CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA

FACULDADE DE TECNOLOGIA DE BOTUCATU

CURSO SUPERIOR DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

SERGIO DI FIORE

Desenvolvimento de um Software de Inteligência Artificial para Reconhecimento de ALGARISMOS ARÁBICOS MANUSCRITOS

Botucatu - SP

Dezembro – 2020

CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA

FACULDADE DE TECNOLOGIA DE BOTUCATU

CURSO SUPERIOR DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

SERGIO DI FIORE

Desenvolvimento de um Software de Inteligência Artificial para Reconhecimento de ALGARISMOS ARÁBICOS MANUSCRITOS

Orientador: Prof. Dr. Gustavo Kimura Montanha

Coorientadora: Profa. Dra. Mônica Regina Gaiotto

Relatório de Iniciação Científica apresentado à FATEC - Faculdade de Tecnologia de Botucatu, como exigência para cumprimento do Trabalho de Conclusão de Curso no Curso Superior de Análise e Desenvolvimento de Sistemas.

Botucatu - SP

Dezembro – 2020

**RESUMO**

O projeto trata do desenvolvimento de um software que visa o reconhecimento de caracteres de algarismos arábicos manuscritos visando a automação de processos de leitura de números manuscritos para diversas finalidades, através de técnicas de Inteligência Artificial, Rede Neural e *Deep Learning*.

**Palavras-chave:** Deep Learning, Inteligência Artificial, Rede Neural,Software.

LISTA DE FIGURAS

[Figura 1. Dois neurônios transmitindo impulso nervoso por meio de sinapse 8](#_Toc64561818)

[Figura 2. Neurônio artificial 8](#_Toc64561819)

[Figura 3. Número 504.192 manuscrito 10](#_Toc64561820)

[Figura 4. Múltiplos diferentes exemplos manuscritos tornam-se exemplares para treinamento de uma rede neural 10](#_Toc64561821)

[Figura 5. Perceptron 11](#_Toc64561822)

[Figura 6. Percepron com mais de uma camada 12](#_Toc64561823)

[Figura 7. O efeito de pequenas alterações em pesos ou bias 13](#_Toc64561824)

[Figura 8. O neurônio sigmóide é muiti próximo ao perceptron 14](#_Toc64561825)

[Figura 9. Gráfico da função sigmoide 15](#_Toc64561826)

[Figura 10. Função degrau ou de Heaviside 15](#_Toc64561827)

LISTA DE EQUAÇÕES

[Equação 1. Cálculo binário de saída de um Perceptron 9](#_Toc64561881)

[Equação 2. Cálculo do valor de saída de um Perceptron 12](#_Toc64561882)

[Equação 3 atualizada 13](#_Toc64561883)

[Equação 4. Cálculo da saída de um neurônio sigmoide 14](#_Toc64561884)

[Equação 5. Cálculo da saída sigmoide considerando pesos e bias 14](#_Toc64561885)

[Equação 6. Cálculo aproximado de ∆saída 16](#_Toc64561886)

Sumário

[1 INTRODUÇÃO 7](#_Toc65079827)

[2 REVISÃO DE LITERATURA 7](#_Toc65079828)

[2.1 Inteligência Artificial, Redes Neurais e *Deep Learning* 7](#_Toc65079829)

[2.2 Usando redes neurais para reconhecer algarismos manuscritos 9](#_Toc65079830)

[2.2.1 Perceptron 11](#_Toc65079831)

[2.2.2 Neurônios Sigmoides 13](#_Toc65079832)

[3 REFERÊNCIAS 17](#_Toc65079833)

# INTRODUÇÃO

A Inteligência Artificial em suas várias escolas tem sido cada vez mais presente na solução de problemas informáticos que necessitam tratar grandes quantidades de dados e a partir deles, prever alguma espécie de resultados. Redes neurais artificiais nada mais são do que a simulação das neurais biológicas por meio de componentes eletrônicos, algoritmos e software e o *Deep Learning* uma extensão das Redes Neurais mais tradicionais pelo aprofundamento de camadas intermediárias e algoritmos mais adequados de aprendizado. Através da técnica de Aprendizado de Máquina que se vale desses recursos supracitados, pretende-se desenvolver um sistema capaz de identificar algarismos arábicos manuscritos, uma aplicação que pode ser empregada para diversos fins práticos de entrada de dados a partir de documentos manuais.

# REVISÃO DE LITERATURA

## Inteligência Artificial, Redes Neurais e *Deep Learning*

Pode-se dizer que a Inteligência Artificial (IA) começou a ser estudada sistematicamente a primeira vez por Hebert Simon e Allen Newell ainda no início da década de 1950 na Universidade de Carnegie Mellon. Em seus primeiros estágios, a IA resolveu problemas que são intelectualmente difíceis para os seres humanos, mas simples para a lógica matemática eletrônica. O desafio se provou na resolução de problemas que resolvemos intuitivamente de uma forma simples, mas que humanos tem dificuldades em descrevê-los.

Inteligência artificial nada mais é do que estatística levada ao seu extremo, onde modelos matemáticos são calculados por estruturas computacionais que se inspiraram no cérebro humano. Isto é, “neurônios” que realizam um determinado tipo de operação, levando o resultado de um para outro, em uma organização que emula a transferência de informação de uma real célula nervosa para outra, no sistema animal (Figuras 1 e 2).

