UNIVERSIDADE FEDERAL FLUMINENSE

ANTONIO CARLOS DO NASCIMENTO CUNHA JUNIOR

TÍTULO DO TRABALHO

Niterói

2019ANTONIO CARLOS DO NASCIMENTO CUNHA JUNIOR

TÍTULO DO TRABALHO

Trabalho de Conclusão de Curso submetido ao Curso de Tecnologia em Sistemas de Computação da Universidade Federal Fluminense como requisito parcial para obtenção do título de Tecnólogo em Sistemas de Computação.

Orientador(a):

ALTOBELLI DE BRITO MANTUAN

NITERÓI

2019

ANTONIO CARLOS DO NASCIMENTO CUNHA JUNIOR

TÍTULO DO TRABALHO

SUBTÍTULO

Trabalho de Conclusão de Curso submetido ao Curso de Tecnologia em Sistemas de Computação da Universidade Federal Fluminense como requisito parcial para obtenção do título de Tecnólogo em Sistemas de Computação.

Niterói, \_\_\_ de \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ de ANO.

Banca Examinadora (provisório):

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. ou Profa. <NOME>, <Título>. – Orientador ou Avaliador

<Sigla da Universidade> - <Nome da Universidade>

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. ou Profa. <NOME>, <Título>. – Orientador ou Avaliador

<Sigla da Universidade> - <Nome da Universidade>

AGRADECIMENTOS

RESUMO

A ser feito.

Palavras-chaves: palavra1, palavra2 e palavra3.

ABSTRACT

A ser feito.

Key words:

LISTA DE ILUSTRAÇÕES

[**Figura 1: Programadoras trabalhando no ENIAC 19**](#__RefHeading___Toc378694305)

**Figura 2: Relação de Aprendizado de Máquina com Mineração de Dados …….19**

**Figura 3: Funcionamento das camadas de uma Rede Neural …………………….19**

**Figura 4: Representação de um Neurônio Artificial …………………………………19**

LISTA DE ABREVIATURAS E SIGLAS

ENIAC – *Electronic Numerical Integrator and Computer*

IA – Inteligência Artificial

ML – *Machine Learning*

RELU – *Rectified Linear Unit*

**2 – F****UNDAMENTAÇÃO TEÓRICA**

Neste capítulo estão apresentados os itens e subitens objetos deste trabalho, bem como suas definições, ou, o que se tem como entendimento de seus conceitos. Estão também abordados, brevemente, eventos importantes na cronologia que nos levou até o presente momento de interesse profundo sobre *Machine Learning* e Redes Neurais.

**2.1 – APRENDIZADO DE MÁQUINA (MACHINE LEARNING)**

Em uma época onde o ENIAC era o único computador disponível, e, sendo operado manualmente funcionava como uma grande calculadora, como mostrado na figura 1, Alan Turing[[1]](#footnote-1) testava a capacidade de uma máquina aprender por meio de comunicação humana. Seus testes não surtiram os efeitos desejados, mas possibilitaram o desenvolvimento de sistemas que conseguissem.



Figura 1 - Programadoras trabalhando no ENIAC

O título de precursor do Machine Learning (e até da própria Inteligência Artificial), ficou por conta de Arthur Samuel que desenvolveu o *Game of Checkers*, jogo de damas em que uma inteligência artificial conseguiu superar a estratégia humana e vencer o jogo. Arthur então definiria *Machine Learning* como ‘um campo de estudos que dá ao computador a capacidade de aprender sem serem explicitamente programados para isso’.

O campo de estudos do MLhoje, foca no desenvolvimento de programas computacionais que possam acessar uma informação e usá-la para ela mesma. O processo de aprendizado começa com observações ou informações, como exemplos, experiências diretas ou instruções, visando, a detecção de padrões nessas informações, e a melhor maneira de tomar decisões no futuro baseando-se em exemplos que lhe foram providos. O objetivo dessa técnica é permitir esse aprendizado sem intervenção ou assistência humana direta, apenas pequenos ajustes muito pontuais.

