KÁSSIO MARQUES TAVARES

PREVISÃO DOS RESULTADOS DO CAMPEONATO BRASILEIRO DE FUTEBOL UTILIZANDO REDES NEURAIS ARTIFICIAIS

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA MECÂNICA
2015

KÁSSIO MARQUES TAVARES

PREVISÃO DOS RESULTADOS DO CAMPEONATO BRASILEIRO DE FUTEBOL UTILIZANDO REDES NEURAIS ARTIFICIAIS

Projeto de Fim de Curso apresentado como exigência parcial para obtenção do Diploma de **BACHAREL EM ENGENHARIA MECATRÔNICA**, pela Universidade Federal de Uberlândia.

Área de concentração: Inteligência Artificial.

Orientador: Prof. PhD. Keiji Yamanaka

Uberlândia - MG
2015

"Do your best, then don't worry, be happy"

Meher Baba

AGRADECIMENTOS

Aos meus pais, Alba Luzia Marques e José Humberto Tavares, por todo o apoio oferecido ao longo de toda minha vida.

À toda minha família, que independentemente da distância, sempre estiveram comigo quando precisei.

À Universidade Federal de Uberlândia, à Faculdade de Engenharia Mecânica e ao curso de Engenharia Mecatrônica pela oportunidade de realizar esse curso.

Ao Professor PhD. Keiji Yamanaka pela possibilidade de realizar este projeto, além de toda a orientação, paciencia e auxílio oferecidos no decorrer do trabalho.

Aos meus amigos, Alexandre Rodrigues, Bruno Pinheiro, Danilo Andrades, Julianne Rodrigues, Lucas Ferraz, Lucas Lopes, Manoel Martins e, principalmente, à minha namorada Natássia Navarro, por toda a ajuda, direta ou indireta, as quais foram essenciais para a conclusão deste projeto.

À todos que em mim depositaram sua confiança e, mesmo não colaborando diretamente, dirigiram suas boas energias, pensamentos, palavras e intensões em favor da concretização desse projeto.

TAVARES, K. M., **Previsão dos resultados do campeonato brasileiro de futebol utilizando redes neurais artificiais.** 2015. xxxxxxxxxxxxxxx f. Monografia de Conclusão de Curso (Bacharel), Universidade Federal de Uberlândia, Uberlândia.

RESUMO

Prever resultados de jogos de futebol é uma tarefa árdua e por vezes dita impossível. Os resultados são altamente instáveis e abordagens estatísticas tem sua efetividade limitada. Tendo isso em vista, o objetivo desse trabalho foi a idealização de um sistema de predição de resultados tendo como base redes neurais artificiais, cujo uso visa eliminar dois problemas da predição de jogos de futebol: A limitação dos métodos estatísticos, ao se utilizar reconhecimento de padrões; e a eliminação de resultados tendenciosos, ao se usar exclusivamente dados concretos de resultados históricos. A linguagem de programação orientada a objetos Java foi utilizada para a montagem da rede neural artificial desse projeto, a qual consiste de um perceptron multicamadas utilizando a regra de treinamento supervisionado de Backpropagation. Dois conjuntos de dados de entrada foram considerados, um visando exprimir o histórico recente dos times e o outro visando exprimir sua estabilidade e constância. Por fim, os resultados foram validados de forma a excluir alguns palpites menos confiáveis, garantindo assim uma taxa de acerto mais elevada nos restantes.

Palavras Chave: Inteligência Artificial, Redes Neurais Artificiais, Backpropagation, Futebol, Predição.

TAVARES, K. M., **Prediction of the Brazilian soccer championship results using artificial neural networks** 2015. xxxxxxxxxxxxxxx f. Monografia de Conclusão de Curso (Bacharel), Universidade Federal de Uberlândia, Uberlândia.

ABSTRACT

The result prediction of soccer games is an arduous task and sometimes said to be impossible. The results are highly unstable and statistical approaches have their effectivity limited. With this in mind, the goal of this work was the idealization of a result prediction system based on artificial neural networks, the use of which aims to eliminate two problems of the soccer games prediction: The limitation of statistical methods, when using pattern recognition; and the elimination of biased results, when using only concrete data from historical results. The object oriented programming language Java was used to assemble the artificial neural network of this project, which consists of a multilayer perceptron using the supervised Backpropagation training rule. Two sets of input data were considered, one seeking to express the recent history of the teams and the other seeking to express their stability and constancy. Finally, the results were validated to exclude some less reliable predictions, thus ensuring a higher success ratio on the others.

Keywords: Artificial Intelligence, Artificial Neural Networks, Backpropagation, Soccer, Prediction.

Lista de Figuras

Lista de Tabelas

Lista de Símbolos

RNA	-	Rede Neural Artificial
POO	-	Programação Orientada a Objetos
IA	-	Inteligência Artificial
NI	-	Neurônio Intermediário
IDE	-	Integrated Development Environment; Ambientes de Desenvolvimento Integrado
GUI	-	Graphical User Interface; Interface Gráfica do Utilizador
API	-	Application Programming Interface; Interface de Programação de Aplicações
UFU	-	Universidade Federal de Uberlândia
$FEMEC$	-	Faculdade de Engenharia Mecânica
CBF	-	Confederação Brasileira de Futebol
α	-	Taxa de aprendizado
P	-	Probabilidade do neurônio de saída
σ	-	Desvio padrão das probabilidades dos neurônios de saída

SUMÁRIO

CAPÍTULO I

INTRODUÇÃO

O trabalho com redes neurais artificiais, também referidas simplesmente como RNA, foi motivado desde a sua criação pelo reconhecimento de que o cérebro humano processa informações de uma maneira completamente diferente dos computadores digitais modernos. O cérebro é um sistema altamente complexo, não linear e que processa paralelamente suas informações. Possui também a capacidade de organizar seus elementos estruturais, conhecidos como neurônios, de forma a realizar certas tarefas (e.g., reconhecimento de padrões, percepção e controle motor) (HAYKIN, 1999a).

Para alcançar uma boa performance, redes neurais empregam uma interconexão massiva de células computacionais simples, que são referenciadas como "neurônios"ou "unidades de processamento". Segundo HAYKIN (1999a) uma rede neural é um processador com distribuição paralela massiva, feito de unidades de processamento simples, as quais possuem uma propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso.

Redes Neurais são ferramentas úteis para a resolução de diversos problemas, os quais podem ser caracterizados como mapeamento (incluindo associação e classificação de padrões), agrupamento, e otimização restrita. Existem diversas redes neurais disponíveis para cada tipo de problema (FAUSSET, 1994). Pela perspectiva de reconhecimento de padrões, redes neurais podem ser consideradas como uma extensão das muitas técnicas convencionais que vêm sendo desenvolvidas ao longo de várias décadas (BISHOP, 1995).

O futebol, por sua vez, tem sido desde a sua criação (no século XIX pelos ingleses)

um dos esportes mais conhecidos e jogados pelo mundo. Há muito, vários países (principalmente Europeus e Sul-americanos) já fizeram do futebol o seu esporte mais popular (SILVA et al., 2002).

O futebol brasileiro é mundialmente conhecido por sua qualidade e pela paixão que os brasileiros sentem pelo esporte. Segundo DA COSTA (2005), o esporte é praticado por aproximadamente 30 milhões de brasileiros e possui 102 milhões de torcedores apenas dentre os 16 maiores clubes (em termos de tamanho de torcida).

Um fato bastante aceito entre os amantes do futebol diz respeito à imprevisibilidade dos resultados das partidas. Os jogos são disputados entre dois times e podem terminar com vitória do time da casa, vitória do time visitante e empate. Para a determinação do resultado de uma partida, encontram-se fatores esperados e imprevistos, ambos os quais têm influência no resultado da disputa (ASLAN e INCEOGLU, 2007).

A abordagem estatística para previsão de resultados já é bastante utilizada e seu aprimoramento já foi bastante estudado, como pode ser visto em ELO e SLOAN (2008), MEHREZ e HU (1995) e SILVA et al. (2002). Apesar dos esforços e estudos, a assertividade desses métodos ainda não chegou a ultrapassar a taxa de 50%.

Por sua vez, as RNAs têm se mostrado uma ferramenta muito poderosa para modelagem matemática, tanto em pesquisas quanto em aplicações práticas. Elas podem efetuar mapeamentos entre espaços de entrada e saída altamente não lineares. Com sua abordagem não-paramétrica, faz-se desnecessária a análise funcional da forma de distribuição dos dados. Assim sendo, se mostram uma ferramenta com grande potencial para a previsão dos resultados de partidas de futebol (CHENG et al., 2003).

