de poucos hiperplanos (SWAIN et al., 2012).

### **Wine**

A base de dados Wine contém a classificação de tipos de vinhos que se dividem em três classes. Os atributos da base Wine são: proline, quantidade de álcool, ácido málico, cinzas, alcalinidade da cinza, magnésio, total de phenols, quantidade de flavonoides, quantidade de phenols não flavonoides, pro antocianinas, intensidade da cor e matiz e OD280/OD315 de vinhos diluídas.

Os dados são o resultado da análise química de vinhos produzidos na mesma região da Itália, porém derivados de três cultivos diferentes. A análise determinou treze constituintes encontrados em cada um dos três tipos de vinhos.

Quando utilizada no contexto de classificação, a base de dados Wine é classificada como uma base de classes bem definidas. Segundo a literatura, é um bom conjunto de dados para o primeiro teste de um novo classificador, mas não possui grandes desafios.

### **Vehicle Silhouettes**

O objetivo desta base de dados é reconhecer o tipo do veículo usando um conjunto de características extraídas dos veículos. A base é composta de valores analisados em quatro tipos de veículo: carros (Saab 9000 e Opel Manta 400), vans (Chevrolet Van) e ônibus (Double decker bus).



Esta combinação particular de veículos foi escolhida com a expectativa de que o ônibus, a van e qualquer um dos dois carros fossem facilmente distinguíveis entre si, mas seria difícil distinguir entre os dois carros. As imagens foram adquiridas por uma câmera olhando para baixo, direcionada para o veículo modelo, com um ângulo fixo de 34.2 graus em relação à horizontal.

Os veículos foram colocados sobre uma superfície de luz de fundo difusa. Os veículos foram pintados de preto fosco para minimizar os reflexos. As imagens foram capturadas com uma câmera CRS4000 com uma resolução espacial de 128x128 pixels com 64 níveis de cinza. As imagens foram processadas de forma a produzir valores binários e foram efetuados processamentos de remoção de ruídos. Esta base de dados possui 18 características, divididas em 4 classes.

### **Pima - Indians Diabetes Database**

A base de dados “Pima” contém o diagnóstico de diabéticos da população Pima, um povo nativo dos Estados Unidos da América vive perto de Phoenix, Arizona, EUA.

O diagnóstico investiga se o paciente apresenta sinais de diabetes, de acordo com critérios da Organização Mundial de Saúde (ou seja, se a glicemia de 2 horas pós-carga foi de pelo menos 200 mg / dl em qualquer exame de pesquisa ou, se for encontrado durante a rotina cuidados médicos).

Os dados foram coletados pelo Instituto Nacional de Diabetes e Doenças Digestivas e Renais dos EUA. Foram utilizados os 532 registros completos depois de deixar os dados (principalmente falta) de insulina no soro.

Várias restrições foram aplicadas sobre a seleção destes casos a partir de uma base de dados maior. Em particular, todos os pacientes eram mulheres, com pelo menos 21 anos de idade, de herança indígena Pima, que viviam perto de Phoenix.

A variável classe tem os valores 0 (saudável) e 1 (diabetes). A base de dados inclui 768 casos, sem valores ausentes. Os conjuntos de treinamento possuem 512 casos, e os conjuntos de testes possuem 256 casos.

A base de dados possui 9 atributos numéricos:

- 1 Número de grávidas.
- 2 Concentração de glicose há 2 horas num teste de tolerância à glicose oral.
- 3 A pressão arterial diastólica (mm Hg).
- 4 Espessura da prega cutânea tricipital (mm).
- 5 Insulina no soro a cada duas horas (mu U / ml).
- 6 Índice de massa corporal (peso em kg / (altura em m) <sup>2</sup>).
- 7 Função de pedigree de diabetes.
- 8 Idade (anos).
- 9 Variável de classe (0 ou 1).

### **Thyroid**

Esta base de dados descreve diagnósticos de pacientes no que se refere à análise de tireoide. Para o diagnóstico há três possíveis resultados: normal ou eutireoidismo, hipertireoidismo ou hipotireoidismo.

Os testes são usados para tentar prever se a tireoide de um paciente pertence a classe normal, hipertireoidismo ou hipotireoidismo. O número de ocorrências em cada uma das três classes é: normal (150 testes), hipertireoidismo (35 testes) e hipotireoidismo (30), totalizando 215 registros.

O diagnóstico, que representa o rótulo da classe, foi baseado em um registro médico completo, incluindo uma entrevista que busca lembrar todos os fatos que se relacionam com a doença, exames, etc.

A base de dados é composta por 5 atributos:

- 1 Teste de captação de resina T3 (percentual);

- 2 Tiroxina total no soro conforme medido pelo método de deslocamento isotópica;
- 3 Tri-iodotironina total no soro conforme medido por radioimunoensaio;
- 4 Hormônio basal estimulante da tireoide (TSH), medido pela técnica de radioimunoensaio;
- 5 Diferença absoluta máxima de valor de TSH após a injeção de 200 micro gramas de hormônio libertador de tirotropina, em comparação com o valor basal.

### **Wisconsin**

A base de dados Wisconsin Breast Cancer Database é uma coleção de informações relacionadas ao cancer de mama no estado de Wisconsin, nos Estados Unidos. Este banco de dados de câncer de mama foi obtido dos Hospitais da Universidade de Wisconsin, pelo Dr. William H. Wolberg.

O objetivo é identificar cada uma das amostras dentre as classes benignas e malignas. Cada amostra é composta de 11 atributos numéricos com valores que variam de 0 a 10. Malignidade é determinada tomando uma amostra de tecido da mama do paciente e de uma biópsia sobre ele. Um diagnóstico benigno é confirmado tanto por biópsia ou por exame periódico, dependendo da escolha do paciente.

As amostras consistem de características nucleares avaliadas visualmente em aspirações de agulhas finas retiradas da mama das pacientes. Para cada amostra foi atribuído um vetor de 9 dimensões (atributos de 3 a 9) pelo Dr. William H. Wolberg. Cada componente está no intervalo de 1 a 10, com o valor 1 representando um estado normal e 10 representando um estado anormal.

Número de atributos: 10 mais o atributo de classe:

- 1 Código numérico da amostra;
- 2 Espessura do agrupamento;
- 3 Uniformidade do tamanho da célula;

- 4 Uniformidade da forma da celular;
- 5 Adesão marginal;
- 6 Tamanho da célula epitelial;
- 7 Envolvimento do núcleo;
- 8 Suavidade da cromatina;
- 9 Normalidade do nucléolo;
- 10 Mitoses;
- 11 Classe.

Os atributos 2 a 10 foram usados para representar as instâncias. Cada instância tem uma das duas classes possíveis: benignos ou malignos.

O atributo 1 (código numérico da amostra) foi removido devido ao fato de ser um campo de identificação da amostra, e não possui nenhum valor classificatório significativo.

## 5.2 Metodologia de Testes

A metodologia de análises visa verificar o generalismo do método proposto para elementos desconhecidos ao conjunto de treinamento utilizado para a extração de regras aplicadas ao Sistema de Inferência Difuso (SID), pois para qualquer elemento utilizado no treinamento o sistema sempre acerta sua classe.

Na metodologia de análises de resultados foram adotados procedimentos que visam verificar a capacidade de generalização do sistema através de métodos estatísticos, visando estudos observacionais das etapas de testes realizadas.

As bases de dados utilizadas nos testes são descritas na seção 5.1.

O banco de dados é dividido em duas partes. A primeira parte é utilizada para treinamento do sistema de inferência difuso. O sistema obtido pelo treinamento é

então testado utilizando a segunda parte da base de dados.

Na Figura 33 é possível visualizar o cenário informado anteriormente.

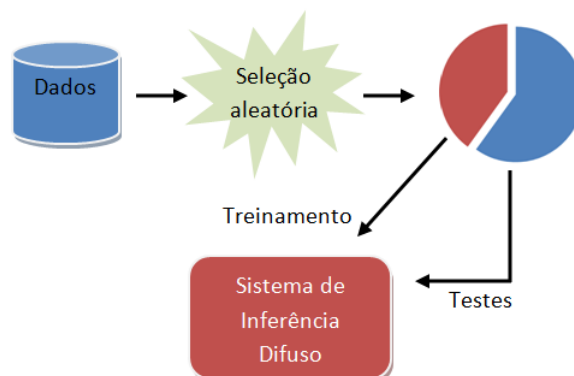


Figura 33: Divisão da base para treinamento e testes

Os percentuais de elementos (ou casos) utilizados para treinamento e testes variam em cada etapa do algoritmo. Para cada etapa de configuração percentual de treinamento e teste são realizados no mínimo 20 sorteios aleatórios para esses dois conjuntos extraídos da lista de casos existentes na base analisada, este número de sorteios é variável dependendo do comportamento dos resultados obtidos nos testes. Quando a dispersão destes resultados torna-se constante os sorteios são finalizados o algoritmo de teste altera a etapa de análise para novos valores percentuais de treinamento e teste. Durante a análise das etapas a variação do percentual foi de 40% a 95% de dados para treinamento e 60% a 5% para casos de testes, com incremento de 5% a cada etapa.

Inicialmente na primeira etapa, 40% da base é separada por sorteio para treinamento do SID enquanto 60% dos dados restantes são utilizados para testes do sistema obtido pelo treinamento. Na Figura 34 é possível visualizar a variação do percentual entre treinamento e testes.

Em cada sorteio, são efetuados os passos necessários para possibilitar o treinamento do SID e posteriormente seu teste.