Essa estrutura de “células” nervosas artificiais e seus algoritmos permitem que padrões existentes em uma série, normalmente bem grande de dados, possam ser aprendidos para seu posterior uso, decodificando padrões. Por exemplo, mostrando diversas imagens de gatos e posteriormente verificando se uma determinada imagem submetida, sem que esse jamais a tenha analisado, se possui ou não uma imagem de um gato. É a chamada Aprendizagem de Máquina (*Machine Learning*). Com ela, pode-se encontrar padrões em quantidade enorme de dados. Esse modelo, que ainda funciona muito bem, mas que mesmo desde esse período inicial, deixava muitas questões sem solução aparente.

Figura 1. Dois neurônios transmitindo impulso nervoso por meio de sinapse

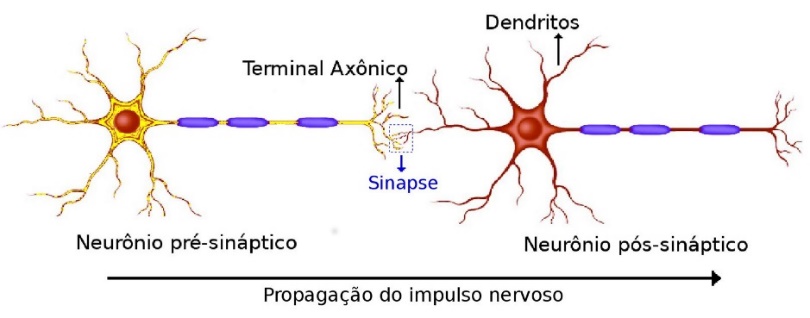
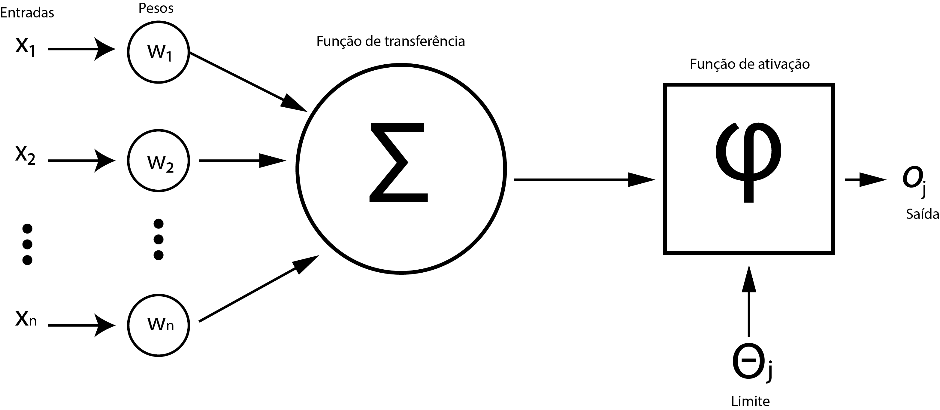


Figura 2. Neurônio artificial



É conveniente neste ponto ressaltar a importância desse aprendizado autônomo. Um sistema para ser entendido como inteligência artificial, deve ser capaz de aprender por si, não dependendo de instruções de programadores.

Em 1969, Marvin Minsky e Saymour Papert publicaram um livro pelo MIT provando matematicamente que esse tipo de redes, chamadas Perceptron (Fórmula 1), poderiam somente executar funções básicas (SOMMERS, 2017). Elas, então, possuíam somente duas camadas de neurônios, uma camada de entrada de dados e uma camada para a saída, e embora, na teoria mais camadas pudessem resolver o problema, ninguém possuía conhecimento suficiente de como treiná-las.

Equação 1. Cálculo binário de saída de um Perceptron

Somente em 1986, Geoffrey Hinton, hoje conhecido como o pai do *Deep Learning*, trouxe novos avanços (HAO, 2018). Ele basicamente, conseguiu desenvolver o treinamento de mais camadas, uma técnica que, iria além do tradicional aprendizado de máquina, permitindo encontrar pequenos padrões, que anteriormente não conseguiriam ser detectados. Em 1990 essa técnica foi reforçada pela assim chamada Máquina de Suporte Vetorial (*Support Vector Machine*) onde estruturas espaciais e funções permitem as transições que simplificam o processamento (WILSON, 2020).

Tal técnica, porém, esbarrava na limitação da capacidade computacional dos sistemas na época e o campo toda da IA ficou por muito tempo parado, até em torno do ano 2000, quando a capacidade de processamento permitiu a retomada das pesquisas (SOMMERS, 2017).

Nesse mesmo período começou-se a empregar a Distribuição Circulares de Autovalores associada a um antigo método matemático conhecido como a Teoria da Matriz Aleatória (*Random Matrix Theory*) minimizando o obstáculo trazidos pelas complexas matrizes que o aprendizado de máquina precisava lidar (WILSON, 2020).

Redes Neurais são um dos mais belos paradigmas de programação jamais inventados. Na abordagem convencional de programação, o computador recebe informações detalhadas do que fazer, dividindo grandes problemas em diversos menores, definem-se precisamente as tarefas que o computador pode executar de maneira simples. Em contraste, em uma Rede Neural nenhuma informação é dada ao computador de como resolver um problema. Ao contrário, ele aprende a partir da observação de dados, identificando a sua própria solução ao problema em questão (Nielsen M. 2015)

Segundo Kelleher J. D. (2019),” *Deep Learning* é um subcampo da Inteligência Artificial que foca em criar grandes modelos de redes neurais que são capazes de fazer previsões precisas a partir de dados. *Deep Learning* é particularmente adequada aos contextos que possuem grandes quantidades de amostragem de dados disponíveis.