1.1 Objetivos

O objetivo geral desse trabalho é desenvolver um programa em Java, capaz de utilizar a técnica de RNAs para realizar a previsão do resultado de uma partida do Campeonato Brasileiro de Futebol.

Os objetivos específicos são:

- Estudar sistemas computacionais estruturados em Redes Neurais Artificiais;

- Estudar a linguagem de programação Java, usada como base para a criação do algoritmo da RNA;
- Estudar dados de partidas de futebol e encontrar informações relevantes que influenciem o resultado das mesmas;
- Criação de um banco de dados estruturado e versátil com os dados de partidas passadas;
- Desenvolvimento do software responsável pela análise e tratamento dos dados e fornecimento dos resultados;
- Análise profunda de dados e da topologia da RNA criada, de forma a otimizar o seu funcionamento.

1.2 Justificativa

O resultado das partidas de futebol é hoje um fator ainda imprevisível. Os estudos já realizados para prever o desfecho de um jogo utilizam quase que exclusivamente métodos estatísticos. Estes por sua vez, apesar das diversas técnicas e abordagens já utilizadas, ainda não conseguiram acertar mais da metade dos resultados. Existem alguns poucos estudos com abordagens hibridas, utilizando estatística e redes neurais artificiais, os quais fornecem uma melhor taxa de acerto. Contudo, essa taxa ainda é baixa e gira em torno de 50%.

A proposta de utilizar somente redes neurais e reconhecimento de padrões é inovadora e promissora. Abdicando dos recursos estatísticos tem-se uma liberdade maior para o reconhecimento de anomalias, que outrora seriam tratadas como ruídos. A instabilidade da relação dados-resultados causada pelos mais diversos fatores do ambiente futebolístico, pode ser contornada pela capacidade inata das redes neurais de mapeamento, percepção e reconhecimento de padrões.

Apesar de o futebol ser o esporte mais seguido no Brasil, estudos de previsão de resultados utilizando dados de jogos disputados com times do Brasil são desconhecidos. O uso inovador de redes neurais e a utilização de dados do Campeonato Brasileiro são, portanto, inéditos. Deste modo, nos deparamos com um campo de estudos promissor e ainda não explorado.

CAPÍTULO II

FUNDAMENTAÇÃO TEÓRICA

Este capítulo trata de alguns assuntos importantes para a compreensão do trabalho, tais como: Redes Neurais Artificiais, Perceptrons, Adaline, Madaline, Regra Delta, funções de ativação, camadas intermediárias, algoritmos de treinamento, Backpropagation, programação orientada a objetos e Java.

2.1 Redes Neurais Artificiais

Toda a base para a realização desse trabalho se encontra no uso uma rede neural artificial (RNA), adequada para a realização do reconhecimento de padrões entre condições anteriores às partidas e o resultado das mesmas. Uma RNA é um sistema de processamento de informações que possui algumas características em comum com redes neurais biológicas. O desenvolvimento da teoria de redes neurais artificiais ocorreu a partir de generalizações matemáticas de modelos de percepção humana. LAURENE FAUSETT. Para isso, algumas hipóteses são levadas em conta: Assume-se que o processamento de informações ocorre em vários elementos simples chamados neurônios; Sinais são passados entre neurônios através de uma conexão; Cada conexão possui um peso associado, o qual, em redes neurais artificiais típicas, multiplicam o sinal transmitido; Cada neurônio possui uma função de ativação, a qual determina o sinal de saída com base na soma dos sinais de entrada. Um modelo completo de um neurônio, com base nas hipóteses acima, está mostrado na Figura 2.1.

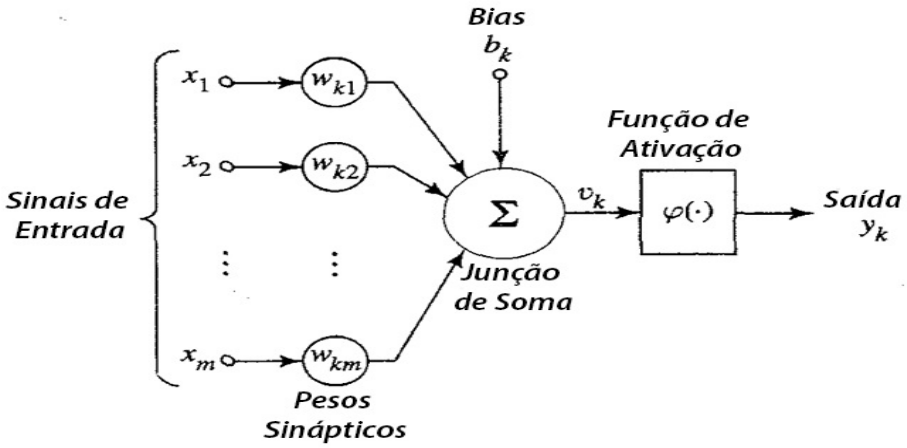

Figura 2.1 – Modelo completo de um neurônio (HAYKIN, 1999a)

2.1.1 McCulloch-Pitts

O primeiro modelo de neurônio artificial foi concebido por Warren McCulloth e Walter Pitts (MCCULLOTH e PITTS, 1943). Algumas das regras de funcionamento desses neurônios são aplicadas ainda hoje nas redes neurais modernas. Podemos descrever as regras que permeiam o funcionamento dos mesmos com base nas seguintes definições:

- A ativação do neurônio é binária. Ou seja, para cada situação de entrada existem apenas duas saídas, ativado ou desativado;
- Os neurônios são conectados diretamente por ligações, que possuem seu respectivo peso único;
- Uma ligação é excitatória se seu peso é positivo, e inibitória se seu peso é negativo. Todas conexões excitatórias possuem o mesmo peso;
- Cada neurônio possui um valor limiar, e sua ativação é função da soma dos pesos de entrada (excitatórios e inibitórios). A ativação ocorre quando a soma dos pesos ultrapassa o valor limiar;
- A definição do limiar ocorre de forma que qualquer entrada não inibitória diferente de zero faça com que o neurônio não seja ativado;
- É necessário um passo temporal para que o sinal passe por um link da conexão.

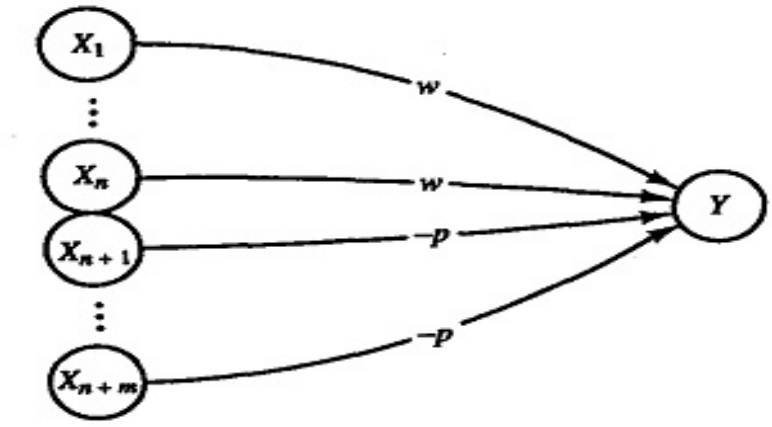

Figura 2.2 – Neurônio de McCulloch-Pitts (FAUSSET, 1994)

Um modelo completo de um neurônio, segundo as definições acima, é mostrado na Figura 2.2.

A função de ativação (degrau) desse neurônio é dado segunda a Equação 2.1, onde Yin é a soma das entradas de cada um dos neurônios, conforme Equação 2.2. O limiar do neurônio é dado pela Equação 2.3.

$$f(y_{in}) = \begin{cases} 1 & \text{se } Y_{in} > \theta \\ 0 & \text{se } Y_{in} \leq \theta \end{cases} \tag{2.1}$$

$$Y_{in} = n \cdot w + m \cdot p \tag{2.2}$$

$$\theta = n \cdot w - p \tag{2.3}$$

2.1.2 Bias

O uso de um limiar para definir a ativação do neurônio pode ser substituído pelo uso de um bias. O bias age exatamente como o peso de uma conexão cuja entrada é sempre 1 (FAUSSET, 1994), conforme mostrado na Figura 2.3.

O uso do bias ou do limiar são equivalentes. Sua importância se mostra durante a análise da separabilidade linear de um problema, conforme mostrado na Subseção 2.1.3, e durante a análise de convergência do erro, não tratada nesse trabalho.

