

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Francisco Felipe Moreira Sousa

**Integração do Mecanismo de Classificação com Opção de Rejeição
em Redes Neurais ARTMAP**

Fortaleza – CE

20 de fevereiro de 2020

Francisco Felipe Moreira Sousa

Integração do Mecanismo de Classificação com Opção de Rejeição em Redes Neurais ARTMAP

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Linha de Pesquisa: Inteligência Artificial

Orientador: Ajalmar Rêgo da Rocha Neto

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Fortaleza – CE

20 de fevereiro de 2020

Dados Internacionais de Catalogação na Publicação
Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

- M835i Moreira Sousa, Francisco Felipe.
 Integração do Mecanismo de Classificação com Opção de Rejeição em Redes Neurais ARTMAP / Francisco Felipe Moreira Sousa. - 2020.
 62 f. : il. color.
- Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Ciência da Computação, Campus Fortaleza, 2020.
 Orientação: Prof. Dr. Ajalmar Rêgo da Rocha Neto.
1. Redes Neurais. 2. ARTMAP. 3. Classificação com Opção de Rejeição. 4. Teoria da Ressonância Adaptativa. I. Título.

CDD 004



Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)
Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Francisco Felipe Moreira Sousa

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, sendo aprovada pela Coordenação do Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará e pela banca examinadora:

Orientador: Ajalmar Rêgo da Rocha Neto

Instituto Federal de Educação, Ciência e Tecnologia do
Ceará (IFCE)

João Fernando Lima Alcântara

Universidade Federal do Ceará (UFC)

**Carlos Maurício Jaborandy de Mattos Dourado
Júnior**

Instituto Federal de Educação, Ciência e Tecnologia do
Ceará (IFCE)

Fortaleza – CE
20 de fevereiro de 2020

*Dedico este trabalho à minha mãe, Francisca Elenice Sousa Moreira,
pelo o apoio incondicional ao longo deste período.*

Agradecimentos

Primeiramente gostaria de agradecer a minha família pelo o total compromisso e carinho que essas pessoas tiveram pela a minha causa quando ingressei no mestrado. Em especial a minha mãe, Elenice Sousa, por ter me dado a vida e dedicou grande parte de seu tempo em facilitar o meu caminho até aqui. Minha Tia, Maria Algediva, por me permitir estar em sua casa durante todo o período do mestrado. As minhas tias, Francisca e Leonice, que durante o início do mestrado pagaram as minhas despesas até que eu tivesse as condições de me manter. Desta forma, pode-se ver que se este trabalho dependesse apenas de mim, ele não estaria nem perto de ser completo. Em virtude disto, meus primeiros agradecimentos vão para as pessoas.

Também devo dizer que agradeço ao Madson Dias, por me ajudar a entrar no mestrado e ao Lucas Matias pela a produção do artigo e da dissertação. Meus agradecimentos também são direcionados para os meus amigos Alberth e Átilla, por serem parceiros de mestrado, de modo que sempre me ajudaram nas dúvidas que surgiam.

Meus agradecimentos também se estendem ao Professor Doutor Ajalmar Rêgo da Rocha Neto, pela sua capacidade de orientar com tanta destreza e apoio durante esses dois anos que foram cruciais para o meu desenvolvimento.

*“O que sabemos é uma gota,
o que ignoramos é um oceano.”
(Isaac Newton)*

Resumo

As redes ARTMAP são técnicas de aprendizado de máquinas voltadas para o aprendizado supervisionado, sendo conhecidas principalmente pela sua capacidade de aprendizado rápido, estável, incremental e online. Outra propriedade das redes ARTMAP é a generalização de problemas de classificação através do uso de formas geométricas dentro de diferentes domínios. No entanto, embora as redes ARTMAP tenham suas vantagens no que concerne ao treinamento, e possuam uma gama de variações que possibilite formas geométricas de generalização diferentes para problemas de classificação, sua taxa de acerto no processo de predição pode ser muito baixa para problemas de classificação muito complexos. Diante dessa questão, a classificação com opção de rejeição se faz oportuna, uma vez que a mesma provê uma forma de trazer maior confiabilidade para um classificador no momento da predição. Tal conceito tem sido extensivamente estudado na literatura de aprendizado de máquinas, de modo que é possível encontrar trabalhos envolvendo classificadores como SVM, MLP e SOM integrados com algum tipo de opção de rejeição, o que acarreta em uma maior taxa de acerto. Contudo, apesar de haver um extensivo estudo voltado para essa área, a classe de redes neurais ARTMAP ainda não foi explorada dentro deste conceito. Neste sentido, é proposto neste trabalho o estudo da adição de classificação com opção de rejeição em redes ARTMAP, deste modo, uma nova família de redes neurais capazes de aprendizado online no contexto de classificação com opção de rejeição.

Palavras-chave: Redes Neurais, Teoria da Ressonância Adaptativa, ARTMAP, Classificação com Opção de Rejeição.

Abstract

ARTMAP networks are machine learning techniques geared towards supervised learning, and are known primarily for their fast, stable, incremental and online learning capabilities. Another property of ARTMAP networks is the generalization of classification problems through the use of geometric shapes within different domains. However, while ARTMAP networks have their advantages in training, and have a range of variations that allow different generalization geometric shapes for classification problems, their prediction process hit rate can be very low for very complex classification problems. Given this issue, rejection classification is timely, as it provides a way to bring greater reliability to a classifier at the time of prediction. Such a concept has been extensively studied in the machine learning literature, so it is possible to find works involving classifiers such as SVM, MLP and SOM integrated with some rejection option mechanism, which results in a higher hit rate. However, although there is an extensive study focused on this area, the ARTMAP neural networks class has not yet been explored within this context. In this sense, it is proposed in this paper to study the addition of rejection classification in ARTMAP networks, thus inserting a new family of neural networks capable of online learning in the context of rejection classification.

Keywords: Neural Networks, Adaptive Resonance Theory, Supervised Learning, ARTMAP, Rejection Option Classification.

Lista de ilustrações

Figura 1	– Dado o vetor de entrada a M -dimensional, F_0 obtém a codificação do complemento $\mathbf{I}$ $2M$ -dimensional. O vetor $2M$ -dimensional $\mathbf{I}$ é o vetor de entrada para o campo F_1 . O campo F_1 envia os sinais de ativação de cada categoria $\mathbf{w}_j$ para F_2 . O campo F_2 seleciona a categoria $\mathbf{w}_j$ com o maior valor de ativação e envia a expectativa para o campo F_1 . O campo F_1 obtém sua saída $\mathbf{x}$ e calcula a taxa de equivalência. O critério de vigilância pode levar a um <i>reset</i> , fazendo com que o processo de busca procure por uma nova categoria em F_2	23
Figura 2	– Representação dos vetores de peso da rede <i>Fuzzy</i> ART em codificação complementar com $M = 2$	26
Figura 3	– Arquitetura da rede <i>Fuzzy</i> ARTMAP	27
Figura 4	– Categorias da rede <i>Fuzzy</i> ARTMAP obtidas de um exemplo: (a) determinada distribuição de dados para um problema binário e (b) as categorias criadas pela a rede.	29
Figura 5	– Superfície de decisão obtida pela rede <i>Fuzzy</i> ARTMAP.	30
Figura 6	– Ilustração dos hiper-retângulos R_1 e R_2	31
Figura 7	– Ilustração dos hiper-retângulos R_1 , R_2 e R_3	32
Figura 8	– Ilustração das regiões de rejeição (V_R) e aceitação (V_A) da regra ótima	35
Figura 9	– Ilustração da curva de compromisso entre o erro e a rejeição.	36
Figura 10	– Ilustração de uma curva hipotética de compromisso entre o erro e a rejeição.	37
Figura 11	– Aplicação da regra de Chow as probabilidades a posteriori "reais" e "estimadas".	38
Figura 12	– Dois diferentes limiares de rejeição T_1 e T_2 aplicados as probabilidades a posteriori estimadas da tarefa de classificação da figura 11	38
Figura 13	– Região de decisão obtida utilizando um classificador independentes para um problema binário.	40
Figura 14	– Região de decisão obtida utilizando dois classificadores independentes para um problema binário.	41
Figura 15	– (a) Conjunto de dados artificial; (b) Categorias criadas pela a rede FAM; Categorias de rejeição criadas usando um valor de limiar de (c) 0.15 e (d) 0.45	44
Figura 16	– Categorias de rejeição usando um valor de limiar de 0.15 (a, c e e) e de 0.45 (b, d e f); Categorias de rejeição criadas usando um único limiar (a e b); Categorias de rejeição criadas para a classe C_{+1} (c e d) e para a classe C_{-1} (e e f)	45

Figura 17 – Conjunto de Dados Sintéticos	46
Figura 18 – Curvas A-R dos modelos FAM-C e FAM-F para cada base de dados artificiais.	50
Figura 19 – Categorias de rejeição do modelo FAM-C (a, c e e) com custo de rejeição w_r igual a (a, b) 0.04, (c, d) 0.24 e (e, f) 0.48 e (b, d e f) superfície de decisão com a região de rejeição em verde.	51
Figura 20 – Curvas A-R dos modelos propostos para os bancos de dados reais. . . .	54

Lista de símbolos

a	Vetor de entrada
b	Rótulo de entrada
x^c	Complemento de um vetor x arbitrário
F_0	Campo de entrada
F_1	Campo de característica
F_2	Campo de representação de categorias
ART_a	Módulo que aprende os padrões de entrada
ART_b	Módulo que aprende os rótulos das classes
ρ	Parâmetro de vigilância
ρ_a	Parâmetro de vigilância do módulo ART_a
ρ_b	Parâmetro de vigilância do módulo ART_b
$\bar{\rho}_a$	Parâmetro de vigilância base do módulo ART_a
ρ_{ab}	Parâmetro de vigilância do módulo inter-ART
ε	Parâmetro de <i>match tracking</i>
α	Parâmetro de escolha
β	Taxa de aprendizado
N_a	Número de categorias no módulo ART_a
N_b	Número de categorias no módulo ART_b
x	Ativação do campo F_1
y	Ativação do campo F_2
w_j	Vetor de pesos
w_J	Categoria ativa
T_j	Valor da ativação

I	Ativação do campo F_0
w_J^{new}	Categoria ativa atualizada
w_J^{old}	Categoria ativa antiga
$\mathbb{R}$	Espaço dos números reais
F_0^a	Campo de entrada do módulo ART_a
F_1^a	Campo de característica do modulo ART_a
F_2^a	Campo de representação de categorias do modulo ART_a
F_0^b	Campo de entrada do módulo ART_b
F_1^b	Campo de característica do modulo ART_b
F_2^b	Campo de representação de categorias do modulo ART_b
I^a	Ativação do campo F_0^a
x^a	Ativação do campo F_1^a
y^a	Ativação do campo F_2^a
w_j^a	Vetor de pesos do módulo ART_a
I^b	Ativação do campo F_0^b
x^b	Ativação do campo F_1^b
y^b	Ativação do campo F_2^b
w_k^b	Vetor de pesos do módulo ART_b
x^{ab}	Ativação do módulo inter-ART
w^{ab}	Campo de mapeamento
t	Limiar de rejeição
W_r	Custo de rejeição
W_c	Custo de acertar
W_e	Custo de errar
$E(t)$	Função de erro
$R(t)$	Função de rejeição

Sumário

1	INTRODUÇÃO	15
1.1	Motivação para a Dissertação	17
1.2	Descrição do Problema	18
1.3	Objetivos Geral e Específicos	19
1.4	Produção Científica	19
1.5	Estrutura da Dissertação	19
2	<i>FUZZY ART E FUZZY ARTMAP</i>	21
2.1	Introdução	21
2.2	<i>Fuzzy ART</i>	22
2.2.1	Função de Escolha	24
2.2.2	Reinício ou Ressonância	24
2.2.3	Aprendizado	25
2.2.4	Normalização do Vetor de Entrada	25
2.2.5	Interpretação Geométrica	26
2.3	<i>Fuzzy ARTMAP</i>	27
2.3.1	Mecanismo de <i>Match Tracking</i>	28
2.3.2	Treinamento da Rede <i>Fuzzy ARTMAP</i>	28
2.3.3	Predição do <i>Fuzzy ARTMAP</i>	30
2.3.4	Proliferação de Categorias	30
3	OPÇÃO DE REJEIÇÃO	33
3.1	Introdução	33
3.2	Regra de Decisão Ótima	34
3.3	Multiplos Limiares	37
3.4	Abordagem de Classificação Binária com Opção de Rejeição	39
3.4.1	Um Classificador Padrão	39
3.4.2	Dois Classificadores Independentes	40
3.4.3	Classificador com Opção de Rejeição Embutida	41
3.5	Proposta para Classificação com Opção de Rejeição	41
3.5.1	Rede Neural PROBART	41
3.5.2	Usando um Único Limiar (Proposta 1)	42
3.5.3	Usando um Limiar Para Cada Classe (Proposta 2)	44
4	SIMULAÇÕES COMPUTACIONAIS	46
4.1	Metodologia	46

4.2	Resultados e Discussões	48
5	CONCLUSÃO	58
5.1	Conclusões	58
5.2	Contribuições deste Trabalho	59
5.3	Trabalhos Futuros	59
	REFERÊNCIAS	60

1 Introdução

A tarefa de reconhecimento de padrões está intimamente relacionada com a descoberta de regularidades em dados através do uso de algoritmos computacionais, utilizando essas regularidades na tomada de decisões (BISHOP, 2006). Nesse tipo de tarefa, um padrão pode ser uma imagem, um sinal de onda ou qualquer tipo de medida que necessite ser descrita ou classificada (THEODORIDIS; KOUTROUMBAS, 2008). Segundo Jain, Duin e Mao (2000), essa tarefa pode ser realizada de diversas formas, sendo uma delas usando as redes neurais artificiais.

As redes neurais artificiais são um tipo de abordagem que utiliza um modelo de grafo orientado disposto em rede, onde os nós representam neurônios artificiais e as arestas orientadas são as conexões entre as entradas e as saídas dos neurônios. O processo de aprendizado envolve a atualização da arquitetura da rede e dos pesos das conexões (JAIN; DUIN; MAO, 2000). Essa atualização é realizada por meio de exemplos obtidos do mundo real definidos como um conjunto de características. Se esse conjunto for previamente rotulado, a aprendizagem é dito supervisionada, caso contrário, é dito não supervisionada.

A Teoria da Ressonância Adaptativa (do inglês, *Adaptive Resonance Theory* - ART), de Grossberg (GROSSBERG, 1976; GROSSBERG, 1987; GROSSBERG, 2013) inspirou novas arquiteturas de redes neurais artificiais para aprendizado supervisionado e não supervisionado. Essa teoria busca descrever como o cérebro, em uma escala de aprendizado rápido e estável, e de maneira totalmente autônoma, é capaz de codificar, reconhecer e prever objetos e eventos em um ambiente que se reorganiza constantemente através do tempo. O *dilema da estabilidade plasticidade* é a principal propriedade da teoria da ressonância adaptativa, no qual descreve uma ligação intrínseca entre os conceitos de estabilidade e plasticidade em um sistema de aprendizado (GROSSBERG, 1987; CARPENTER; GROSSBERG, 1987b). Considerando um sistema de aprendizado estável, o mesmo deve possuir mecanismos internos capazes de amenizar possíveis fontes de instabilidades. Por outro lado, a plasticidade do sistema de aprendizado traz consigo a capacidade de responder de forma adaptativa a qualquer nova excitação decorrente do ambiente exterior.

No contexto da ART, as redes neurais não supervisionadas foram os primeiros modelos a serem introduzidos na literatura. Como por exemplo, o ART 1 (CARPENTER; GROSSBERG, 1987b), ART 2 (CARPENTER; GROSSBERG, 1987a), ART 3 (CARPENTER; GROSSBERG, 1990) e o conhecido *Fuzzy ART* (CARPENTER; GROSSBERG; ROSEN, 1991) para tarefas de aprendizado não supervisionados. Em poucas palavras, o ART 1 auto-estabiliza códigos de reconhecimento em resposta a padrões arbitrários de entrada binária, enquanto o ART 2 é capaz de aprender de forma estável a categorizar

dados binários e analógicos. O ART 3 possui algumas características da rede ART 2, mas difere pelo fato de incorporar as características de um sistema ART em uma rede hierárquica. Por outro lado, o *Fuzzy ART* é uma arquitetura que utiliza operações *fuzzy* em sua concepção e é capaz de processar padrões multidimensionais, sejam eles contínuos ou binários.

A abordagem de aprendizado supervisionado, baseada na teoria de Grossberg, foi introduzida na arquitetura ARTMAP (CARPENTER; GROSSBERG; REYNOLDS, 1991). Para resolver esse tipo de tarefa, o ARTMAP usa dois módulos ART 1, conhecidos como ART_a (o módulo do lado A) e ART_b (o módulo do lado B) para dados binários. Junto com esses módulos, há um módulo inter ART responsável por mapear as categorias ART_a para as categorias ART_b em uma abordagem muitos-para-um. Como uma melhoria, foi proposto o *Fuzzy ARTMAP* para lidar com dados analógicos (CARPENTER et al., 1992).

O *Fuzzy ARTMAP* pode mapear um conjunto de dados multidimensionais arbitrário criando hiper retângulos para o padrão de entrada e o rótulo de entrada nos módulos ART_a e ART_b , respectivamente. Algumas vantagens do *Fuzzy ARTMAP* são notáveis, como aprendizado rápido e estável, a necessidade de poucas épocas para alcançar a estabilidade, a capacidade de aprender dados novos de maneira rápida e estável sem esquecer os dados passados, sendo essa uma característica de destaque em relação aos outros métodos de treinamento existentes na literatura de aprendizado de máquinas. Apesar dessas vantagens, o *Fuzzy ARTMAP* é muito sensível à ordem de apresentação dos padrões de entrada e aos dados de ruído.

Sabendo que o treinamento da rede *Fuzzy ARTMAP* é baseado em uma forma de aprendizado biológico, a rede em si herda uma característica de sobreajuste sobre o problema a ser aprendido pelo sistema. Como se sabe, o sobreajuste nos dados de treinamento leva o sistema a ter uma generalização ineficiente para os padrões de entrada que ainda são desconhecidos. Uma forma de diminuir a taxa de erro de tal sistema pode ser alcançada através da adição de um mecanismo de classificação com opção de rejeição.

Devido a incertezas e ruídos inerentes a qualquer tarefa de reconhecimento de padrões, os erros geralmente são inevitáveis. A opção de rejeição é introduzida para proteger o sistema contra o excesso de reconhecimento (CHOW, 1957). O sistema converte o possível erro de reconhecimento em rejeição. No entanto, a troca entre os erros e as rejeições é raramente de um para um. Sempre que a opção de rejeição é exercida, alguns possíveis reconhecimentos corretos também são convertidos em rejeições.

Um sistema de aprendizado que possua um mecanismo de opção de rejeição pode evitar tomar alguma decisão em relação a algum padrão de aspecto ambíguo, sendo posteriormente examinado de maneira minuciosa por algum especialista na área de interesse. Desse modo, inserir a opção de rejeição em tal sistema é uma forma de resguardar contra erros que ocorrem excessivamente diante de decisões difíceis a serem tomadas.

1.1 Motivação para a Dissertação

O problema de performance degradada para problemas de classificação complexos devido ao sobreajuste de redes ARTMAP foi estudado e assim novas arquiteturas foram propostas, como a proposição da arquitetura ARTMAP-IC (CARPENTER; MARKUZON, 1998), em que a predição é realizada de forma distribuída por parte dos neurônios da rede. Em 2014, foi proposta uma modificação na rede Fuzzy ARTMAP chamada de *Threshold and Posterior Probability Fuzzy ARTMAP* (TPPFAM) (ZHANG; JI; ZHANG, 2014), que utiliza a predição distribuída e probabilidade a posteriori para se obter maior eficácia durante a predição de padrões desconhecidos.