## Usando redes neurais para reconhecer algarismos manuscritos

Qualquer ser humano reconhece com facilidade os dígitos:

Figura 3. Número 504.192 manuscrito



A figura 3 exibe o número 504.192 manuscrito por uma pessoa e subsequentemente digitalizado em uma resolução não muito alta como pode-se perceber. O cérebro humano, possui em cada um de seus hemisférios o córtex visual primário, também conhecido como V1 contendo 140 milhões de neurônios, com dezenas de bilhões de conexões entre eles. Ainda assim, a visão não depende somente do V1, mas de toda uma série de córtices: V2, V3, V4 e V5 cada um fazendo progressivamente mais do que o anterior. Milhões de anos de evolução garantiram que o cérebro humano se tornasse um supercomputador para o qual a identificação de dígitos, como os representados, extremamente eficiente.

A dificuldade do reconhecimento de padrões visuais torna-se clara quando tenta-se escrever um programa convencional de computador para fazê-lo. Como reconhecer um “9”? Uma bolinha em cima e uma perna curvada. Expressar isso em termos de algoritmos é extremamente complexo, e ainda, essa representação pode ter uma infinidade de variáveis.

Redes neurais abordam esse mesmo problema de uma forma totalmente distinta. A ideia é tomar uma grande quantidade de algarismos manuscritos conhecidos como exemplares de treino (Figura 4) e desenvolver um sistema que pode aprender a partir destes. Ou seja, a rede neural usa os exemplos para ela própria inferir regras para reconhecer os algarismos manuscritos.

Figura 4. Múltiplos diferentes exemplos manuscritos tornam-se exemplares para treinamento de uma rede neural



Além disso, o incremento do número de exemplos propícia um melhor aprendizado à rede, melhorando o seu desempenho no reconhecimento. Assim, enquanto os 100 dígitos acima refletem um certo potencial de acerto nas inferências quando o sistema analisa dados desconhecidos, usando-se milhares, talvez milhões de exemplos melhoram imensamente essa capacidade.

O primeiro objetivo deste presente projeto é o de atingir uma meta de acerto de 96% sem a intervenção humana. Dentro deste primeiro objetivo, serão desenvolvidos os conhecimentos de dois importantes tipos de neurônios artificiais: o Perceptron e o Neurônio Sigmoide, além de um algoritmo básico para as redes neurais conhecido como Gradiente Descendente Estocástico.

## Perceptron

Desenvolvido entre os anos 1950 e 1960 por Frank Rosenblatt, com base em trabalhos anteriores de Watten McCulloch e Walter Pitts, Rosenblatt apresentou um algoritmo que nos trina anos subsequentes foram alvos de inúmeros trabalhos e estudos acadêmicos mostrando sua eficácia, mas sem que ele mesmo tenha de alguma forma justificado o resultado ao qual chegou.

O Perceptron toma diversas entradas binárias para produzir uma saída também binária como a figura 5 mostra:

Figura 5. Perceptron



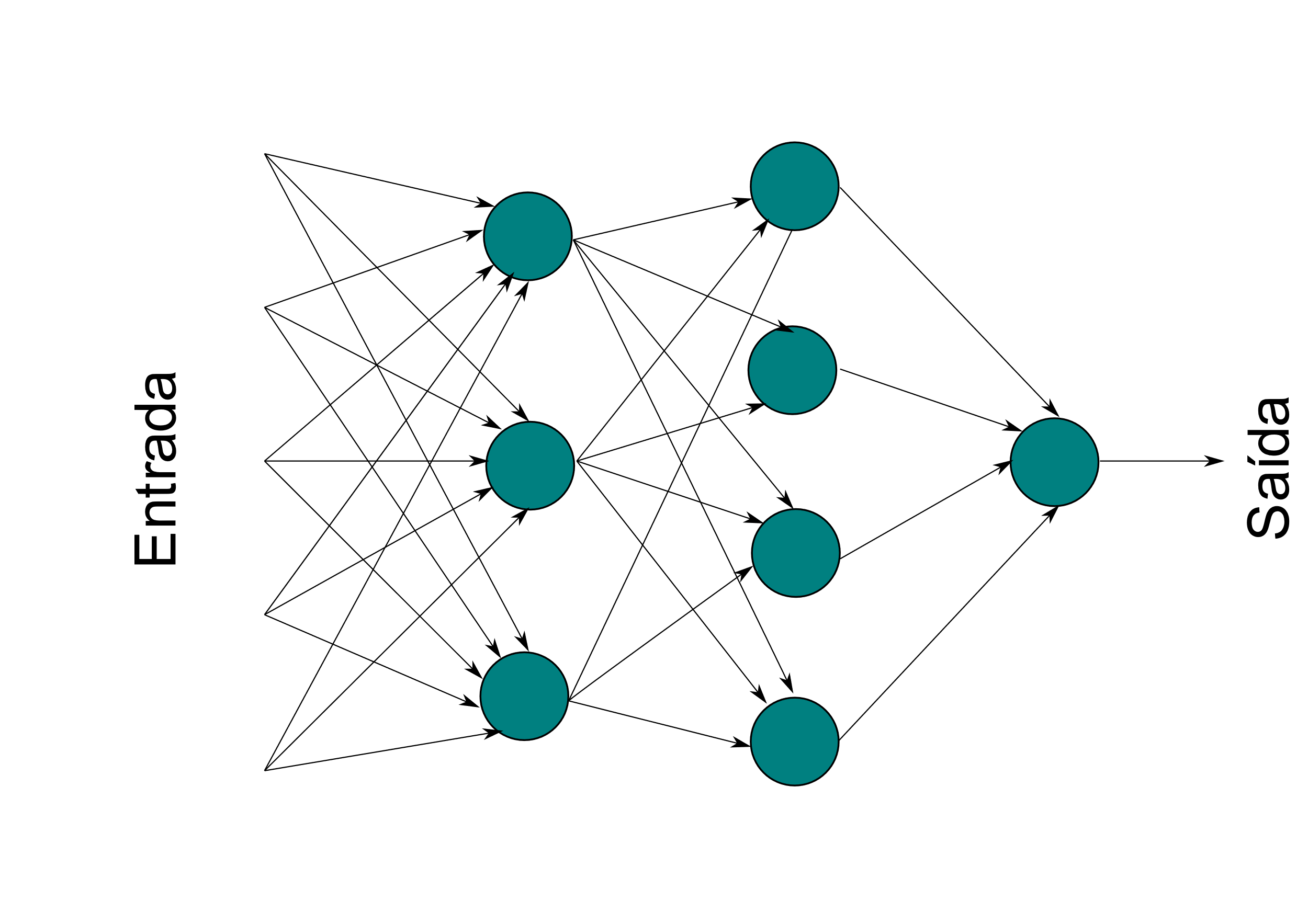
No exemplo, n entradas binárias são processadas para produzir uma saída também binária. Rosenblatt propôs pesos, *w*1, *w* 2, ..., *w* n, números reais indicando a importância de cada entrada. A saída é calculada pela soma do valor de entrada com o seu peso (*w*. *x*), para todas as entradas () atribuindo-se 0 se for menor ou igual que um determinado valor limite e 1 se for maior, como mostrado pela equação 2.