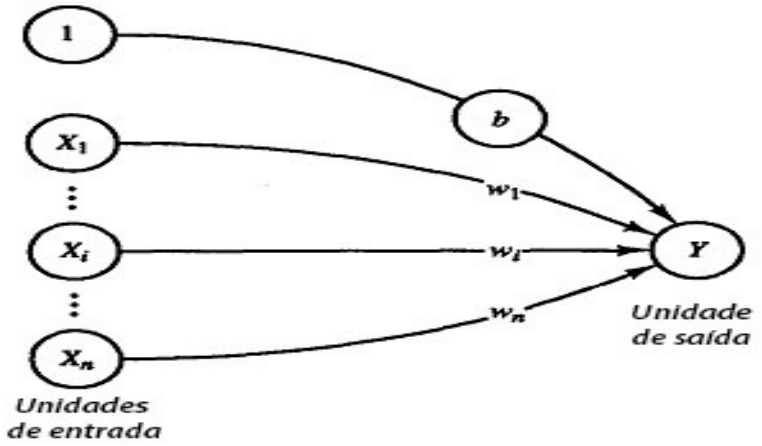

Figura 2.3 – Bias "b" em uma RNA

A diferença de aplicação entre as duas técnicas é dada pela função de ativação do neurônio, que com o uso do bias fica escrita conforme a Equação 2.4.

$$f(y_{in}) = \begin{cases} 1 & \text{se } Y_{in} > 0 \\ 0 & \text{se } Y_{in} \leq 0 \end{cases} \tag{2.4}$$

2.1.3 Separabilidade Linear

Para uma determinada rede, a resposta está confinada entre dois valores (1 ou 0, sim ou não, -1 ou 1, etc...). Existe uma fronteira, chamada *fronteira de decisão* Figura 2.4 que delimitada as regiões em que a resposta assume um ou outro valor.

A Equação 2.5 mostra a relação que determina essa fronteira para funções de ativação do tipo degrau. Dependendo do número de entradas na rede, essa equação pode representar uma linha, um plano ou um hiperplano.

$$b + \sum_i x_i \cdot w_i = 0 \tag{2.5}$$

Diz-se que um problema é linearmente separável caso existam pesos (e bias) suficientes para que todos vetores de entrada do conjunto de treinamento sejam separados pela fronteira de decisão em dois grupos, de forma que os vetores cujas saídas sejam positivas fiquem todos em um grupo, e os vetores com saídas negativas fiquem em outro. Essas duas regiões são geralmente

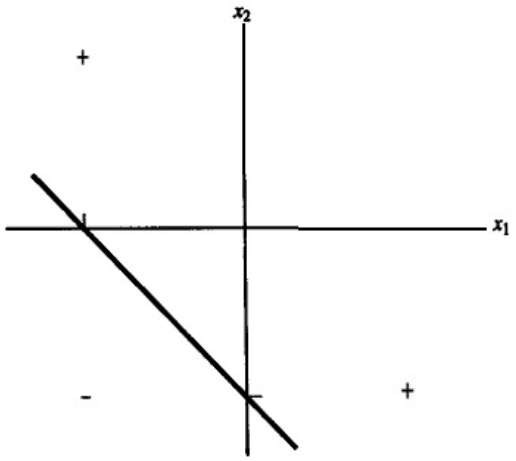

Figura 2.4 – Fronteira e Regiões de decisão para a função lógica *OU* (FAUSSET, 1994)

chamadas de *regiões de decisão* (Figura 2.4).

MINSKY e PAPERT (1969) mostraram que problemas não linearmente separáveis, são impossíveis de serem resolvidos por redes de apenas uma camada. É possível demonstrar também que redes de várias camadas, com funções de ativação lineares, também são incapazes de resolver tais problemas.

A inclusão de um bias pode transformar um problema linearmente inseparável (e portanto incapaz de ser resolvido com os métodos já apresentados) em um problema que pode ser resolvido. Isso é possível graças a adição de uma nova conexão de entrada, que será usada como valor constante para o bias. Essa nova conexão aumenta o conjunto de entradas, fornecendo mais um grau de liberdade para a rede.

2.1.4 Representação dos dados

No início dos estudos de redes neurais, a representação de dados se dava na forma binária, ou seja, utilizando 1 para reforço positivo e 0 para reforço negativo. Essa representação inicial não é, de modo geral, recomendada.

A forma bipolar é uma alternativa bastante eficiente. Essa representação utiliza 1 para reforço positivo e -1 para reforço negativo. Assim como a inclusão de um bias, a simples mudança da representação de dados para a forma bipolar pode transformar um problema sem solução em um problema que pode ser resolvido.

Caso o problema proposto requeira da rede uma generalização dos dados de entrada, o uso da forma bipolar é altamente indicado. A representação bipolar permite que dados não disponíveis sejam adotados com valor 0 de entrada, não fornecendo assim reforço positivo ou negativo.

Ao contrário do que acontece na forma binária, a representação bipolar também é mais eficiente do ponto de vista do treinamento. Por exemplo, em uma rede treinada pelo método de Hebb (Subseção 2.1.5), para conjuntos de sinais entrada cujo resultado da rede seja negativo (0 para binário ou -1 para bipolar), o uso da primeira representação impossibilita completamente um aprendizado da rede.

2.1.5 Aprendizado de Hebb

Das regras de treinamento de redes neurais, a regra de Hebb é a mais antiga, mais simples e mais conhecida de todas. Hebb propôs seu método de treinamento tendo como base um contexto neurobiológico. STENT (1973) e CHANGEUX e DANCHIN (1976) reescreveram o pensamento de Hebb de uma forma mais simples e concisa, como se segue:

- Se dois neurônios em ambos os lados de uma sinapse (conexão) são ativados simultaneamente (i.e. de forma síncrona), então a força dessa sinapse é seletivamente aumentada.
- Se dois neurônios em ambos lados de uma sinapse são ativados de forma assíncrona, então essa sinapse é seletivamente enfraquecida ou eliminada.

A afirmação original de Hebb não incluía a segunda parte. Da mesma forma, ela também comentava apenas a respeito de neurônios sendo ativados ao mesmo tempo.

McCLELLAND e RUMELHART (1988) estenderam a regra de Hebb para que o reforço positivo da sinapse também ocorra caso os neurônios sejam ambos desativados simultaneamente. Podemos, então, complementar o pensamento anterior:

- Se dois neurônios em ambos os lados de uma sinapse são DESATIVADOS simultaneamente, então a força dessa sinapse é seletivamente aumentada.

Estamos considerando inicialmente redes de apenas uma camada, onde cada conexão possui, em cada lado, apenas um neurônio de entrada e um neurônio de saída. O valor do neurônio

de entrada X é multiplicado pelo peso W dessa conexão para se obter o valor do neurônio de saída Y. Seguindo essa nomenclatura e representando os dados na forma bipolar, podemos representar a atualização do peso pela regra de Hebb segundo a Equação 2.6.

$$\Delta w_{ij} = x_i \cdot y_i \tag{2.6}$$

Se utilizarmos a representação binária, essa equação não distingue conexões que possuem uma entrada "positiva" e uma saída "negativa" de conexões onde tanto a entrada quanto a saída são "negativas". Essa é uma das vantagens mais facilmente observáveis da utilização de dados bipolares (ao invés de binários) para a representação dos dados.

Técnica de aprendizado

Existem várias técnicas de implementação das regras de aprendizado existentes. A técnica abordada aqui nada mais é do que uma das formas mais simples e genéricas de se aplicar a regra de Hebb. O entendimento dessa técnica é crucial para a posterior compreensão das outras técnicas aplicadas nesse trabalho, que podem ser tomadas como simples variações mais elaboradas desta.

O aprendizado é dado na forma iterativa e segue o seguinte algoritmo:

1. Os pesos de todas conexões são inicializados com o valor 0.

2. Para todos os conjuntos de entradas e saídas, repetem-se os passos 3-4.

    3. Para cada conexão, calculamos a variação do seu respectivo peso pela equação 6.

    4. Cada peso é atualizado conforme calculado no passo 3.

Para esclarecer um pouco mais o algoritmo acima mostrado, segue um exemplo de sua aplicação em uma rede real com representação bipolar, onde usaremos como parâmetros de entrada a função lógica E. Buscaremos com essa rede descobrir os valores dos pesos das conexões que determinam a fronteira de decisão dessa função.