2.1.1 – ORIGEM NA INTELIGÊNCIA ARTIFICIAL

O grande sonho da Inteligência Artificial, desde o surgimento de sua ideia inicial nos idos dos anos cinquenta, era proporcionar máquinas que pudessem pensar de forma que consideramos inteligente. O conceito presente na IA impulsionou grandes produções científicas tanto no cinema como na literatura, mas é fácil perceber que estamos bem distantes ainda de convivermos com ciborgues ou androides. Uma máquina que pensasse de forma tão inteligente quanto um ser humano, ou até mais, seria chamada de IA Forte, que segundo John Searle, “deveria ter um cérebro propriamente dito, de tal maneira que entendesse e possuísse estados cognitivos”[1].

Enquanto esse modelo não se torna realidade, convivemos habitualmente com modelos limitados de inteligência artificial (IA Fraca) que fazem tarefas específicas, como lidar com reconhecimento facial ou dirigir carros autônomos. Sistemas de inteligência artificial então, podem ser abastecidos com dados explícitos que o levam a executar uma tarefa com precisão invejável. E a técnica que uma IA utiliza é que diferenciará a sua classificação, uma dessas técnicas de aplicação é o *Machine Learning.*

2.1.2 – ÁREAS DE UTILIZAÇÃO E APLICAÇÃO

Aos nos darmos conta que convivemos com IA’s por toda nossa volta, passamos as vezes a nos perguntar de que forma tal aplicação foi instruída para sua funcionalidade. Assistentes virtuais, como dos sistemas operacionais dos *smartphones*, coletam e refinam dados de cada interação nossa com eles. Sistemas de tráfego, utilizam informações em tempo real do trânsito e da demanda, para calcular rotas e até preços variáveis. Recomendação de produtos gravam algumas informações de nossas buscas e interesses para oferecer algo que tenhamos maior chance de comprar.

Esses exemplos utilizam aplicações mesmo que pequenas de ML, assim como o fazem sistemas de detecção de fraudes, sistemas de busca, suporte ao usuário, serviço de controle de spam, redes sociais e sistemas de vigilância.

O campo de estudos do Aprendizado de Máquina então, está presente dentro do de Inteligência Artificial, como uma subárea, onde se torna como um método no qual se viabiliza a evolução dessa tecnologia, que necessita lidar com o crescente número de dados gerados e disponíveis para análise.

2.1.3 – RELAÇÃO COM MINERAÇÃO DE DADOS

O grande problema desses dados, é a capacidade que nós temos, como seres humanos e limitados que somos em relação a nossa própria velocidade e racionalidade, de lidar com eles. Tornou-se necessário então, aprimorar a capacidade de instrucionalidade de programas, para que se pudesse acompanhar essa aceleração na geração de dados sem ser necessária uma intervenção humana a todo momento.

A mineração de dados aborda essa questão, na procura por padrões consistentes e na separação daquilo que é importante para o analista, daquilo que não é. Consiste em um processo que utiliza algoritmos do ML combinado com estatística para ajudar a colher informações importantes de banco de dados com uso de uma computação de baixo custo, essa relação entre áreas é demonstrada na figura 2. A mineração é viável tanto para a pesquisa científica, como para o ramo empresarial e de serviços, e é eficiente tanto para dados estruturados, como para dados não estruturados.

****

Figura 2 - Relação com Mineração de Dados

2.1.4 – DADOS ESTRUTURADOS E NÃO ESTRUTURADOS

Quando gerados e armazenados através de uma estrutura previamente definida para esta finalidade, os dados são considerados estruturados [2]. São fáceis de serem localizados e possuem ligações com outros dados que os relacionam a algo. Planilhas, formulários e principalmente banco de dados são exemplos de dados estruturados. Já quando não há essa organização e nem uma indexação prévia, os dados se tornam mais difícies de serem localizados, desta forma, são definidos como não estruturados. Para este caso podem ser colocados como exemplo os arquivos de texto, imagens e vídeo em geral assim como informações de redes sociais.