Apesar das arquiteturas ARTMAP-IC e TPPFAM oferecerem maior taxa de acertos para determinados problemas de classificação, a taxa de erro ainda pode ser alta para problemas de classificação muito complexos. Desse modo, o uso de mecanismo de classificação com opção de rejeição se faz proveitoso. A classificação com opção de rejeição teve início de fato com o trabalho de Chow (CHOW, 1957), envolvendo a classificação de caracteres. Chow descreve então de forma mais completa o problema de classificação com opção de rejeição em seu trabalho de 1970 (CHOW, 1970).

A opção de rejeição é inserida para proteger o sistema contra erros excessivos frutos das tomadas de decisões difíceis. Assim, o sistema deve converter potenciais erros em rejeição. Desta forma, pode-se atingir uma maior eficácia na aprendizagem, definindo a região de sobreposição como uma nova classe chamada de classe de rejeição.

Considerando um problema binário, é atraente o desenvolvimento de classificadores capazes de criar uma terceira classe, a classe de rejeição, pois a principal vantagem dessa técnica é a proteção do sistema contra um eventual número de decisões erradas em situações mais difíceis nas quais se exige uma alta confiabilidade, ou seja, os padrões que são atribuídos à classe de rejeição não são classificados e são deixados para um posterior exame.

A classificação com opção de rejeição, em uma definição resumida, é um conjunto de técnicas que buscam melhorar a confiabilidade de classificação de sistemas de aprendizado. Esse aumento considerável do desempenho dos classificadores com mecanismos de rejeição foi demonstrado no trabalho de Friedel, Ruckert e Krameret (FRIEDEL; RUCKERT; KRAMER, 2006). Diante disso, classificadores como MLP, SVM, LVQ e SOM tiveram suas propriedades exploradas em uma integração com a opção de rejeição no momento de predição.

Proposições do uso de classificação com opção de rejeição em redes MLP podem ser encontrados nos trabalhos (CORDELLA et al., 1995b; FUMERA; ROLI; GIACINTO, 2000; LIU; SAKO; FUJISAWA, 2002). Para o classificador SVM, Fumera e Roli incorporaram a capacidade de rejeitar no processo de treinamento (FUMERA; ROLI, 2002).

Outros trabalhos que utilizam o SVM com classificação com opção de rejeição podem ser encontrados nos trabalhos de Sousa et al. (SOUSA; MORA; CARDOSO, 2009a; SOUSA et al., 2015) e Rocha Neto et al. (NETO et al., 2011). Com relação à rede LVQ, a adição da opção de rejeição pode ser encontrada nos trabalhos de Cordella et al. (CORDELLA et al., 1995a), e Suutala et al. (SUUTALA et al., 2004). Por último, o trabalho de Sousa et al. (SOUSA et al., 2015) propõe o uso da rede SOM com a opção de rejeição para tarefas de classificação.

Embora a classificação com opção de rejeição seja um tópico recorrentemente explorado ao longo dos anos na literatura de aprendizado de máquina, não é possível encontrar nenhum trabalho que envolva a inserção de tal mecanismo em redes ARTMAP. Essa ausência pode ser explicada devido ao problema largamente conhecido dessa rede como o problema de proliferação de categorias. No qual a maioria dos trabalhos encontrados buscam amenizar esse problema.

Sabendo disso, nesta dissertação é proposta uma variação da rede *Fuzzy ARTMAP* para se comportar como um classificador com a opção de rejeição. A proposta se baseia em uma única rede *Fuzzy ARTMAP* com a modificação do campo de mapeamento de uma outra neural chamada PROBART. A ideia principal é criar uma região de rejeição, observando o local onde ocorre a proliferação das categorias. A construção dessa variação da rede *Fuzzy ARTMAP* é focada em problemas de classificação binária capazes de decidir, após o treinamento, se um categoria é ou não uma categoria de rejeição. Como se pode esperar, as categorias de rejeição não são categorias que mapeiam uma classe, mas categorias que retêm as amostras críticas e as rejeitam.

1.2 Descrição do Problema

A rede *Fuzzy ARTMAP* possui vantagens como a capacidade de aprendizagem rápido e estável, a necessidade de pouca época para convergir e atingir a estabilidade e a habilidade de aprender de maneira *online*, novos dados sem esquecer padrões já aprendidos anteriormente. Entretanto, essas vantagens trazem consigo a característica de que a rede *Fuzzy ARTMAP* aprende a relação que existe entre os dados e de forma excessiva, acarretando em sobreajuste nos dados de treinamento. Muitas vezes a capacidade de generalização da rede *Fuzzy ARTMAP* é deteriorada por conta dessa ocorrência de sobreajuste durante o treinamento.

Durante o treinamento da rede *Fuzzy ARTMAP* os erros preditivos tendem a ocorrer em bases de dados muito ruidosas, isto é, com sobreposição de uma ou mais classes. Esses erros preditivos são tratados por um mecanismo interno chamado mecanismo de *match tracking*, conhecido também como *reset* lateral, que determina uma quantidade mínima de generalização necessária para codificar o padrão de entrada sem que ele leve a

rede a cometer outro erro preditivo interno. Através desse mecanismo o foco do subsistema de atenção do módulo ART_a é restringido e direcionado para categorias com tamanhos menores que sejam capazes de codificar o novo padrão de entrada dado um novo valor do parâmetro de vigilância. Essa restrição no foco do sistema, recorrentemente, faz com que muitas categorias sejam criadas na região ruidosa dos dados e muitas vezes a capacidade de generalização da rede *Fuzzy ARTMAP* é deteriorada.

1.3 Objetivos Geral e Específicos

O objetivo principal dessa dissertação consiste em integrar a capacidade de classificação com opção de rejeição na rede neural *Fuzzy ARTMAP* e avaliar o seu desempenho em determinados problemas de classificação binária.

Em seu sentido mais particular, os seguintes objetivos específicos são

- Realizar um levantamento sobre a rede *Fuzzy ARTMAP* e os diferentes métodos na literatura que empreguem a classificação com opção de rejeição;
- Integrar métodos de classificação com opção de rejeição à rede *Fuzzy ARTMAP* sob estudo;
- Realizar testes de desempenho da rede *Fuzzy ARTMAP* com a opção de rejeição para diferentes problemas de classificação;
- Fornecer uma comparação de desempenho entre as redes *Fuzzy ARTMAP* com e sem a opção de rejeição;

1.4 Produção Científica

Durante a realização deste trabalho, o seguinte artigo foi produzido:

- **SOUSA, F. F. M.**; MATIAS, A. L. S.; NETO, A. R. da R. Classification with rejection option using the fuzzy artmap neural network. In: ROJAS, I.; JOYA, G.; CATALA, A.(Ed.).Advances in Computational Intelligence. Cham: Springer International Publishing, 2019. p. 568–578. ISBN 978-3-030-20518-8

1.5 Estrutura da Dissertação

Esta dissertação está organizada da seguinte maneira. No Capítulo 2, é primeiramente apresentado as dinâmicas e as equações que regem o funcionamento da rede *Fuzzy ART*, como a forma geométrica hiper retangular das categorias é alcançada através da codificação complementar. Na seção 2.2, a arquitetura da rede *Fuzzy ARTMAP* é descrita, explicando como os módulos ART_a e ART_b são conectados e como é o funcionamento do mecanismo de *match tracking*.

O Capítulo 3 trata da classificação com opção de rejeição. Neste capítulo são apresentados a regra de decisão ótima e o uso de múltiplos limiares para a escolha do limiar de rejeição. Também são apresentadas as abordagens para a classificação com opção de rejeição usando (i) um classificador padrão; (ii) dois classificadores independentes; e (iii) um classificador com opção de rejeição embutida. Ainda neste capítulo, são propostos dois métodos para fins de classificação com opções de rejeição baseada na rede *Fuzzy* ARTMAP. Ambas as propostas utilizam apenas uma única rede *Fuzzy* ARTMAP com a modificação do campo de mapeamento de uma outra neural chamada PROBART, porém a primeira proposta cria categorias de rejeições para todas as classes já a segunda cria categorias de rejeições específicas para cada classe.

Já no Capítulo 4, a metodologia utilizada, resultados de simulações computacionais e discursões referentes a estes são apresentados. Tabelas com métricas de desempenho obtidas a partir das simulações para os métodos propostos e para os métodos usados para a comparação são apresentadas, assim como gráficos relacionando o desempenho obtido.

No Capítulo 5, são apresentadas as considerações finais e uma conclusão geral, destacando as principais contribuições deixadas durante este trabalho de mestrado. Por fim, são alencadas sugestões para trabalhos futuros.

2 *Fuzzy* ART e *Fuzzy* ARTMAP

Este capítulo tem como objetivo detalhar o funcionamento das redes neurais *Fuzzy* ART e *Fuzzy* ARTMAP. Primeiramente na seção 2.1 é dada uma breve introdução sobre os subsistemas que estão presentes na rede. Na seção 2.2 é descrita a arquitetura da rede *Fuzzy* ART apresentando a sua composição dividida em campos e as suas equações seguindo as nomenclaturas convencionais da sua literatura. Na subseção 2.2.1 é apresentada a função de escolha, na subseção 2.2.2 é explicado a ressonância, na subseção 2.2.3 é descrito o processo de aprendizado, na subseção 2.2.4 é apresentado uma breve descrição da codificação complementar. E na subseção 2.2.5 é explicado como a rede, através da codificação complementar, utiliza as categorias com geometria hiper-retangular para generalizar o espaço de características. Na seção 2.3 é apresentada a arquitetura da rede *Fuzzy* ARTMAP, que é formada por dois módulos *Fuzzy* ART e um módulo inter-ART (campo de categoria). Na subseção 2.3.1 é apresentado a função de escolha, na subseção 2.3.2 é descrito o processo de aprendizado e a sua predição na subseção 2.3.3. E na subseção 2.3.4 é descrito o problema da proliferação de categorias que ocorre nas redes *Fuzzy* ARTMAP.

2.1 Introdução

A teoria da ressonância adaptativa descreve propriedades no cérebro que buscam caracterizar como o mesmo, em uma escala de aprendizado rápido e estável, e de maneira totalmente autônoma, é capaz de codificar, reconhecer e prever objetos e eventos em um ambiente que se reorganiza constantemente através do tempo. Do ponto de vista da teoria da ressonância adaptativa, a característica fundamental para sistemas que desejam simular o aprendizado biológico é a capacidade de alternar apropriadamente entre os estados plásticos e estável durante o aprendizado.

Considerando um sistema de aprendizado estável, o mesmo deve possuir mecanismos internos capazes de atenuar possíveis fontes de instabilidade, prevenindo que códigos internos já estabelecidos pelo sistema sejam afetados por entradas arbitrárias consideradas irrelevantes. Por outro lado, a plasticidade em um sistema de aprendizado traz consigo a capacidade de gerar códigos adaptativos em resposta a qualquer nova excitação provinda do ambiente exterior. Em outras palavras, o aprendizado do sistema deve ser capaz de responder a eventos significantes, e ainda assim, se manter estável quando eventos rotineiros, ou irrelevantes, ocorrem frequentemente.

Diante dessa questão, o uso de múltiplos subsistemas se faz necessário para que o sistema como um todo (rede neural) reaja de forma adaptativa a novos eventos externos

mantendo sua estabilidade interna. Desse modo, tem-se os subsistemas de atenção e orientação que permite a rede neural processar tanto padrões de entradas que são familiares, como padrões de entradas não familiares. Isso é possível devido o subsistema de atenção que é capaz de estabelecer representações internas em resposta a eventos que já foram codificados pela rede, enquanto o subsistema de orientação redireciona o foco do subsistema de atenção quando um evento ainda não conhecido pela rede ocorre.

Com essa dinâmica que busca por similaridades entre padrões externos e categorias internas, se torna possível distinguir padrões que são familiares de padrões que não são familiares à rede. Quando uma rede ART processa um evento que lhe é familiar, ocorre o que é chamado de estado de ressonância, um estado que surge quando há um consenso com relação ao evento ou objeto que está no mundo externo e que tipo de evento ou objeto é esperado internamente pelo sistema.

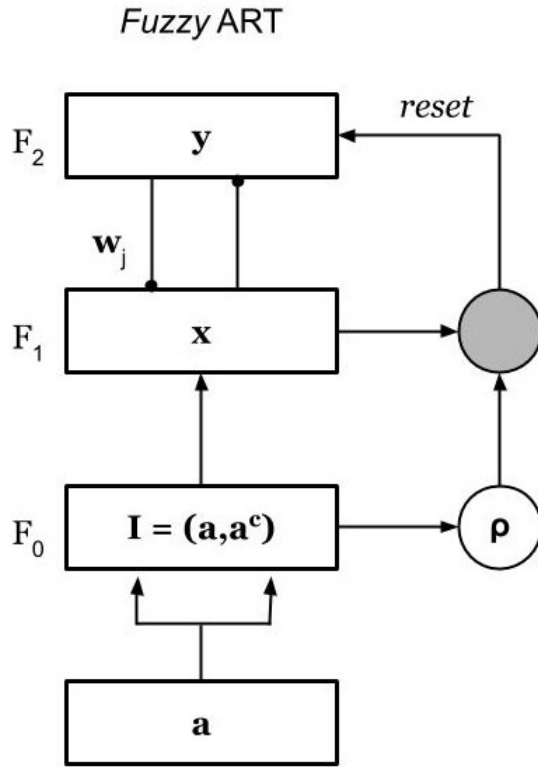
Em particular, um estado ressonante pode surgir através de um teste de hipótese interno responsável por verificar se a combinação entre a entrada e a expectativa é boa o suficiente. Se a resposta do teste for “sim”, então pelo menos uma categoria interna está familiarizada com o padrão externo e o estado de ressonância ocorre. No entanto, se a resposta do teste for “não”, o subsistema de orientação busca por uma outra categoria interna já estabelecida. Quando o sistema alcança o estado de ressonância, ele foca sua atenção em características críticas comuns ao padrão externo e às expectativas internas.

2.2 Fuzzy ART

A rede *Fuzzy ART* (Figura 1) é uma arquitetura que utiliza operações *fuzzy* em sua concepção e é composta por três camadas: F_0 , F_1 e F_2 . A camada F_0 é o campo de entrada da rede e cada nó dessa camada representa um vetor de entrada. É nessa camada que ocorre todos os pré-processamentos nos vetores de entrada. A camada F_1 é o campo de representação de características e recebe as informações de ambas as camadas F_0 e F_2 . A comunicação $F_0 \rightarrow F_1 \rightarrow F_2$ ocorre no sentido *bottom-up*, já a comunicação $F_2 \rightarrow F_1$ ocorre no sentido *top-down*. Por último, a camada F_2 consiste no campo de categorias, onde estão armazenados os protótipos aprendidos.

As camadas que formam a rede *Fuzzy ART* possuem seu próprio vetor de ativação (saída). Na camada F_0 a atividade é indicada pelo vetor de entrada $\mathbf{I} = (I_1, \dots, I_M)$ onde $I_i \in [0, 1]$, tal que $i = 1, \dots, M$. Na camada F_1 é indicada pelo o vetor de saída $\mathbf{x} = (x_1, \dots, x_M)$ e na camada F_2 é indicada pelo o vetor de saída $\mathbf{y} = (y_1, \dots, y_N)$. Vale ressaltar que cada nó em $j = 1, \dots, N$ armazenado em F_2 é uma categoria que representa um conjunto de padrões, e que cada categoria j é denotada por um vetor de pesos $\mathbf{w}_j = (w_{j1}, \dots, w_{jM})$. Mais detalhes da interação entre as camadas da arquitetura podem ser vistos na Figura 1.

Figura 1 – Dado o vetor de entrada a M -dimensional, F_0 obtém a codificação do complemento $\mathbf{I}$ 2M-dimensional. O vetor 2M-dimensional $\mathbf{I}$ é o vetor de entrada para o campo F_1 . O campo F_1 envia os sinais de ativação de cada categoria $\mathbf{w}_j$ para F_2 . O campo F_2 seleciona a categoria $\mathbf{w}_j$ com o maior valor de ativação e envia a expectativa para o campo F_1 . O campo F_1 obtém sua saída $\mathbf{x}$ e calcula a taxa de equivalência. O critério de vigilância pode levar a um *reset*, fazendo com que o processo de busca procure por uma nova categoria em F_2



Fonte: Autor.

Durante o treinamento, as categorias em F_2 podem assumir dois estados distintos: comprometido e não-comprometido. Uma categoria é dita comprometida quando o vetor de peso associado a ela já codificou algum vetor de entrada. Já uma categoria é dita não-comprometida quando o vetor de peso de tal categoria ainda não codificou nenhum vetor de entrada. No início do treinamento, os vetores de pesos das categorias estão em estados de não comprometidos e são inicializados como $\mathbf{w}_j = (I_1, \dots, I_M)$. Além do parâmetro de vigilância ρ , os outros parâmetros que definem a dinâmica da rede *Fuzzy ART* são o parâmetro de escolha $\alpha > 0$ e a taxa de aprendizagem $\beta \in [0, 1]$.

A rede *Fuzzy ART* atribui vetores de entrada às categorias através de dois critérios de distância distintos, *escolha* e *equivalência*. O critério de escolha utiliza o parâmetro de escolha α , enquanto o segundo utiliza o parâmetro de vigilância ρ para a realização do teste de hipótese (CARPENTER; GROSSBERG; ROSEN, 1991).

2.2.1 Função de Escolha

A função de escolha (ou função de ativação) é definida por

$$T_j(\mathbf{I}) = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|} \quad (2.1)$$

em que $\mathbf{I}$ é um vetor de entrada, $\mathbf{w}_j$ é uma categoria e $\mathbf{I} \wedge \mathbf{w}_j$ é a operação *fuzzy* AND componente-a-componente, que é dado por

$$\mathbf{I} \wedge \mathbf{w}_j = (\min(I_1, w_{1j}), \dots, \min(I_M, w_{Mj})) \quad (2.2)$$

e $|\cdot|$, é a norma dado por

$$|\mathbf{a}| = \sum_{i=1}^M a_i \quad (2.3)$$

A escolha da categoria, indexada por J , é obtida por

$$T_J = \max \{T_j : j = 1, \dots, N\} \quad (2.4)$$

sendo que T_j usado no lugar de $T_j(\mathbf{I})$ apenas por questão de simplicidade de notação.

Se mais de uma categoria possui um valor de T_j maximal, a categoria que possuir o menor índice j é a escolhida. Assim, pode-se obter a saída $\mathbf{y}$ de F_2 de tal maneira que a componente

$$y_j = \begin{cases} 1 & \text{se } j = J \\ 0 & \text{se } j \neq J \end{cases} \quad (2.5)$$

A saída $\mathbf{y}$ do campo de categorias em F_2 representa a expectativa *top-down* dado um padrão de entrada $\mathbf{I}$. Dessa maneira, a saída da camada F_1 é dado por

$$\mathbf{x} = \begin{cases} \mathbf{I} & \text{se } F_2 \text{ esta inativo} \\ \mathbf{I} \wedge \mathbf{w}_j & \text{se } F_2 \text{ esta ativo} \end{cases} \quad (2.6)$$

2.2.2 Reinício ou Ressonância

A ressonância ocorre quando a função de equivalência, $|\mathbf{x}| / |\mathbf{I}|$, para uma categoria J escolhida, atende ao critério de vigilância. O *critério de vigilância* é descrito por

$$\frac{|\mathbf{x}|}{|\mathbf{I}|} = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{|\mathbf{I}|} \geq \rho \quad (2.7)$$

Quando uma categoria $\mathbf{w}_j$ possui equivalência com o vetor de entrada $\mathbf{I}$ o aprendizado ocorre. Porém isso só acontece quando a equação acima é satisfeita. Do contrário, um novo processo de busca acontece e o valor de ativação T_j é atenuado para prevenir a seleção repetitiva da mesma categoria para o padrão de entrada atual. Esse processo de busca é repetido até que uma categoria $\mathbf{w}_j$ satisfaça o parâmetro de vigilância ou até que uma nova categoria seja comprometida (criada).