A função lógica E retorna um valor "positivo" apenas se ambas as entradas possuírem também um valor "positivo". Assim sendo, montaremos a tabela verdade dessa função conforme a Tabela 2.1. A representação gráfica dessa rede está mostrada na Figura 2.5.

Tabela 2.1 – Tabela Verdade para a função lógica E

Entradas			Saída
Bias	X_1	X_2	Y
1	1	1	1
1	1	-1	-1
1	-1	1	-1
1	-1	-1	-1

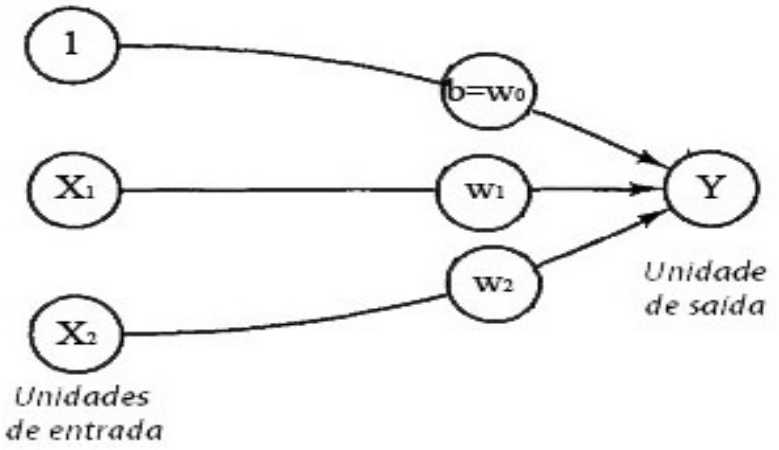

Figura 2.5 – RNA para a função lógica E

Iniciamos o treinamento conforme o passo 1 do algoritmo. Fazemos então os valores de $w_0 = 0, w_1 = 0$ e $w_2 = 0$.

Utilizaremos o primeiro par de entradas (1 ,1). Assim, calculamos o valor da variação dos pesos conforme a Equação 2.6. O resultado é mostrado na Tabela 2.2.

Tabela 2.2 – Variação dos pesos para o primeiro par de entradas

	Entrada		Saída desejada	Variação dos pesos (Δw)		
1	X_1	X_2	Y	Δw_0	Δw_1	Δw_2
1	1	1	1	1	1	1

Na sequência, atualizamos os pesos com os resultados da Equação 2.7.

$$\begin{cases} w_0 = 0 + \Delta w_0 = 1 \\ w_1 = 0 + \Delta w_1 = 1 \\ w_2 = 0 + \Delta w_2 = 1 \end{cases} \tag{2.7}$$

Repetimos o procedimento para o segundo par de entradas (1, -1). Assim, calculamos o valor da variação dos pesos conforme a Equação 2.6. O resultado é mostrado na Tabela 2.3.

Tabela 2.3 – Variação dos pesos para o segundo par de entradas

	Entrada		Saída desejada	Variação dos pesos (Δw)		
1	X_1	X_2	Y	Δw_0	Δw_1	Δw_2
1	1	-1	-1	-1	-1	1

Na sequência, atualizamos os pesos com os resultados da Equação 2.8.

$$\begin{cases} w_0 = 1 + \Delta w_0 = 0 \\ w_1 = 1 + \Delta w_1 = 0 \\ w_2 = 1 + \Delta w_2 = 2 \end{cases} \tag{2.8}$$

Repetimos o procedimento para o terceiro par de entradas (-1, 1). Calculamos o valor da variação dos pesos conforme a Equação 2.6. O resultado é mostrado na Tabela 2.4.

Tabela 2.4 – Variação dos pesos para o terceiro par de entradas

	Entrada		Saída desejada	Variação dos pesos (Δw)		
1	X_1	X_2	Y	Δw_0	Δw_1	Δw_2
1	-1	1	-1	-1	1	-1

Na sequência, atualizamos os pesos com os resultados da Equação 2.9.

$$\begin{cases} w_0 = 0 + \Delta w_0 = -1 \\ w_1 = 0 + \Delta w_1 = 1 \\ w_2 = 2 + \Delta w_2 = 1 \end{cases} \tag{2.9}$$

Repetimos uma última vez o procedimento para o quarto par de entradas (-1, -1).

Calculamos o valor da variação dos pesos conforme a Equação 2.6. O resultado é mostrado na Tabela 2.5.

Tabela 2.5 – Variação dos pesos para o quarto par de entradas

	Entrada		Saída desejada	Variação dos pesos (Δw)		
1	X_1	X_2	Y	Δw_0	Δw_1	Δw_2
1	-1	-1	-1	-1	1	1

Na sequência, atualizamos os pesos com os resultados da Equação 2.10.

$$\begin{cases} w_0 = -1 + \Delta w_0 = -2 \\ w_1 = 1 + \Delta w_1 = 2 \\ w_2 = 1 + \Delta w_2 = 2 \end{cases} \tag{2.10}$$

Terminamos assim o treinamento e a rede resultante é mostrada na Figura 2.6.

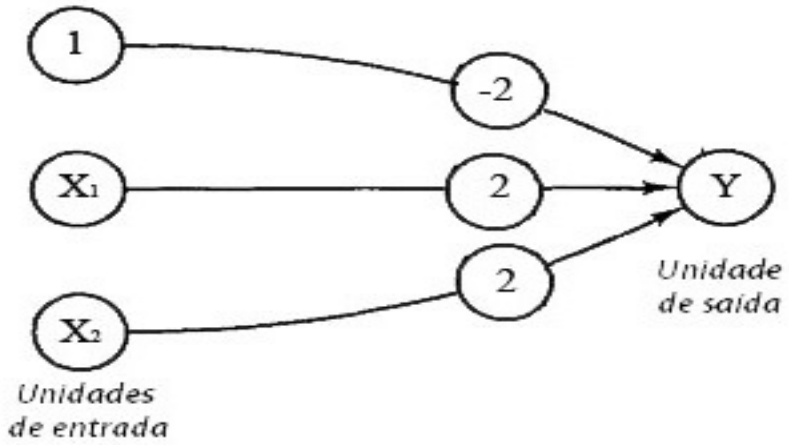

Figura 2.6 – RNA resolvida para a função lógica *E*

Essa configuração de pesos nos fornece uma fronteira de decisão em forma de reta, que pode ser encontrada utilizando a Equação 2.11. A equação da reta resultante do exemplo acima é dada pela Equação 2.12 e está mostrada na Figura 2.7.

$$1 + X_1 + X_2 = 0 \tag{2.11}$$

$$X_2 = -X_1 + 1 \tag{2.12}$$

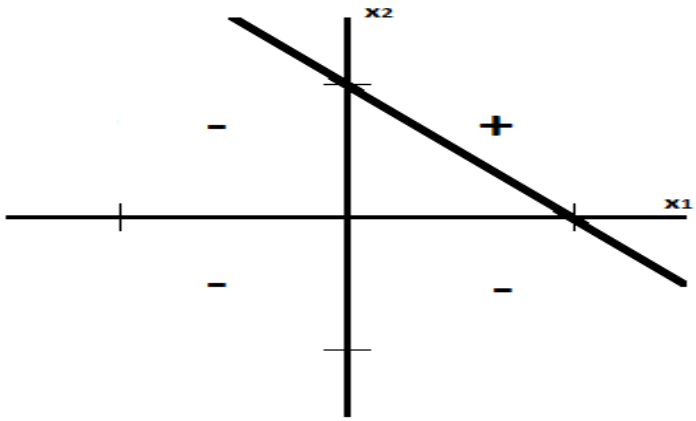

Figura 2.7 – Fronteira de decisão da função lógica *E*

Podemos notar que apesar de os pesos terem sido alterados pelo último loop, a equação da reta (e consequentemente a fronteira de decisão) não é alterada após o mesmo. Isso significa que o treinamento já havia encontrado a solução do problema antes do fim do algoritmo. Como será mostrado mais adiante, outras técnicas de treinamento podem continuar indefinidamente, repetindo os vetores de entrada, refazendo os cálculos sem fim. Nesses casos, uma condição de parada deverá ser adotada.

2.1.6 Perceptron

Redes com um modelo de treinamento iterativo e supervisionado foram propostas por Frank Rosenblatt em 1962. Ele as chamou de Perceptrons e elas tiveram um grande impacto nos estudos de redes neurais artificiais. Isso se deve ao fato de que a regra de convergência dos perceptrons é mais poderosa do que sua antecessora, a regra de Hebb. Além disso, sob determinadas condições, é possível provar matematicamente a convergência de seu procedimento de aprendizado iterativo. Basicamente, ela consiste de um simples neurônio com pesos sinápticos e bias ajustáveis.