Apesar das divergências no que é defendido como a relação entre estruturados e não estruturados, sabe-se que há muito mais dados não estruturados do que estruturados [3]. Até recentemente, apenas os estruturados eram utilizados em análises e buscas, mas aplicações com algoritmos de ML tornaram possível a mineração de dados não estruturados.

2.1.5 – MÉTODOS DE APRENDIZADO

O método de aprendizado é a forma como o algoritmo vai aprender a interagir com as informações que encontrar durante sua aplicação. É considerado supervisionado, de acordo com Jason Brownlee, “quando existem variáveis de entrada e saída e há o uso de um algoritmo que aprende a função de mapeamento entre as duas.” [4, p.16]. Em um exemplo simples, Mostra-se ao programa uma foto (entrada), e diz que aquilo que corresponde a foto é um cachorro ou gato (saída). Através desse aprendizado, ele saberá ao analisar cachorros e gatos, diferencia-los. Quanto mais treinado na diferença entre cachorros e gatos, mais caracteristicas ele aprenderá sobre cada um, e mais perfeito se tornará.

No caso de aprendizado não supervisionado, segundo Brownlee, ‘não há uma saída correspondente a entrada, e o objetivo é modelar a estrutura ou distribuição subjacente nos dados para saber mais sobre eles. O algoritmo é deixado por conta própria para descobrir e apresentar a estrutura interessante nos dados’. [4, p.17]. Saberá diferenciar os exemplos de entrada, mas sem saber que são gatos e cachorros. O método de aprendizado a ser escolhido depende muito do tipo de dados e do objetivo que se almeja.

**2.2 – REDES NEURAIS ARTIFICIAIS**

Entre as entradas e saídas de um algoritmo voltado ao aprendizado, há um vasto caminho a ser percorrido pela informação. Se imaginarmos a entrada como um ponto A e a saída (quando existe), como um ponto B, então passaremos a pensar na implementação desse caminho entre eles. Uma rede neural é uma forma de implementação, onde múltiplas camadas podem ser inseridas nesse caminho, com nós funcionando como neurônios e diferente conexões interligando os neurônios das diversas camadas, vide figura 3 abaixo.

Um modelo primitivo, embora prático, do que viria a se transformar nos estudos de redes neurais, foi apresentado pela primeira vez em 1958 por *Frank Rosenblatt*, e foi chamado de *Perceptron* [5].



Figura - Funcionamento das camadas de uma Rede Neural

2.2.1 – ARQUITETURA E CAMADAS

Além da regra de aprendizado, o que definirá o modelo de estrutura de uma rede neural é o número de camadas e a forma como os neurônios dessas camadas se comunicam [6, p. 46]. Podemos, por exemplo, ter uma rede de camada única ou de múltiplas camadas. No primeiro modelo, os nós da entrada se conectam diretamente aos de saída, sendo que somente na camada de saída há qualquer tipo de computação, por isso chamada de única. No segundo modelo, múltiplas camadas ocultas participam do processo entre a entrada e a saída [6, p. 47].

2.2.2 – NÓS (NEURÔNIOS) E PESOS DAS CONEXÕES

Funcionando como pequenas unidades computacionais, os neurônios realizam a soma entre os valores de entradas, fornecendo um valor de saída. As diversas conexões provenientes de neurônios da camada anterior possuem pesos (ou força). Esse peso, funciona como um multiplicador daquela conexão, reforçando a sua importância para a soma a ser realizada pelo neurônio de destino.

Três neurônios de uma determinada camada *A*, por exemplo, possuem pesos diferentes em suas conexões com um neurônio de uma camada *B*. Entretanto, para um outro nó da mesma camada *B*, as conexões com os nós da camada *A* serão diferentes, e por consequência, o valor do peso de cada uma dessas conexões também diferirá.

2.2.4 – BIAS

Em redes de múltiplas camadas, os neurônios intermediários, possuem uma constante atrelada a sua função de ativação, que recebe o nome de Bias [7, cap. 4]. Se os pesos servem pra aumentar ou diminuir a importância das conexões entre neurônios, a Bias é referente ao próprio Neurônio. Portanto, o valor do peso afetará a conexão mas a Bias afetará o valor computado em um neurônio e consequentemente, todas as conexões originárias nele.