Na etapa, cada sorteio é repetido no mínimo 20 vezes, de modo a permitir,

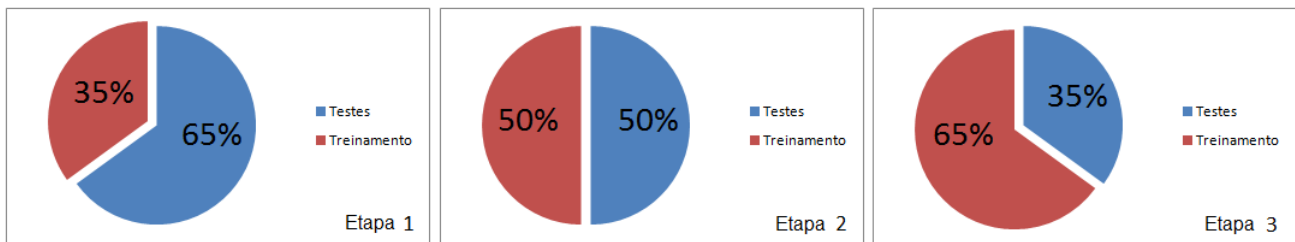


Figura 34: Variação do percentual treinamento e testes

através da seleção aleatória de elementos que diversos cenários de uma única etapa sejam verificados. Em cada um desses sorteios são obtidos: a média dos elementos para teste classificados corretamente, o menor e o maior índice de acerto do sorteio na classificação e a variação da média atual de acertos em relação a média do sorteio anterior. Este índice obtido de variação entre médias informa a dispersão dos sorteios, indicando a troca de etapa ao algoritmo.

Após o sorteio de cada etapa por no mínimo 20 vezes, o algoritmo passa a verificar a dispersão da média de acertos, sendo critério de parada um valor menor do que 5%. Quando os valores que satisfazem as condições de parada das iterações são alcançados, as informações referentes a execução da etapa são coletadas e os percentuais devem ser alterados para a próxima etapa do algoritmo. Os passos efetuados para cada etapa são exibidos no Algoritmo 4.

---

**Algoritmo 4:** Passos executados em cada variação de modelo
 

---

```

begin
  for modelo ← 40 to 95% do
    while (num iterações > 20) e (variação média < 5%) do
      - Preparação dos dados para treinamento do SID -
      * seleciona aleatoriamente os elementos para testes
      * utiliza o restante dos elementos para treinamento
      * verifica e soluciona as colisões entre as hipercaixas envolventes
      * verifica e soluciona as hipercaixas não envolvidas pela hipercaixa sigma
      * efetua o treinamento do sistema de inferência difuso

      - Utilizando a base de testes -
      * realiza a classificação dos elementos
      * obtém o número de elementos classificados corretamente
      * obtém o número de elementos classificados incorretamente
      * calcula a média do número de elementos classificados corretamente no modelo atual
      * calcula a variação da média atual em relação à média da iteração anterior
    end while
  end for

  - Coleta dos resultados do modelo atual -
  * obtém o menor número de elementos classificados corretamente
  * obtém o maior número de elementos classificados corretamente
  * obtém o número de iterações efetuadas no modelo
  * obtém a média do número de elementos classificados corretamente
end

```

---

### 5.3 Medidas de Avaliação

As medidas mais comuns para avaliar a qualidade de um sistema de busca e recuperação de informação são conhecidas como revocação e precisão.

Considere:

$Q$  : uma busca pelas palavras chaves 'Extração de Regras'.

$D$  : o conjunto de artigos relacionados à busca existentes na base de dados.

$N$  : o conjunto resposta, gerado pela consulta.

$D \cap N$  : artigos relevantes na resposta.

De acordo com (SALTON; MCGILL, 1983), as medidas de revocação e precisão são definidas da seguinte maneira:

**Revocação** ou **Recall** é a fração de artigos relevantes (conjunto  $D$ ) que foram

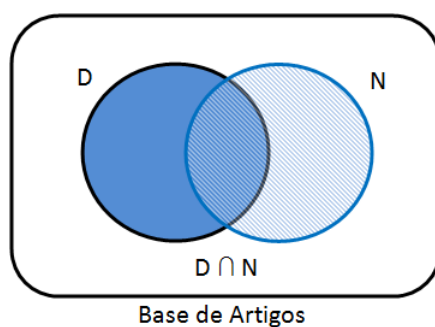


Figura 35: Revocação e Precisão

recuperados na resposta (conjunto N).

$$Recall = \frac{|D \cap N|}{|D|} \quad (5.1)$$

**Precisão** ou **Precision** é a fração de artigos recuperados (conjunto N) que são relevantes.

$$Precision = \frac{|D \cap N|}{|N|} \quad (5.2)$$

A **Medida-F**, também conhecida como **F-score** ou **F1-score**, é derivada da precisão e revocação, e consiste em outra forma de avaliar a qualidade de um sistema de busca. A Medida-F é definida como a média ponderada entre precisão e revocação. Seu valor varia entre 0 e 1, sendo 0 o pior resultado e 1 o melhor resultado.

$$F = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5.3)$$

## 5.4 Testes

Todas as bases de dados mencionadas foram testadas com a metodologia de análise apresentada.



Na Figura 36 são exibidos os resultados obtidos utilizando a base de dados Iris. O eixo X representa o percentual da base de dados utilizado para treinamento do sistema de inferência difuso. O eixo Y representa a classificação correta dos testes realizados com a etapa de treinamento, indicando respectivamente o menor, o maior e a média do número de acertos obtidos durante a execução da etapa.

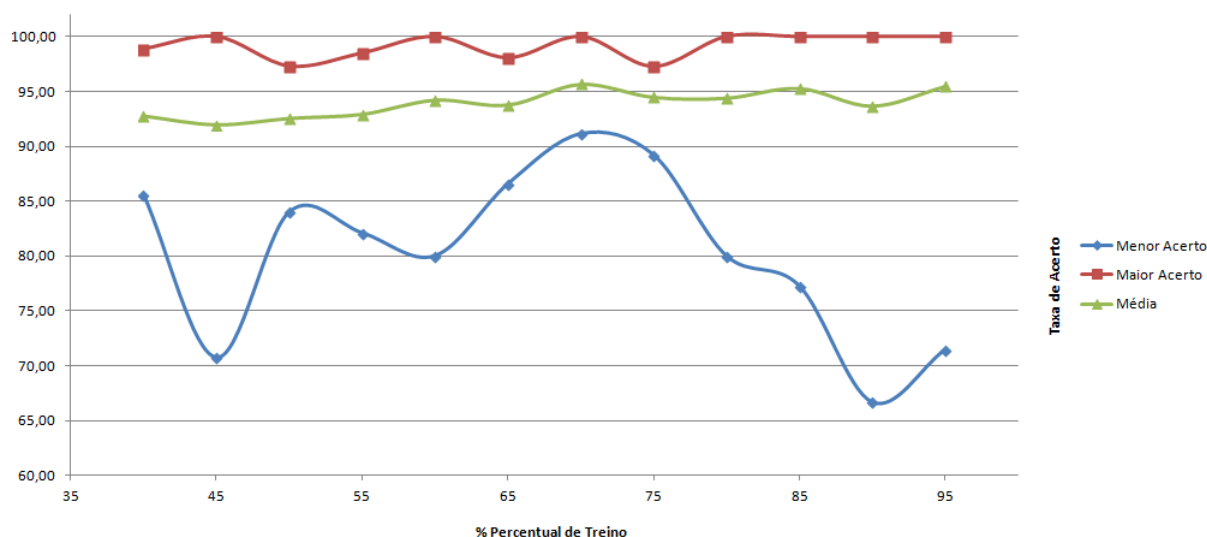


Figura 36: Resultados obtidos para a base de dados Iris

Pelos resultados mostrados na Figura 36 é possível visualizar a efetividade da técnica em realizar corretamente a classificação mesmo com um baixo percentual de elementos utilizados para o treinamento do sistema. A média de acertos na generalização obtida em todos os modelos foi superior a 90%. Lembrando que todos os dados de testes não são conhecidos pelo modelo de treinamento.

Devido a pouca quantidade de casos existentes na base *Iris*, em torno de 150 casos, quando o percentual de elementos utilizados para testes é menor do que 25% do total, o que representa menos de 37 casos, um único sorteio desastroso reduz drasticamente o índice mínimo de acerto da etapa, e esse efeito pode ser verificado na Figura 36 a partir de 75% de casos utilizados para treinamento.

No Apêndice A são exibidos os resultados obtidos para as demais bases de dados descritas.

Na Tabela 10 é exibido um resumo evidenciando os melhores resultados obtidos com todas as bases de dados descritas. A tabela representa o melhor índice de valor médio de acertos relacionados a um percentual de casos utilizados para treinamento. É possível verificar que para bases com poucos atributos o índice de acerto é bem superior comparado com bases com muitos atributos. Quando o número de atributos é elevado, as separações de hipercaixas tornam-se extremamente complexas. Pois os planos de separação representam os atributos, prejudicando a qualidade das regras difusas extraídas da análise e reduzindo a capacidade de generalizar causado pelo aumento abrupto de espaço de busca.

Base de Dados	Casos	Atributos	% Treino	Acertos
breast-cancer-wisconsin	699	9	75	87,91
iris	150	4	70	95,66
new-thyroid	215	5	85	95,88
pima-indians-diabetes	768	8	95	74,59
vehicle	846	18	90	73,27
wine	178	13	95	86,90

Tabela 10: Resultados obtidos com a técnica proposta

Foram efetuados testes comparativos entre a técnica proposta e o algoritmo *J48*, utilizando as medidas de avaliação mencionadas anteriormente. Nas Figuras 37, 38 e 39 são exibidos os gráficos comparativos utilizando *Precisão*, *Revocação* e *Medida-F* respectivamente.