2.2.3 Aprendizado

Quando o processo de busca termina e o estado de ressonância ocorre, o vetor de pesos do $\mathbf{w}_j$ é atualizado de acordo com:

$$\mathbf{w}_j^{new} = \beta(\mathbf{I} \wedge \mathbf{w}_j^{old}) + (1 - \beta)\mathbf{w}_j^{old} \quad (2.8)$$

em que β , como já mencionado, é a taxa de aprendizado. Como por ser visto, o $\mathbf{w}_j^{new}$ é uma soma ponderada entre os vetores de peso antigo e a saída da camada F_1 quando ativada.

A rede *Fuzzy ART* possui duas estratégias de aprendizado: o lento e o rápido. Na estratégia de aprendizado lento, $\beta < 1$, os valores dos pesos são ajustados lentamente durante o ciclo de treinamento. Já na estratégia de aprendizado rápido o valor de $\beta = 1$. Com essa estratégia a função de aprendizado (Eq. 2.8) pode ser simplificada para:

$$\mathbf{w}_j^{new} = \mathbf{I} \wedge \mathbf{w}_j^{old} \quad (2.9)$$

2.2.4 Normalização do Vetor de Entrada

A codificação complementar é uma regra alternativa para o pré-processamento dos vetores de entrada, pois atinge a normalização enquanto preserva a amplitude das informações. A normalização por codificação complementar de um dado vetor de entrada $\mathbf{a} \in \mathbb{R}^M$, tal que $\mathbf{a} \in [0, 1]$, é um vetor de entrada $\mathbf{I} \in \mathbb{R}^{2M}$, tal que $\mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$ em que $\mathbf{a}^c = 1 - \mathbf{a}$. Note que:

$$|\mathbf{I}| = |(\mathbf{a}, \mathbf{a}^c)| = \sum_{i=1}^M a_i + (M - \sum_{i=1}^M a_i) = M \quad (2.10)$$

Desse modo, utilizando a codificação complementar nos vetores de entrada estes são automaticamente normalizados. Vale ressaltar que ao normalizar os vetores de entrada por codificação complementar, os vetores de pesos armazenados na camada F_2 também são alterados, ou seja, um vetor de peso $\mathbf{w}_j$ M -dimensional passa a ser um vetor de peso $2M$ -dimensional $\mathbf{w}_j = (\mathbf{u}_j, \mathbf{v}_j^c)$, em que os vetores $\mathbf{u}$ e $\mathbf{v} \in \mathbb{R}^M$.

2.2.5 Interpretação Geométrica

Ao utilizar a normalização por codificação complementar nos vetores de entrada, o *Fuzzy ART* é induzido a generalizar os dados através do uso de hiper retângulos que são posicionados no espaço de característica. Se considerarmos um vetor de entrada $\mathbf{a}$ bidimensional a codificação complementar desse vetor é

$$\mathbf{I} = (\mathbf{a}, \mathbf{a}^c) = (a_1, a_2, 1 - a_1, 1 - a_2) \quad (2.11)$$

Como já mencionado, quando o vetor de entrada é normalizado por codificação complementar, os vetores de pesos armazenados em F_2 ganham uma nova notação. Considerando um problema bidimensional em que $M = 2$, tem-se que:

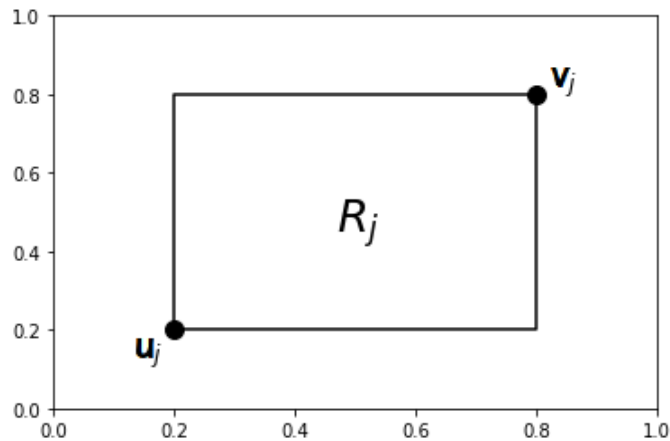
$$\mathbf{w}_j = (\mathbf{u}_j, \mathbf{v}_j^c) = (u_{j1}, u_{j2}, 1 - v_{j1}, 1 - v_{j2}) \quad (2.12)$$

em que os vetores $\mathbf{u}_j$ e $\mathbf{v}_j$ são, respectivamente, o ponto inferior e ponto superior do hiper retângulo R_j que representa a categoria $\mathbf{w}_j$. Sabendo disso, o tamanho deste hiper retângulo R_j pode ser calculado como:

$$|R_j| = |\mathbf{v}_j - \mathbf{u}_j| \quad (2.13)$$

Na Figura 2 é demonstrado como o vetor de pesos bidimensional em codificação do complemento cobre o espaço de característica. Vale lembrar que cada vetor de peso $\mathbf{w}_j$ possui uma representação geométrica R_j de um hiper-retângulo, cujo os pontos inferior e superior são denotados, respectivamente por $\mathbf{u}_j$ e $\mathbf{v}_j$

Figura 2 – Representação dos vetores de peso da rede *Fuzzy ART* em codificação complementar com $M = 2$.

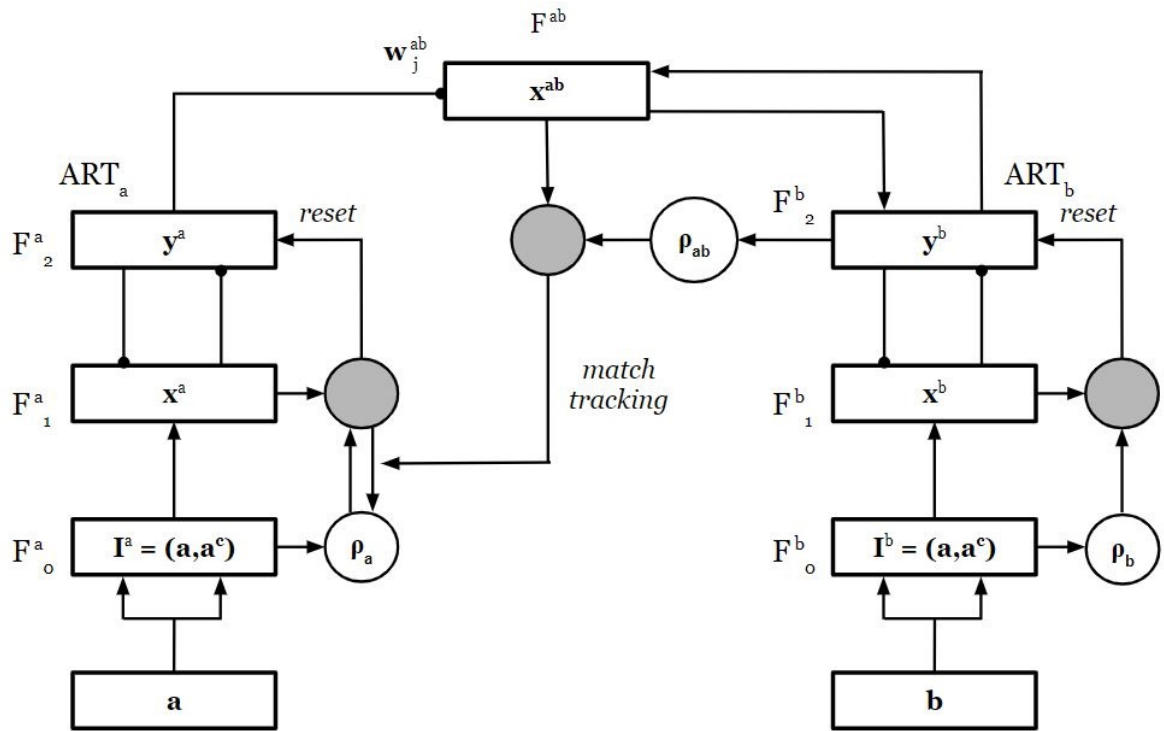


Fonte: Autor.

2.3 Fuzzy ARTMAP

O *Fuzzy ARTMAP* é uma rede neural proposta para atuar em modo de aprendizado supervisionado. Essa rede é constituída por dois módulos ART, sendo um responsável por codificar os atributos, denominado ART_a , e o outro é responsável por codificar os rótulos dos padrões de entrada, denominado ART_b . E um módulo inter-ART ou campo de mapeamento (*map-field*) que define o mapeamento entre as categorias do ART_a e as categorias que representam as classes no ART_b .

Figura 3 – Arquitetura da rede *Fuzzy ARTMAP*



Fonte: Autor.

Os módulos ART_a e ART_b tem os vetores de entrada $\mathbf{I}^a = (\mathbf{a}, \mathbf{a}^c)$ para atributos e $\mathbf{I}^b = (\mathbf{b}, \mathbf{b}^c)$ para os rótulos, respectivamente. Assim, o vetor de entrada $\mathbf{I}^a$ age como ativação do campo F_o^a e o vetor de entrada $\mathbf{I}^b$ como ativação do campo F_o^b . A ativação dos campos F_1^a e F_2^a é dado por $\mathbf{x}^a = (x_1^a, \dots, x_{2M_a}^a)$ e $\mathbf{y}^a = (y_1^a, \dots, y_{N_a}^a)$, respectivamente, enquanto os pesos das categorias do módulo ART_a são denotados por $\mathbf{w}_j^a = (w_{j1}^a, \dots, w_{j2M_a}^a)$. De modo semelhante, para o módulo ART_b , $\mathbf{x}^b = (x_1^b, \dots, x_{2M_b}^b)$ define a ativação do campo F_1^b , $\mathbf{y}^b = (y_1^b, \dots, y_{N_b}^b)$ define a ativação do campo F_2^b , enquanto que os pesos das categorias do módulo ART_b são denotados por $\mathbf{w}_k^b = (w_{k1}^b, \dots, w_{k2M_b}^b)$. O módulo inter-ART possui um campo F^{ab} cuja a ativação é indicada $\mathbf{x}^{ab} = (x_1^{ab}, \dots, x_{N_b}^{ab})$ e as categorias são indicadas pelos vetores de pesos $\mathbf{w}_j^{ab} = (w_{j1}^{ab}, \dots, w_{j2N_b}^{ab})$.

A ativação do campo de mapeamento é computada da seguinte forma

$$\mathbf{x}^{ab} = \begin{cases} \mathbf{y}^b \wedge \mathbf{w}_j^{ab} & \text{se o } J - \text{ésimo nó em } F_2^a \text{ esta ativo e } F_2^b \text{ esta ativo,} \\ \mathbf{w}_j^{ab} & \text{se o } J - \text{ésimo nó em } F_2^a \text{ esta ativo e } F_2^b \text{ esta inativo,} \\ \mathbf{y}^b & \text{se } F_2^a \text{ esta inativo e } F_2^b \text{ esta ativo, e} \\ 0 & \text{se } F_2^a \text{ esta inativo e } F_2^b \text{ esta inativo.} \end{cases} \quad (2.14)$$

Observe que a ativação $\mathbf{x}^{ab}$ do campo de F^{ab} só é ativado quando um ou ambos os campos de categorias de F_2^a e F_2^b estão ativos. Quando nenhum é ativado, ocorre um erro de predição e o mecanismo de *match tracking* é engatilhado e o subsistema de atenção no módulo ART_a é direcionado para categorias que necessitem de menos generalização para incluir o padrão de entrada. A arquitetura do Fuzzy ARTMAP é detalhada na Figura 3.

2.3.1 Mecanismo de Match Tracking

A cada padrão de entrada que é apresentado à rede, o valor do parâmetro de vigilância ρ_a é o mesmo que o valor do parâmetro de vigilância base $\overline{\rho}_a$. O valor de vigilância base $\overline{\rho}_a$ é inicializado antes do treinamento começar. O módulo inter-ART também possui um parâmetro de vigilância denominado ρ_{ab} . Desta forma, um erro de predição de ART_a para ART_b acontece quando:

$$|\mathbf{x}^{ab}| < \rho_{ab} |\mathbf{y}^b| \quad (2.15)$$

Quando um erro de predição acontece, o parâmetro de vigilância ρ_a é temporariamente aumentado, sendo calculado por

$$\frac{|\mathbf{I}^a \wedge \mathbf{w}_J^a|}{|\mathbf{I}^a|} + \varepsilon \quad (2.16)$$

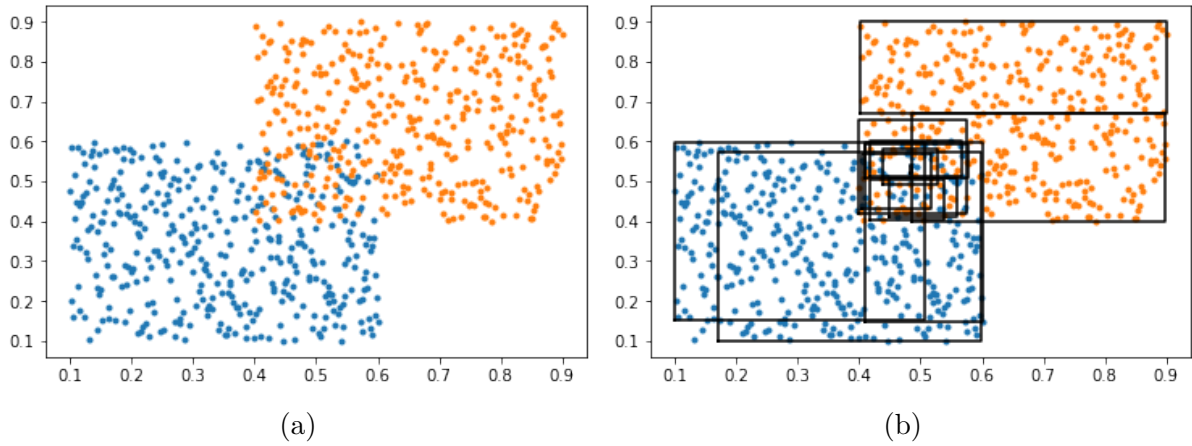
em que $0 < \varepsilon < 1$. Dessa forma uma nova categoria J é ativada em F_2^a caso $|\mathbf{I}^a \wedge \mathbf{w}_J^a| \geq \rho_a |\mathbf{I}^a|$. Esse processo é repetido até que um nó de F_2^a acerte a predição dada pela ativação do campo F_2^b ou que uma nova categoria seja criada no módulo ART_a . Essa nova categoria será vinculada à predição dada pela ativação no campo de categorias de F_2^b através do campo de mapeamento F^{ab} do módulo inter-ART.

2.3.2 Treinamento da Rede Fuzzy ARTMAP

Os parâmetros de treinamento iniciais da rede Fuzzy ARTMAP são o parâmetro de escolha α , o parâmetro de vigilância dos módulos ART_a , ART_b e inter-ART, ρ_a , ρ_b , ρ_{ab} , o parâmetro de *match tracking* ε e a taxa de aprendizado β .

Com os padrões de entrada já normalizados por codificação complementar, eles são apresentados à rede *Fuzzy ARTMAP*. A rede calcula o valor de ativação das categorias $\mathbf{w}_j^a$ e $\mathbf{w}_k^b$ para as respectivas entradas $\mathbf{I}^a$ e $\mathbf{I}^b$. Após o cálculo, as categorias que possuem as maiores ativações T_j e T_k são escolhidas. O critério de vigilância de cada um dos módulos ART_a e ART_b são verificados e em caso de *reset* em algum dos módulos, o processo de busca irá procurar novas categorias que satisfaçam o critério de vigilância. Quando nenhuma é encontrada, uma nova categoria é criada. Lembrando que os processos realizados nos módulos ART_a e ART_b são independentes, ou seja, pode ocorrer um *reset* no ART_a mas não necessariamente no ART_b e vice-versa.

Figura 4 – Categorias da rede *Fuzzy ARTMAP* obtidas de um exemplo: (a) determinada distribuição de dados para um problema binário e (b) as categoria criadas pela a rede.



Fonte: Autor.

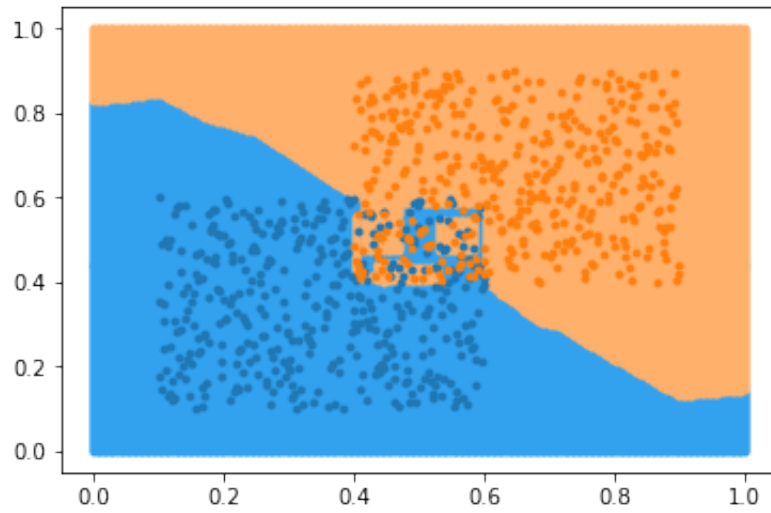
Quando ambos os módulos ART_a e ART_b alcançam o estado de ressonância, a ativação $\mathbf{x}^{ab}$ do módulo inter-ART é computada e o critério de vigilância deste módulo é verificado. Caso o critério não seja atendido acontece o *reset* lateral, o mecanismo de *match tracking* é engatilhado, o valor do parâmetro ρ_a é temporariamente aumentado e o módulo ART_a inicia a busca por uma categoria $\mathbf{w}_j^a$ que acerte a classe $\mathbf{w}_k^b$ ou irá criar uma nova categoria. Vale ressaltar que o *reset* lateral do inter-ART ativa apenas o processo de busca no módulo ART_a , nunca no módulo ART_b . No final desse processo, as categorias $\mathbf{w}_j^a$ e $\mathbf{w}_k^b$ são atualizadas, bem como o campo de mapeamento $\mathbf{w}_j^{ab} = \mathbf{x}^{ab}$.

O treinamento da rede *Fuzzy ARTMAP* só termina quando a estabilidade da rede é alcançada, ou até que alcance um número de épocas pré-definido. Um exemplo de como as categorias do módulo ART_a criadas pelo *Fuzzy ARTMAP* se alocam no espaço de características pode ser visto na Figura 4.

2.3.3 Predição do Fuzzy ARTMAP

Quando um padrão desconhecido $\mathbf{I}^a$ é apresentado à rede, o valor de ativação T_j de todas as categorias $\mathbf{w}_j$ é computada através da Equação. (2.1). A categoria com o maior valor de ativação T_j é escolhida pela Equação. (2.4) e a ativação do campo F_2^a , $\mathbf{y}^a$, é obtida pela Equação. (2.5). A classe indicada ao padrão $\mathbf{I}^a$ é obtida utilizando a ativação do campo F_2^a , $\mathbf{y}^a$, que aponta para o índice J da categoria atualmente ativa no sistema. Esse índice J é pareado com o vetor de pesos do campo de mapeamento e o $\mathbf{w}_j^{ab}$ é obtido. Em $\mathbf{w}_j^{ab}$ está o mapeamento para a K -ésima categoria no módulo ART_b , ou seja, a classe no qual será atribuído ao padrão $\mathbf{I}^a$. A superfície de decisão obtida pela rede Fuzzy ARTMAP para o problema ilustrado na Figura 4, pode ser vista na Figura 5.

Figura 5 – Superfície de decisão obtida pela rede Fuzzy ARTMAP.



Fonte: Autor.