No caso de todas as entradas de um neurônio qualquer em uma camada qualquer tiverem valor nulo, independente do valor dos pesos, sem uma Bias, a função soma no neurônio será nula. Com a Bias, é adicionado um valor após a soma, que ajuda a aproximar o valor resultante do desejado [7, cap. 4]. É extremamente necessário em redes que não funcionam com propagação direta, e portanto, necessitam de correções que são implementadas com base no erro do resultado obtido em relação ao esperado.

2.2.5 – FUNÇÃO DE ATIVAÇÃO

Para evitar que uma pequena alteração em peso ou bias afete toda a rede além do ponto ou segmento esperado, existe a função de ativação. Elas funcionam com base no valor de saída de um neurônio, como que decidindo o comportamento daquela saída com base no definido na função.

No modelo mais simples, o binário, a função de ativação simplesmente decide se aquele neurônio vai ser ativado ou não. Ela verifica, por exemplo, se o valor de saída de um neurônio é maior que um determinado valor xis, e, caso positivo, ativa o nó de forma que seu valor prossiga na rede como um dos valores de entrada de um ou mais nós da camada seguinte. No caso negativo, o neurônio é considerado desimportante e seu valor é ignorado [8, cap. 8].

Entretanto, esta função de ativação não é efetiva em redes com o mínimo de complexidade, de forma que se torna mais útil a escolha de um tipo de ativação de categoria não linear, como Sigmóide, Tanh ou ReLU.

2.2.6 – REPRESENTAÇÃO MATEMÁTICA

Como pode ser observado na figura 4, um neurônio recebe valores diversos de até sendo n o número de entradas, com pesos sinapticos até respectivos aos sinais. A função de soma calcula seu valor aplicado a um bias  observando sua função de ativação φ para definir seu sinal de saída .



Figura 4 - Representação de um Neurônio Artificial

A função de ativação age em cima do valor que pode ser definido de forma simplicada como : .

Esse valor recebido pode ser representado em relação ao resultado de sua função de ativação. Conforme demostrado na figura 5, uma função linear simples define f(x) como o próprio x. Nesse caso, o *output* do neurônio não estaria limitado em um arco com um valor máximo e um mínimo, podendo propagar um valor qualquer desde menos infinito a mais infinito.



Figura 5 – Exemplo de Função de Ativação Linear

Uma função não linear, propaga um valor dentro de uma delimitação, dependendo daquilo recebido do neurônio. Sua representação foge da forma de linha e suas subclassificações são feitas com base na curva formada. Um exemplo de função de ativação não linear pode ser visto na figura 6.



Figura 6 – Exemplo de representação de Função de Ativação não linear

**2.3 – BACKPROPAGATION (RETRO PROPAGAÇÃO)**

Entre as formas de treinamento por aprendizado supervisionado de uma rede neural, o uso do algoritmo de Backpropagation é o mais comum. Sua eficácia consiste nas repetitivas iterações dentro da rede a partir do erro no resultado final em relação ao que se esperava.

Corrige-se os valores dos pesos desde a última camada em direção a primeira, visando diminuir o erro. A cada correção (iteração), o resultado obtido pela rede pode sofrer uma pequena alteração, se aproximando, assim, do desejado. Para a correta aplicação desse algoritmo, são esperados pares (*xi,yi*) de entradas da rede e suas consecutivas desejadas saídas. Também é esperado que a rede seja do tipo *feed-forward* [[2]](#footnote-2)e que possua uma função de erro.

2.3.1 – INICIANDO UMA REDE

Podemos iniciar uma rede de teste, como da figura 7, usando as seguintes configurações:

* Entradas (*inputs*) **i1** e **i2** com valores 0.05 e 0.10 respectivamente.
* Pesos (*weights*) variados conectando neurônios de camdas distintas.
* Camadas ocultas (*hidden*) **h1** e **h2** com valores a serem definidos.
* Camadas de saída (*outputs*) **o1** e **o2** com esperados valores 0.01 e 0.99.
* Bias **b1** de valor 0.35 e **b2** de valor 0.60 correspondendo respectivamente a camada oculta e a camada de saída.