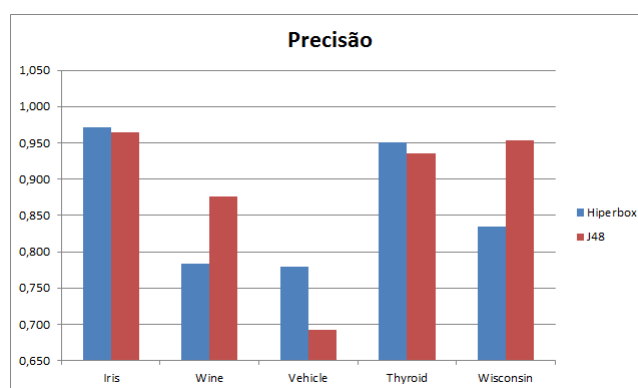


Figura 37: Comparativo utilizando a Precisão

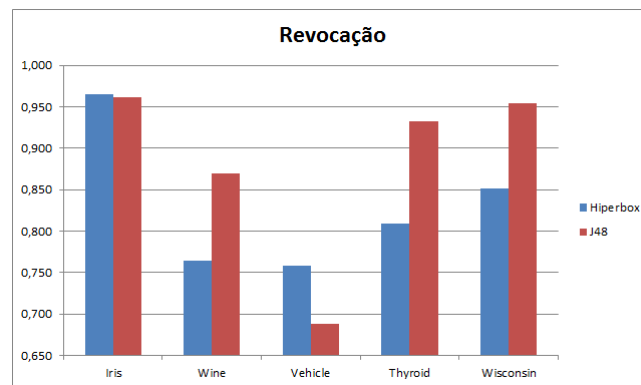


Figura 38: Comparativo utilizando a Revocação

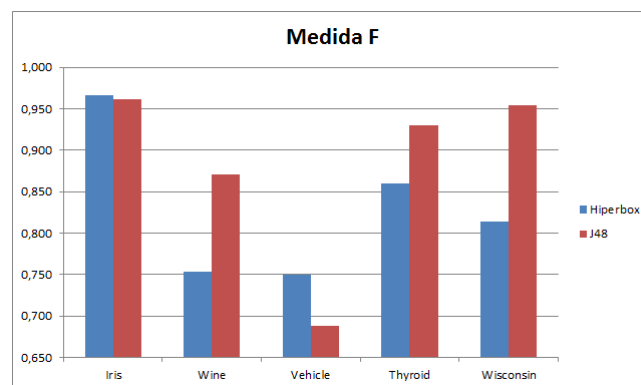


Figura 39: Comparativo utilizando a Medida F

Pelos gráficos exibidos é possível visualizar que embora a técnica proposta não alcance os mesmos resultados que o algoritmo *J48*, possui resultados aproximados. Melhorias nas regras de quebra das hipercaixas difusas podem melhorar o desempenho do algoritmo de hipercaixas difusas.

## 6 Conclusão

O trabalho apresenta o desenvolvimento de uma abordagem híbrida inovadora, onde a extração de regras difusas é realizada pela análise gráfica em casos classificados. Este estudo gráfico é baseado na combinação entre técnicas de colisão de objetos utilizadas em jogos com regiões limítrofes determinadas por funções difusas. Assim aliam-se as vantagens da separação distinta das classes por conjuntos delimitados pela técnica de colisão, modificada para hiperespaços amostrais, com as regiões de incerteza, localizadas entre os conjuntos, modeladas com funções difusas, permitindo incorporar a capacidade de generalização à metodologia.

A metodologia de análise sobre as bases descritas permitiu identificar a capacidade de generalização da técnica sobre casos desconhecidos ao extrator de regras difusas, aumentando o índice de acerto de classificação da técnica sem prejudicar a velocidade de teste dos casos com as regras.

Ainda durante a fase de análises com as bases, a fator de quebra de hipercaixas com colisão mostrou-se importante na eficiência na busca de conjuntos distintos de classes. Nos dois métodos apresentados, com maior e menor índice de colisão, sua eficiência variou diretamente com o número de casos e atributos contidos nos bancos de dados. Quando uma hipercaixa está contida dentro da outra faz aumentar a eficácia do método de quebra por menor índice de colisão, indicando o algoritmo romper a maior caixa, comum para bases com poucos atributos. Já nas situações que envolvem contato superficial entre as hipercaixas, o método de quebra por maior índice torna-se eficaz, reduzindo o número de passos para obtenção

das regras, cenário presente em bases com muitos atributos (dimensões).

A escalabilidade foi outro ponto positivo detectado na abordagem, pois a análise gráfica realizada pelo detector de colisões, ao tratar os casos como hipercaixas distintas, utiliza em seu cálculo apenas os valores limites dos hiperplanos. Logo, a hipercaixa poderá conter um ou milhares de casos, mas a velocidade de análise será pouco alterada.

O paralelismo pode ser aplicado na metodologia desenvolvida devido ao grupo de análises de colisão entre pares de hipercaixas serem isolados, não realizando troca de informações entre esses grupos durante a detecção de planos de contato.

O processo de extração de regras difusas apresentou deficiências quanto a bases com grande número de atributos, pois o processo de detecção de colisão necessita identificar contato em cada plano associado a um atributo. Portanto o desempenho da extração de regras e a qualidade de generalização são prejudicados pelo aumento do número de atributos causado pela expansão do espaço de busca analisado.

É importante destacar que esta metodologia por enquanto obtém resultados apenas com bases com atributos numéricos. Isto é devido ao estudo gráfico do espaço não permitir a comparação ou o cálculo de distância entre atributos simbólicos, como exemplo expressões, variáveis linguísticas, grupos sequenciais ou índices.

Por fim, a metodologia proposta se apresenta como substituta para os métodos tradicionais de extração de regras devido a sua escalabilidade, generalização e facilidade do emprego de técnicas de paralelismo, tornando-a uma alternativa robusta e eficiente quando os métodos convencionais tem dificuldades em trabalhar com bases com grande número de casos.

## 7 Trabalhos Futuros

A técnica de detecção de colisão Axis Aligned Bounding Box (AABB) é utilizada na metodologia desenvolvida, mas como melhoria poderia-se empregar técnicas de detecção mais complexas como a Oriented Bounding Box (OBB), possibilitando o aumento de eficiência na quebra das hipercaixas. Como o OBB se ajusta melhor ao conjunto de casos deslocando-se os planos que formam as hipercaixas, permite solucionar as colisões com um menor número de iterações de quebras.

Nas duas técnicas de análise de quebra de hipercaixas com colisão, seria possível a melhoria da árvore de busca criada, pela inserção de um sistema especialista neste processo para a escolha automática da técnica mais eficiente com o tipo de região de contato entre as hipercaixas.

No processo de estudo para casos com atributos simbólicos, a implementação de um classificador Bayesiano criará uma métrica entre os símbolos através da similaridade do contexto dos mesmos.

As bases que apresentam um grande número de atributos poderiam ser simplificadas aplicando a técnica de redução, presente como etapa parcial dos conjuntos aproximados, a fim de reduzir a dimensionalidade, enviando para a metodologia desenvolvida apenas atributos importantes da base.

## Referências

AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. N. Mining association rules between sets of items in large databases. In: BUNEMAN, P.; JAJODIA, S. (Ed.). *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*. Washington, D.C.: [s.n.], 1993. p. 207–216.

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules in large databases. In: *International Conference On Very Large Data Bases (VLDB '94)*. San Francisco, Ca., USA: Morgan Kaufmann Publishers, Inc., 1994. p. 487–499. ISBN 1-55860-153-8.

ALMEIDA, R. M. A. d. *Reconhecimento de Padrões Utilizando o Teorema do Eixo de Separação e Redes Neurais Artificiais*. Dissertação (Mestrado) — Universidade Federal de Itajubá, UNIFEI, 2010.

CAUDILL, M. Neural networks primer, part 1. *AI Expert*, v. 2, n. 12, p. 46–52, dez. 1987.

CODD, E. F. Relational completeness of database sublanguages. In: RUSTIN, R. (Ed.). *Data Base Systems*. New Jersey: Prentice-Hall, 1972.

EBERLY, D. H. *3D game engine design - a practical approach to real-time computer graphics*. [S.l.]: Morgan Kaufmann, 2001. I–XXVIII, 1–560 p. ISBN 978-1-55860-593-0.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI Magazine*, v. 17, n. 3, p. 37–54, 1996.

FENG, S. Supervised classification algorithms based on artificial immune. In: *ICNC*. [S.l.]: IEEE, 2012. p. 879–882.

FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, v. 7, p. 179–188, 1936.

GOTTSCHALK, S. *Separating Axis Theorem*. [S.l.], 1996.

GOTTSCHALK, S. *Collision queries using oriented bounding boxes*. Tese (Doutorado) — University of North Carolina at Chapel Hill, 2000.

