

Deep Learning

Advanced Institute for Artificial Intelligence – AI2

<https://advancedinstitute.ai>

Agenda

- ☐ Inteligência Artificial
- ☐ Deep Learning
- ☐ Redes Neurais
- ☐ Arquitetura de Redes Neurais
- ☐ Otimização
- ☐ Suporte de Software e Hardware a álgebra linear

- ☐ Inteligência Artificial é uma grande área da ciência
- ☐ Diversas vertentes da IA buscam avançar a ciência com reapeito a automatizar tarefas intelectuais normalmente realizadas por seres humanos
- ☐ Atualmente, um campo estratégico para desenvolvimento de tecnologia

O que é Inteligência?

- ☐ Habilidade de aprender, contemplar, pensar e raciocinar
- ☐ Mentalidade, senso
- ☐ Discernimento, julgamento, sabedoria

O que é Inteligência?

Nossas mentes contêm processos que nos capacitam a solucionar problemas que consideramos difíceis.

Inteligência é o nome que damos a qualquer um destes processos que ainda não compreendemos

—Marvin Minsky

As abordagens para o estudo de IA se dividem em 4 categorias:

- ☐ Sistemas que pensam como seres humanos
 - Modelagem cognitiva
 - Programa de computador capaz de implementar o raciocínio humano
- ☐ Sistemas que pensam racionalmente
 - Utilização de lógica formal para modelar a forma de resolver problemas
- ☐ Sistemas que agem como seres humanos
 - Teste de Turing
- ☐ Sistemas que agem racionalmente

Comportamento racional

- ☐ Agir corretamente = fazer o que é esperado para atingir seus objetivos, dada a informação disponível
- ☐ Não necessariamente envolve pensamentos (raciocínios lógicos)
- ☐ O raciocínio lógico deve ser usado para alcançar um objetivo.

Inteligência Artificial

Aprendizagem de Máquina

Deep Learning

Aprendizagem profunda

- Buscar representações úteis para os dados de entrada Em um espaço pré-definido de possibilidades (Pode ser um espaço grande)
- Usando como critério o sucesso de realizar a representação
- Tais passos permitem resolver diversas tarefas intelectuais, desde o reconhecimento de fala até a condução autônoma

O Deep learning é um subcampo específico do aprendizado de máquina

- ❑ Abordagem de representação de aprendizado a partir de dados que enfatiza o aprendizado de camadas sucessivas de representações (Redes Neurais)
- ❑ Outras abordagens ao aprendizado de máquina tendem a se concentrar no aprendizado de apenas uma ou duas camadas de representações dos dados (shallow learning)
- ❑ No aprendizado profundo, essas representações em camadas são (quase sempre) aprendidas através de modelos chamados redes neurais, estruturadas em camadas literais empilhadas

De modo geral, uma rede neural é representada por: $f_{\Theta} : x \rightarrow y$