Equação 2. Cálculo do valor de saída de um Perceptron

Uma forma conveniente de entender o Perceptron é que é um dispositivo capaz de tomar decisões pesando evidências.

A figura 6 traz um perceptron com uma camada adicional às de entrada e de saída. A primeria camada pesa as entradas e toma alguma decisão. Neste exemplo, a saída dessas entradas são alimentadas numa nova camada ( a segunda exemplo) que com os seus próprios pesos toma alguma outra decisão e encaminha para mais uma camada (no exemplo, a terceira) que repete o processo, para só então determinar um resultado final como saída. Dessa forma um perceptron de muitas camadas pode decisões sofisticadas.

Figura 6. Percepron com mais de uma camada



Pode-se simplificar a descrição dos perceptrons. Inicialmente alterando , mudando para o produto escalar onde e são vetores respectivamente dos pesos e entradas. A segunda mudança é mudar o limite para o outro lado da inequalidade, substituindo-o pelo que é conhecido como tendência (o termo é mais usado em inglês: *bias*). Assim, podemos reescrever a equação 2 pela equação 3:

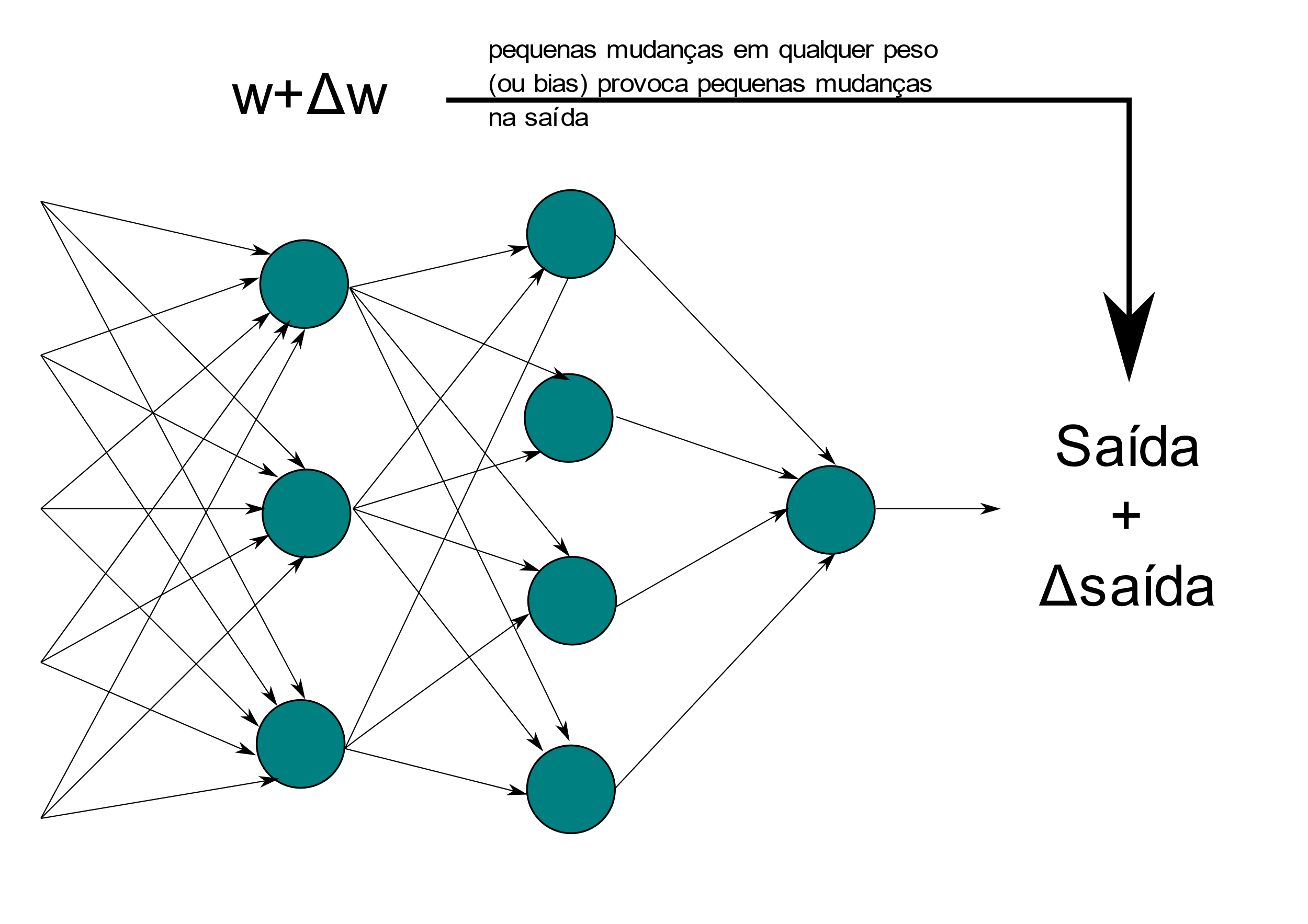
Equação 3 atualizada

Pode-se pensar no *bias* como quão facilmente um perceptron pode atingir o valor 1. Para *bias* com valores altos, a chance de se chegar ao valor 1 é alta, já se o *bias* for um grande valor negativo, isto será difícil. Deste ponto em diante do trabalho nós só usaremos o *bias* e não mais o limite.

## Neurônios Sigmoides

Suponha-se uma rede de perceptrons que é alimentada com o resultado bruto da digitalização de algarismos escritos à mão. É desejável que a rede aprenda pesos e *bias* de tal forma que na saída tenha-se os dígitos classificads de forma correta. Para entender