- GOTTSCHALK, S.; LIN, M.; MANOCHA, D. OBBTree: A hierarchical structure for rapid interference detection. In: *Proceedings of the ACM Conference on Computer Graphics*. New York: ACM, 1996. p. 171–180. ISBN 0-201-94800-1.
- HAN, J.; FU, Y. *Discovery of Multiple-Level Association Rules from Large Databases*. [S.l.], mar. 1995. 35 p.
- HAN, J.; KAMBER, M. *Data Mining: Concepts and Techniques*. [S.l.]: Morgan Kaufmann, 2000. ISBN 1-55860-489-8.
- JIMÉNEZ, P.; THOMAS, F.; TORRAS, C. 3D collision detection: a survey. *Computers & Graphics*, v. 25, n. 2, p. 269–285, 2001.
- LIANG, C. et al. An unsupervised vegetation classification algorithm based immune. In: *ICNC*. [S.l.]: IEEE, 2010. p. 2842–2846. ISBN 978-1-4244-5961-2.
- LIANG, S.-C.; LEE, Y.-C.; LEE, P.-C. The application of ant colony optimization to the classification rule problem. In: HONG, T.-P. et al. (Ed.). *GrC*. [S.l.]: IEEE, 2011. p. 390–392. ISBN 978-1-4577-0372-0.
- LIPPMANN, R. P. An introduction to computing with neural nets. *IEEE ASSP Magazine*, v. 4, n. 2, p. 4–22, abr. 1987.
- MIAO, D. et al. Rough set based hybrid algorithm for text classification. *Expert Syst. Appl*, v. 36, n. 5, p. 9168–9174, 2009. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2008.12.026>>.
- OTERO, F. E. B.; FREITAS, A. A.; JOHNSON, C. G. A new sequential covering strategy for inducing classification rules with ant colony algorithms. *IEEE Trans. Evolutionary Computation*, v. 17, n. 1, 2013.
- PARPINELLI, R. S.; LOPES, H. S.; FREITAS, A. A. Data mining with an ant colony optimization algorithm. *IEEE Transactions on Evolutionary Computation*, v. 6, n. 4, p. 321–332, ago. 2002. ISSN 1089-778X.
- PAWLAK, Z. Rough sets. *International Journal of Computer and Information Sciences*, v. 11, n. 5, p. 341–356, 1982.
- PAWLAK, Z. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Dordrecht: Kluwer Academic Publishers, 1991.
- PAWLAK, Z.; SKOWRON, A. Rudiments of rough sets. *Inf. Sci*, v. 177, n. 1, p. 3–27, 2007.



- PIATETSKY-SHAPIRO, G.; FRAWLEY, W. J. (Ed.). *Knowledge discovery in databases*. Menlo Park, CA/Cambridge, MA: AAAI Press/MIT Press, 1991.
- QUINLAN, R. Induction of decision trees. *Machine Learning*, v. 1, p. 81–106, 1986.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning internal representations by error propagation. In: RUMELHART, D. E.; MCCLELLAND, J. L.; GROUP. the PDP research (Ed.). *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundations*. [S.l.]: MIT Press, 1986.
- SALTON, G.; MCGILL, M. J. *Introduction to Modern Information Retrieval*. [S.l.]: McGraw Hill, 1983.
- SELINGER, P. G. et al. Access path selection in a relational database management system. In: *Proceedings of the ACM SIGMOD Annual Conference on Management of Data*. [S.l.: s.n.], 1979.
- SOTIROPOULOS, D. N.; TSIHRINTZIS, G. A. Artificial immune system-based classification in class-imbalanced image classification problems. In: TSIHRINTZIS, G. A. et al. (Ed.). *IIH-MSP*. [S.l.]: IEEE, 2012. p. 138–141. ISBN 978-1-4673-1741-2.
- SWAIN, M. et al. An approach for IRIS plant classification using neural network. Academy & Industry Research Collaboration Center (AIRCC), 2012. ISSN 22297103.
- TABOADA, K. S. et al. Fuzzy association rule mining and classifier with chi-squared correlation measure using genetic network programming. In: *ICRAS & SICE International Joint Conference, ICCAS-SICE, 2009*. Fukuoka: IEEE, 2009. p. 3863–3869.
- WANG, X.; WANG, M.; LI, C. Research on collision detection algorithm based on AABB. In: WANG, H. et al. (Ed.). *ICNC*. [S.l.]: IEEE Computer Society, 2009. p. 422–424. ISBN 978-0-7695-3736-8.
- WASSERMAN, P. D. *Advanced methods in neural computing*. [S.l.]: Van Nostrand Reinhold, 1993. I–VIII, 1–255 p. (VNR computer library). ISBN 978-0-442-00461-3.
- WIKIPÉDIA. *Desvio padrão — Wikipédia, a enciclopédia livre*. 2013. [Online; accessed 6-junho-2013 at 21:53].

ZADEH, L. Fuzzy sets as a basis for possibility. *Fuzzy Sets and Systems*, v. 1, p. 3–28, 1978.

## APÊNDICE A - Gráficos dos resultados obtidos

Os gráficos mostrados a seguir apresentam o grau de acerto das regras difusas obtidas pela metodologia híbrida desenvolvida. Cada gráfico exibe a característica generalista da técnica para uma determinada base, onde o eixo X representa o percentual de casos para o treinamento e o eixo Y indica o percentual de acerto para os demais casos restantes, isto é, os casos de teste desconhecidos no treinamento. As curvas plotadas exibem o grau de acerto máximo, médio e mínimo para um conjunto de sorteios de seleção de casos para cada etapa de percentual de treinamento, onde o valor mínimo de sorteios foi igual a vinte, e o ponto de parada dos sorteios na etapa foi referente a um indicador de estabilização do valor médio de acertos para cada novo sorteio. É importante salientar que, em alguns gráficos, o valor mínimo de acerto tende a piorar conforme se aumenta o número de casos utilizados para treinamento, pois isso resulta em um número reduzido de casos restantes para teste, principalmente em algumas bases pequenas. Assim, com poucos casos para testes, um único caso classificado errado pela metodologia resulta em um baixo grau de acerto (valor mínimo) no conjunto utilizado para teste.

Na Figura 40 é possível verificar que com o percentual inicial de 40% utilizado para treinamento do modelo, a técnica apresentou um bom índice de acertos, aproximando-se dos 85%. No entanto, o valor do menor acerto aproximou-se dos 25%. Esse valor se deve tanto ao fato de que um baixo percentual da base analisada foi utilizado para treinamento quanto ao fato que em algum dos sorteios a maioria dos elementos que representam uma determinada classe podem ter sido removidos,

e com isso elementos utilizados para testes que pertençam àquela classe não puderam ser classificados corretamente. Também é possível verificar que a média de acertos se estabilizou entre 60 e 75%, o que demonstra que mesmo com a variação do percentual de treinamento a técnica conseguiu realizar a correta classificação dos dados.

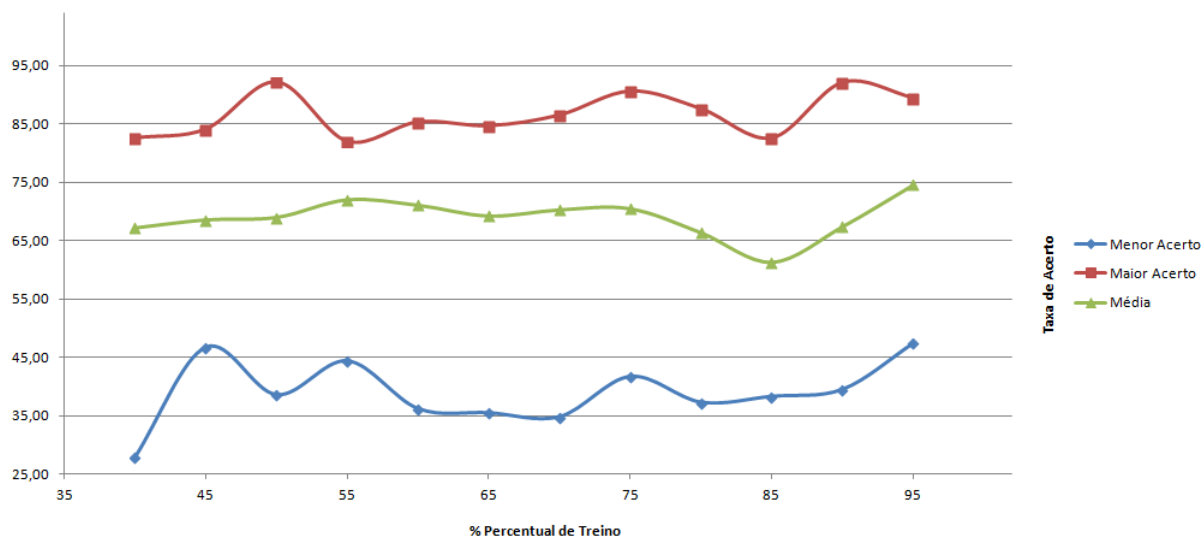


Figura 40: Base Pima

Na Figura 41 o índice médio de acertos aumentou gradativamente a medida que o percentual utilizado para treinamento também aumentou. A técnica alcançou resultados acima de 95% mesmo com um baixo percentual de treinamento. Com o percentual de treinamento em 95% o menor acerto sofreu uma queda, chegando a 70%. Este comportamento se deve ao fato de que com apenas 5% de casos utilizados para testes, um único caso classificado incorretamente afeta consideravelmente os resultados.