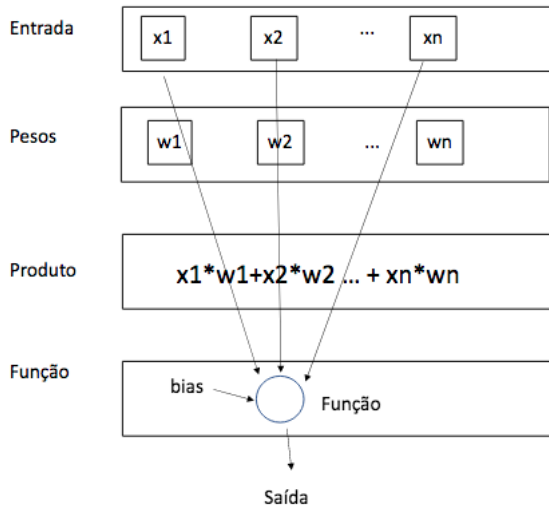
Entrada: $x \in \mathbb{R}^m$

Saída: $y \in \mathbb{R}^n$

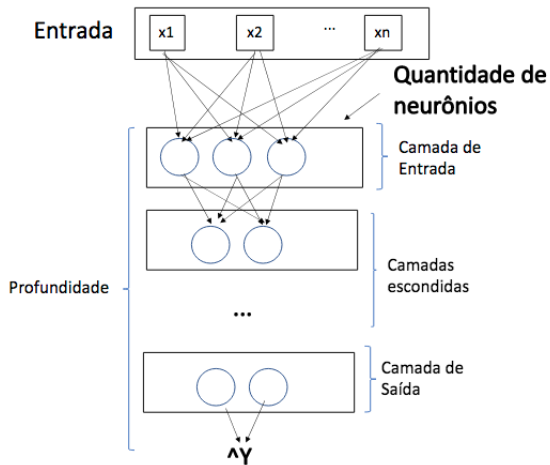
parametrizada por: $\Theta \in \mathbb{R}^p$

Exemplo: $f_{\Theta}(x) = \Theta * x$

Modelo de Neurônio



Representação geral de uma Rede Neural

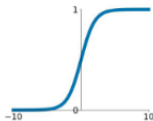


Funções de ativação

- ☐ Linear
- ☐ Sigmoid
- ☐ Softmax
- ☐ Rectified Linear Unit (ReLU)
- ☐ Hyperbolic Tangent

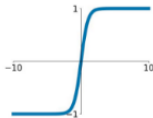
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



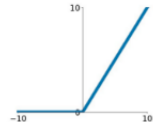
tanh

$$\tanh(x)$$



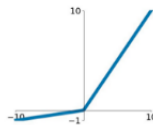
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

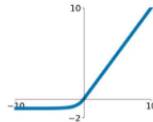


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

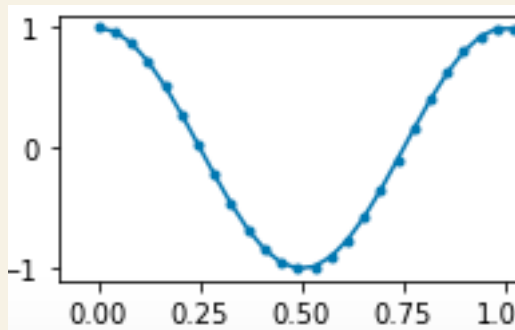


Fonte: <https://hsf-training.github.io/hsf-training-ml-webpage/03-nn/index.html>

Funções de perda (loss):

- Recebe o conjunto de valores conhecidos para a entrada e o conjunto de valores preditos, para calcular a distância entre o valor predito e a resposta correta
- Funções de perda estão relacionadas com a natureza do problema, regressão, classificação, etc)
 - binary cross entropy: classificação binária
 - cross entropy: classificação múltiplas classes
 - squared error: regressão

Otimização: encontrar parâmetros que minimizem uma função matemática



Em uma rede neural o processo de otimização consiste em encontrar o valor de cada peso da rede de forma que função de perda apresente o menor valor

- ☐ Uma forma simplista de fazer isso congelar todos os pesos exceto um, e testar o valor da função de perda.
- ☐ Após encontrar o valor ideal, congelar e tentar outro, até otimizar todos os parâmetros
- ☐ Inviável para redes com muitos parâmetros, que é o mais comum

- A função a ser definida pela rede neural é diferenciável, portanto sua derivada pode ser calculada
- Dessa forma, é possível utilizar o método de gradiente da perda para minimizar a função
- Os coeficientes podem ser movidos em direção oposta ao gradiente provocando a minimização da função perda
- Esse método é mais eficiente e efetivo para descobrir o mínimo da função