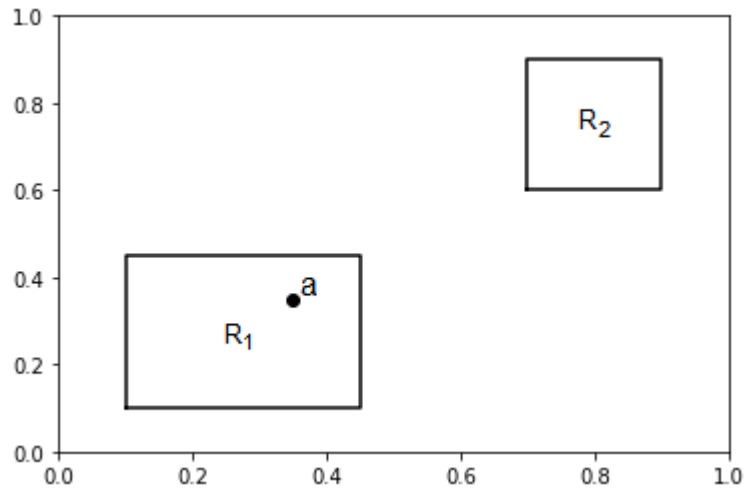
2.3.4 Proliferação de Categorias

A proliferação de categorias ocorre em geral devido à sobreposição dos dados de duas ou mais classes. Esse problema é muito conhecido e visitado na literatura de Fuzzy ARTMAP (CARPENTER, 1996; CARPENTER, 1997; SANCHEZ et al., 2000; ZHONG et al., 2007; POURPANAH; LIM; SALEH, 2016). Durante o treinamento, essa sobreposição ativa constantemente o mecanismo de *match tracking* a fim de corrigir os erros preditivos internos. A correção desse erro se dá através do aumento temporário do parâmetro de vigilância ρ_a , o que estimula a criação de novas categorias menores dentro das categorias maiores no processo de busca no ART_a .

Considere uma rede Fuzzy ARTMAP com os seguintes parâmetros, $\alpha = 0.001$ e com aprendizado rápido, ou seja $\beta = 1$, além disso $\overline{\rho}_a = 0.6$ e $\varepsilon = 0.0001$. Tomando como

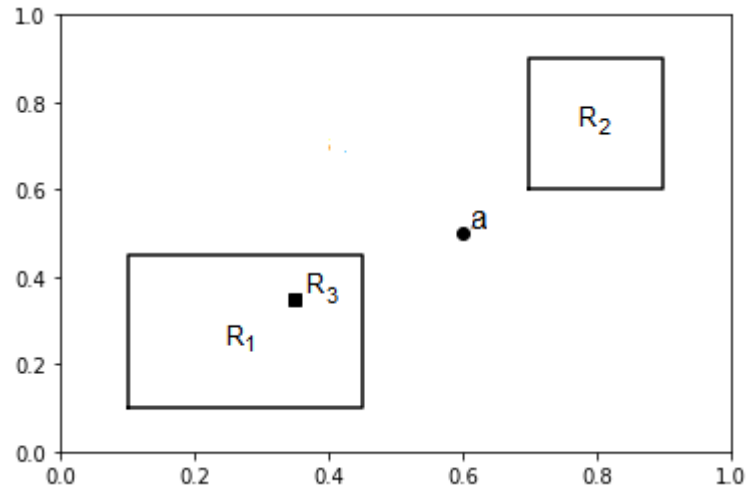
exemplo a Figura 6, o hiper-retângulo $R_1 = (0.7, 0.6, 0.9, 0.9)$ da categoria $\mathbf{w}_1^a = (0.7, 0.6, 0.1, 0.1)$ é mapeado em $\mathbf{w}_1^{ab}$ para predizer a classe C_1 ; já o hiper retângulo $R_2 = (0.1, 0.1, 0.45, 0.45)$ da categoria $\mathbf{w}_2^a = (0.1, 0.1, 0.55, 0.55)$ é mapeado em $\mathbf{w}_2^{ab}$ para predizer a classe C_2 . Quando um padrão de entrada $\mathbf{a}$ (em codificação complementar $\mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$) da classe C_1 é apresentado à rede, a função de escolha calcula os valores pela Equação (2.1), tal que $T_1 = 0.6129$ e $T_2 = 0.9992$. Pelo critério de escolha (Eq.(2.4)), a categoria $\mathbf{w}_2^a$ é escolhida. Como o $\bar{\rho}_a = 0.6$, a categoria $\mathbf{w}_2^a$ passa pelo o critério de vigilância Equação (2.7), porém, a predição de $\mathbf{a}$ aponta para a classe C_2 . Neste caso a Equação (2.15) não é satisfeita e o mecanismo de *match tracking* será engatilhado e o sinal da categoria $\mathbf{w}_2^a$ é atenuado ($T_2 = -1$). O *match tracking* aumenta o valor do parâmetro de vigilância ρ_a , tal que $\rho_a = 0.6501$ (Equação.(2.16)). Já que o sinal da categoria $\mathbf{w}_2^a$ foi atenuado, a categoria $\mathbf{w}_1^a$ é escolhida, porém não passa pelo o critério de vigilância. Perceba que, mesmo que a categoria $\mathbf{w}_2^a$ aponte para a mesma classe do padrão de entrada $\mathbf{a}$, por conta do critério de vigilância ela não foi capaz de codificar o ponto $\mathbf{a}$. Assim, uma nova categoria $\mathbf{w}_3^a$ será comprometida e seu mapeamento $\mathbf{w}_3^{ab}$ será apontado para a classe C_1 .

Figura 6 – Ilustração dos hiper-retângulos R_1 e R_2



Fonte: Autor.

Agora suponha que um novo padrão de entrada $\mathbf{a}$ da classe C_2 é apresentado à rede. Como pode ser visto na Figura 7, com o padrão de entrada anterior codificado pela rede, tem-se agora três categorias, das quais $\mathbf{w}_1^{ab}$ e $\mathbf{w}_3^{ab}$ mapeia a classe C_1 e $\mathbf{w}_2^{ab}$ mapeia a classe C_2 . Pela a Eq. (2.1) são obtidos os valores de ativação de modo que $T_1 = 0.8660$, $T_2 = 0.8455$ e $T_3 = 0.7996$. Como a categoria $\mathbf{w}_1^a$ possui o maior valor de ativação, ela é escolhida para o teste do critério de vigilância no modulo ART_a . Uma vez que $\rho_a = \bar{\rho}_a = 0.6$, o critério de vigilância no modulo ART_a é satisfeito. Entretanto, como $\mathbf{w}_1^{ab}$ mapeia a classe C_1 e $\mathbf{a}$ pertence à classe C_2 , ocorre um erro de predição. O mecanismo

Figura 7 – Ilustração dos hiper-retângulos R_1 , R_2 e R_3 

Fonte: Autor.

de *match tracking* é acionado e o valor do parâmetro de vigilância é aumentado de modo $\rho_a = |\mathbf{I}^a \wedge \mathbf{w}_j^a| / M + \varepsilon = 0.6501$ e o sinal da categoria $\mathbf{w}_1^a$ é atenuado ($T_1 = -1$).

Assim, a próxima categoria escolhida é a $\mathbf{w}_2^a$. Perceba que a categoria $\mathbf{w}_2^a$ mapeia a mesma classe que o padrão de entrada, mas, ela não atende ao novo critério de vigilância do módulo ART_a . Dessa forma, seu sinal é atenuado ($T_2 = -1$) e faz com que o processo de busca do módulo ART_a a escolher a categoria $\mathbf{w}_3^a$; entretanto, como a predição de $\mathbf{w}_3^a$ e $\mathbf{a}$ são diferentes, o módulo inter-art não irá ressoar e uma nova categoria $\mathbf{w}_4^a$ será comprometida mapeando a classe C_1 . Por esses cenários é possível perceber que, quando uma região muito ruidosa é aprendida pelo o *Fuzzy ARTMAP*, muitas categorias são criadas com o objetivo de corrigir erros preditivos durante o treinamento.

3 Opção de Rejeição

Neste capítulo são descritos os conceitos relacionados a classificação com opção de rejeição. Neste contexto, são apresentadas a regra de Chow e do Fumera, bem como as abordagens de classificação com opção de rejeição. Neste capítulo também são apresentadas as propostas desenvolvidas durante este trabalho.

3.1 Introdução

Para um dado vetor de entrada $\mathbf{x}$ e considerando que o problema é binário, o classificador tem a tarefa de atribuir uma classe entre as duas disponíveis, com base em estimativas de probabilidade a posteriori $P(C_i|\mathbf{x})$. Essa atribuição acontece mesmo se a diferença entre as probabilidades a posteriori das classes sejam pequenas. Quando a decisão entre as probabilidades a posteriori se encontram próximas do limiar, a automatização das decisões se torna um trabalho difícil, podendo a conduzir predições erradas e como consequência a elevação do erro de classificação.

Em certos ambientes é atraente que tais sistemas sejam capazes de rejeitar casos críticos, deixando para um especialista analisar o caso. Neste ponto de vista, uma nova classe pode ser definida, chamada de classe de rejeição, tal que seus padrões se encontram na região entre as duas classes. O trabalho de Chow (CHOW, 1957) que envolveu a tarefa de reconhecimento de caracteres foi o pioneiro no que se refere-se à classificação com opção de rejeição. Entretanto, foi em um trabalho seguinte que Chow (CHOW, 1970) descreveu de forma mais completa o problema de classificação com opção de rejeição, sendo este considerado como o trabalho clássico inicial na área. Friedel, Ruckert e Kramere (FRIEDEL; RUCKERT; KRAMER, 2006) demonstrou que a inclusão da opção de rejeição aumenta consideravelmente a acurácia dos classificadores.

A utilização da opção de rejeição serve para evitar erros excessivos como consequência de tomadas de decisão difíceis. Assim o sistema converte potenciais erros em rejeição. No entanto, quando a opção de rejeição é utilizada, decisões potencialmente corretas podem vir a ser rejeitadas. Vale ressaltar que o custo de rejeitar e errar raramente estão na proporção um-para-um. Assim, é necessário analisar o desempenho do classificador para diferentes custos de rejeição, no qual é escolhido o classificador com base na acurácia desejada ou em uma taxa de rejeição aceitável para o tipo de problema a ser abordado.

3.2 Regra de Decisão Ótima

A regra de decisão para classificação com opção de rejeição é considerada ótima se para uma determinada taxa de rejeição, ou probabilidade de rejeição, também se minimiza a taxa de erro (ou probabilidade de erro). A regra da decisão ótima deve recusar um determinado padrão quando o maior valor das probabilidades a posteriori é menor do que algum limiar (CHOW, 1970).

Explicitamente, a regra da decisão ótima $\delta(d_k|\mathbf{x})$ é dado por:

$$(a) \quad \delta(d_k|\mathbf{x}) = 1 \quad (k \neq 0) \quad (3.1)$$

$$(b) \quad \delta(d_o|\mathbf{x}) = 1 \quad (k = 0) \quad (3.2)$$

em que $\mathbf{x} \in \mathbb{R}^N$ representa o padrão a ser analisado, d_o representa a decisão de rejeitar o padrão $\mathbf{x}$ e d_k representa a decisão de aceitar o padrão $\mathbf{x}$ e atribuí-lo a classe $k = 1, \dots, c$, tal que c representa o número de classes do problema.

A Equação (3.1) considera que o padrão $\mathbf{x}$ é aceito, se:

$$p_k F(\mathbf{x}|C_k) \geq p_j F(\mathbf{x}|C_j), \quad \forall j = 1, \dots, c \quad (3.3)$$

e

$$p_k F(\mathbf{x}|C_k) \geq (1 - t) \sum_{i=1}^c p_i F(\mathbf{x}|C_i) \quad (3.4)$$

em que C_i representa a i -ésima classe, p_i representa a probabilidade a priori da i -ésima classe, $F(\mathbf{x}|C_k)$ descreve a função de verossimilhança dos dados da classe k e o limiar t é uma constante entre 0 e 1 ($0 < t < 1$).

A Equação (3.2), considera que se rejeite o padrão $\mathbf{x}$, se

$$\max_i \{p_i F(\mathbf{x}|C_i)\} < (1 - t) \sum_{j=1}^c p_j F(\mathbf{x}|C_j) \quad (3.5)$$

Definido $m(\mathbf{x})$, tal que

$$m(\mathbf{x}) = \frac{p_i F(\mathbf{x}|C_i)}{\sum_{i=1}^c p_i F(\mathbf{x}|C_i)} \quad (3.6)$$

Assim, a regra ótima de classificação pode ser simplificada e entendida como

1. aceitar o padrão $\mathbf{x}$, se:

$$m(\mathbf{x}) \geq (1 - t) \quad (3.7)$$

2. rejeita o padrão $\mathbf{x}$, se:

$$m(\mathbf{x}) < (1 - t) \quad (3.8)$$

Portanto, a classificação com opção de rejeição considera a rejeição do padrão $\mathbf{x}$ sempre que o maior valor entre as probabilidades a posteriori for pequeno, ou seja, menor que $(1 - t)$ (CHOW, 1957)

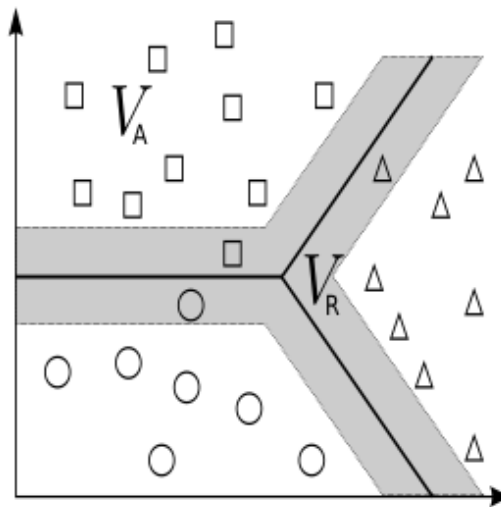
O parâmetro t é denominado limiar de rejeição e para qualquer valor deste limiar, a regra de decisão o divide o espaço de entrada em dois conjuntos (regiões) disjuntos V_A e V_R , tal que

$$V_A(t) = \{\mathbf{x} | m(x) \geq (1 - t)\} \quad (3.9)$$

$$V_R(t) = \{\mathbf{x} | m(x) < (1 - t)\} \quad (3.10)$$

Na Figura 8 são ilustradas as regiões dos conjuntos disjuntos V_A e V_R , que são chamados de região de aceitação e de rejeição, respectivamente. Uma região de rejeição é um espaço no qual todos os exemplos dentro dessa região são rejeitados pelo classificador. Observando a Figura 8, os padrões que devem ser rejeitados encontram-se na região sombreada.

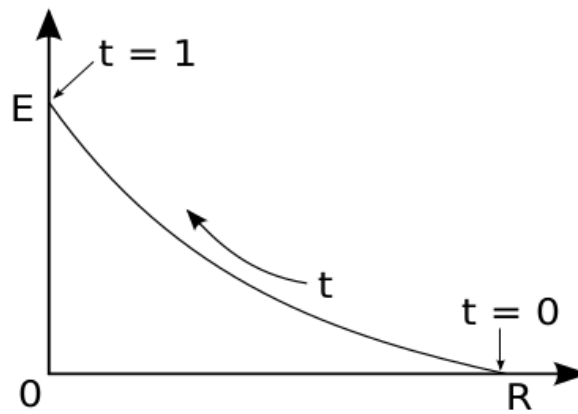
Figura 8 – Ilustração das regiões de rejeição (V_R) e aceitação (V_A) da regra ótima



Fonte: (NETO, 2011)

Como já mencionado, uma descrição completa do desempenho de um sistema de reconhecimento é dada pela relação de compromisso entre o erro e a rejeição. A curva típica para tal relação pode ser vista na Figura 9. Analisando a Figura 9, pode-se dizer que à medida que a rejeição aumenta a taxa de erro diminui e o inverso também pode ser dito. Desde que o erro (E) e a rejeição (R) de um sistema de reconhecimento ótimo são funções monotônicas do limiar de rejeição t , pode-se calcular o compromisso entre a taxa de erro e a taxa de rejeição para um determinado limiar.

Figura 9 – Ilustração da curva de compromisso entre o erro e a rejeição.



Fonte: (NETO, 2011)

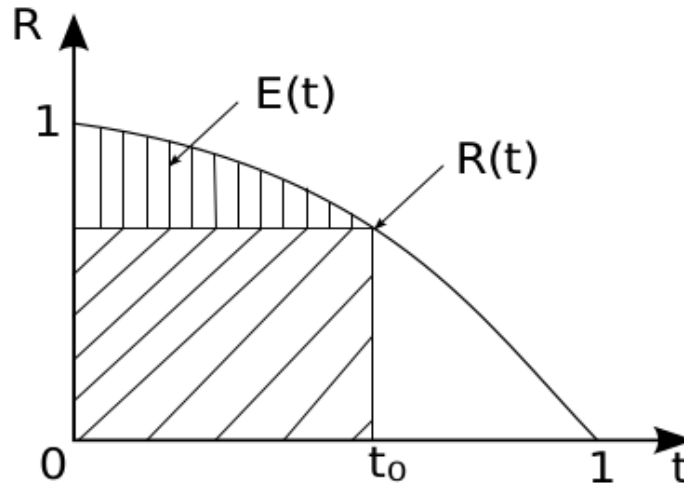
No trabalho de Chow (1970), mostra-se ainda que a taxa de rejeição em função do limiar de rejeição é suficiente para descrever o desempenho do sistema de reconhecimento. Na Figura 10 é mostrado um caso típico deste tipo de curva. Nessa figura, as somas das duas áreas hachuradas representam a taxa de erro quando nenhuma rejeição é permitida, ou então a área da hachura superior é a região que representa a taxa de erro quando são permitidas rejeições.

O limiar de rejeição pode então ser relacionado aos custos como segue (CHOW, 1957):

$$t = \frac{W_r - W_c}{W_e - W_c} \quad (3.11)$$

em que W_r , W_e e W_c são os custos de rejeitar, de errar e de acertar respectivamente. Se os custos associados aos erros, aos acertos e as rejeições são iguais, a regra de decisão ótima representada nas Equações (3.1) a (3.5) é também uma regra de risco mínimo. Usualmente, tem-se que $W_c < W_r < W_e$. Considerando $W_c = 0$ e $W_e = 1$, pode-se inferir que o custo de rejeição W_r é igual ao limiar de rejeição t ($W_r = t$). Assim, o risco mínimo para um

Figura 10 – Ilustração de uma curva hipotética de compromisso entre o erro e a rejeição.



Fonte: (NETO, 2011)

determinado limiar é dado por

$$r(t) = E(t) + tR(t) = \int_0^{t_0} R(t)dt \quad (3.12)$$

o que corresponde à área hachurada na Figura 10. Enquanto o risco empírico para uma função de decisão e um determinado conjunto de treinamento é definido por

$$r_{emp}(\alpha) = E(\alpha) + W_r R(\alpha) \quad (3.13)$$

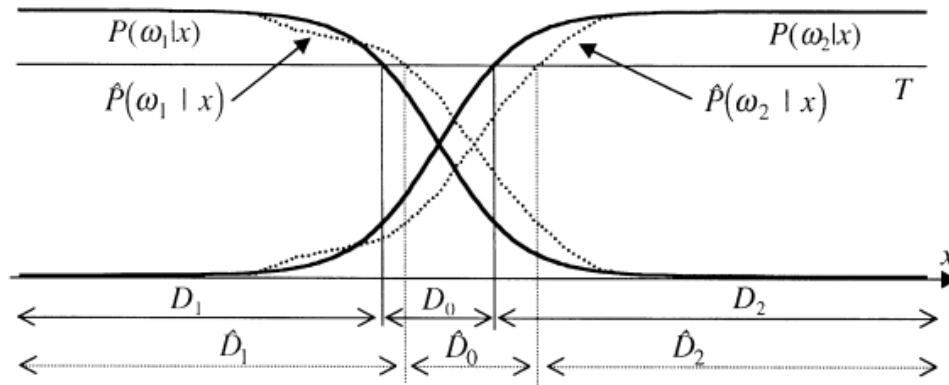
em que α representa o vetor de parâmetros do modelo, dentre eles o limiar de rejeição.