Na Figura 42 é possível visualizar que a média de acertos se manteve praticamente estável até o percentual de treinamento de aproximadamente 60%. Entre os percentuais de 60% e 75% houve um aumento na média de acertos assim como no maior acerto. Após este percentual de treinamento, a média se manteve praticamente estável, enquanto o maior acerto alcançou resultados próximos a 95%. O

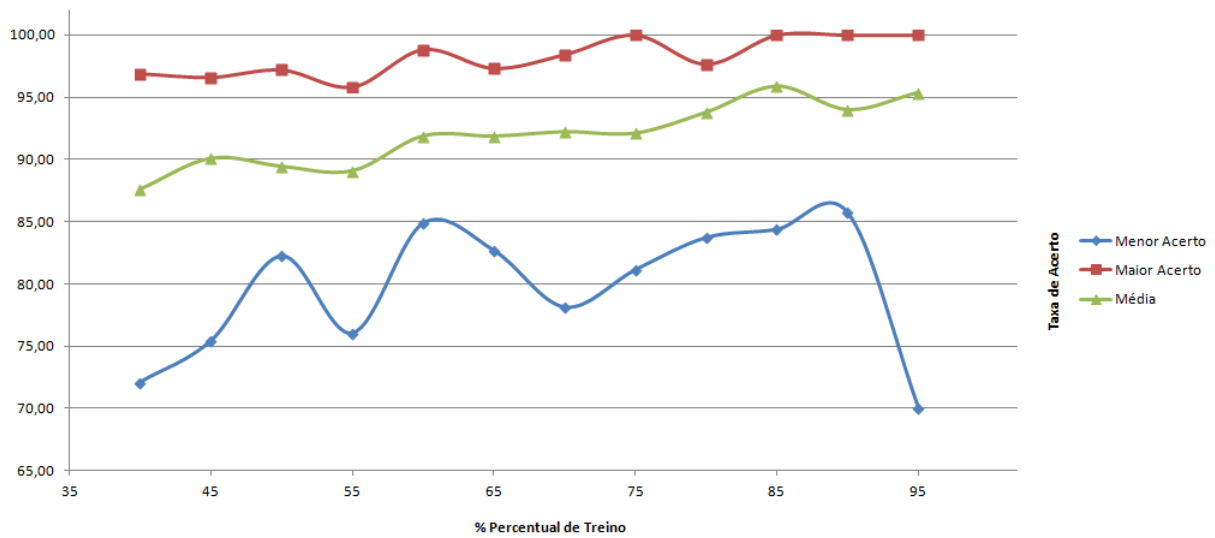


Figura 41: Base Thyroid

melhor resultado para o menor acerto foi alcançado em 90%, onde provavelmente todos os elementos foram classificados corretamente.

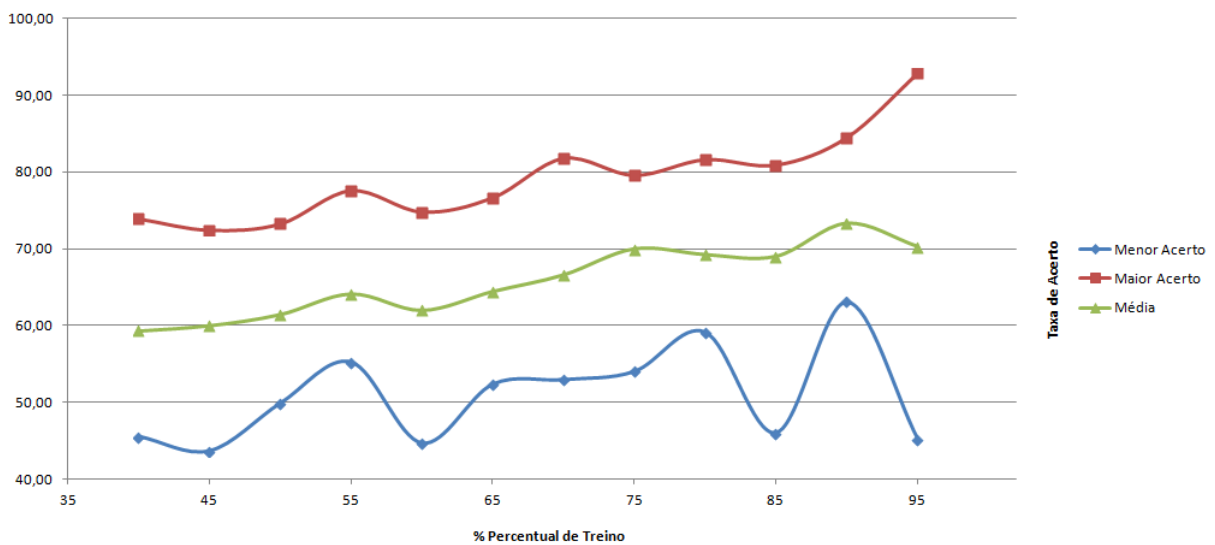


Figura 42: Base Vehicle

Na Figura 43 verifica-se que a base Wine foi a que obteve um dos melhores resultados de menor acerto, variando entre aproximadamente 57% e 73%, embora tenha sofrido variações durante o aumento do percentual de treinamento. Os

melhores resultados de menor acerto foram obtidos entre 75% e 85%, sendo que neste intervalo o percentual da média e do maior acerto se mantiveram estáveis. A partir de 85% do percentual de treinamento, o índice de maior acerto aumentou até alcançar 100%.

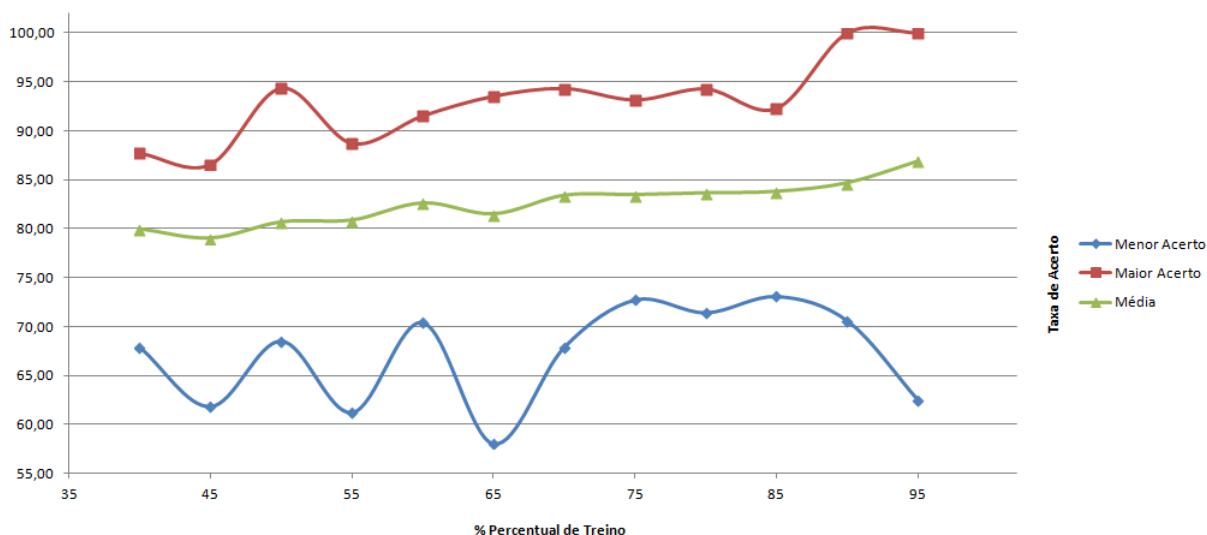


Figura 43: Base Wine

Na Figura 44 são exibidos os resultados que demonstram que a base Wisconsin foi a que apresentou a maior dispersão entre os índices de menor acerto e acerto médio. Isto demonstra que a base de dados, apesar de apresentar bons resultados de acerto médio e de maior acerto, possui dados que podem estar muito aproximados, não sendo possível distinguir as classes quando muitos dos elementos que a representam são removidos. Nos casos analisados, o índice de acertos mínimo reflete as ocasiões onde os sorteios realizados removem muitos elementos de uma classe. Os melhores resultados de maior acerto foram obtidos com a base Wisconsin, e este foi alcançado quando os sorteios removiam o mesmo número de elementos de cada uma das classes, o que possibilitava a correta classificação dos casos de testes.

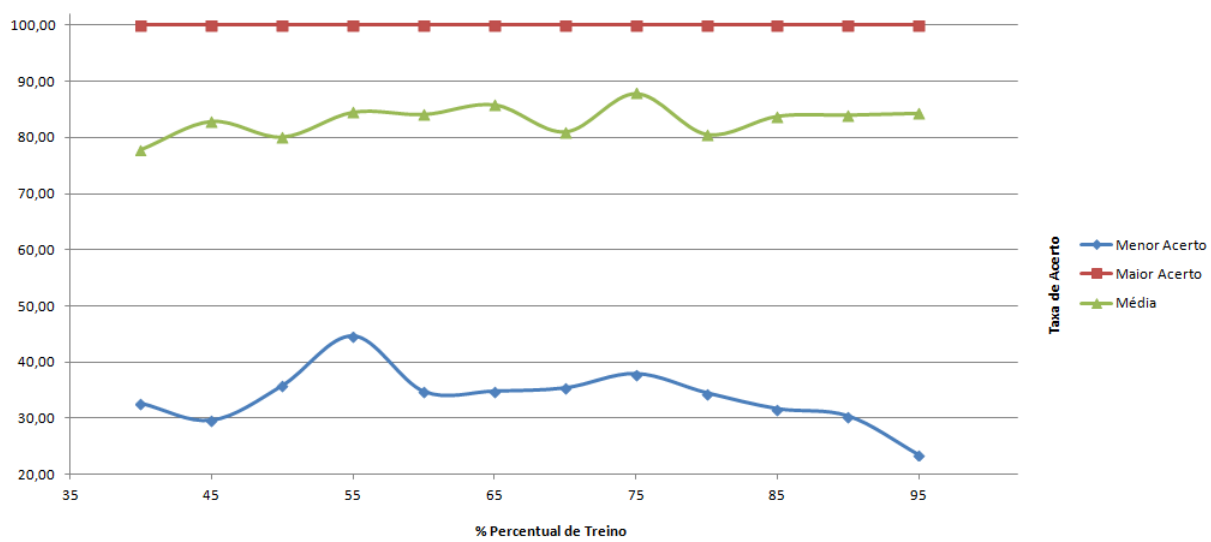


Figura 44: Base Wisconsin

## APÊNDICE B - Exemplo de Regras Difusas para Base Wisconsin

A seguir será mostrado, como exemplo, um dos modelos difusos de regras para classificação da base Wine. Os hiperplanos serão responsáveis pela etapa de fuzzificação dos parâmetros de entrada dos casos, transformando-os em graus de pertinência para os conjuntos difusos através das funções trapezoidais difusas encontradas em seus limites. As regras difusas são obtidas pela união de todos os hiperplanos de uma hipercaixa envolvente, sendo esta hipercaixa a caracterização da classe a ser representada na regra difusa. Devido às quebras realizadas nas hipercaixas a fim de solucionar as colisões, cada classe pode ser representada por  $N$  hipercaixas, e com isso, várias regras podem representar apenas uma classe.