como o processo de aprendizagem pode funcionar, imagina-se (Figura 7) conduzir pequenas mudanças em alguns dos pesos (ou dos *bias*) da rede. Deseja-se que às pequenas mudanças referidas somente pequenas mudanças ocorram correspondentemente na saída. Como será visto na sequência, é justamente essa propriedade que torna possível o processo de aprendizagem.

Figura 7. O efeito de pequenas alterações em pesos ou bias



Se fosse verdadeiro que pequanas modificações em pesos (ou *bias*) provocam somente pequenas modificaçõas no valor de saída, poderia-se usar essa fato para a rede se comportar mais como se deseja. Por exemplo, se é esperado um resultado com o valor 8 e está sendo obtendido o 9, pequenas alterações nos pesos e *bias* levariam mais próximos ao resultado desejado, e repetindo esse processo a precisão melhoraria.

Mas, não é o que acontece com os perceptrons. É possível que uma pequena alteração em um peso ou *bias* cause a mudança completa de estado de um perceptron, por exemplo, de 0 para 1.

Esse problema pode ser evitado com o uso do neurônio sigmóide (Figura 8). Semelhante ao perceptron mas permite justamente que pequenas mudanças em pesos e *bias* reflitam em pequenas mudanças no valor de saída.

Figura 8. O Neurônio sigmóide é muiti próximo ao perceptron



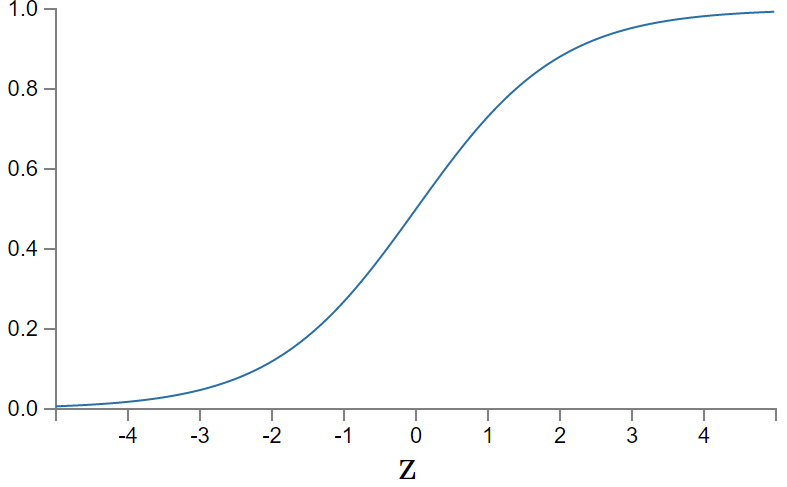
Como o perceptron, o sigmóide também tem pesos para cada entrada, w1, w2, ..., wn. A saída, ao invés de 0 ou 1 esta é calculada pela função sigmoide como ilustrado pela equação 4:

Equação 4. Cálculo da saída de um neurônio sigmoide

De forma mais explicita, agora com pesos e *bias* tem-se:

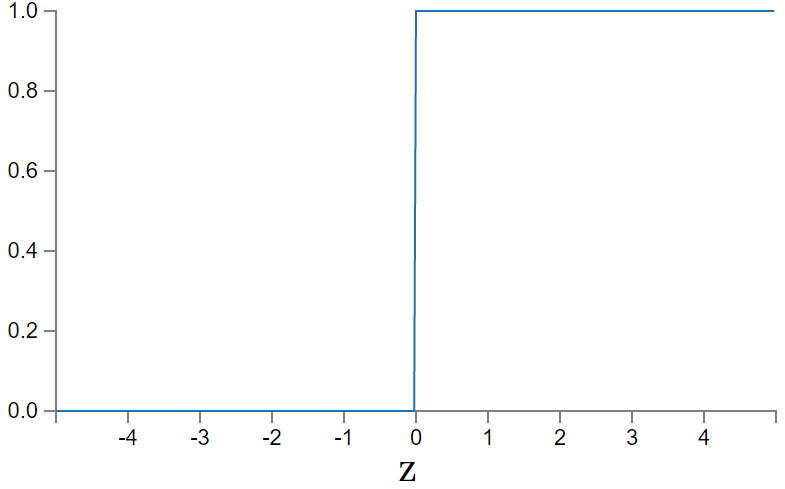
Equação 5. Cálculo da saída sigmoide considerando pesos e bias