Considerando um problema de classificação binária, a condição para rejeitar, descrita na equação (3.5), pode não ser satisfeita quando o valor de t for maior que 0.5, uma vez que a taxa de rejeição é sempre zero se o limiar de rejeição t exceder o 0.5. Esta é a razão pelo o qual o valor de W_r estar sempre no intervalo $[0;0,5]$.

3.3 Múltiplos Limiares

Como mencionado anteriormente, a regra de Chow fornece um limiar ótimo de erro e rejeição. A Figura 11 mostra uma simples tarefa de classificação unidimensional com duas classes de dados w_1 e w_2 caracterizadas por duas distribuições gaussianas. Os termos $P(w_i|x)$ e $\hat{P}(w_i|x)$, $i = 1, 2$, indicam as probabilidades “reais” e “estimadas” a posteriori, respectivamente.

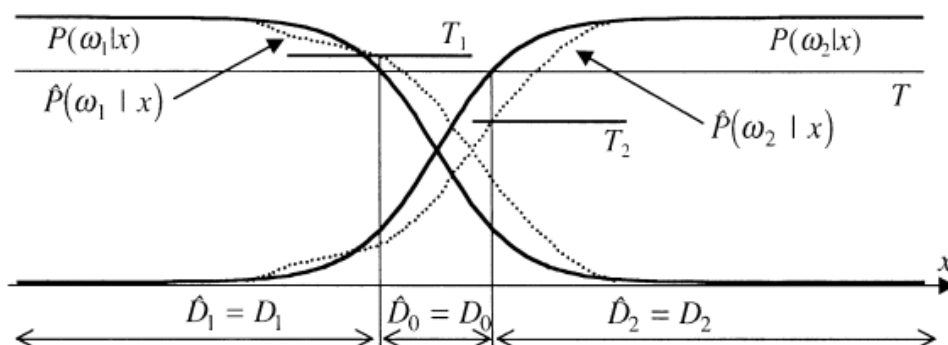
Figura 11 – Aplicação da regra de Chow as probabilidades a posteriori "reais" e "estimadas".



Fonte: (FUMERA; ROLI; GIACINTO, 2000)

Na Figura 11, as regiões de decisão e rejeição ideais fornecidas pela a regra de Chow aplicadas às probabilidades reais são indicadas pelos termos D_1 , D_2 e D_0 e o termo T indica o limite de rejeição de Chow. Analogicamente, os termos $\hat{D}_1$, $\hat{D}_2$ e $\hat{D}_0$ significam as regiões de decisão e de rejeitar que são fornecidas pela a regra de Chow aplicada às probabilidades estimadas. É fácil ver que nenhum valor do limiar de T aplicado às probabilidades estimadas podem fornecer tanto as regiões de decisão ótimas quanto a região de rejeição ótima.

Uma análise cuidadosa da Figura 11 sugere uma abordagem diferente da regra de Chow para obter o limiar ótimo de erro e rejeição. O uso de múltiplos limiares de rejeição para as diferentes classes de dados para obter as regiões ótimas de decisão e rejeição (FUMERA; ROLI; GIACINTO, 2000). A Figura 12 mostra o uso de dois limiares de rejeição distintos T_1 e T_2 para a tarefa de classificação da Figura 11. É fácil ver que tais limiares aplicados às probabilidades estimadas permitem obter tanto as regiões de decisão ótimas quanto a região de rejeição. Portanto, esse exemplo sugere que o uso de N -limiares de rejeição relacionados à classe pode fornecer um erro melhor do que o limiar de Chow.

Figura 12 – Dois diferentes limiares de rejeição T_1 e T_2 aplicados as probabilidades a posteriori estimadas da tarefa de classificação da figura 11

Fonte: (FUMERA; ROLI; GIACINTO, 2000)

Desta forma, a equação (3.5) é alterada para uma tarefa de classificação com N classes de dados da seguinte maneira. Um padrão é rejeitado se:

$$\max_i \{p_i F(\mathbf{x}|C_i)\} < (1 - t_i) \sum_{j=1}^c p_j F(\mathbf{x}|C_j) \quad (3.14)$$

em que t_i representa o limiar da i -ésima classe. E como já foi explicado o valor de t_i não pode ser maior que 0.5, caso seja maior que este valor a condição de rejeitar não seria satisfeita.

3.4 Abordagem de Classificação Binária com Opção de Rejeição

Para a classificação com opção de rejeição existem três abordagens diferentes de implementar a opção de rejeição: (i) usando apenas um único classificador; (ii) usando dois classificadores independentes e (iii) usando um classificador com opção de rejeição incorporada na regra de aprendizado. As abordagens (ii) e (iii) são enquadradas na categoria das técnicas que incorporam o custo de rejeitar no aprendizado dos classificadores. Essas abordagens são descritas em mais detalhes a seguir.

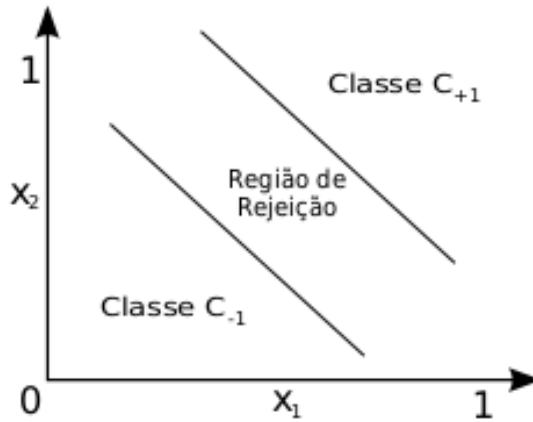
3.4.1 Um Classificador Padrão

Como o próprio nome sugere, nesta abordagem é utilizado apenas um único classificador. A rejeição acontece quando o máximo das probabilidades a posteriori for menor que um determinado limiar. Quando o classificador não fornece saídas probabilísticas pode-se ainda definir um limiar de rejeição desde que a função de saída seja monotônica. Logo após a fase de treinamento padrão do classificador é definida a região de rejeição, definido valores adequados sobre a saída do classificador. Neste processo busca-se minimizar

$$\min(E(\alpha) + W_r R(\alpha)) \quad (3.15)$$

para se obter o limiar ótimo para um valor específico de W_r . Esta abordagem leva a uma situação em que não há interseção entre as superfícies de decisão, conforme ilustrado na Figura 13 a seguir.

Figura 13 – Região de decisão obtida utilizando um classificador independentes para um problema binário.



Fonte: (SOUSA; MORA; CARDOSO, 2009b)

3.4.2 Dois Classificadores Independentes

Nessa abordagem, são utilizados dois classificadores indicados como M_1 e M_2 . No treinamento do classificador M_1 são penalizados os erros do tipo falso negativo. Desta forma, espera-se que quando o classificador M_1 classifique um padrão como sendo da classe C_{+1} , esse padrão deve pertencer realmente à classe C_{+1} . Por outro lado, no treinamento do classificador M_2 , são penalizados os erros do tipo falso positivo. Deste modo, quando o classificador M_2 classificar um padrão como sendo da classe C_{-1} , este padrão é verdadeiramente da classe C_{-1} .

Segundo Bounsiar et al. (2007) a regra de aceitação ou rejeição de um padrão para essa estratégia é dado por

$$\text{Se } P_1(C_{+1}|\mathbf{x}) \geq 0.5 \text{ e } P_2(C_{-1}|\mathbf{x}) < 0.5 \text{ então Classe } C_{+1}. \quad (3.16a)$$

$$\text{Se } P_1(C_{+1}|\mathbf{x}) < 0.5 \text{ e } P_2(C_{-1}|\mathbf{x}) \geq 0.5 \text{ então Classe } C_{-1}. \quad (3.16b)$$

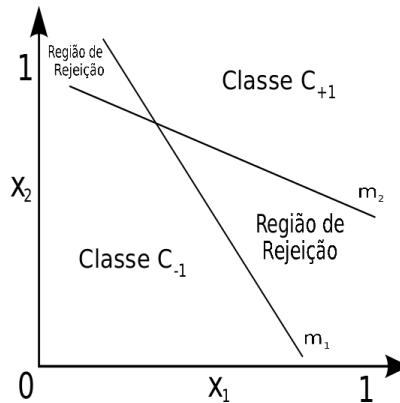
$$\text{Se } P_1(C_{+1}|\mathbf{x}) \geq 0.5 \text{ e } P_2(C_{-1}|\mathbf{x}) \geq 0.5 \text{ então Classe } \textit{rejeitar}. \quad (3.16c)$$

$$\text{Se } P_1(C_{+1}|\mathbf{x}) < 0.5 \text{ e } P_2(C_{-1}|\mathbf{x}) < 0.5 \text{ então Classe } \textit{rejeitar} \quad (3.16d)$$

em que $P_1(C_{+1}|\mathbf{x})$ e $P_2(C_{-1}|\mathbf{x})$ são as probabilidades a posteriori dos classificadores M_1 e M_2 , respectivamente.

Essa abordagem pode fazer com que ocorra uma interseção entre as superfícies de decisão das classes, levando à ocorrência de regiões em que a tomada de decisão é bastante fraca (SOUSA; MORA; CARDOSO, 2009b). Essa situação pode ser vista na Figura 14 a seguir

Figura 14 – Região de decisão obtida utilizando dois classificadores independentes para um problema binário.



Fonte: (SOUSA; MORA; CARDOSO, 2009b)

3.4.3 Classificador com Opção de Rejeição Embutida

Nesta abordagem o classificador é projetado de forma que traga embutido no processo de otimização de sua função-custo a capacidade de classificação com opção de rejeição. Dessa forma, não é mais necessário adicionar mais etapas para obter um valor de limiar de rejeição ótima, como também não é necessária a especialização de classificadores em determinada classe como ocorre na abordagem com dois classificadores.

Na opção de rejeição embutida, o modelo aprende a decidir entre as três saídas possíveis C_{+1} , C_{-1} e a rejeição. O trabalho desenvolvido por Fumera e Roli (2002) foi o primeiro trabalho a propor um classificador único com opção de rejeição embutido. Alguns outros trabalhos relacionados a essa abordagem são apresentados em Bounsiar et al. (2008) e em Sousa et al. (2009b).

3.5 Proposta para Classificação com Opção de Rejeição

3.5.1 Rede Neural PROBART

A rede *Fuzzy ARTMAP* na sua arquitetura convencional previne os erros preditivos no módulo inter-ART através do critério de vigilância. Devido a essa propriedade o mapeamento entre as categorias dos módulos ART_a e ART_b armazenados em w^{ab} seguem uma orientação de muitos-para-um. Devido a isso, a rede generaliza de forma excessiva o conjunto de treinamento até que seu erro de treinamento chegue a zero, porém isso não significa que a predição dos padrões que ainda não foram apresentados à rede seja aprimorada.

Outro problema subjacente à rede *Fuzzy ARTMAP* é que o mapeamento um-para-muitos não é possível. Isso significa que apenas uma categoria ART_b pode ser associada a cada categoria ART_a . Em condições ruidosas, um conjunto de entradas normalmente associadas a uma única categoria pode acionar mais de uma categoria em F_b . No entanto, esse conjunto de categorias não pode ser associado a uma única categoria em F_a . A incapacidade de formar mapeamentos um-para-muitos faz com que apenas uma única categoria em ART_b seja prevista para cada categoria ART_a e, portanto, as informações pertencentes a todas as outras categorias ART_b no conjuntos são perdidas

Ao contrário do *Fuzzy ARTMAP*, o processo de *match tracking* é omitido na rede neural *PROBART*. Em vez disso, ele opera coletando informações probabilísticas na forma de contagens de frequência de associações entre as várias categorias em F_a e F_b . Essas informações probabilísticas são armazenadas no campo de mapeamento. A rede neural *PROBART* foi proposta por Marriott e Harrison (1995) para aproximar funções de maneira incremental. Essa rede foi projetada principalmente para melhorar os recursos de manipulação de ruído do *Fuzzy ARTMAP*. No entanto, a rede *PROBART* não possui recursos de generalização muito bons. A sua ativação x^{ab} é dada pela substituição do operador *fuzzy AND* ($\wedge$) da Equação (2.14) pelo operador de adição (+), tal que

$$\mathbf{x}^{ab} = \begin{cases} \mathbf{y}^b + \mathbf{w}_j^{ab} & \text{se o } J\text{-ésimo nó em } F_2^a \text{ esta ativo e } F_2^b \text{ esta ativo,} \\ \mathbf{w}_j^{ab} & \text{se o } J\text{-ésimo nó em } F_2^a \text{ esta ativo e } F_2^b \text{ esta inativo,} \\ \mathbf{y}^b & \text{se } F_2^a \text{ esta inativo e } F_2^b \text{ esta ativo, e} \\ 0 & \text{se } F_2^a \text{ esta inativo e } F_2^b \text{ esta inativo.} \end{cases} \quad (3.17)$$

É importante perceber que a matriz de pesos do campo de mapeamento agora possui as informações sobre a frequência com o qual os pares das categorias ART_a e ART_b estão associados. Esse modo de coleta probabilística de informações permite que ocorra muitas associações entre os módulos ART_a e ART_b na rede *PROBART*. Em condições ruidosas, o mapeamento um-para-muitos permite o cálculo de uma saída estimada de mais de uma categoria ART_b . Assim, a formação de mapeamentos um-para-muitos combinada com a remoção do *match tracking* reduz bastante a sensibilidade do *PROBART* a sinais de ruído, porque categorias desnecessárias não são criadas no F_a que representam associações ruidosas.

3.5.2 Usando um Único Limiar (Proposta 1)

A primeira proposta deste trabalho para a classificação com opção de rejeição é baseada em uma única rede *Fuzzy ARTMAP* com a modificação do campo de mapeamento presente no *PROBART*, ou seja, a rede *Fuzzy ARTMAP* continua operando com as mesmas

fórmulas que foram apresentadas, porém seu campo de mapeamento também armazena a frequência das associações entre os módulos ART_a e ART_b .

Essa modificação é construída em cima da premissa de que a proliferação de categorias ocorre na sobreposição entre as classes. Além disso, em alguns casos, a sobreposição entre classes pode deteriorar a performance da rede *Fuzzy* ARTMAP, principalmente quando o conjunto de dados é bastante ruidoso. Diante disso, essa modificação busca se aproveitar da proliferação de categorias do *Fuzzy* ARTMAP e desta forma obter informações sobre a frequência de associações entre os módulos ART_a e ART_b usando o campo de mapeamento do PROBART.

Após o treinamento da rede FAM, a matriz de pesos do campo de mapeamento agora possui as informações sobre a frequência com o qual os pares das categorias ART_a e ART_b estão associados. Desta forma, com a implementação do campo de mapeamento do PROBART, se $w_{jk}^{ab} = l$, então l representa a quantidade de vezes que os padrões $\mathbf{I}^a$ foram associados à j -ésima categoria do módulo ART_a , de modo que todos os padrões são vinculados a k -ésima categoria do módulo ART_b e $\mathbf{w}_k^b$ é a classe desses padrões.

Com essas frequências armazenadas, é possível calcular a medida de impureza (ou heterogeneidade) da distribuição de classes das categorias do módulo ART_a e assim determinar quais categorias podem se tornar categorias de rejeição. A medida utilizada para calcular a impureza das categorias é o índice de Gini que representa a desigualdade de uma distribuição. O índice de Gini adaptado para a proposta é expresso por

$$\text{Índice de Gini} = IG_i = 1 - \sum_{j=1}^c [P(C_j|w_i)]^2 \quad (3.18)$$

em que c representa o número de classes e $P(C_j|w_i)$ representa a probabilidade da j -ésima classe para a i -ésima categoria. A probabilidade $P(C_j|w_i)$ é estimado como

$$P(C_j|w_i) \approx \frac{n_{ij}}{n_i} \quad (3.19)$$

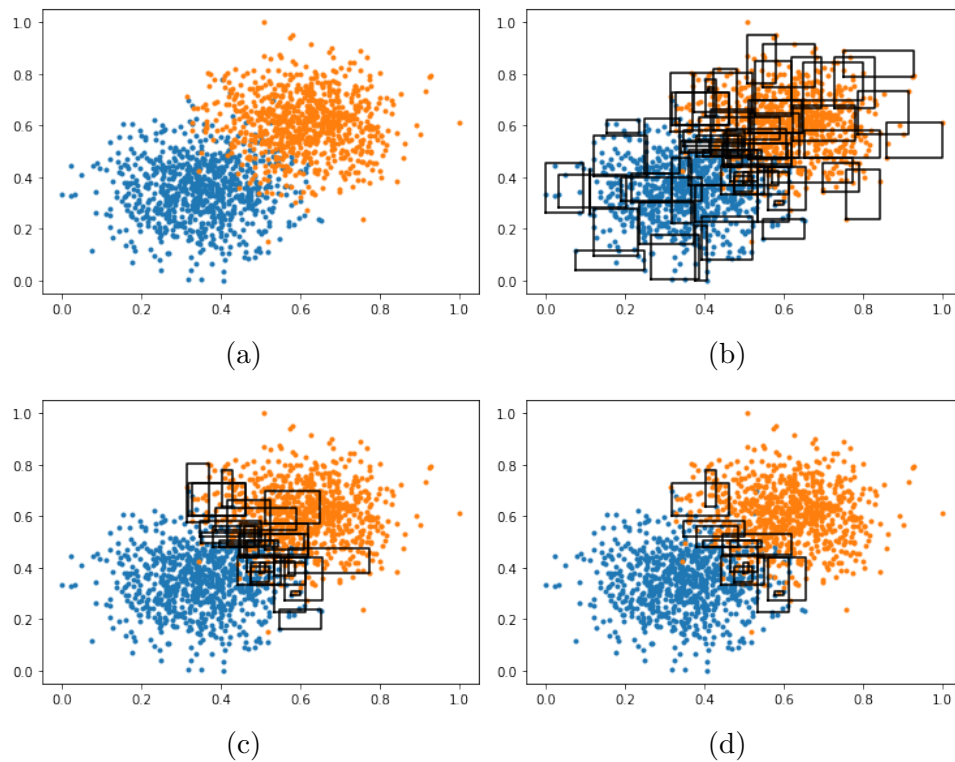
em que n_{ij} é o número de exemplos da j -ésima classe mapeados na i -ésima categoria e n_i é o número total de exemplos mapeados na i -ésima categoria. Por exemplo, supondo que uma categoria mapeou 40 padrões, sendo que 10 para a classe C_{+1} e 30 para a classe C_{-1} . Neste caso, $P(C_{+1}|w_i) = 10/40 = 0.25$, $P(C_{-1}|w_i) = 30/40 = 0.75$ e $IG = 1 - [(0.25^2) + (0.75^2)] = 0.375$.

Para o índice de Gini, o valor 0 significa que todos os rótulos dos padrões mapeados por uma determinada categoria são iguais, por outro lado o valor 1 indica que todos os rótulos são diferentes. O limiar de impureza (ou rejeição) é definido em um processo de otimização que busca minimizar a Equação (3.15). Esse limiar é o responsável por transformar categorias padrões em categorias de rejeição.

Para essa primeira proposta é otimizado apenas um único limiar (regra de Chow) para todas as classes. Desta forma, uma categoria é dita categoria de rejeição se o índice de Gini for maior que esse limiar. Na Figura 15 a seguir, temos um exemplo da influência desse limiar de rejeição. Analisando a imagem é possível vê que quando um limiar é baixo (Fig. 15.c) a quantidade de categoria de rejeição é maior do que um limiar alto (Fig. 15.d).

De maneira geral, a primeira proposta se baseia em utilizar uma única rede *Fuzzy* ARTMAP com a modificação do campo de mapeamento do PROBART e após a etapa de treinamento padrão da rede aplicar a opção de rejeição, buscando minimizar a Equação (3.15) e desta forma encontrar o melhor limiar para a criação das categorias de rejeição.