### Funções Difusas

A Tabela 11 apresenta informações sobre cada uma das hipercaixas geradas pelo SID. Cada atributo dos dados analisados é representado por um plano na hipercaixa difusa, e para cada plano quatro valores são necessários para representar a respectiva função de pertinência. Na tabela são exibidos os quatro valores para cada um dos planos da base de dados Wine. Foram obtidas várias hipercaixas difusas, devido às quebra realizadas, e os valores para cada uma dessas hipercaixas são exibidos na tabela. A coluna *v1* demonstra o limite inferior da função de pertinência, onde a representatividade inicia em zero e tende a próximo a um.



As colunas  $v_2$  e  $v_3$  demonstram a área de maior representatividade, onde valores que estejam nesse intervalo possuem representatividade igual a um. A coluna  $v_4$  demonstra o limite superior da função de pertinência, onde a representatividade inicia em um e tende a zero. Os valores dessa tabela são utilizados na avaliação das regras exibidas na Tabela 12.

Tabela 11: Funções difusas da base Wine

Hipercaixa	Hiperplano	$v_1:\mu(0)$	$v_2:\mu(1)$	$v_3:\mu(1)$	$v_4:\mu(0)$
1	Proline_1	882,77	1120	1680	1736,62
	Alcohol_1	12,65	13,05	14,75	15,1
	Malic_acid_1	1,12	1,43	2,36	2,44
	Ash_1	1,99	2,04	2,72	3,01
	Alcalinity_of_ash_1	9,03	11,4	22,5	24,81
	Magnesium_1	74,08	89	132	138,07
	Total_phenols_1	1,82	2,2	3,88	4,11
	Flavanoids_1	1,94	2,43	3,93	4,41
	Nonflavanoid_phenols_1	0,07	0,17	0,5	0,53
	Proanthocyanins_1	0,59	1,25	2,96	3,48
	Color_intensity_1	3,09	4,8	8,9	9,64
	Hue_1	0,76	0,86	1,28	1,45
	OD280/OD315_1	2,2	2,57	3,58	3,82

Tabela 11 – *Continuação*

Hipercaixa	Hiperplano	v1: $\mu(0)$	v2: $\mu(1)$	v3: $\mu(1)$	v4: $\mu(0)$
2	Proline_2	932,79	970	1105	1151,1
	Alcohol_2	12,23	12,85	14,83	15,36
	Malic_acid_2	-0,43	1,35	4,04	4,55
	Ash_2	1,85	2,1	2,59	2,77
	Alcalinity_of_ash_2	10,71	11,2	18,9	21,96
	Magnesium_2	74,99	90	127	131,45
	Total_phenols_2	1,89	2,4	3,25	3,63
	Flavanoids_2	1,86	2,33	3,56	3,87
	Nonflavanoid_phenols_2	0,11	0,17	0,42	0,43
	Proanthocyanins_2	0,72	1,25	2,38	2,85
	Color_intensity_2	1,97	3,7	7,22	8,15
	Hue_2	0,72	0,87	1,23	1,34
	OD280/OD315_2	2,01	2,51	4	4,53
3	Proline_3	647,24	760	920	1027,05
	Alcohol_3	12,88	13,48	14,22	14,64
	Malic_acid_3	-0,44	1,63	3,99	5
	Ash_3	1,95	2,28	2,8	3,04
	Alcalinity_of_ash_3	8,91	13,2	20,5	25,52
	Magnesium_3	75,85	96	128	149,86
	Total_phenols_3	2,2	2,53	3,15	3,52
	Flavanoids_3	2,38	2,61	3,29	3,65
	Nonflavanoid_phenols_3	0,05	0,17	0,37	0,48
	Proanthocyanins_3	1,15	1,66	2,34	2,7
	Color_intensity_3	2,51	3,52	6,13	7,51
	Hue_3	0,68	0,89	1,25	1,41
	OD280/OD315_3	2,99	3,36	3,82	4,06

Tabela 11 – *Continuação*

Hipercaixa	Hiperplano	v1: $\mu(0)$	v2: $\mu(1)$	v3: $\mu(1)$	v4: $\mu(0)$
4	Proline_4	549,34	680	885	1043,99
	Alcohol_4	12,71	12,93	13,28	13,55
	Malic_acid_4	-0,44	1,64	3,98	5,71
	Ash_4	1,43	2,1	3,22	3,89
	Alcalinity_of_ash_4	8,83	15,5	25	29,37
	Magnesium_4	84,54	102	124	136,8
	Total_phenols_4	2,08	2,41	3	3,28
	Flavanoids_4	2,12	2,41	3	3,25
	Nonflavanoid_phenols_4	0,1	0,25	0,47	0,58
	Proanthocyanins_4	1,04	1,36	2,03	2,55
	Color_intensity_4	2,97	3,58	5,04	5,83
	Hue_4	0,63	0,82	1,13	1,36
	OD280/OD315_4	2,3	2,78	3,52	3,96
5	Proline_5	985	985	985	985
	Alcohol_5	12,99	12,99	12,99	12,99
	Malic_acid_5	1,67	1,67	1,67	1,67
	Ash_5	2,6	2,6	2,6	2,6
	Alcalinity_of_ash_5	30	30	30	30
	Magnesium_5	139	139	139	139
	Total_phenols_5	3,3	3,3	3,3	3,3
	Flavanoids_5	2,89	2,89	2,89	2,89
	Nonflavanoid_phenols_5	0,21	0,21	0,21	0,21
	Proanthocyanins_5	1,96	1,96	1,96	1,96
	Color_intensity_5	3,35	3,35	3,35	3,35
	Hue_5	1,31	1,31	1,31	1,31
	OD280/OD315_5	3,5	3,5	3,5	3,5

Tabela 11 – *Continuação*

Hipercaixa	Hiperplano	v1: $\mu(0)$	v2: $\mu(1)$	v3: $\mu(1)$	v4: $\mu(0)$
6	Proline_6	159,83	345	750	886,8
	Alcohol_6	10,58	11,41	13,67	13,91
	Malic_acid_6	-0,05	0,74	2,68	2,84
	Ash_6	1,3	1,71	2,92	3,21
	Alcalinity_of_ash_6	9,78	14,8	26	29,7
	Magnesium_6	47,13	70	136	140,5
	Total_phenols_6	1,29	1,6	2,48	2,65
	Flavanoids_6	0,96	1,09	2,09	2,55
	Nonflavanoid_phenols_6	0,07	0,24	0,63	0,73
	Proanthocyanins_6	-0,17	0,41	2,76	2,85
	Color_intensity_6	0,46	1,74	4,68	5,72
	Hue_6	0,85	1,06	1,45	1,56
	OD280/OD315_6	1,11	1,67	3,48	4,06
7	Proline_7	-2,7	380	714	1040,7
	Alcohol_7	12,23	12,7	13,05	13,41
	Malic_acid_7	-2,16	1,81	5,8	9,81
	Ash_7	1,82	2,13	2,4	2,66
	Alcalinity_of_ash_7	14,71	18,8	23	27,49
	Magnesium_7	65,02	86	101	116,98
	Total_phenols_7	1,59	2,2	2,83	3,51
	Flavanoids_7	2,38	2,53	2,65	2,77
	Nonflavanoid_phenols_7	0,06	0,26	0,43	0,6
	Proanthocyanins_7	1,54	1,77	2,01	2,28
	Color_intensity_7	0,75	2,57	3,9	5,3
	Hue_7	0,25	0,73	1,19	1,8
	OD280/OD315_7	3,06	3,1	3,14	3,19