Um perceptron particularmente simples utilizava valores binários de entrada e saída, contudo a função de ativação do neurônio de saída poderia retornar valores de -1, 0 ou 1 durante o aprendizado. A possibilidade de 3 diferentes valores só ocorre graças ao novo tipo de função de ativação, onde o valor do limiar passou a delimitar uma região intermediária, e não apenas um valor limítrofe entre saídas positivas e negativas. Essa função é descrita pela Equação 2.14.

$$Y_{in} = \sum_i x_i \cdot w_i \tag{2.13}$$

$$f(Y_{in}) = \begin{cases} 1 & \text{se } Y_{in} > \theta \\ 0 & \text{se } -\theta \leq Y_{in} \leq \theta \\ -1 & \text{se } Y_{in} < \theta \end{cases} \tag{2.14}$$

Para se atualizar os pesos das ligações durante o treinamento usamos a equação 11. Tal equação apresenta um novo parâmetro, chamado de taxa de aprendizado, que é um valor constante determinado antes do início do treinamento.

$$\Delta w_i = \alpha \cdot x_i \cdot t \tag{2.15}$$

Observando a função de ativação de saída, que retorna valores de -1, 0 ou 1, e a saída real (Target t) que deve ter valores de -1 ou 1, podemos concluir alguns pontos importantes relacionados a esse perceptron.

- A rede não diferencia situações onde, por exemplo, a saída real é -1 e a saída calculada é 0 ou 1. Em ambos os casos os pesos são atualizados, de maneira igual, e o sinal do erro informa se os mesmos devem aumentar ou diminuir.
- Apenas conexões de entrada com valores diferentes de 0 possuem seus pesos atualizados.
- Padrões de entrada que não produzem erro na saída também não têm seus pesos atualizados.
- Como o treinamento ocorre até que os pesos sejam tais que a saída seja correta para quaisquer padrões de entrada, e padrões sem erro não resultam em treinamento da rede, quanto mais treinada a rede estiver, menor será o seu aprendizado.
- O valor do limiar não mais fornece uma equação, delimitando uma fronteira de decisão, mas sim uma inequação, fornecendo uma região de "indecisão", separando a região de respostas positivas e a região de respostas negativas. Dessa forma, o valor do limiar se torna mais significativo para o processo de aprendizagem.

O treinamento do perceptron continua até que os pesos das conexões sejam tais que para todos os padrões de entrada a resposta obtida pela rede seja correta. Segundo o teorema de convergência dos perceptrons, caso tais pesos existam, eles serão encontrados em um número finito de passos do treinamento.

Ao contrário das redes mostradas até agora, no perceptron o uso do limiar e do bias não são equivalentes. Como agora eles possuem funções diferentes, ambos são necessários para que o treinamento funcione corretamente.

Técnica de aprendizado

O algoritmo de aprendizado do perceptron está apresentado a seguir. Sua estrutura é semelhante à do treinamento de Hebb, porém já mais avançado e mais parecido com o algoritmo utilizado para a realização desse trabalho.

1. Inicializar pesos e bias (podem ser setados em 0, por simplicidade)
2. Definir taxa de aprendizado $0 < \alpha < 1$
3. Enquanto a condição de parada não for atingida, repetem-se os passos 4-7
    4. Para todos os conjuntos de entradas e saídas, repetem-se os passos 5-6
        5. Calcula-se o valor da saída de acordo com a Equação 2.13 e a Equação 2.14
        6. Caso a saída calculada seja diferente da saída correta, atualizam-se os pesos de acordo com a Equação 2.15
        7. Caso nenhum peso tenha sido alterado no passo 6, pare. Caso contrário, continue

2.1.7 Adaline

WIDROW e HOFF Jr. (1960) apresentaram um novo modelo de redes chamado Adaline, que é uma redução para Adaptive Linear Neuron (Neurônio Linear Adaptativo, em tradução livre). A inovação apresentada por eles se destaca pela regra de treinamento utilizada, a Regra Delta, também conhecida por Regra dos Mínimos Quadrados Médios (LMS) ou Regra de

Widrow-Hoff. A Regra Delta se mostra como uma inovação diante das opções anteriores à ela devido ao fato de que é capaz de fornecer correções pequenas e variáveis ao longo do treinamento. Seu desenvolvimento se deu com a ideia de que a mudança dos pesos da conexão deve minimizar a diferença entre a saída calculada Yin e o valor alvo t, minimizando o erro ao longo de todos os padrões de entrada. Isso é obtido se reduzindo o erro em cada padrão de entrada, um por vez. Para se atualizar os pesos das ligações usamos a Equação 2.17. Tal equação, assim como nos perceptrons, apresenta um parâmetro chamado de taxa de aprendizado, que é um valor constante determinado antes do início do treinamento.

$$Y_{in} = \sum_i x_i \cdot w_i \tag{2.16}$$

$$\Delta w_i = \alpha \cdot x_i \cdot (t - Y_{in}) \tag{2.17}$$

Semelhantemente as redes já apresentadas, um Adaline tipicamente usa ativação bipolar para os sinais de entrada e saída, possui pesos ajustáveis nas conexões e também possui bias. Apesar da Regra Delta poder ser utilizada em redes com múltiplas saídas e/ou camadas, um Adaline é um caso especial onde existe apenas uma saída e uma camada.

A arquitetura de uma rede Adaline é mostrada na Figura 2.8.

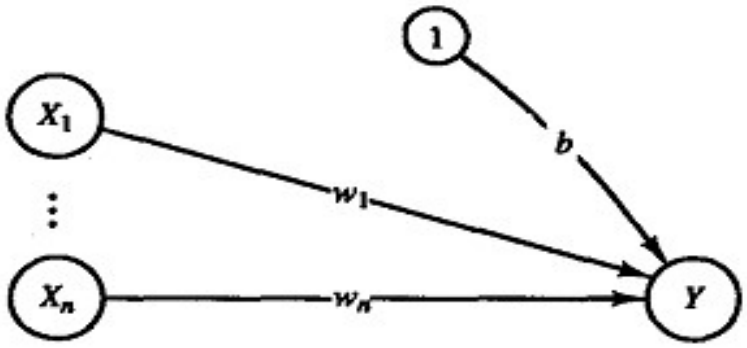

Figura 2.8 – Arquitetura de uma Adaline (FAUSSET, 1994)

Técnica de aprendizado

O algoritmo de aprendizado do Adaline está apresentado a seguir. Sua estrutura não é muito diferente à do perceptron, contudo, possui uma mudança singela: Os pesos aqui são alterados em todos loops de treinamento e o controle de parada se dá por um limiar máximo de alteração dos pesos. Como pode ser visto abaixo.

1. Inicializar pesos e bias (podem ser setados em 0, por simplicidade)
2. Definir taxa de aprendizado $0 < \alpha < 1$
3. Enquanto a condição de parada não for atingida, repetem-se os passos 4-7
    4. Para todos os conjuntos de entradas e saídas, repetem-se os passos 5-6
        5. Calcula-se o valor da saída de acordo com a Equação 2.16
        6. Atualizam-se os pesos de acordo com a Equação 2.17
    7. Caso a maior alteração de peso que ocorreu no passo 6 seja menor que um valor determinado, pare. Caso contrário, continue.

A taxa de aprendizado deve ser determinada com cuidado. Um valor muito pequeno pode fazer com que a rede seja treinada muito lentamente, e de forma oposta, um valor muito grande pode fazer com que o treinamento não convirja. HECHT-NIELSEN (1990) chegou a uma equação para limitar o valor da taxa de aprendizado. Porém, é mais comum simplesmente a escolha de um valor inicial pequeno (0,1 por exemplo).

Um outro ponto importante desse treinamento, se deve ao fato de que a aplicação da rede em situações que a saída deve ser bivalente não pode ser apenas uma repetição de parte do algoritmo de treinamento, pois o mesmo retornaria valores quebrados. Assim, a aplicação do Adaline em tais situações requer uma função de ativação para poder ser usada. Comumente a função degrau é utilizada para tais aplicações.