Figura 15 – (a) Conjunto de dados artificial; (b) Categorias criadas pela a rede FAM; Categorias de rejeição criadas usando um valor de limiar de (c) 0.15 e (d) 0.45



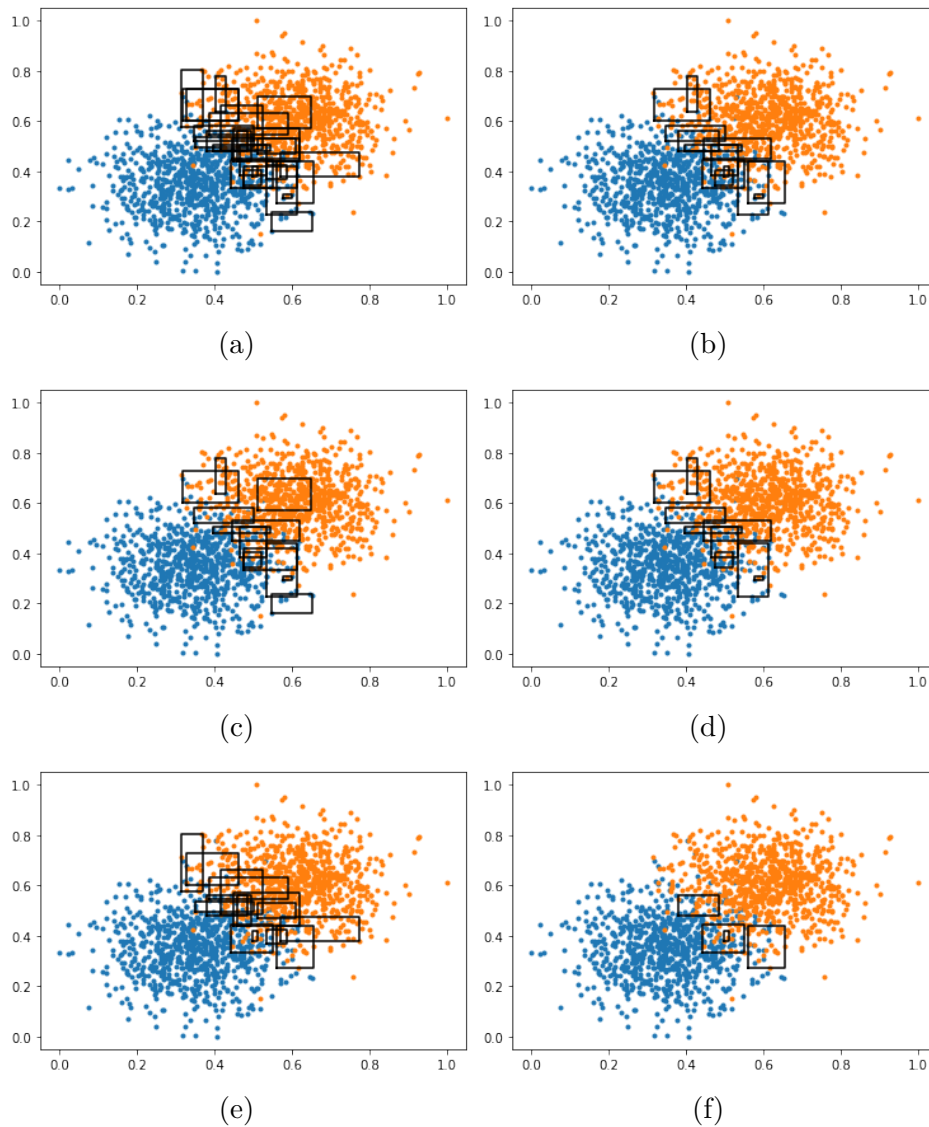
Fonte: Autor.

3.5.3 Usando um Limiar Para Cada Classe (Proposta 2)

A segunda proposta deste trabalho para a classificação com opção de rejeição na rede *Fuzzy* ARTMAP segue a mesma ideia da primeira proposta. A diferença está no final do processo, pois para esta segunda proposta são usados um limiar para cada classe, ou seja, em vez de usar um único limiar para todas as classes, são utilizados limiares independentes para cada classe. Pois como foi visto na seção 3.3 ao utilizar um limiar de rejeição para cada classe pode-se obter um melhor valor de limiar de rejeição.

Dessa forma, o processo de otimização busca definir o melhor limiar para cada classe do problema. Assim, a rede FAM cria categorias de rejeição específicas para cada classe. Utilizando o mesmo exemplo da Figura 15, a Figura 16 a seguir mostra a diferença entre usar um único limiar e um limiar para cada classe. A Figura 16.a e 16.b apresentam categorias de rejeição usando apenas um único limiar para todas as classes. As figuras 16.c e 16.e mostram categorias de rejeição para as classes C_{+1} e C_{-1} , respectivamente, usando um valor de limiar de 0.15. Já as figuras 16.d e 16.f mostram categorias de rejeição para as classes C_{+1} e C_{-1} respectivamente, para um valor de limiar de 0.45.

Figura 16 – Categorias de rejeição usando um valor de limiar de 0.15 (a, c e e) e de 0.45 (b, d e f); Categorias de rejeição criadas usando um único limiar (a e b); Categorias de rejeição criadas para a classe C_{+1} (c e d) e para a classe C_{-1} (e e f)



Fonte: Autor.

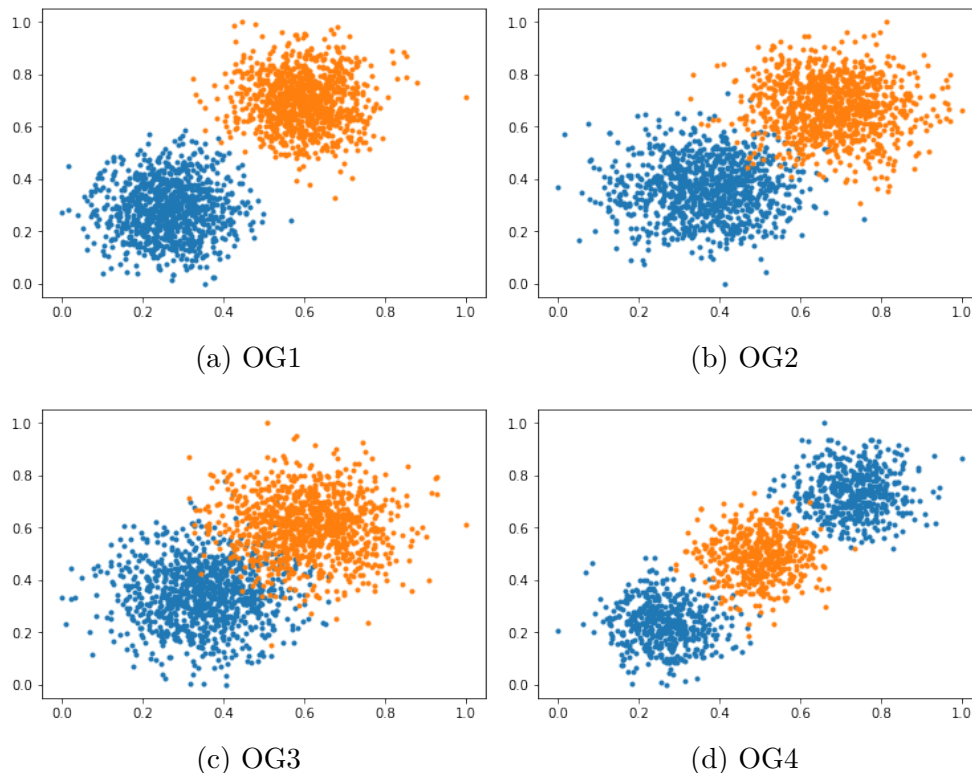
4 Simulações Computacionais

Neste capítulo são apresentadas as características dos conjuntos de dados escolhidos para validação do desempenho dos modelos propostos em relação à arquitetura original. O modelo usando a regra de Chow foi chamado de FAM-C e o do Fumera foi chamado de FAM-F. Também são apresentados os resultados das simulações computacionais realizadas. Para finalizar, é realizada uma análise sobre os resultados, bem como uma comparação entre os métodos propostos e os classificadores MLP, SVM e SOM com opção de rejeição.

4.1 Metodologia

Como prova de conceito, foram realizadas simulações utilizando bases de dados criadas artificialmente. Essas bases foram projetadas com diferentes graus de sobreposição. Desta forma, cada conjunto apresenta distintos níveis de dificuldade para as redes ARTMAP em estudo. Foram gerados quatro conjuntos de dados artificiais compostos por distribuições Gaussianas 2-dimensionais. A disposição dos dados de cada conjunto pode ser vista na figura a seguir.

Figura 17 – Conjunto de Dados Sintéticos



Fonte: Autor.

Os conjuntos de dados OG1, OG2 e OG3, são formados por duas classes com médias iguais a $\mu_1 = (0.3, 0.3)$ e $\mu_2 = (0.7, 0.7)$, ambas com 500 instâncias cada. Os valores da diagonal principal da matriz de covariância foram gradativamente aumentados, estabelecendo dessa forma os graus de sobreposição das classes. Para os conjuntos de dados OG1, OG2 e OG3 os valores utilizados na diagonal principal da matriz de covariância foram de 0.01, 0.02 e 0.03, respectivamente.

No conjunto artificial OG4, a sobreposição entre as classes encontra-se em dois lugares diferente no espaço de característica. Uma classe possui duas distribuições Gaussianas com médias $\mu_{11} = (0.2, 0.2)$ e $\mu_{12} = (0.7, 0.7)$, enquanto a outra classe possui média $\mu_2 = (0.5, 0.5)$. O valor da diagonal principal da matriz de covariância para todas as distribuições foi de 0.02.

Também foram realizadas simulações com os conjuntos de dados reais binários, que estão disponíveis no repositório de aprendizado de máquinas da UCI (LICHMAN, 2013). De todos os conjuntos de dados disponíveis, apenas 7 foram utilizados, sendo eles: *Heart Disease*, *Vertebral Column Pathology*, *Statlog Heart*, *Ionosphere*, *Ripley*, *Pima Indian Diabetes*, e *Sonar*. A Tabela 1 apresenta a denominação, abreviação, número de padrões e o número de características dos conjuntos de dados utilizados nos experimentos.

Tabela 1 – Conjuntos de dados utilizados neste trabalho.

Conjunto de dados	Abreviação	Padrões	Características
<i>Heart Disease</i>	HEA	270	13
<i>Vertebral Column Pathology</i>	VCP	310	6
<i>Statlog Heart</i>	SLH	270	13
<i>Ionosphere</i>	ION	351	34
<i>Ripley</i>	RIP	1250	2
<i>Pima Indian Diabetes</i>	PID	768	8
<i>Sonar</i>	SON	208	60

Todos os conjuntos de dados foram submetidos a uma etapa de pré-processamento, conhecida como normalização, fazendo com que valores medidos em escalas diferentes sejam ajustados para a mesma escala. Existem diferentes tipos de normalização, como por exemplo a normalização *min-max*, normalização de média zero e a normalização por escalonamento decimal. Nos experimentos realizados utilizou-se a normalização *min-max*, gerando valores correspondentes aos originais dentro de um intervalo entre 0 e 1, colocando todos os valores existentes em uma mesma escala.

Em todos os experimentos com as redes FAM, o aprendizado rápido foi empregado, ou seja, $\beta = 1$. Para os módulos ART_b e o inter-ART foram atribuídos os valores padrões ρ_b e $\rho_{ab} = 1$ e o valor de $\eta = 0.0001$. Já o valor do parâmetro de vigilância no módulo ART_a é encontrado utilizando a busca em grade (*grid search*) combinada com a técnica de validação cruzada (*cross-validation*).

Na validação cruzada utilizou-se dos métodos *holdout* e *k-fold* nos experimentos. O método *holdout* com uma divisão de 80% dos dados para o conjunto de treinamento e os outros 20% restante para o conjunto de teste. O método *k-fold*, com cinco divisões ($k = 5$) foi utilizado na otimização do parâmetro de vigilância ρ_a no intervalo de 0 até 0.97, a partir dos 80% de dados do subconjunto de treinamento.

4.2 Resultados e Discussões

O objetivo deste primeiro experimento é avaliar as estratégias propostas de classificação com opção de rejeição, em relação à rede FAM original. Os resultados deste experimento pode ser visto na Tabela 2.

Tabela 2 – Resultados para os conjuntos de dados artificiais.

MODELOS	METRICA	OG1	OG2	OG3	OG4
FAM	ACC	99.75±00.20	96.94±00.77	92.43±01.03	97.24±00.68
FAM-C _{w_r=0.04}	ACC	99.79±00.19	99.12±00.75	98.73±00.80	98.77±00.60
	REJ	00.35	12.98	25.48	11.22
FAM-F _{w_r=0.04}	ACC	99.78±00.20	99.23±00.42	98.54±00.93	98.69±00.65
	REJ	00.42	14.12	27.66	12.64
FAM-C _{w_r=0.24}	ACC	99.77±00.20	98.27±00.67	97.15±01.27	98.17±00.75
	REJ	00.18	03.57	12.26	02.43
FAM-F _{w_r=0.24}	ACC	99.77±00.20	98.33±00.69	97.18±01.13	98.15±00.69
	REJ	00.19	03.89	12.50	02.29
FAM-C _{w_r=0.48}	ACC	99.76±00.21	97.82±00.55	95.48±01.13	98.11±00.71
	REJ	00.08	02.15	06.19	01.74
FAM-F _{w_r=0.48}	ACC	99.76±00.21	97.86±00.54	95.60±01.14	98.11±00.71
	REJ	00.08	02.22	06.46	01.80

Como pode ser visto na Tabela 2, os métodos propostos com opção de rejeição apresentaram um melhor desempenho com relação ao original, ou seja, melhores taxas de acurácia. Para o conjunto de dados OG1, não ocorre uma variação significativa na acurácia dos modelos, visto que esse conjunto não apresenta um alto grau de sobreposição de dados. Desta forma, ambos os modelos apresentam uma acurácia similar, porém, com acurácias maiores que o modelo original. Já para os conjuntos de dados OG2 e OG3, os modelos apresentam uma vantagem superior ao modelo original. Para o conjunto de dados OG2, o modelo original apresentou uma acurácia de 96.94%, já para os modelos FAM-C e FAM-F com um custo de rejeição baixo ($w_r = 0.48$) os valores das acurácias foram de 97.82% e 97.86%, respectivamente. Já aumentando o custo de rejeição ($w_r = 0.04$) esses valores aumentam para 99.12% e 99.23%, respectivamente. Para o conjunto OG3, o modelo original apresentou uma acurácia de 92.43% e os modelos FAM-C e FAM-F com um custo de rejeição baixo apresentaram uma acurácia respectivamente de 95.48% e 95.60%.

Já usando um custo de rejeição alto, os valores de acurácia foram de 98.73% e 98.54% respectivamente. Para o conjunto de dados OG4, no qual a sobreposição se encontra em dois lugares diferentes no espaço de características, o modelo original apresentou uma acurácia de 97.24%. Já os modelos FAM-C e FAM-F, ambos apresentaram uma acurácia de 98.11% com um custo de rejeição baixo e acurácias de 98.77% e 98.69%, respectivamente, com um custo de rejeição alto.

Em casos de classificação com opção de rejeição, a acurácia depende da taxa de rejeição, ou seja, diferentes taxas de rejeição tem efeito diferente no desempenho do classificador. Desta forma, a avaliação do desempenho do classificador com capacidade de rejeição pode ser descrito através de uma curva de acurácia-rejeição (*Accuracy-Rejection Curve* - ARC) que é uma função que representa a taxa de acerto de um classificador em função da sua taxa de rejeição.

Dessa forma, foram plotadas as curvas A-R dos modelos propostos para cada conjunto de dados artificiais. Cada ponto em destaque na curva corresponde a um determinado valor de W_r . Os valores de W_r considerados foram de 0.04, 0.24 e 0.48. Na figura a seguir (Figura 18), são apresentadas as curvas A-R dos modelos propostos.

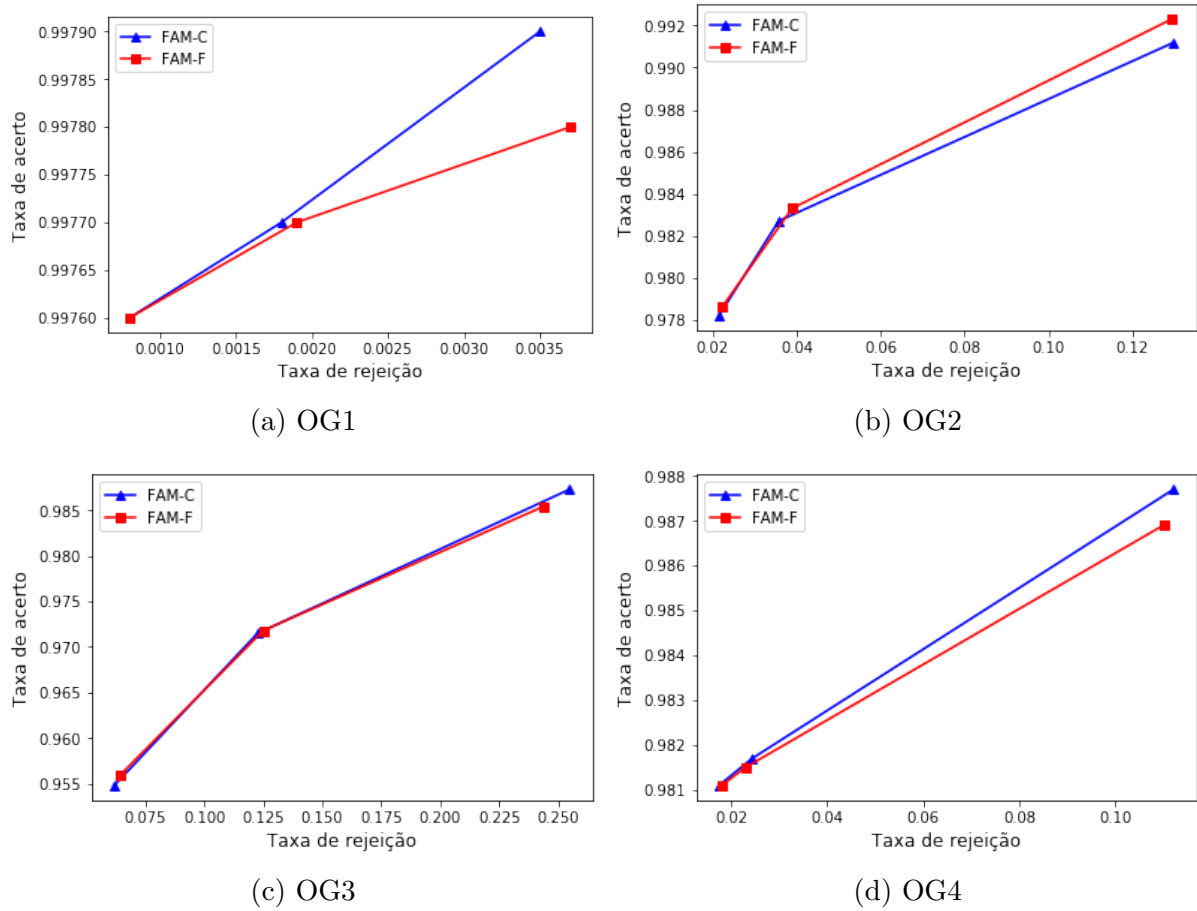
Pode ser observado na Figura 18 que para valores elevados da taxa de rejeição, também se verifica valores elevados da acurácia. Pode ser observado também que os modelos avaliados obtêm resultados similares, porém apesar da similaridade entre os resultados dos modelos FAM-C e FAM-F, o primeiro apresenta resultados um pouco melhores, pois em geral a sua curva esta posicionada sobre a curva do modelo FAM-F.