Tabela 11 – *Continuação*

Hipercaixa	Hiperplano	v1: $\mu(0)$	v2: $\mu(1)$	v3: $\mu(1)$	v4: $\mu(0)$
8	Proline_8	-61,43	278	886	1048,83
	Alcohol_8	10,77	11,03	12,08	12,81
	Malic_acid_8	-0,75	0,89	3,74	4,33
	Ash_8	1,23	1,7	2,78	3,24
	Alcalinity_of_ash_8	10,95	17,5	28,5	30,35
	Magnesium_8	69,22	84	107	116,58
	Total_phenols_8	1,32	2,13	3,38	3,92
	Flavanoids_8	1,77	2,14	2,65	2,82
	Nonflavanoid_phenols_8	-0,13	0,13	0,58	0,77
	Proanthocyanins_8	-0,4	1,04	3,58	3,99
	Color_intensity_8	1,62	1,9	3,3	4,14
	Hue_8	0,18	0,75	1,71	1,98
	OD280/OD315_8	2,26	2,44	3,28	3,7
9	Proline_9	-175,05	325	937	1178,77
	Alcohol_9	12,16	12,29	12,52	12,63
	Malic_acid_9	-1,19	1,21	4,43	5,5
	Ash_9	1,57	1,98	2,73	3,11
	Alcalinity_of_ash_9	9,93	16	26,5	31,81
	Magnesium_9	22,63	85	162	183,37
	Total_phenols_9	1,96	2,2	2,56	2,9
	Flavanoids_9	1,74	2,11	2,65	2,94
	Nonflavanoid_phenols_9	0,16	0,26	0,43	0,52
	Proanthocyanins_9	-0,15	1,22	3,28	3,95
	Color_intensity_9	0,02	2	4,6	5,43
	Hue_9	0,49	0,8	1,23	1,54
	OD280/OD315_9	1,78	2,3	3,38	3,86

Tabela 11 – *Continuação*

Hipercaixa	Hiperplano	v1: $\mu(0)$	v2: $\mu(1)$	v3: $\mu(1)$	v4: $\mu(0)$
10	Proline_10	267,96	352	502	570,61
	Alcohol_10	10,19	11,56	13,86	14,81
	Malic_acid_10	-1,54	1,01	4,31	5,74
	Ash_10	0,95	1,7	3,23	3,8
	Alcalinity_of_ash_10	7,65	15	28,5	34,63
	Magnesium_10	48,96	78	119	129,9
	Total_phenols_10	2,24	2,74	3,5	3,78
	Flavanoids_10	1,02	2,86	5,08	5,66
	Nonflavanoid_phenols_10	-0,05	0,19	0,47	0,68
	Proanthocyanins_10	0,75	1,77	2,91	3,64
	Color_intensity_10	0,04	2,3	6	8,01
	Hue_10	0,21	0,69	1,42	1,93
	OD280/OD315_10	2,07	2,83	3,69	4,27
11	Proline_11	571,48	625	680	738,52
	Alcohol_11	10,34	11,45	12,37	13,28
	Malic_acid_11	-0,5	1,07	2,4	3,71
	Ash_11	1,51	2,1	2,7	3,31
	Alcalinity_of_ash_11	16,9	18,5	20	22,1
	Magnesium_11	80,18	88	96	105,16
	Total_phenols_11	1,82	2,74	3,52	4,29
	Flavanoids_11	1,59	2,79	3,75	4,72
	Nonflavanoid_phenols_11	0,16	0,24	0,32	0,4
	Proanthocyanins_11	1,04	1,83	2,49	3,14
	Color_intensity_11	0,64	2,65	4,5	6,3
	Hue_11	0,57	0,8	1,04	1,3
	OD280/OD315_11	2,16	2,77	3,39	4,12

Tabela 11 – *Continuação*

Hipercaixa	Hiperplano	v1: $\mu(0)$	v2: $\mu(1)$	v3: $\mu(1)$	v4: $\mu(0)$
12	Proline_12	128,89	315	680	870,11
	Alcohol_12	10,24	11,64	13,49	14,23
	Malic_acid_12	-0,59	1,51	4,3	5,58
	Ash_12	2,04	2,23	2,46	2,63
	Alcalinity_of_ash_12	15,72	18	24	27,48
	Magnesium_12	55,76	80	112	123,9
	Total_phenols_12	1,11	1,45	2,1	2,48
	Flavanoids_12	0,83	1,25	1,84	2,4
	Nonflavanoid_phenols_12	0,08	0,27	0,66	0,84
	Proanthocyanins_12	0,53	0,95	1,63	2,05
	Color_intensity_12	1,47	2,6	3,74	4,56
	Hue_12	0,65	0,79	1,05	1,23
	OD280/OD315_12	1,92	2,52	3,3	3,61
13	Proline_13	174,37	312	672	800,03
	Alcohol_13	11,28	11,82	12,77	13,32
	Malic_acid_13	-0,77	0,94	3,43	4,46
	Ash_13	1,3	1,36	2,12	2,54
	Alcalinity_of_ash_13	9,04	10,6	20,8	26,9
	Magnesium_13	69,89	80	100	104,51
	Total_phenols_13	0,99	1,45	2,5	2,97
	Flavanoids_13	0,23	0,57	2,04	2,86
	Nonflavanoid_phenols_13	0,14	0,28	0,53	0,59
	Proanthocyanins_13	-0,29	0,42	2,08	2,93
	Color_intensity_13	-0,94	1,28	5,75	6,51
	Hue_13	0,63	0,7	1,05	1,27
	OD280/OD315_13	0,68	1,59	3,57	4,69

Tabela 11 – *Continuação*

Hipercaixa	Hiperplano	v1: $\mu(0)$	v2: $\mu(1)$	v3: $\mu(1)$	v4: $\mu(0)$
14	Proline_14	96,66	488	870	1150,54
	Alcohol_14	11,63	12,29	13,34	14,01
	Malic_acid_14	-1,42	0,94	3,86	5,29
	Ash_14	2,11	2,21	2,36	2,46
	Alcalinity_of_ash_14	13,87	17	22,5	27,37
	Magnesium_14	52,68	80	110	132,12
	Total_phenols_14	-0,04	1,1	2,53	3,26
	Flavanoids_14	0,58	1,02	1,76	2,27
	Nonflavanoid_phenols_14	0,23	0,37	0,61	0,81
	Proanthocyanins_14	-0,22	0,42	1,63	2,92
	Color_intensity_14	1,23	3,05	4,8	5,72
	Hue_14	0,71	0,84	1,02	1,13
	OD280/OD315_14	1,37	1,82	2,42	2,73
15	Proline_15	295,44	470	880	986,74
	Alcohol_15	11,67	12,36	14,34	14,9
	Malic_acid_15	0,2	1,29	5,19	6,28
	Ash_15	1,89	2,1	2,86	3,04
	Alcalinity_of_ash_15	15,23	18,5	27	28,6
	Magnesium_15	63,88	86	123	136,12
	Total_phenols_15	0,64	1,35	2,8	2,91
	Flavanoids_15	-0,01	0,51	1,57	1,74
	Nonflavanoid_phenols_15	0,12	0,22	0,63	0,79
	Proanthocyanins_15	0,03	0,64	2,7	2,71
	Color_intensity_15	5,16	7,5	13	13,6
	Hue_15	0,37	0,48	0,85	0,85
	OD280/OD315_15	1,14	1,3	1,96	2,14



Tabela 11 – *Continuação*

Hipercaixa	Hiperplano	v1: $\mu(0)$	v2: $\mu(1)$	v3: $\mu(1)$	v4: $\mu(0)$
16	Proline_16	272,05	425	780	919,77
	Alcohol_16	11,62	12,2	13,73	14,57
	Malic_acid_16	0,24	1,24	4,36	5,65
	Ash_16	1,91	2,19	2,75	2,91
	Alcalinity_of_ash_16	13,98	17,5	25,5	28,38
	Magnesium_16	68,47	85	116	127,71
	Total_phenols_16	0,56	1,15	2,2	2,62
	Flavanoids_16	-0,21	0,47	1,28	1,67
	Nonflavanoid_phenols_16	0,02	0,24	0,6	0,85
	Proanthocyanins_16	0,11	0,64	1,56	1,9
	Color_intensity_16	3,92	5,4	7,3	8,05
	Hue_16	0,44	0,61	0,96	1,05
	OD280/OD315_16	0,69	1,33	2,47	2,76
17	Proline_17	551,36	600	630	678,64
	Alcohol_17	12,44	12,7	12,86	13,12
	Malic_acid_17	-2,22	1,35	3,55	7,12
	Ash_17	2,26	2,32	2,36	2,42
	Alcalinity_of_ash_17	12,33	18	21,5	27,17
	Magnesium_17	80,06	106	122	147,94
	Total_phenols_17	1,2	1,51	1,7	2,01
	Flavanoids_17	1,12	1,2	1,25	1,33
	Nonflavanoid_phenols_17	0,11	0,17	0,21	0,27
	Proanthocyanins_17	0,68	0,84	0,94	1,1
	Color_intensity_17	2,64	4,1	5	6,46
	Hue_17	0,73	0,76	0,78	0,81
	OD280/OD315_17	1,29	1,29	1,29	1,29

Tabela 11 – *Continuação*

Hipercaixa	Hiperplano	v1: $\mu(0)$	v2: $\mu(1)$	v3: $\mu(1)$	v4: $\mu(0)$
18	Proline_18	182,81	415	830	1014,33
	Alcohol_18	11,34	12,25	13,88	14,7
	Malic_acid_18	0,67	2,81	5,51	7,27
	Ash_18	1,82	2,15	2,72	3,2
	Alcalinity_of_ash_18	16,48	20	25	27,95
	Magnesium_18	71,53	80	102	117,33
	Total_phenols_18	0,35	0,98	2,32	2,81
	Flavanoids_18	0,24	0,34	0,6	0,78
	Nonflavanoid_phenols_18	0,28	0,4	0,63	0,73
	Proanthocyanins_18	0,08	0,55	1,3	1,63
	Color_intensity_18	3,13	3,85	5	5,91
	Hue_18	0,38	0,58	0,89	1,14
	OD280/OD315_18	0,57	1,27	2,31	2,99
19	Proline_19	347,33	550	675	877,67
	Alcohol_19	11,89	12,96	13,62	14,69
	Malic_acid_19	1,02	3,45	4,95	7,38
	Ash_19	2,35	2,35	2,35	2,35
	Alcalinity_of_ash_19	16,07	18,5	20	22,43
	Magnesium_19	69,3	92	106	128,7
	Total_phenols_19	0,4	1,39	2	2,99
	Flavanoids_19	0,54	0,7	0,8	0,96
	Nonflavanoid_phenols_19	0,29	0,4	0,47	0,58
	Proanthocyanins_19	0,81	0,94	1,02	1,15
	Color_intensity_19	2,97	4,4	5,28	6,71
	Hue_19	0,31	0,68	0,91	1,28
	OD280/OD315_19	1,26	1,75	2,05	2,54