2.1.8 Redes Multicamadas

Muitas vezes, uma rede neural de apenas uma camada, como as mostradas até agora, não é capaz de solucionar o problema proposto. Tais limitações foram um dos fatores que reduziram o interesse nas redes neurais artificiais nos anos 70. Para tentar solucionar esses casos, foram desenvolvidas as redes multicamadas (Figura 2.9). Nesse trabalho serão expostas as redes Madaline e Multilayer Perceptron com BackPropagation, que são extensões multicamadas das redes Adaline e Perceptron, respectivamente.

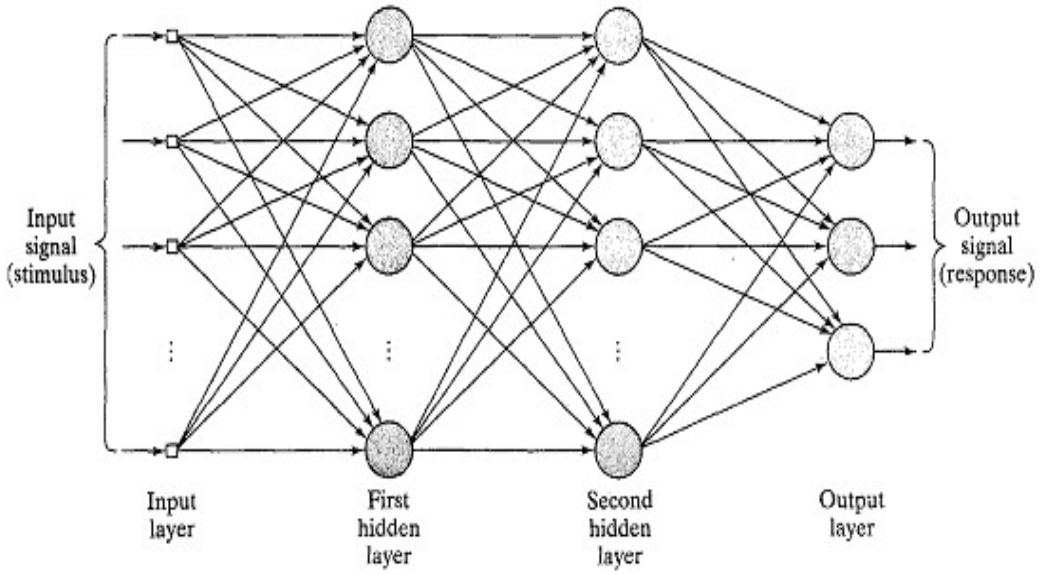

Figura 2.9 – Arquitetura de uma rede multicamadas com duas camadas intermediárias (HAYKIN, 1999a)

2.1.9 Madaline

O uso das redes Adaline em uma configuração multicamadas é possível, e chamado Madaline - Many Adaptive Linear Neurons (Muitos Neurônios Lineares Adaptativos, em tradução livre).

Um modelo simples de uma Madaline pode ser visto na Figura 2.10. Nela, temos dois neurônios de entrada (X1 e X2), duas Adalines resultando dois neurônios intermediários (Z1 e Z2) e uma terceira Adaline, que utiliza os neurônios intermediários (aqui chamados NI) como entradas e resulta a saída Y. Podemos ver que cada uma das três Adalines intermediarias precisam de um bias.

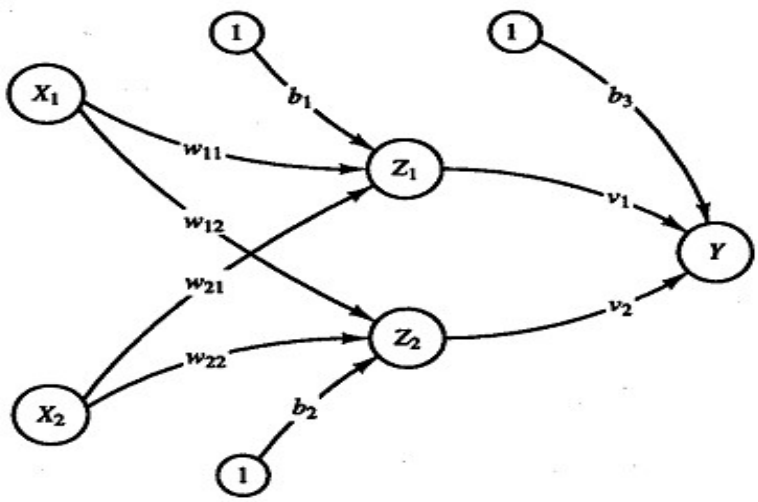

Figura 2.10 – Arquitetura de uma Madaline com uma camada intermediária de dois neurônios (FAUSSET, 1994)

O processo de treinamento de uma Madaline não sofre grandes alterações em relação à sua versão mais simples, de uma camada. Basicamente, a rede é uma combinação de várias Adalines, assim como seu processo de treinamento também é a combinação do processo de treinamento de várias Adalines. Assim sendo, o mesmo algoritmo já mostrado é apenas repetido para cada pedaço da rede multicamadas.

Devemos, entretanto, levar em conta que as entradas do sistema são utilizadas apenas pelas Adalines que resultam a primeira camada intermediária. Por sua vez, a primeira camada intermediária fornece as entradas para as Adalines que resultam a segunda camada intermediária, e assim sucessivamente, até que se chegue na saída do sistema.

2.1.10 Backpropagation

O uso de perceptrons multicamadas, com a regra de treinamento supervisionado de Backpropagation (retro propagação de erros) ou Regra Delta Generalizada, tem sido amplo e eficaz na resolução de problemas, por vezes complexos, em diversas áreas. Apesar de sua importância atual, não se diz que o Backpropagation possui um criador, pois foi desenvolvido simultaneamente por diversos pesquisadores. Esse método visa reduzir o gradiente do erro quadrático total da saída da rede ao longo do treinamento, ou seja, minimizando o erro da saída calculado pelo sistema.

O algoritmo utilizado nesse trabalho usa a regra de Backpropagation. O objetivo é treinar a rede para alcançar um balanço entre a habilidade de responder corretamente a padrões de entrada (utilizados para treinamento) e a habilidade de responder bem a entradas similares, mas não necessariamente idênticas, às utilizadas no treinamento.

Podemos dizer que o Backpropagation utiliza o melhor de cada um dos modelos de redes neurais já apresentados. Em sua base temos uma arquitetura multicamadas, baseada em Perceptrons (Figura 2.11). A regra de treinamento utilizada é a Regra Delta Generalizada, que é uma modificação da Regra Delta. Além disso, possui a retro propagação do erro, que visa uma redução gradual do erro ao longo do treinamento.

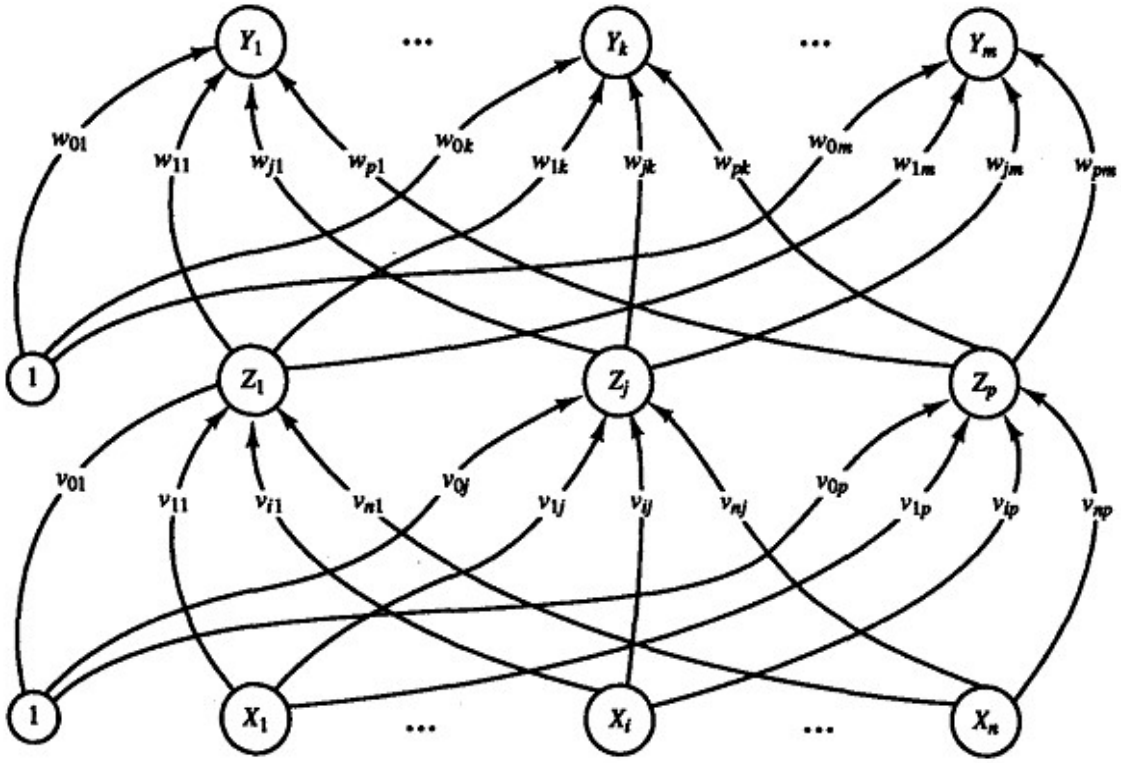

Figura 2.11 – Arquitetura de uma rede Backpropagation com uma camada intermediária de três neurônios. (FAUSSET, 1994)