Também foi verificado a quantidade de categorias criadas por cada modelo para os diferentes valores do custo de rejeição, dessa forma, na Tabela 3 é apresentada a quantidade média de categorias criadas por cada modelo (N_c) e a quantidade média de categorias de rejeição (N_r) para as simulações com os conjuntos de dados artificiais.

Tabela 3 – Categorias criadas pelos os modelos. Média de categorias criadas (N_c) e a média de categorias de rejeição (N_r)

Modelos	OG1		OG2		OG3		OG4	
	N_c	N_r	N_c	N_r	N_c	N_r	N_c	N_r
FAM	32.80	-	36.30	-	40.13	-	23.13	-
FAM-C _{$w_r=0.04$}	31.20	1.60	30.37	05.93	08.53	31.60	16.26	06.87
FAM-C _{$w_r=0.24$}	32.13	0.67	34.17	02.13	24.66	15.57	18.53	04.60
FAM-C _{$w_r=0.48$}	32.30	0.50	34.63	01.67	31.80	08.33	20.73	02.40
FAM-F _{$w_r=0.04$}	30.57	2.23	30.23	06.07	10.33	29.80	17.23	05.90
FAM-F _{$w_r=0.24$}	31.54	1.26	33.30	02.67	28.06	12.07	18.53	04.60
FAM-F _{$w_r=0.48$}	32.21	0.59	34.63	01.67	35.19	04.94	21.73	01.14

Figura 18 – Curvas A-R dos modelos FAM-C e FAM-F para cada base de dados artificiais.



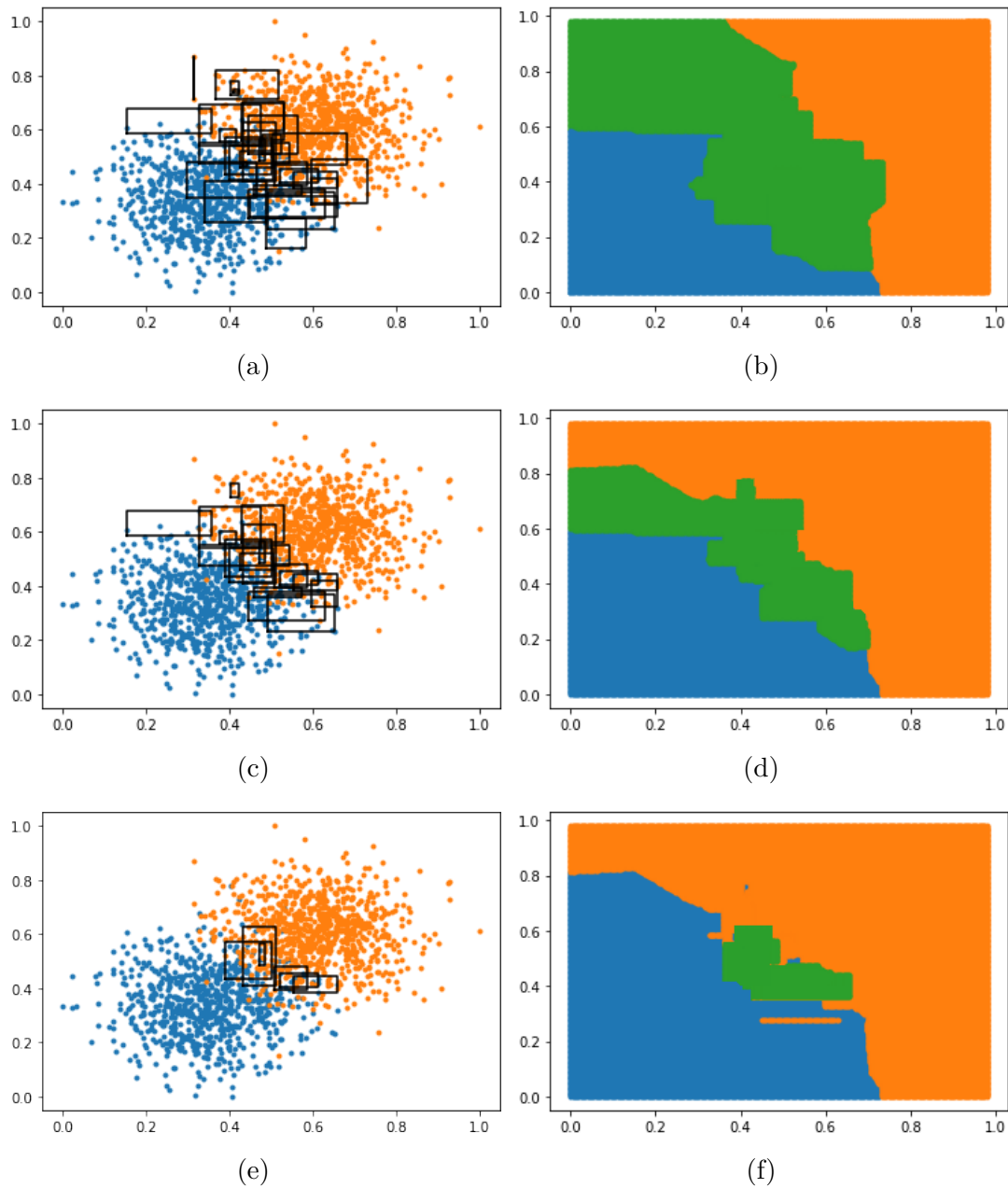
Fonte: Autor.

Observando a Tabela 3, é possível notar que para os conjuntos de dados OG1, OG2 e OG3, a quantidade média de categorias criadas (N_c) pelo modelo FAM aumenta à medida que o grau de sobreposição entre os dados aumenta. Observe também que para os modelos propostos, à medida que o valor do custo de rejeição aumenta a quantidade de categorias médias criadas (N_c) diminui e as categorias de rejeição (N_r) cresce.

Colocando em números, para os conjuntos de dados OG2, OG3 e OG4 a quantidade média de categorias criadas pela rede FAM é de 36.60, 40.13 e 23.13 respectivamente. No modelo FAM-C usando um custo de rejeição alto, a quantidade média de categorias criadas foram de 30.37, 8.53 e 16.26 e usando um custo de rejeição baixa, a média de categorias foram de 34.63, 31.80 e 20.73. Já para o modelo FAM-F, foram criadas 30.23, 10.33 e 17.23 usando um custo de rejeição baixa e usando um valor alto, foram criadas 34.63, 35.19 e 21.73. Para o conjunto de dados OG1, o número de categorias criadas é praticamente o mesmo para todos os modelos e para os diferente valores de custo de rejeição, visto que esse conjunto não apresenta um alto grau de sobreposição.

Dado que o modelo FAM-C obteve um desempenho um pouco melhor que o modelo FAM-F, na Figura 19 temos a superfície de decisão deste modelo para a base de dados OG3. Através dessa figura, fica claro perceber as regiões de decisão do modelo. Observe que variando o custo de rejeição, a quantidade de categorias de rejeição presentes na região de sobreposição também varia. Devido a essa variação na quantidade de categorias, a região de rejeição (cor verde) aumenta ou diminui.

Figura 19 – Categorias de rejeição do modelo FAM-C (a, c e e) com custo de rejeição w_r igual a (a, b) 0.04, (c, d) 0.24 e (e, f) 0.48 e (b, d e f) superfície de decisão com a região de rejeição em verde.



Fonte: Autor.

As Tabelas 4, 5 e 6 apresentam os resultados das simulações dos modelos para os valores de $w_r = 0.04$, 0.24 , e 0.48 , respectivamente. Esses resultados são dos conjuntos de dados reais binários que foram apresentados na Tabela 1. A avaliação de desempenho envolve a acurácia e a taxa de rejeição, que para este experimento foram obtidas através do cômputo de 30 realizações independentes, em que as amostras de treinamento e de teste foram escolhidas aleatoriamente.

Observando a Tabela 4, pode-se inferir que os modelos propostos apresentaram as melhores taxas de acerto do que a rede FAM em todos os conjuntos de dados. Para esse valor de custo de rejeição, a rede FAM-C apresentou as melhores taxas de acerto do que a rede FAM-F nos conjuntos de dados HEA, RIP e SON. Já a rede FAM-F apresentou as melhores taxas de acertos nos conjuntos de dados SLH, ION e PID. E ambos os modelos apresentaram a mesma taxa de acerto para o conjunto de dados VCP.

Tabela 4 – Resultados das simulações com $W_r = 0.04$

DATASET	METRICAS	FAM	FAM-C	FAM-F
HEA	ACC	78.364±04.36	84.40±07.28	80.64±21.88
	REJ	-	70.91	77.94
VCP	ACC	78.36±04.17	96.84±03.79	96.84±03.79
	REJ		77.39	77.39
SLH	ACC	77.94±04.22	87.05±13.98	88.79±10.74
	REJ		74.30	76.61
ION	ACC	90.28±03.61	94.42±06.85	95.04±06.17
	REJ		70.19	69.77
RIP	ACC	88.92±01.87	96.54±02.11	96.46±02.17
	REJ		51.90	51.75
PID	ACC	70.41±03.72	85.71±09.66	86.30±10.72
	REJ		83.23	84.33
SON	ACC	81.47±05.55	87.03±11.53	85.35±15.85
	REJ		66.88	76.67

Na Tabela 5, estão os resultados das simulações com um custo de rejeição de 0.24 . Observando essa tabela, é possível dizer que o modelo FAM-F obteve as melhores taxas de acertos, perdendo apenas nos conjuntos de dados HEA e RIP para o modelo FAM-C. Já na Tabela 6 estão os resultados das simulações com um custo de rejeição de 0.48 . Nessa tabela, o modelo FAM-C apresentou as melhores taxas de acertos para os conjuntos de dados SLH e ION. O modelo FAM-F apresentou as melhores taxas de acertos para os conjuntos de dados VCP e RIP. E ambos os modelos apresentaram a mesma taxa de acertos para os conjuntos de dados HEA, PID e SON. De uma maneira geral, os modelos propostos apresentaram taxas de acurácias melhores do que a rede FAM e ambos se mostraram bastantes similares em seus resultados.

Tabela 5 – Resultados das simulações com $W_r = 0.24$

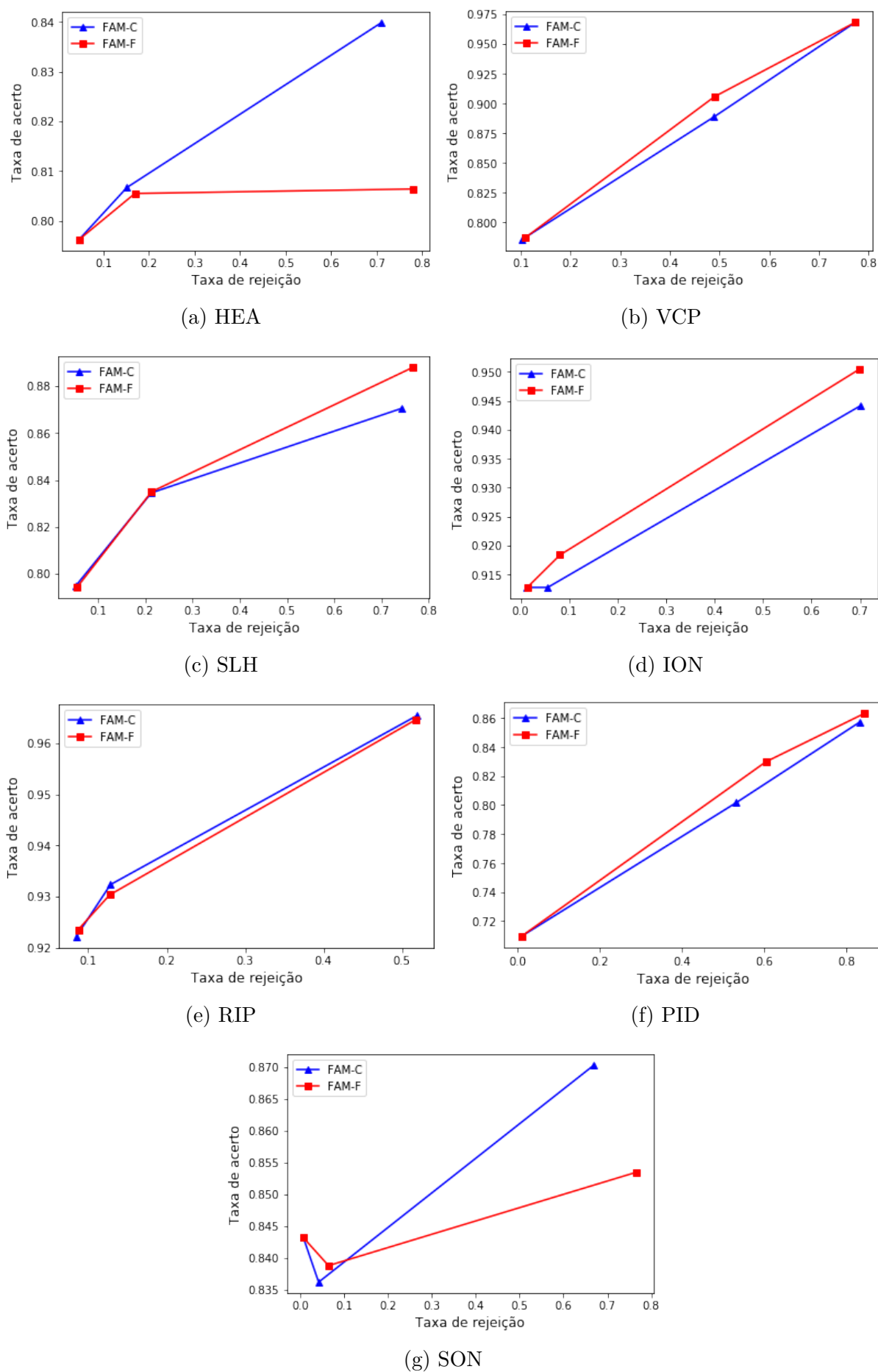
DATASET	METRICAS	FAM	FAM-C	FAM-F
HEA	ACC	78.36±04.36	80.67±04.60	80.55±04.62
	REJ	-	15.15	16.91
VCP	ACC	78.36±04.17	88.88±07.82	90.58±07.17
	REJ		48.89	48.99
SLH	ACC	77.94±04.22	83.45±05.82	83.50±06.16
	REJ		21.27	21.33
ION	ACC	90.28±03.61	91.28±04.36	91.84±04.57
	REJ		05.59	08.12
RIP	ACC	88.92±01.87	93.23±01.64	93.03±01.67
	REJ		12.83	12.74
PID	ACC	70.41±03.72	80.16±10.63	82.97±08.89
	REJ		53.18	60.37
SON	ACC	81.47±05.55	83.62±05.77	83.88±05.88
	REJ		04.26	06.51

Tabela 6 – Resultados das simulações com $W_r = 0.48$

DATASET	METRICAS	FAM	FAM-C	FAM-F
HEA	ACC	78.36±04.36	79.62±05.16	79.62±05.16
	REJ	-	04.61	04.61
VCP	ACC	78.36±04.17	78.56±04.89	78.71±05.12
	REJ		10.21	10.85
SLH	ACC	77.94±04.22	79.46±05.62	79.44±05.62
	REJ		05.15	05.58
ION	ACC	90.28±03.61	91.28±03.80	91.28±03.84
	REJ		01.22	01.36
RIP	ACC	88.92±01.87	92.21±01.55	92.35±01.50
	REJ		08.46	08.73
PID	ACC	70.41±03.72	70.95±03.78	70.95±03.78
	REJ		00.90	00.90
SON	ACC	81.47±05.55	84.32±05.86	84.32±05.86
	REJ		00.78	00.78

Como já mencionado, diferentes taxas de rejeição têm efeitos diferentes no desempenho do classificador e como foi visto nas tabelas anteriores (Tabelas 4, 5 e 6), os modelos propostos apresentaram taxas de acertos similares. Dessa forma, utilizando a curva de acurácia-rejeição, é possível distinguir qual classificador possui o melhor desempenho. Na Figura 18, pode ser visto as curvas A-R dos modelos propostos para cada conjunto de dados reais binários. Observando as imagens da Figura 20, é possível perceber que o modelo FAM-F apresentou uma leve vantagem com relação ao modelo FAM-C, pois a sua curva está posicionada sobre a curva do modelo FAM-C (observe as Figuras b, c, d e f) na maioria dos casos.

Figura 20 – Curvas A-R dos modelos propostos para os bancos de dados reais.



Fonte: Autor.

Assim como foi feito nos conjuntos de dados artificiais, também foi verificada a quantidade de categorias criadas por cada modelo para os conjuntos de dados reais. Na Tabela 7, é mostrada a quantidade média de categorias criadas por cada modelo e a quantidade média de categorias de rejeição.

Tabela 7 – Média das categorias criadas para os conjuntos de dados reais

Dataset	Metrica	FAM	FAM-C			FAM-F		
			W_r			W_r		
			0.04	0.24	0.48	0.04	0.24	0.48
HEA	N_c	76.47	56.67	65.80	68.80	48.73	69.34	71.14
	N_r	-	19.80	10.67	07.67	27.74	07.13	05.33
VCP	N_c	97.20	73.20	80.33	80.53	73.40	80.23	80.67
	N_r	-	24.00	16.87	16.53	23.80	16.97	16.53
SLH	N_c	46.47	30.24	39.94	40.27	30.57	39.33	40.20
	N_r	-	16.23	06.53	06.20	15.90	07.14	06.27
ION	N_c	161.93	140.46	159.00	160.13	141.86	159.10	160.06
	N_r	-	21.47	02.93	01.80	22.07	02.83	01.87
RIP	N_c	61.27	15.87	40.30	56.77	17.54	37.67	58.01
	N_r	-	45.40	20.97	04.50	43.73	23.60	03.26
PID	N_c	167.73	127.53	142.86	143.46	124.53	142.81	143.47
	N_r	-	40.20	24.87	24.27	43.20	24.92	24.26
SON	N_c	97.87	76.98	92.24	97.57	82.31	94.17	97.57
	N_r	-	20.89	05.63	00.30	15.56	03.70	00.30

Analisando a Tabela 7, pode ser visto o mesmo efeito observado nos conjuntos de dados artificiais: quando o valor do custo de rejeição é baixo, mais categorias são criadas; e quando o valor do custo de rejeição é alto, menos categorias são criadas. O inverso é visto nas categorias de rejeição: quanto maior o custo de rejeitar, mais categorias de rejeição são criadas; e quanto menor o custo de rejeitar, menos categorias de rejeição são criadas.

Vamos agora comparar a quantidade média de categorias criadas pela rede FAM com os outros modelos propostos: quando o custo de rejeição é alto, os modelos propostos apresentaram os menores valores e à medida que o custo de rejeição diminui, a quantidade média de categorias criadas aumenta, porém se mantendo abaixo do valor da rede FAM. Por exemplo, para o conjunto de dados ION, a rede FAM criou uma média de 161.93 categorias enquanto os modelos FAM-C e FAM-F, usando um custo de rejeição alto, criaram uma média de 140.46 e 141.86 categorias, respectivamente. Já usando um custo de rejeição baixo, os modelos apresentaram uma média de 160.13 e 160.06 categorias, respectivamente.

Também foi executado um teste estatístico de hipótese; tal teste visa estabelecer os limites além dos quais duas amostras devem ou não ser consideradas pertencentes a uma mesma população. Dito isso, foi adotado um teste não paramétrico chamado teste de

Friedman, que não depende de qualquer suposição sobre a distribuição dos dados. Dado que a hipótese nula significa dizer que todos os algoritmos são equivalentes e a hipótese alternativa diz exatamente o contrário; se o teste fornecer um nível de significância menor ou igual ao valor limite escolhido (neste caso, 5%), o teste sugere que os dados observados são inconsistentes com a hipótese nula e, desta forma, a hipótese nula é rejeitada.