## Regras Difusas

A Tabela 12 apresenta as regras difusas obtidas pela técnica desenvolvida. As regras geradas são utilizadas para classificar os novos casos, desconhecidos ao modelo, dentre as classes existentes. Cada atributo de um novo caso desconhecido é comparado ao seu respectivo valor em cada uma das regras difusas obtidas. Por exemplo o atributo *Proline* de um novo caso é comparado ao valor *Proline\_1* da Regra 1, ao valor *Proline\_2* da Regra 2, e assim por diante. *Proline\_1* é representado por uma função de pertinência, que por sua vez é representada por quatro valores (exibidos na Tabela 11), logo se o valor deste atributo no novo caso estiver dentre os valores da função de pertinência, este atributo possui alguma representatividade na classe que a regra representa. A análise é feita para cada atributo em todas as regras a fim de obter um grau de pertinência do novo caso em cada uma das classes existentes.

Tabela 12: Regras difusas da base Wine

<b>Regra difusa</b>	
Regra 1	<i>Se</i> (Proline $\acute{E}$ Proline <sub>1</sub> ) <i>E</i> (Alcohol $\acute{E}$ Alcohol <sub>1</sub> ) <i>E</i> (Malic_acid $\acute{E}$ Malic_acid <sub>1</sub> ) <i>E</i> (Ash $\acute{E}$ Ash <sub>1</sub> ) <i>E</i> (Alcalinity_of_ash $\acute{E}$ Alcalinity_of_ash <sub>1</sub> ) <i>E</i> (Magnesium $\acute{E}$ Magnesium <sub>1</sub> ) <i>E</i> (Total_phenols $\acute{E}$ Total_phenols <sub>1</sub> ) <i>E</i> (Flavanoids $\acute{E}$ Flavanoids <sub>1</sub> ) <i>E</i> (Nonflavanoid_phenols $\acute{E}$ Nonflavanoid_phenols <sub>1</sub> ) <i>E</i> (Proanthocyanins $\acute{E}$ Proanthocyanins <sub>1</sub> ) <i>E</i> (Color_intensity $\acute{E}$ Color_intensity <sub>1</sub> ) <i>E</i> (Hue $\acute{E}$ Hue <sub>1</sub> ) <i>E</i> (OD280/OD315 $\acute{E}$ OD280/OD315 <sub>1</sub> ) <i>Ent\~{a}o</i> (Classe $\acute{E}$ 1)

<b>Regra difusa</b>	
Regra 2	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>2</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>2</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>2</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>2</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>2</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>2</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>2</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>2</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>2</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>2</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>2</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>2</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>2</sub>)  <i>Ent\~ao</i> (Classe <math>\acute{E}</math> 1)</p>
Regra 3	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>3</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>3</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>3</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>3</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>3</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>3</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>3</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>3</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>3</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>3</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>3</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>3</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>3</sub>)  <i>Ent\~ao</i> (Classe <math>\acute{E}</math> 1)</p>

<b>Regra difusa</b>	
Regra 4	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>4</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>4</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>4</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>4</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>4</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>4</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>4</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>4</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>4</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>4</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>4</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>4</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>4</sub>)  <i>Ent\~ao</i> (Classe <math>\acute{E}</math> 1)</p>
Regra 5	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>5</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>5</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>5</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>5</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>5</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>5</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>5</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>5</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>5</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>5</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>5</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>5</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>5</sub>)  <i>Ent\~ao</i> (Classe <math>\acute{E}</math> 2)</p>

<b>Regra difusa</b>	
Regra 6	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>6</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>6</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>6</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>6</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>6</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>6</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>6</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>6</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>6</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>6</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>6</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>6</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>6</sub>)  <i>Ent\~{a}o</i> (Classe <math>\acute{E}</math> 2)</p>
Regra 7	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>7</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>7</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>7</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>7</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>7</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>7</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>7</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>7</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>7</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>7</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>7</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>7</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>7</sub>)  <i>Ent\~{a}o</i> (Classe <math>\acute{E}</math> 2)</p>

<b>Regra difusa</b>	
Regra 8	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>8</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>8</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>8</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>8</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>8</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>8</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>8</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>8</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>8</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>8</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>8</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>8</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>8</sub>)  <i>Ent\~ao</i> (Classe <math>\acute{E}</math> 2)</p>
Regra 9	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>9</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>9</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>9</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>9</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>9</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>9</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>9</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>9</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>9</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>9</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>9</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>9</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>9</sub>)  <i>Ent\~ao</i> (Classe <math>\acute{E}</math> 2)</p>



<b>Regra difusa</b>	
Regra 10	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>10</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>10</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>10</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>10</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>10</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>10</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>10</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>10</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>10</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>10</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>10</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>10</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>10</sub>)  <i>Então</i> (Classe <math>\acute{E}</math> 2)</p>
Regra 11	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>11</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>11</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>11</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>11</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>11</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>11</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>11</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>11</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>11</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>11</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>11</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>11</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>11</sub>)  <i>Então</i> (Classe <math>\acute{E}</math> 2)</p>

<b>Regra difusa</b>	
Regra 12	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>12</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>12</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>12</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>12</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>12</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>12</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>12</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>12</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>12</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>12</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>12</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>12</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>12</sub>)  <i>Ent\~ao</i> (Classe <math>\acute{E}</math> 2)</p>
Regra 13	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>13</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>13</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>13</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>13</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>13</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>13</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>13</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>13</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>13</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>13</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>13</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>13</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>13</sub>)  <i>Ent\~ao</i> (Classe <math>\acute{E}</math> 2)</p>

<b>Regra difusa</b>	
Regra 14	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>14</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>14</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>14</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>14</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>14</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>14</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>14</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>14</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>14</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>14</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>14</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>14</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>14</sub>)  <i>Então</i> (Classe <math>\acute{E}</math> 2)</p>
Regra 15	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>15</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>15</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>15</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>15</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>15</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>15</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>15</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>15</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>15</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>15</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>15</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>15</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>15</sub>)  <i>Então</i> (Classe <math>\acute{E}</math> 3)</p>

<b>Regra difusa</b>	
Regra 16	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>16</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>16</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>16</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>16</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>16</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>16</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>16</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>16</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>16</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>16</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>16</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>16</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>16</sub>)  <i>Ent\~{a}o</i> (Classe <math>\acute{E}</math> 3)</p>
Regra 17	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline<sub>17</sub>)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol<sub>17</sub>)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid<sub>17</sub>)  <i>E</i> (Ash <math>\acute{E}</math> Ash<sub>17</sub>)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash<sub>17</sub>)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium<sub>17</sub>)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols<sub>17</sub>)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids<sub>17</sub>)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols<sub>17</sub>)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins<sub>17</sub>)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity<sub>17</sub>)  <i>E</i> (Hue <math>\acute{E}</math> Hue<sub>17</sub>)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315<sub>17</sub>)  <i>Ent\~{a}o</i> (Classe <math>\acute{E}</math> 3)</p>

<b>Regra difusa</b>	
Regra 18	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline_18) <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol_18)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid_18)  <i>E</i> (Ash <math>\acute{E}</math> Ash_18)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash_18)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium_18)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols_18)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids_18)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols_18)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins_18)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity_18)  <i>E</i> (Hue <math>\acute{E}</math> Hue_18)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315_18)  <i>Ent\~{a}o</i> (Classe <math>\acute{E}</math> 3)</p>
Regra 19	<p><i>Se</i> (Proline <math>\acute{E}</math> Proline_19)  <i>E</i> (Alcohol <math>\acute{E}</math> Alcohol_19)  <i>E</i> (Malic_acid <math>\acute{E}</math> Malic_acid_19)  <i>E</i> (Ash <math>\acute{E}</math> Ash_19)  <i>E</i> (Alcalinity_of_ash <math>\acute{E}</math> Alcalinity_of_ash_19)  <i>E</i> (Magnesium <math>\acute{E}</math> Magnesium_19)  <i>E</i> (Total_phenols <math>\acute{E}</math> Total_phenols_19)  <i>E</i> (Flavanoids <math>\acute{E}</math> Flavanoids_19)  <i>E</i> (Nonflavanoid_phenols <math>\acute{E}</math> Nonflavanoid_phenols_19)  <i>E</i> (Proanthocyanins <math>\acute{E}</math> Proanthocyanins_19)  <i>E</i> (Color_intensity <math>\acute{E}</math> Color_intensity_19)  <i>E</i> (Hue <math>\acute{E}</math> Hue_19)  <i>E</i> (OD280/OD315 <math>\acute{E}</math> OD280/OD315_19)  <i>Ent\~{a}o</i> (Classe <math>\acute{E}</math> 3)</p>