Treinamento e Aplicação

Treinar uma rede com Backpropagation pode ser uma tarefa lenta mas, uma vez treinada, os resultados da rede são produzidos rapidamente. Para realizar o treinamento temos 3 estágios: Feedforward (Alimentação); Backpropagation (Retro Propagação); Ajuste de Pesos.

Na Alimentação, cada neurônio da primeira camada recebe seu valor de entrada e reproduz seu resultado para todos neurônios da camada seguinte, de acordo com sua função de ativação. Os neurônios das camadas intermediárias seguem o mesmo comportamento, e propagam o resultado da rede através de todas camadas existentes até os neurônios de saída.

Na Retro Propagação, é calculado o erro obtido entre o resultado da etapa de alimentação e o valor correto esperado. Esse erro é então retro propagado para toda a rede de acordo com um fator específico.

No Ajuste de Pesos, o erro que foi retro propagado é então utilizado para se calcular o ajuste que deve ser feito no peso de cada uma das conexões. Com as correções calculadas, todos os pesos das conexões da rede são atualizados simultaneamente.

A aplicação da rede, por sua vez, se torna tarefa simples após o treinamento. Com os valores finais dos pesos das conexões, basta aplicarmos a primeira etapa do treinamento (Feedforward) com os valores de entrada desejados.

Técnica de Aprendizado

A técnica de aprendizado Backpropagation está apresentada a seguir. Sua estrutura é a mais complexa e avançada presente nesse trabalho. Também é importante citar que o algoritmo mostrado a seguir foi utilizado para se montar a parte prática desse trabalho.

As equações utilizadas no algoritmo mostrado abaixo são descritas a seguir. A nomenclatura utilizada por ser vista na Figura 2.11, para uma melhor ilustração de sua representação. Uma explicação do que cada equação representa é dada em sequência.

$$Y_{in} = \sum_i x_i \cdot w_i \tag{2.18}$$

$$Y_{out} = f(Y_{in}) \tag{2.19}$$

$$\delta_z = (t - Z_{out}) \cdot f'(Z_{in}) \tag{2.20}$$

$$\Delta w_{yz} = \alpha \cdot Y_{out} \cdot \delta_z \tag{2.21}$$

$$\delta_{in} = \sum_i \delta_i \cdot w_i \tag{2.22}$$

$$\delta_y = \delta_{in} \cdot f'(Y_{in}) \tag{2.23}$$

$$\Delta w_{xy} = \alpha \cdot X_{out} \cdot \delta_y \tag{2.24}$$

$$w_{novo} = w_{velho} + \Delta w \tag{2.25}$$

A Equação 2.18 fornece o valor que chega a cada um dos neurônios, sendo o resultado da somatória dos valores fornecidos por cada conexão que "chega" no neurônio. O valor fornecido por cada conexão é o resultado do valor de seu neurônio de origem multiplicado pelo peso da conexão.

A Equação 2.19 fornece o valor que "sai" de cada neurônio. Tal valor é obtido ao se aplicar a função de ativação utilizada ao valor obtido na Equação 2.18.

A Equação 2.20 fornece o "fator de informação do erro" dos neurônios da última camada. Esse fator é obtido a partir da diferença entre o valor de saída esperado "t" e o valor de saída obtido "Z_{out}". Tal resultado é então multiplicado pelo valor de entrada do neurônio aplicado à derivada da função de ativação utilizada.

A Equação 2.21 fornece o "fator de correção do erro" das conexões que chegam aos neurônios da última camada. Esse valor é obtido através da multiplicação da taxa de aprendizado "α" com o valor de saída do neurônio anterior da conexão "Y_{out}" com o fator de informação do erro do neurônio posterior da conexão "δ_z".

A Equação 2.22 fornece o valor retro propagado de erro que "chega" ao neurônio intermediário. Esse valor é obtido fazendo a somatória da multiplicação do fator de informação do erro "δ_i" com o fator de correção do erro "w_i" através de todas as conexões posteriores ao neurônio.

A Equação 2.23 fornece o "fator de informação do erro" dos neurônios intermediários. Esse fator é obtido a partir da multiplicação do valor retro propagado de erro que "chega" ao neurônio "δ_{in}" e o valor de entrada do neurônio "Y_{in}" aplicado à derivada da função de ativação utilizada.

A Equação 2.24 fornece o "fator de correção do erro" das conexões que chegam aos neurônios intermediários. Esse valor é obtido através da multiplicação da taxa de aprendizado "α" com o valor de saída do neurônio anterior da conexão "X_{out}" com o fator de informação do erro do neurônio posterior da conexão "δ_y".

A Equação 2.25 fornece as diretrizes de atualização dos pesos de cada conexão. O valor do peso atualizado w_{novo} corresponde à soma do peso antigo w_{velho} com o fator de correção do erro da conexão.

Segue o algoritmo:

1. Inicializar pesos e bias

2. Enquanto a condição de parada não for atingida, repetem-se os passos 3-10.

    3. Para todos pares de entradas e saídas, repetem-se os passos 4-9

        (FeedForward - Alimentação)

    4. Os neurônios da primeira camada recebem o valor de entrada e os propagam para as camadas intermediárias.

    5. Os neurônios da segunda camada recebem os valores da camada anterior e calcula-se seus valores de entrada de acordo com a Equação 2.18.

    6. A saída dos neurônios da segunda camada é então calculada de acordo com sua função de ativação, conforme a Equação 2.19. O cálculo segue para todas as camadas intermediárias seguintes até que se compute a saída do sistema.

        (Backpropagation - Retro Propagação)

    7. Cada neurônio de saída recebe um fator de informação do erro, calculado de acordo com a Equação 2.20. Recebe também um fator de correção do erro, calculado

de acordo com a Equação 2.21.

8. Cada neurônio intermediário calcula seu valor de entrada do erro retro propagado de acordo com a Equação 2.22. Tal valor é multiplicado pela derivada de sua função de ativação para obter seu fator de informação de erro, conforme Equação 2.23. Seu fator de correção do erro é calculado de acordo com a Equação 2.24.

(Ajuste de Pesos)

9. Cada conexão tem seu peso atualizado, de acordo com a Equação 2.25.

10. Testar condição de parada

Mais detalhes sobre os parâmetros utilizados e as escolhas que devem ser feitas para o treinamento, dentre eles o número de NIs e a função de ativação utilizada, serão mostrados no capítulo de Materiais e Métodos.

2.2 Programação Orientada a Objetos

Para a realização desse trabalho, foi necessária a criação de um programa de computador capaz de realizar os cálculos determinados pelo algoritmo de Backpropagation. A Programação Orientada a Objetos (POO) foi escolhida como paradigma para se escrever o código. Nela, nos expressamos no código em termos de objetos, suas propriedades, métodos e estados.

Essa forma ou paradigma de programação foi escolhido por fornecer algumas vantagens essenciais necessárias ao longo do desenvolvimento do projeto. Entre elas podemos citar:

- Possui uma base conceitual no campo de estudo da cognição, ou seja, tenta representar o mundo da forma mais real possível. Da mesma forma, a Inteligência Artificial também busca representar a inteligência humana da forma mais real possível. Criando assim um vínculo conceitual entre ambas, o que facilita o desenvolvimento.

- Suas representações são facilmente visíveis no mundo real. Como todas as representações são feitas em forma de objetos, a compreensão dos mesmos se torna mais clara e perceptível, tanto para o programador quanto para um usuário.