Figura 7 – Exemplo de valores iniciados em uma RNA.

Podemos passar a calcular o valor de entrada nos neurônios ocultos da rede, observando as conexões e os valores referentes a elas que afetam cada um. O neurônio **h1** receberá o valor de entrada **i1** multiplicado pelo peso **w1**, somando com o valor de entrada **i2** multiplicado pelo peso **w3** e somando esse resultado ainda com o bias da camada 1 (**b1**). O mesmo raciocínio usaremos para o neurônio h2. Assim, definindo:

2.3.2 – USANDO ATIVAÇÃO LOGÍSTICA

Nesta rede, é utilizada como função de ativação dos neurônios, a função logística (ou função Sigmoide). Isto é, a função existirá dentro de um intervalo compreendido entre 0 e 1, é não-linear e diferenciável. Ideal para casos de predição de probabilidade. Sua representação gráfica e sua fórmula[[3]](#footnote-3) matemática podem ser observadas na figura 8. Para mais exemplos de funções de ativação, verifique o Anexo A.



Figura 8 – Fórmula e representação da função Logística.

Aplicando a fórmula da função de ativação com a saída do neurônio intermediário h1, teremos .

Aplicando o mesmo em relação a h2, temos .

O processo se repete em relação aos neurônios da camada de saída, agora as entradas serão justamente os valores calculados de saída da camada anterior. Assim, definindo:

Repetindo o processo de encontrar o valor de saída (*output*) com a função de ativação, teremos:  e .

2.3.2 – CÁLCULO DO ERRO

Para os ajustes serem feitos iteração após iteração, calcula-se então o valor desse erro obtido. Em um dos métodos mais comuns para isso, utiliza-se uma fórmula conhecida na estatística como erro quadrático médio (MSE), onde para cada saída, tira-se a diferença entre o resultado verdadeiro e o resultado obtido, elevando ao quadrado. Repete-se o processo para todo par de saída da rede/saída desejada.

Conforme definido acima, a função de erro (também chamada de função de custo) poderia ser representada como:

Ou seja, o erro total sendo igual a soma da metade do target (saída esperada) menos o output (saída encontrada) ao quadrado. A fração ½ que divide o resultado entre parenteses ao meio, serve para cancelar expoentes futuramente e facilitar o processo. Essa inserção não interfere na equação, já que é uma constante e o resultado da fórmula será multiplicado por uma taxa de aprendizado futuramente.

Para e , temos então respectivamente:

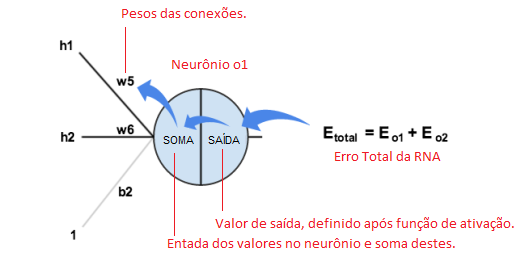
E para o erro total da rede, fazemos:

2.2.5 – O GRADIENTE E A APLICAÇÃO DA REGRA DA CADEIA

Conforme dito no item 2.3, precisamos adaptar o valor dos pesos para diminuir o valor da perda/erro. Isso é feito para cada neurônio, de trás para frente. Em um primeiro momento, se calcula então o erro dos pesos das conexões da camada de saída, ou seja, de , , , e .

Para tal, precisamos por exemplo saber o tanto que uma mudança no valor de afeta o valor de erro total. Representamos isso como:

Podemos interpretar como a derivada parcial do erro total em relação a , ou, simplesmente, o gradiente em relação a Aplicamos aqui nesta fórmula, a regra da cadeia[[4]](#footnote-4), conforme ilustrado na figura 9.



Como temos que encontrar o gradiente do erro total em relação a ,

Encaixar Uso do Gradient Descend, obrigatorio para backprop.

Encaixar explicação que a rede é uma feed foadward MLP, obrigatorio pra backprop.