Figura 9. Gráfico da função sigmoide



ertghyrPara entender a similaridade entre perceptron e sigmoide, suponha-se sendo um grande número positivo. Então e por isso, . Ou seja, quando é um número grande positivo, a saida da função sigmoide é aproximadamente 1, como ocorreria com um perceptron. Quando é um valor de grande magnitude negativa, então , causando, . Isto é, para valores grandes negativos, o comportamento da função sigmoide também parece a do perceptron. Somente quando possuir valores moderadamente intermediários é que vai haver um desvio considerável do perceptron.

Figura 10. Função degrau ou de Heaviside



Se fosse de fato uma função degrau, o neurônio sigmoide seria um perceptron. Pelo uso da função obtem-se um perceptron “suavizado”. E justamente essa suavização que é o fator crucial. Ela significa que pequenas alterações dos pesos ∆wj e das *bias* ∆b produziram uma pequena alteração na saída ∆saída. Pode-se saber por meio da disciplina do cálculo que ∆saída será aproximadamente:

Equação 6. Cálculo aproximado de ∆saída

onde a soma é entorno de todos os pesos, wj, e denotam as derivadas parciais a saída com relação a wj e b respectivamente. Portanto, ∆saída é uma função linear das mudanças que ocorrem com ∆wj e ∆b.

Como a neurônios sigmoides trabalham com qualquer valor entre 0 e 1, isso é útil, por exemplo, para se considerar o valor de saída como a intensidade média de um pixel da imagem que estamos trabalhando.

## Arquitetura das redes neurais

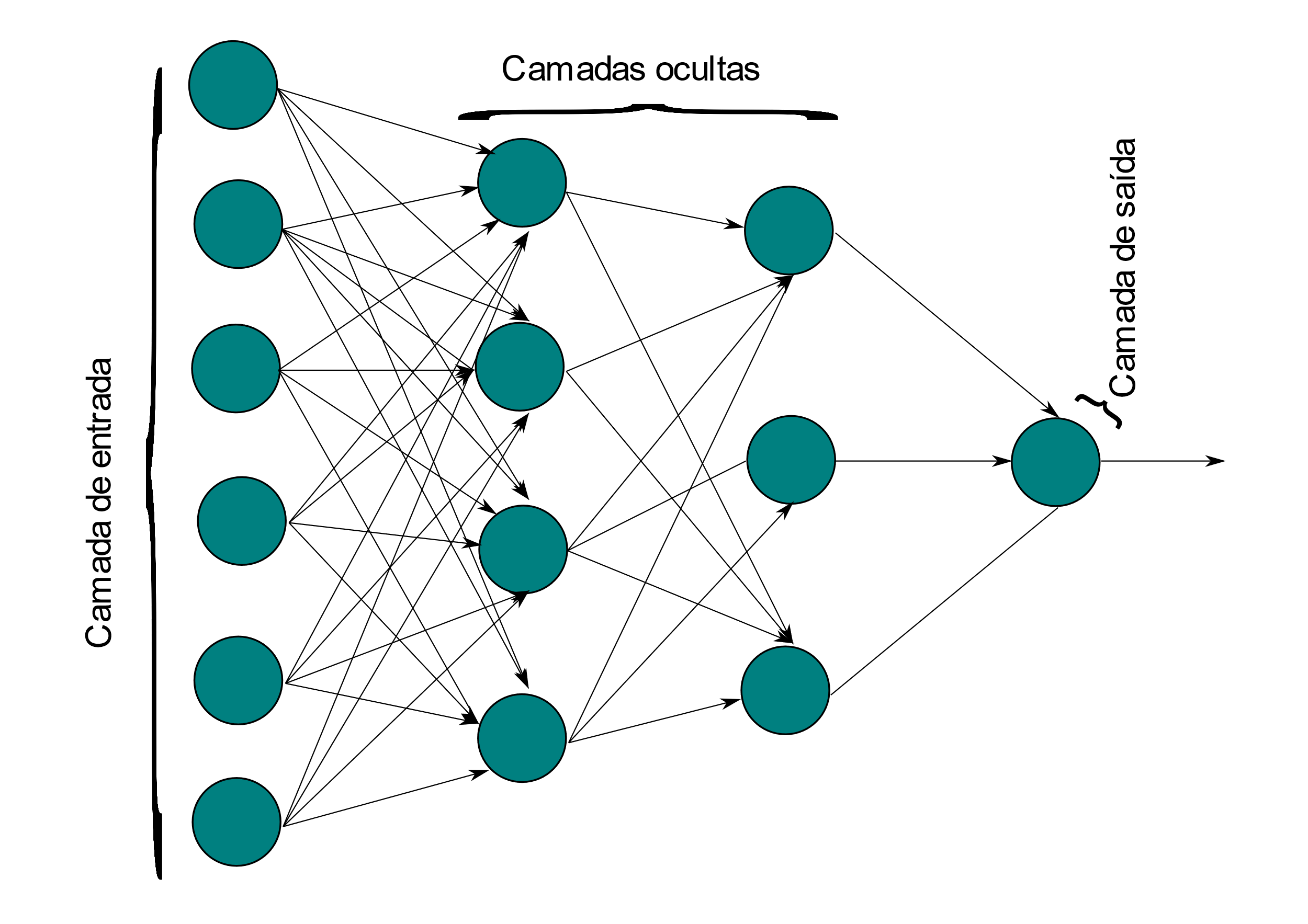
Suponha-se a rede da fixura 11 abaixo:

Figura 11. Uma rede neural genérica



Como anteriormente mencionado, a parte esquerda chamada de camada de entrada contém os denominados neurônios de entrada. A parte direita chamada camada de saída possui os chamados neurônios de saída (note que a figura mostra somente um, mas isso não é necessário). O meio é denominado de camadas escondidas *(hidden layer*) uma vez que não é de entrada nem de saída. A rede da figura 11 possui somente uma camada de saída, mas isso não é necessário, como é o exemplo da figura 12.

Figura 12. Rede neural com mais de uma camada intermediária



Vale ressaltar que, apesar da figura 12 exibir duas camadas ocultas, não existe um número predeterminado e portanto, pelo menos teoricamente, para a sua quantidade também.

Dando, com bom motivo, margem a discussões, essas redes multicamadas são muitas vezes denominadas perceptrons multicamadas (*multilayer perceptrons –* MLPs) apesar de serem constituídas de neurônios sigmoides. Esta terminologia não é usada no presente trabalho.

As camadas de entrada e saída são bem fáceis de entender. Imagine-se que se está tentando identificar se uma imagem é o algarismo “9”. Codificam-se os pixels da imagem da camada de entrada de acordo com a intensidade deste, por exemplo em uma escala de tons de cinza. Se a imagem for uma matriz de 64 por 64 pixels tem-se 4.096 neurônios de entrada com valores que variam entre 0 e 1. A camada de saída contém um único neurônio onde valores inferiores a 0,5 são entendidos com “não é um 9” e valores entre 0,5 e 1 como confirmação de que o algarismo é um “9”.

Já as camadas ocultas têm no seu funcionamento uma verdadeira arte. É extremamente complexo resumi-las em alguma regra simples e muitas heurísticas para tal foram desenvolvidas. Um exemplo considera a quantidade das camadas ocultas versus o tempo necessário para treinar a rede. Algumas dessas heurísticas serão discutidas neste trabalho.

Até o momento foram discutidas redes na qual as informações se propagam em um único sentido, da camada de entrada para a camada de saída. São as chamadas redes *feedforward*, onde não existe nenhuma espécie de laço dos dados. Caso houvesse ter-se-ia a entrada para a função dependendo da saída, o que não faz muito sentido.

No entanto deve-se ressaltar que existem redes neurais na qual laços de retroalimentação ocorrem. O conceito é o de se ter neurônios que se ativam por um período limitado antes de se tornarem quiescente. A sua atividade estimula outros que podem se tornar ativos um pouco depois e assim, ao longo do tempo tem-se uma cascata de atividade dos neurônios e, neste modelo, laços não afetam o funcionamento uma vez que a entrada será dada em algum momento posterior.

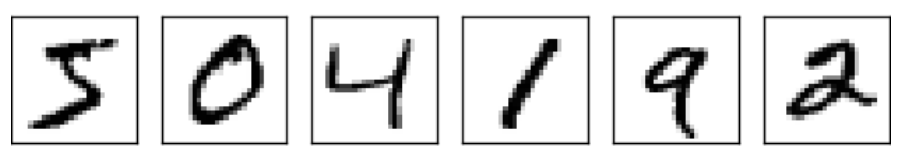
## Uma rede simples para classificar algarismos manuscritos

Inicialmente divide-se o problema em unidades menores, por exemplo, toma-se o número manuscrito, tomando-se por exemplo o úmero 504.192 que pode ser manuscrito como o exemplo da figura 13 e divide-se em nível dos dígitos como exibido pela figura 14, onde separou-se cada um dos dígitos compondo seis unidades distintas.