- Fácil manutenção. A POO possui a capacidade de encapsulamento dos objetos, tornando-os entidades únicas, com propriedades e funções únicas. O encapsulamento no código reduz drasticamente a chance de que uma parte do sistema interfira no funcionamento de outra, reduzindo assim uma das mais comuns fontes de erros em programas.

- Facilidade de evolução. O encapsulamento fornece também modularidade, facilitando o desenvolvimento do código através de reuso e extensão de componentes (módulos) já existentes.

2.2.1 Java

Java é uma linguagem de programação desenvolvida pela Sun Microsystems em 1995. Existem atualmente inúmeras aplicações, em incontáveis plataformas, utilizando essa linguagem, que é definida pela sua desenvolvedora da seguinte forma: "Java é uma linguagem simples, orientada a objetos, distribuída, interpretada, robusta, segura, independente de arquitetura, portável, de alto desempenho, suportando multithreads e dinâmica."

Das características citadas acima, podemos destacar algumas vantagens que foram determinantes para que Java fosse escolhida como a linguagem de programação a ser usada nesse trabalho.

- Simplicidade: Java é considerada uma linguagem simples, com uma sintaxe baseada em C e C++, porém com diversas vantagens e melhorias em relação a essas.

- Orientada a Objetos: Utiliza o paradigma da POO, o que encaixa com o objetivo do projeto e permite uma analogia do mundo real muito mais clara e perceptível.

- Robusto: Por ser destinada para escrever programas que devem ser confiáveis numa variedade de maneiras, a linguagem Java coloca muita ênfase na verificação precoce de possíveis problemas, em checagens dinâmicas em tempo de execução e na eliminação de situações que são propensas a erros.

- Multithread: Multithreading é um modo de criação de aplicativos com vários segmentos ou processos em execução simultânea. Tal capacidade melhora a capacidade de resposta

interativa do sistema e seu comportamento em tempo real, resultando em uma velocidade de execução aprimorada.

2.3 Futebol

O futebol é um esporte de grande popularidade no mundo em geral, e principalmente no Brasil. Trata-se de um esporte coletivo com partidas que duram 90 minutos entre dois times de 11 jogadores cada.

Ele foi escolhido como tema desse trabalho por sua popularidade e por se tratar de um esporte considerado imprevisível.

Contudo, um certo padrão pode ser observado nos times mais vencedores. A quantidade de gols marcados e sofridos, o retrospecto jogando em "casa" e "fora de casa", a relação entre número de vitórias, empates e derrotas, entre outros. Nesse trabalho busca-se reconhecer esses padrões e, a partir deles, determinar com certa assertividade o resultado de jogos futuros.

2.3.1 Campeonato Brasileiro de Futebol

O Brasileirão, como também é conhecido o Campeonato Brasileiro de Futebol, é o principal torneio entre clubes de futebol do Brasil. O torneio foi iniciado em 1971 e ocorre anualmente desde então.

O campeonato historicamente não tinha uma padronização no sistema de disputa, o que gerava mudanças frequentes de regras e de número de participantes. Contudo, a partir de 2003 o sistema de pontos corridos foi adotado e a partir de 2006 o número de times participantes foi reduzido a 20. Esse formato de pontos corridos com 20 times foi determinado como o "formato definitivo" pela Confederação Brasileira de Futebol (CBF) e assim vem sendo utilizado ao longo dos anos.

O formato da competição é determinado de forma que durante o decorrer da temporada (de maio a dezembro), cada clube joga duas vezes contra os outros (em um sistema de pontos corridos), uma vez em seu estádio e a outra no de seu adversário, em um total de 38 jogos.

As equipes recebem três pontos por vitória e um por empate. Não são atribuídos pontos para derrotas.

As equipes são classificadas pelo total de pontos, depois pelo saldo de gols e, em seguida, pelos gols marcados. Em caso de empate entre dois ou mais clubes, os critérios de desempate são os seguintes: maior número de vitórias; maior saldo de gols; maior número de gols pró; confronto direto; menor número de cartões vermelhos recebidos; menor número de cartões amarelos recebidos.

O Campeonato Brasileiro de Futebol foi escolhido como tema de estudo por algumas de suas características, entre as quais podemos destacar:

- Longa duração. Por se tratar de um campeonato longo, é possível obter uma maior quantidade de informações para serem usadas no treinamento e/ou aplicação da Rede Neural Artificial, resultando em uma aplicação mais eficiente e confiável.

- Pontos corridos. O sistema de classificação em pontos permite uma melhor comparação entre as performances que os times vêm apresentando.

- Duas rodadas. Além de aumentar a quantidade de jogos para análise, a dupla rodada permite adicionarmos mais um parâmetro de entrada, sendo esse a condição de partida "em casa" ou "fora de casa".

- Vários Times. A quantidade de times permite diversas opções de confrontos, o que eleva as oportunidades de análise e aprendizado que a Rede Neural realiza.

- Proximidade e familiaridade. Por ser um campeonato de conhecimento geral, com regras simples e conhecidas, e fontes de dados facilmente acessíveis, o Campeonato Brasileiro de Futebol se mostra como um campo de estudo "sem mistérios".

- Sistema de disputa padronizado. Desde 2006 o formato do campeonato é o mesmo, sem previsões de alterações, o que facilita a extensão da abordagem de um ano para o outro e a comparação dos resultados entre anos diferentes.

Referências Bibliográficas

S. R. ALPERT, S. W. WOYAK, H. J. SHROBE, and L. F. ARROWOOD. Guest editor's introduction: Object-oriented programming in ai. *IEEE Expert: Intelligent Systems and Their Applications*, 5:6–7, 12 1990.

B. G. ASLAN and M. M. INCEOGLU. A comparative study on neural network based soccer result prediction. In *Seventh International Conference on Intelligent Systems Design and Applications*. IEE Computer Society, 2007.

C. M. BISHOP. *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.

J. P. CHANGEUX and A. DANCHIN. Selective stabilization of developing synapses as a mechanism for the specification of neuronal networks. *Nature*, 264:705–712, 1976.

T. CHENG, D. CUI, Z. FAN, J. ZHOU, and S. LU. A new model to forecast the results of matches based on hybrid neural networks in the soccer rating system. In *Fifth International Conference on Computational Intelligence and Multimedia Application*. IEE Computer Society, 2003.

L. P. DA COSTA. *Atlas do esporte no Brasil: Atlas do esporte, educação física e atividades de saúde e lazer no Brasil*. Shape Editora e Promoções Ltda., 2005.

A. E. ELO and S. SLOAN. *The Rating of Chess Players, Past and Present*. Ishi Press International, 2008.

L. V. FAUSSET. *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Prentice-Hall, 1994.

A. M. FLITMAN. Towards probabilistic footy tipping: a hybrid approach utilising genetically defined neural networks and linear programming. *Computers & Operations Research*, 33:2003–2022, 7 2006.

S. S. HAYKIN. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999a.

S. S. HAYKIN. *Neural Networks and Learning Machines*. Prentice Hall, 1999b.

R. HECHT-NIELSEN. *Neurocomputing*. Addison-Wesley Publishing Company, 1990.

J. L. McCLELLAND and D. E. RUMELHART. *Explorations in Parallel Distributed Processing*. Cambridge, MA: MIT Press, 1988.

W. S. MCCULLOTH and W. PITTS. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

A. MEHREZ and M. Y. HU. Predictors of the outcome of a soccer game - a normative analysis illustrated for the israeli soccer league. *Zeitschrift für Operations Research*, 42:361–372, 10 1995.

W. T. MINSKY and S. A. PAPERT. *Perceptrons, Expanded Edition*. Cambridge, MA: MIT Press, 1969.

M. C. PURUCKER. Neural network quarterbacking: How different training methods perform in calling the games. *Potentials, IEEE*, 15:9–15, 8 1996.

F. ROSENBLATT. *Principles of Neurodynamics*. New York: Spartan, 1962.

C. F. SILVA, E. S. GARCIA, and E. SALIBY. Soccer championship analysis using monte carlo simulation. In *2002 Winter Simulation Conference*, volume 1-2, pages 2011–2016, 2002.

G. S. STENT. A physiological mechanism for hebb's postulate of learning. *Proceedings of the National Academy of Science of the U.S.A.*, 70:997–1001, 1973.

M. WATSON. *Practical Artificial Intelligence Programming With Java*. Mark Watson, 2008.

B. WIDROW and M. E. HOFF Jr. Adaptive switching circuits. In *IRE WESCON Convention Record*, volume 4, pages 96–104, 1960.