Encaixar a taxa de aprendizado, com isso consigo explicar backpropagation.

2.3.2 – FÓRMULA GERAL

**REFERÊNCIAS BIBLIOGRÁFICAS**

1. SEARLE, John. *Minds, brains, and programs. Behavioral and Brain Sciences*, 1980.
2. TAYLOR, Christine. *Structured vs. Unstructured Data*. Acessado em 06/10/2019.

Disponível em <https://www.datamation.com/big-data/structured-vs-unstructured-data.html> .

1. *Unstructured Data and the 80 percent rule*. 2008. Acessado em 08/09/2019. Disponível em <http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/>.
2. BROWNLEE, Jason. *Master Machine Learning Algorithms - Discover How They Work and Implement Them from Scratch*, 2016.
3. ROSENBLATT, Frank. *The Perceptron--a perceiving and The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain*, 1958.
4. HAYKIN, S. Redes neurais: princípios e prática. Porto Alegre: Bookman, 2001
5. *Deep Learning Book* – Em Português, Online e Gratuito. Acessado em 01/10/2019.

Disponível em <http://deeplearningbook.com.br/o-neuronio-biologico-e-matematico/> .

1. *Deep Learning Book* – Em Português, Online e Gratuito. Acessado em 01/10/2019. Disponível em <http://deeplearningbook.com.br/funcao-de-ativacao/> .

**APÊNDICE A – ANALOGIA PARA COMPREENSÃO DO PROBLEMA DO BIG DATA**

(REESCREVER, NÃO ESTÁ PRONTO, SÓ SE DER TEMPO)

É como se um grande galpão, de tamanho aparentemente ilimitado, recebesse várias caixas de destinatários diferentes por hora, e a capacidade de analisar e separar as caixas, de acordo com seu conteúdo e importância para serem enviados a seu destino, estivesse abaixo do necessário para lidar com a quantidade de caixas que chegam. Que fazemos então? A princípio, pensamos em minimizar o problema, para então controlá-lo. Organizamos corredores nesse galpão, passamos a desenvolver empilhadeiras e esteiras para ajudar na movimentação das caixas. Nesta analogia, podemos entender que as empilhadeiras e esteiras são os programas convencionais, que, em suas linguagens de programação, ajudam os computadores a interpretar instruções e realizar tarefas cada vez mais complexas. Entretanto, como podemos perceber na computação, escrever comandos para eventos e tarefas não elimina a necessidade do protagonismo humano, e isso é até bom em muitas sub áreas da computação. Já no galpão, as esteiras e empilhadeiras não eliminam o fato de cada caixa ter que ser conferida e separada e embora as máquinas ajudem em partes da tarefa, elas não o fazem por completo. A popularização da internet e a disponibilidade de dados via multiplataforma nos leva então ao caos completo da informação, passamos a dobrar o volume de dados disponíveis na rede a cada dois anos, e nosso galpão não parece chegar a seu limite de armazenamento tão cedo, pois cada vez mais caixas passam a chegar por hora, e cada vez mais a relação de caixas recebidas por caixas enviadas aumenta. O protagonismo humano aqui torna-se tão inútil nessa situação que se torna algo como desesperador ver o acúmulo de caixas numa velocidade maior que se poderia prever. Esse acúmulo, essa quantidade inexplorada de informação crescendo em pilhas e pilhas pelo galpão, na computação, chamamos de Big Data.

**ANEXO A – PLANILHA DE FUNÇÕES DE ATIVAÇÃO E CASOS DE USO**

(A SER DESENVOLVIDO)



1. **Alan Turing** (1912 – 1954):Considerado o pai da ciência da computação. [↑](#footnote-ref-1)
2. Uma rede onde todos os neurônios de cada camada se conectam com todos os neurônios da camada posterior mas não ocorrem conexões entre neurônios de uma mesma camada. [↑](#footnote-ref-2)
3. O “e” presente na fórmula refere-se a constante matemática, também conhecida como número de Euller cujo valor aproximado é 2,718. [↑](#footnote-ref-3)
4. [↑](#footnote-ref-4)