- Na prática, uma função de rede neural consiste em muitas operações de tensor encadeadas. cada uma com uma derivada.
- Por exemplo Uma rede f composta por três operações tensoriais, a , b e c , com as matrizes de peso $W1$, $W2$ e $W3$
 - $f(W1, W2, W3) = a(W1, b(W2, c(W3)))$
- Essa derivada pode ser calculada por um método chamado regra da cadeia
- O algoritmo backpropagation utiliza essa técnica para otimizar os pesos em redes com múltiplas camadas

Estratégias de Otimização:

- ☐ mini-batch stochastic gradient descent (mini- batch SGD)
 - Seleciona aleatoriamente uma parte do dataset para otimizar o peso em cada iteração do algoritmo
- ☐ batch SGD
 - Em cada iteração um o dataset inteiro é utilizado
 - O dataset pode ser dividido em lotes (batches)

Atualização dos pesos:

- ☐ O peso pode ser iniciado e calculado em todas as iterações
- ☐ O peso da iteração anterior pode ser utilizado como base para o cálculo do peso atual

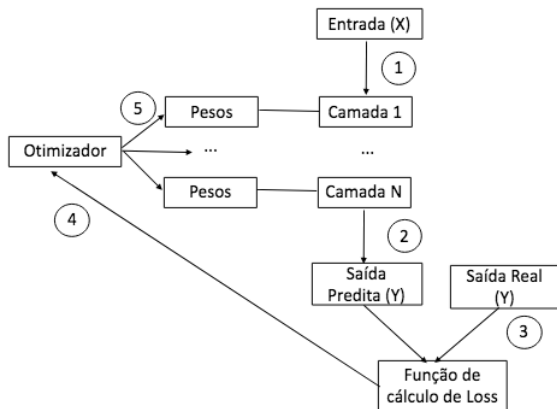
Treinamento do modelo

- ❑ Inicialmente os neuronios da rede recebem valores aleatórios
- ❑ o processo de ajustes de peso é feito por um algoritmo iterativo.
- ❑ Em cada passo são executadas as seguintes funções
 - Entrada é submetida a rede neural e um valor predito é gerado para cada item da base de dados
 - Função loss calcula a distância entre valor predito e real de cada item da base de entrada
 - Função otimizadora utiliza o valor retornado na função loss e otimizar os parâmetros (pesos)
- ❑ A cada passo do algortimo o retorno da função loss é minimizado, e os pesos começam a representar de forma mais efetiva os dados de entrada

A rotina iterativa de atualização dos pesos também é chamada de época

Passos de uma época

- ❑ 1) Dados submetidos as camadas (X)
- ❑ 2) Respostas geradas (y)
- ❑ 3) Função para comparar a perda entre predito e real
- ❑ 4) Diferença é enviada a função otimizadora
- ❑ 5) Otimizador atualiza os pesos de cada camada



Diversos parâmetros podem ser definidos na rotina de aprendizagem

- ☐ Número de camadas
 - Quantidade de neurônios por camada
 - Função de ativação de cada camada
 - Parâmetros de cada função de ativação
- ☐ Tamanho do Batch
- ☐ Critério de parada do laço de treinamento
- ☐ otimizador
- ☐ Função de perda

O que impulsiona deep learning:

- ☐ Hardware
 - Arquiteturas dedicadas a deep learning: GPU, TPU vetorização, entre outros
- ☐ Conjuntos de dados e benchmarks: disponibilidade de dados, desafios em diversas áreas, facilidade de manipular dados
- ☐ Algoritmos
 - Melhores funções de ativação para camadas neurais
 - Melhores esquemas de inicialização de peso
 - Melhores esquemas de otimização, como RMSProp e Adam
- ☐ Frameworks: keras, Pytorch, tensorflow, etc

- Bibliotecas otimizadas para Algebra Linear expoe toda a potencialidade de um hardware para operações comuns de algebra linear
- Algumas bibliotecas comuns para esse fim Numpy Tensorflow Theano
- Tais bibliotecas são construidas com suporte de bibliotecas de mais baixo nível, otimizadas para diferentes hardware, implementadas por fabricantes chamadas BLAS