Na Tabela 8, é apresentado os resultados do teste estatístico que confronta os valores da taxa de acerto com o resultado obtido pelo outros classificadores com a opção de rejeição. Em relação aos parâmetros dos modelos, eles foram otimizados através de validação cruzada em *k-folds* com $k = 5$. Para o classificador MLP, foi realizado sobre o número de neurônios ocultos que devem compor a rede. Os valores analisados foram de 5 até 25 neurônios com passo de 5 na única camada oculta. Para o classificador SVM foi usado o *kernel* rbf e a validação cruzada sobre os parâmetros C e gama no intervalo 2^{-5} até 2^5 com passos de 1 e 0.1, respectivamente. Já para o classificador SOM, para obter o número de neurônios dentre os mapas 4x4 até 14x14 com passo de 2. A taxa de aprendizado e a função de vizinhança possuem um decaimento exponencial em função das iterações de treinamento. O valor inicial e final é de 0.5 até 0.001, respectivamente. A mesma estratégia de opção de rejeição usada nas propostas foi adotada para os classificadores MLP, SVM e SOM, contudo apenas a regra Chow foi usada dando origem a MLP-C, SVM-C e SOM-C.

Através da análise da Tabela 8, pode-se inferir que os métodos se mostraram suficientemente estáveis em termos de taxa de acerto, já que as propostas obtêm valores médios de taxa de acerto mais próximos aos métodos com os maiores valores para os conjuntos de dados VCP, ION e RIP usando um custo de rejeição de 0.04 e para os conjuntos de dados ION, RIP e SON usando um custo de rejeição de 0.24 e 0.48.

Olhando para os resultados estatísticos dessa tabela, pode-se inferir que ambos os modelos propostos são equivalentes entre si para todos os conjuntos de dados usando os diferentes valores de custo de rejeição. Também é possível notar que ambos os modelos propostos são equivalentes ou melhores em comparação a rede SOM-C (veja os resultados para os conjunto de dados HEA, VCP, SLH e SON). Em comparação com o modelo MLP-C, o modelo FAM-C é equivalente para os conjuntos de dados ION e SON e o modelo FAM-F é equivalente apenas para o conjunto de dados SON. No entanto, o desempenho de ambos os modelos são, em geral, piores que os resultados obtidos pelo modelo SVM-C.

Ainda baseado no teste de hipótese, ambos os modelos apresentaram alguns desempenhos estatisticamente diferentes dos outros métodos; em particular, o modelo FAM-C para os conjuntos de dados RIP e PID usando um custo de rejeição 0.24 e 0.48, respectivamente. Já o modelo FAM-F para os conjuntos de dados RIP usando um custo de rejeição de 0.24 e para o conjunto de dados PID usando um custo de rejeição de 0.04 e 0.48.

Tabela 8 – Taxa de acerto (ACC); Custo de rejeição (W_r); Modelo FAM-C (M1); Modelo FAM-F (M2); e resultados dos testes estatísticos. Os símbolos ✓ e ✗ significam equivalência ou não equivalência, respectivamente.

Dataset	Modelos	ACC $W_r=0.04$	M1	M2	ACC $W_r=0.24$	M1	M2	ACC $W_r=0.48$	M1	M2
HEA	FAM-C	83.98±14.69	-	✓	80.67±04.60	-	✓	79.62±05.16	-	✓
	FAM-F	80.64±21.88	✓	-	80.55±04.62	✓	-	79.62±05.16	✓	-
	MLP-C	96.43±17.95	✗	✗	87.87±04.10	✗	✗	85.56±04.33	✗	✗
	SVM-C	97.83±04.51	✗	✗	87.31±04.27	✗	✗	85.04±04.07	✗	✗
	SOM-C	92.48±07.82	✓	✓	82.47±05.64	✓	✓	81.21±05.57	✓	✓
VCP	FAM-C	96.84±03.79	-	✓	88.88±07.82	-	✓	78.56±04.89	-	✓
	FAM-F	96.84±03.79	✓	-	90.58±07.17	✓	-	78.71±05.12	✓	-
	MLP-C	99.16±02.57	✓	✓	96.61±04.56	✗	✓	93.72±05.88	✗	✗
	SVM-C	96.89±03.42	✓	✓	90.96±04.54	✓	✓	85.70±04.42	✗	✗
	SOM-C	95.05±05.59	✓	✓	87.87±07.15	✓	✓	80.38±06.20	✓	✓
SLH	FAM-C	87.05±13.98	-	✓	83.45±05.82	-	✓	79.46±05.62	-	✓
	FAM-F	88.79±10.74	✓	-	83.50±06.16	✓	-	79.44±05.62	✓	-
	MLP-C	97.76±04.75	✗	✓	93.35±06.48	✗	✗	90.50±06.99	✗	✗
	SVM-C	96.52±06.59	✗	✓	87.59±03.37	✗	✗	84.03±03.15	✗	✗
	SOM-C	88.11±08.35	✓	✓	81.60±05.35	✓	✓	81.09±04.78	✓	✓
ION	FAM-C	94.42±06.85	-	✓	91.28±04.36	-	✓	91.28±03.80	-	✓
	FAM-F	95.04±06.17	✓	-	91.84±04.57	✓	-	91.28±03.84	✓	-
	MLP-C	94.15±03.65	✓	✓	92.31±04.07	✓	✗	91.49±03.98	✓	✓
	SVM-C	97.36±02.39	✗	✓	96.60±02.88	✗	✗	95.27±03.02	✗	✗
	SOM-C	96.68±04.78	✗	✓	90.18±03.61	✓	✓	90.34±03.09	✓	✓
RIP	FAM-C	96.54±02.11	-	✓	93.23±01.64	-	✓	92.21±01.55	-	✓
	FAM-F	96.46±02.17	✓	-	93.03±01.67	✓	-	92.35±01.50	✓	-
	MLP-C	99.56±00.88	✗	✗	94.76±01.65	✗	✗	90.40±01.58	✗	✗
	SVM-C	96.91±01.13	✓	✓	95.62±01.23	✗	✗	91.77±01.45	✓	✓
	SOM-C	98.27±01.23	✗	✗	94.82±01.73	✗	✗	92.86±01.94	✓	✓
PID	FAM-C	85.71±09.66	-	✓	80.16±10.63	-	✓	70.95±03.78	-	✓
	FAM-F	86.30±10.72	✓	-	82.97±08.89	✓	-	70.95±03.78	✓	-
	MLP-C	93.73±11.73	✗	✗	91.44±08.82	✗	✓	86.69±10.06	✗	✗
	SVM-C	93.23±06.85	✓	✗	86.42±04.15	✓	✓	79.59±02.82	✗	✗
	SOM-C	96.61±03.75	✗	✗	96.42±04.00	✗	✗	77.06±03.85	✗	✗
SON	FAM-C	87.03±11.73	-	✓	83.62±05.77	-	✓	84.32±05.86	-	✓
	FAM-F	85.35±15.85	✓	-	83.88±05.88	✓	-	84.32±05.86	✓	-
	MLP-C	86.17±06.99	✓	✓	84.75±06.55	✓	✓	83.73±06.53	✓	✓
	SVM-C	89.35±04.52	✓	✓	88.73±04.97	✗	✗	88.03±05.44	✗	✗
	SOM-C	92.21±06.49	✗	✗	89.07±05.58	✗	✗	86.64±06.17	✓	✓

5 Conclusão

Neste capítulo são apresentadas as conclusões, as contribuições científicas dos trabalhos realizados nesta dissertação e os trabalhos futuros que podem ser alcançados através de uma visão mais aprofundada dos métodos aqui propostos.

5.1 Conclusões

Nesta dissertação de mestrado foi apresentado a rede neural *Fuzzy ARTMAP* que tem como base a teoria da ressonância adaptativa. A rede *Fuzzy ARTMAP* foi detalhada em sua metodologia de aprendizado e como o mesmo aprende um determinado conjunto de dados utilizando uma síntese geométrica hiper-retangular pré-definida pela sua arquitetura. Também foi detalhado como ocorre a classificação com opção de rejeição e os dois tipos de limiares que foram utilizados para criar a região de rejeição. Além disso, foram mostradas as diferentes abordagens usadas para incorporar a opção de rejeição nos classificadores.

A principal contribuição desta dissertação é o desenvolvimento de uma nova variante da rede *Fuzzy ARTMAP*, com a técnica de opção de rejeição. O método proposto aumenta a performance da rede *Fuzzy ARTMAP* à medida que a taxa de rejeição aumenta. A proposta consiste em utilizar apenas uma única rede *Fuzzy ARTMAP* com uma modificação no campo de mapeamento e logo após o treinamento aplicar mais um passo para encontrar um limiar de rejeição ótimo. O primeiro método utiliza a regra de Chow, que utiliza apenas um único limiar para todas as classes, enquanto que o segundo método é baseado no do Fumera que utiliza um limiar para cada classe.

Esses dois métodos foram usados a fim de obter o melhor limiar para cada conjunto de dados e dessa forma criar categorias de rejeição na rede *Fuzzy ARTMAP*. As propostas foram aplicadas a conjuntos de dados artificiais e em conjuntos de dados reais, disponíveis no repositório de aprendizado de máquinas da UCI. Em um primeiro experimento, os métodos propostos foram comparados à rede FAM sem a opção de rejeição e em um segundo experimento, comparadas aos classificadores MLP, SVM e SOM no qual foram incorporadas a opção de rejeição.

A análise dos resultados do primeiro experimento indica que as propostas apresentadas funcionam de forma bastante adequada, uma vez que alcançam resultados superiores à rede FAM sem a opção de rejeição. Já os resultados do segundo experimento demonstraram que ambos os modelos são equivalentes ou melhores quando comparados à rede SOM; em alguns casos, equivalente a rede MLP e piores que os resultados obtidos pela rede SVM.

5.2 Contribuições deste Trabalho

Do ponto de vista geral deste trabalho, as contribuições foram apresentadas na forma de duas propostas idealizadas a partir do objetivo de aplicar opção de rejeição na rede *Fuzzy ARTMAP*. Dessa forma, criando uma nova variação da rede capaz de gerar uma nova classe: a classe de rejeição.

O detalhamento da rede *Fuzzy ARTMAP* no Capítulo 2 dá ao leitor o arcabouço necessário para a compreensão do funcionamento da rede. Além disso, foram mostrados exemplos de como a rede aprende os dados e como as suas categorias são criadas, bem como sua superfície de decisão. No Capítulo 3 são apresentados como ocorre o processo da classificação com opção de rejeição, bem como a incorporação na rede *Fuzzy ARTMAP*.

Os resultados obtidos em termos de taxa de acerto (acurácia) no Capítulo 4, através de simulações computacionais utilizando conjuntos de dados sintéticos e reais, confirmam que as propostas apresentadas atendem aos objetivos definidos inicialmente.

5.3 Trabalhos Futuros

Como trabalhos futuros, pode-se ainda abordar a seguinte situação: apesar da proposta atender ao objetivo definido, o mesmo é muito sensível aos parâmetros de vigilância do módulo ART_a . Durante os experimentos realizados nesta dissertação, esse parâmetro foi otimizado através da validação cruzada. Entretanto um estudo mais aprofundado para compreender bem o comportamento desse parâmetro de forma genérica se faz necessário para construir uma função geral que compute esse valor antes do treinamento.

Outro trabalho sugerido é um estudo mais aprofundado sobre a aplicação da opção de rejeição para a resolução do problema da proliferação de categorias, pois como foi visto nos resultados, quando se utilizava um custo de rejeição alto, a quantidade de categorias criadas - desconsiderando as categorias de rejeição - eram menores.

Sugere-se ainda a modificação dos métodos propostos no intuito de torná-los classificadores multiclasse. Atualmente, manipula-se apenas problemas binários. Entretanto é interessante estender a abordagem desses modelos para que problemas com mais de duas classes possam ser contemplados.

Para finalizar, outro trabalho a ser considerado é a aplicação da opção de rejeição em outras redes *Fuzzy ARTMAP*, pois como já foi dito, existem diferentes formas geométricas, além dos hiper-retângulos, que são alocadas no espaço de características.

Referências

BISHOP, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN 0387310738. Citado na página 15.

BOUNSIAR, A.; GRALL, E.; BEAUSEROY, P. A kernel based rejection method for supervised classification. In: . [S.l.: s.n.], 2007. Citado na página 40.

CARPENTER, G. A. Distributed art networks for learning, recognition, and prediction. In: *Proceedings of the World Congress on Neural Networks (WCNN'96)*. [S.l.: s.n.], 1996. p. 333–344. Citado na página 30.

CARPENTER, G. A. Distributed learning, recognition, and prediction by art and artmap neural networks. *Neural networks*, Elsevier, v. 10, n. 8, p. 1473–1494, 1997. Citado na página 30.

CARPENTER, G. A.; GROSSBERG, S. Art 2: self-organization of stable category recognition codes for analog input patterns. *Appl. Opt.*, OSA, v. 26, n. 23, p. 4919–4930, Dec 1987. Disponível em: <<http://ao.osa.org/abstract.cfm?URI=ao-26-23-4919>>. Citado na página 15.

CARPENTER, G. A.; GROSSBERG, S. A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, v. 37, p. 54–115, 1987. Citado na página 15.

CARPENTER, G. A.; GROSSBERG, S. Art 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, v. 3, n. 2, p. 129 – 152, 1990. ISSN 0893-6080. Disponível em: <<http://www.sciencedirect.com/science/article/pii/089360809090085Y>>. Citado na página 15.

CARPENTER, G. A. et al. Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, v. 3, n. 5, p. 698–713, Sep. 1992. ISSN 1941-0093. Citado na página 16.

CARPENTER, G. A.; GROSSBERG, S.; REYNOLDS, J. H. Artmap: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, v. 4, n. 5, p. 565 – 588, 1991. ISSN 0893-6080. Disponível em: <<http://www.sciencedirect.com/science/article/pii/089360809190012T>>. Citado na página 16.

CARPENTER, G. A.; GROSSBERG, S.; ROSEN, D. B. Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, v. 4, n. 6, p. 759 – 771, 1991. ISSN 0893-6080. Disponível em: <<http://www.sciencedirect.com/science/article/pii/089360809190056B>>. Citado 2 vezes nas páginas 15 e 23.

- CARPENTER, G. A.; MARKUZON, N. Artmap-ic and medical diagnosis: Instance counting and inconsistent cases. *Neural networks : the official journal of the International Neural Network Society*, v. 11, p. 323–36, 04 1998. Citado na página 17.
- CHOW, C. On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*, v. 16, n. 1, p. 41–46, January 1970. ISSN 0018-9448. Citado 4 vezes nas páginas 17, 33, 34 e 36.
- CHOW, C. K. An optimum character recognition system using decision functions. *IRE Trans. Electronic Computers*, v. 6, p. 247–254, 1957. Citado 5 vezes nas páginas 16, 17, 33, 35 e 36.
- CORDELLA, L. P. et al. An adaptive reject option for lvq classifiers. In: *ICIAP*. [S.l.: s.n.], 1995. Citado na página 18.
- CORDELLA, L. P. et al. A method for improving classification reliability of multilayer perceptrons. *IEEE transactions on neural networks*, v. 6, p. 1140–7, 1995. Citado na página 17.
- FRIEDEL, C. C.; RUCKERT, U.; KRAMER, S. Cost curves for abstaining classifiers. *Proceedings of the ICML 2006 workshop on ROC Analysis in Machine Learning, Pittsburgh*, p. 33–40, 2006. Citado 2 vezes nas páginas 17 e 33.
- FUMERA, G.; ROLI, F. Support vector machines with embedded reject option. In: LEE, S.-W.; VERRI, A. (Ed.). *Pattern Recognition with Support Vector Machines*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. p. 68–82. ISBN 978-3-540-45665-0. Citado 2 vezes nas páginas 17 e 41.
- FUMERA, G.; ROLI, F.; GIACINTO, G. Reject option with multiple thresholds. *Pattern Recognition*, v. 33, p. 2099–2101, 2000. Citado 2 vezes nas páginas 17 e 38.
- GENERAL solution and learning method for binary classification with performance constraints. *Pattern Recognition Letters*, v. 29, n. 10, p. 1455 – 1465, 2008. ISSN 0167-8655. Citado na página 41.
- GROSSBERG, S. Adaptive pattern classification and universal recoding: Ii. feedback, expectation, olfaction, illusions. *Biological Cybernetics*, v. 23, n. 4, p. 187–202, Dec 1976. ISSN 1432-0770. Disponível em: <<https://doi.org/10.1007/BF00340335>>. Citado na página 15.
- GROSSBERG, S. Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, v. 11, n. 1, p. 23 – 63, 1987. ISSN 0364-0213. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0364021387800253>>. Citado na página 15.
- GROSSBERG, S. Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Netw.*, Elsevier Science Ltd., Oxford, UK, UK, v. 37, p. 1–47, jan. 2013. ISSN 0893-6080. Disponível em: <<http://dx.doi.org/10.1016/j.neunet.2012.09.017>>. Citado na página 15.
- JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 22, p. 4–37, 2000. Citado na página 15.

- LICHMAN, M. *UCI Machine Learning Repository*. 2013. Disponível em: <<http://archive.ics.uci.edu/ml>>. Citado na página 47.
- LIU, C.-L.; SAKO, H.; FUJISAWA, H. Performance evaluation of pattern classifiers for handwritten character recognition. *International Journal on Document Analysis and Recognition*, v. 4, p. 191–204, 2002. Citado na página 17.
- MARRIOTT, S.; HARRISON, R. F. A modified fuzzy artmap architecture for the approximation of noisy mappings. *Neural Networks*, v. 8, p. 619–641, 1995. Citado na página 42.
- NETO, A. R. da R. et al. Diagnostic of pathology on the vertebral column with embedded reject option. In: *IbPRIA*. [S.l.: s.n.], 2011. Citado na página 18.
- NETO, A. R. R. Sinpatco ii: Novas estratégias de aprendizado de máquina para classificação de patologias da coluna vertebral. tese (doutorado). UNIVERSIDADE FEDERAL DO CEARÁ, 2011. Citado 3 vezes nas páginas 35, 36 e 37.
- POURPANAH, F.; LIM, C. P.; SALEH, J. M. A hybrid model of fuzzy artmap and genetic algorithm for data classification and rule extraction. *Expert Systems with Applications*, Elsevier, v. 49, p. 74–85, 2016. Citado na página 30.
- SANCHEZ, E. G. et al. Microartmap: use of mutual information for category reduction in fuzzy artmap. In: . [S.l.: s.n.], 2000. v. 6, p. 47 – 52 vol.6. ISBN 0-7695-0619-4. Citado na página 30.
- SOUSA, R. G.; MORA, B.; CARDOSO, J. S. An ordinal data method for the classification with reject option. *2009 International Conference on Machine Learning and Applications*, p. 746–750, 2009. Citado na página 18.
- SOUSA, R. G.; MORA, B.; CARDOSO, J. S. An ordinal data method for the classification with reject option. *2009 International Conference on Machine Learning and Applications*, p. 746–750, 2009. Citado 2 vezes nas páginas 40 e 41.
- SOUSA, R. G. et al. Robust classification with reject option using the self-organizing map. *Neural Computing and Applications*, v. 26, p. 1603–1619, 2015. Citado na página 18.
- SUUTALA, J. et al. Reject-optional lvq-based two-level classifier to improve reliability in footstep identification. In: *Pervasive*. [S.l.: s.n.], 2004. Citado na página 18.
- THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition, Fourth Edition*. 4th. ed. Orlando, FL, USA: Academic Press, Inc., 2008. ISBN 1597492728, 9781597492720. Citado na página 15.
- ZHANG, Y.; JI, H.; ZHANG, W. Tppfam: Use of threshold and posterior probability for category reduction in fuzzy artmap. *Neurocomputing*, v. 124, p. 63–71, 01 2014. Citado na página 17.
- ZHONG, M. et al. Experiments with safe artmap : Effect of the network parameters on the network performance. *Neural networks : the official journal of the International Neural Network Society*, v. 20, p. 245–59, 04 2007. Citado na página 30.