Figura . Exemplo de problema a ser reconhecido



Figura . O problema dividido em subproblemas: os dígitos separados



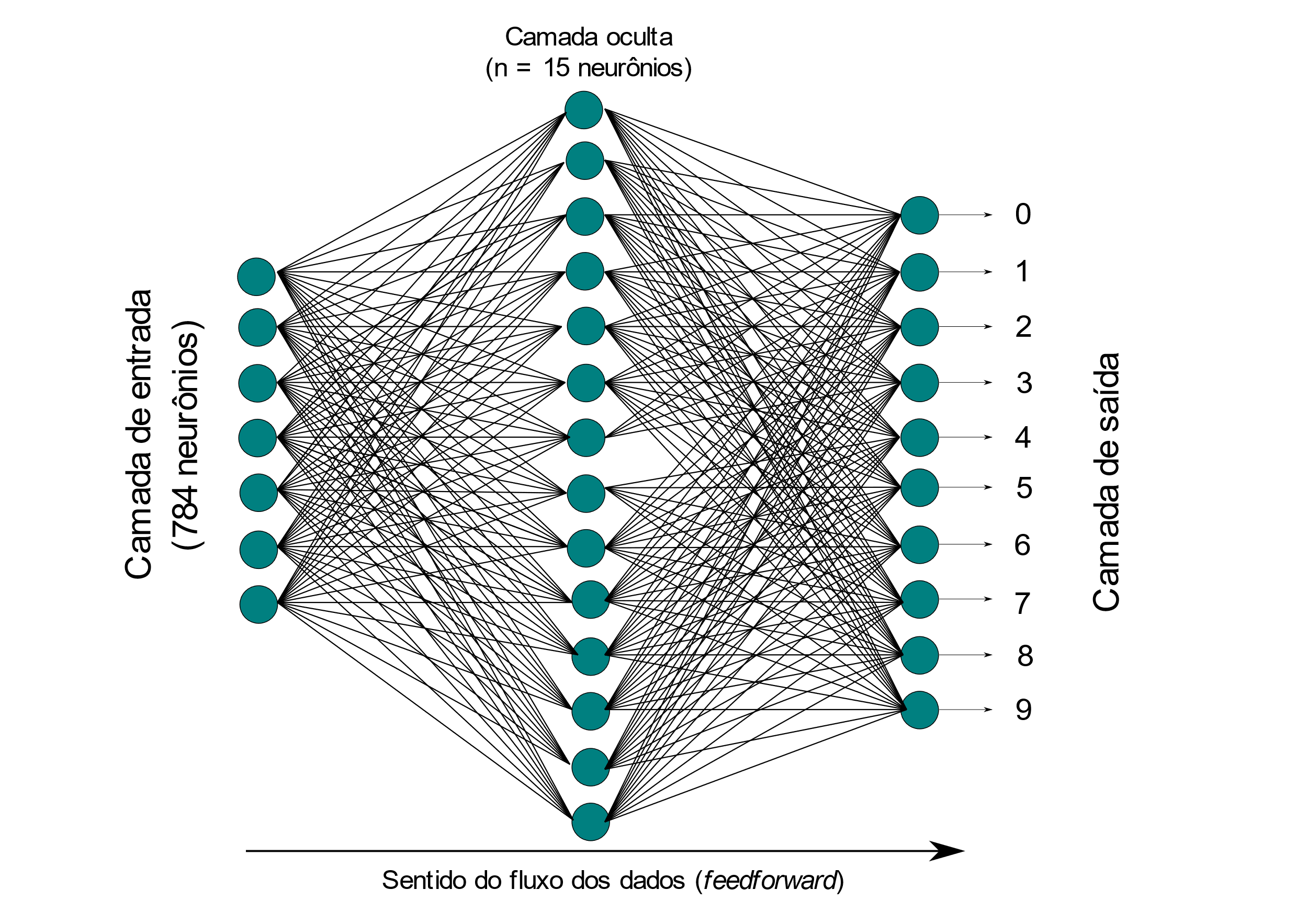
Humanos fazem essa segmentação de forma extremamente simples e natural, mas isso se constitui um verdadeiro desafio em termos computacionais. Uma vez feita essa divisão, o prgrama precisa classificar cada componente (dígito) individualmente. Tome-se por exemo o dígito que fácilmente um humano reconhece como o valor 5, à figura 15.

Figura . O algarismo 5 individualizado



Neste momento será dado enfoque em escrever um programa que resolva esse segundo problema, isto é, classificar os dígitos separadamente. Existem muitas abordagens que resolvem esse problema. A intensão é desenvolver uma rede neural que possa reconhecer os dígitos manuscritos individualmente. Será empregada uma rede neural com três camadas, como exibido pela figura 16.

Figura . Rede neural a ser empregada para reconhecimento dos dígitos individuais



# REFERÊNCIAS

Nielsen, Michael A., **Neural Network and Deep Learning**, Determination Press, 2019, <http://neuralnetworksanddeeplearning.com/>.

Goodfellow et al, **Deep Learning**, MIT Press, 2016, <http://www.deeplearningbook.org>.

**6.036 Introduction to Machine Learning**, MIT Open Learning Library, Massachusetts Institute of Technology, 2021

Kelleher, John D., Deep Learning (The MIT Press Essential Knowledge series) (p. 3). MIT Press. Kindle Edition.

Glassner, Andrew., **Deep Learning, Vol. 1: From Basics to Practice** (p. a). Kindle Edition.

Glassner, Andrew., **Deep Learning, Vol. 2: From Basics to Practice** (p. a). Kindle Edition.

Mining, Ethem., **Machine Learning: 4 Books in 1: Basic Concepts + Artificial Intelligence + Python Programming + Python Machine Learning. A Comprehensive Guide to Build Intelligent Systems Using Python Libraries**. Kindle Edition.

**18.01 Single Variable Calculus**, MIT Open Learning Library, Massachusetts Institute of Technology, 2021

**18.06 Linear Algebra**, MIT Open Learning Library, Massachusetts Institute of Technology, 2021

Marc Peter Deisenroth Et al., **Mathematics for Machine Learning**. Pre-publication version to be published by Cambridge University Press.

Dawani, Jay., **Hands-On Mathematics for Deep Learning: Build a solid mathematical foundation for training efficient deep neural networks**. Packt Publishing. Kindle Edition.

Tom Pratt Et al., **Docs/.NET/C# guide/Language reference/C# 6.0 draft specification**, [Introduction - C# language specification | Microsoft Docs](https://docs.microsoft.com/en-us/dotnet/csharp/language-reference/language-specification/introduction?source=docs);

Price, Mark J. **C# 9 and .NET 5 – Modern Cross-Platform Development: Build intelligent apps, websites, and services with Blazor, ASP.NET Core, and Entity Framework Core using Visual Studio Code** (p. 1). Packt Publishing. Kindle Edition.

Sam Wood, **A Brief History of Python**, Packt Books, [A Brief History of Python | Packt Hub (packtpub.com)](https://hub.packtpub.com/brief-history-python/);

Sanderson, Grant, **Neural Networks**, 3Blue1Brown, [(2799) Neural networks - YouTube](https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi).