

Aprendizagem Supervisionada Aula 4

Redes Neurais

Bernard da Silva
Orientador: Rafael S. Parpinelli
30/11/2022

Aula Passada

- Pré-processamento de dados
- Dicas para seleção de algoritmos
- Conceito Redes Neurais

Objetivo dessa aula

- Escalonamento de dados
- Separação Base de Dados
- Técnica de validação cruzada
- Parâmetros de configuração Redes Neurais
- Entendimento Overfitting e Underfitting

Escalonamento de dados

- O que é escalonamento de dados?
- Quando realizar escalonamento de dados?
- Quais técnicas são comumente utilizadas?

Existem diversas técnicas já implementadas no Sklearn como:

1. Normalizer
2. MinMaxScaler
3. StandardScaler
4. RobustScaler
5. QuantileTransformer
6. PowerTransformer

Escalonamento de dados

1. Normalizer

É uma boa escolha quando você sabe que a distribuição dos seus dados não é normal/gaussiana ou quando você não sabe qual é o tipo de distribuição dos seus dados.

Normaliza as amostras por linha.

2. MinMaxScaler

Essa técnica funciona melhor se a distribuição dos dados não for normal e se o desvio padrão for pequeno, além disso o MinMaxScaler não reduz de forma eficaz o impacto de outliers e também preserva a distribuição original.

Normaliza as amostras por coluna.

Escalonamento de dados

3. StandardScaler

Subtrai do valor em questão a média da coluna e divide o resultado pelo desvio padrão.

Esse método trabalha melhor em dados com distribuição normal

Normaliza as amostras por coluna.

4. RoubustScaler

O diferencial deste método é a combinação com o uso de quartis o que garante um bom tratamento dos outliers.

Subtrai a média do valor em questão e então divide o resultado pelo segundo quartil.

Normaliza as amostras por coluna.

Escalonamento de dados

5. QuantileTransformer

Trata os outliers com uso de quartis. Este método transforma os valores de tal forma que a distribuição tende a se aproximar de uma distribuição normal.

Normaliza as amostras por coluna.

6. RoubustScaler

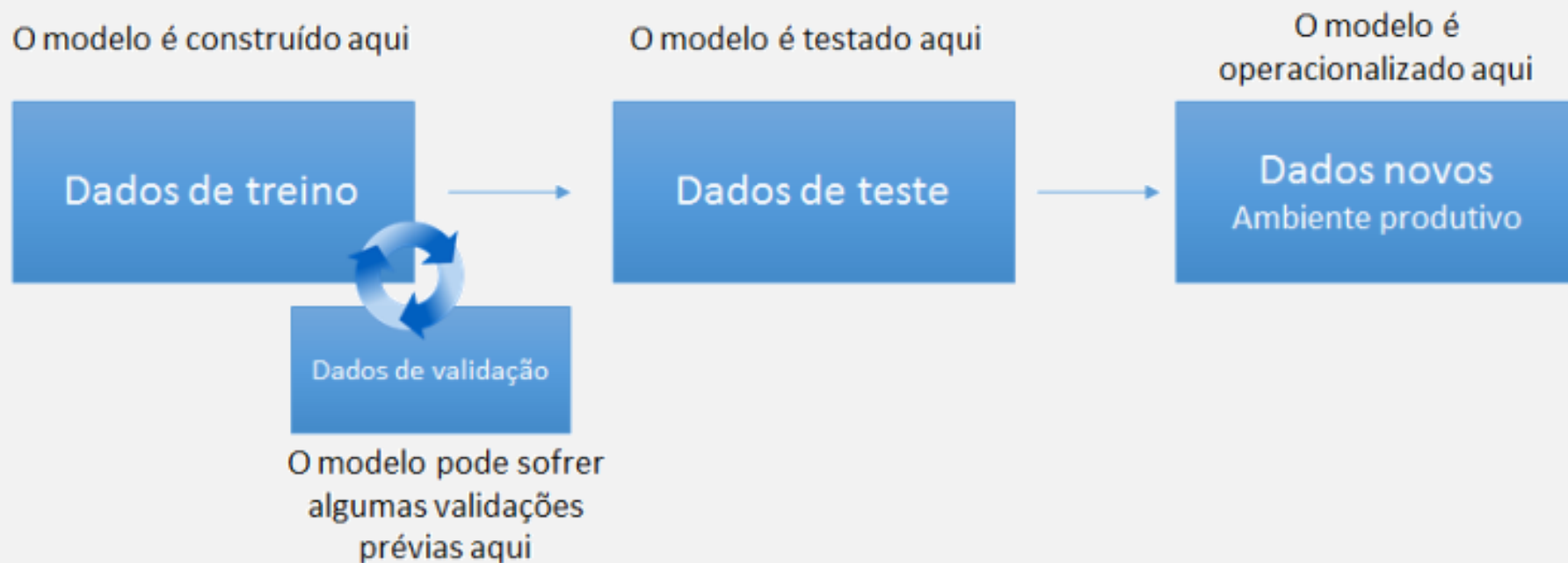
Procura transformar os valores em uma distribuição mais normal, sendo indicado em situações onde uma distribuição normal é desejada para os dados.

Normaliza as amostras por coluna.

Escalonamento de dados

Método		Dados em distribuição normal	Dados não estão em distribuição normal	É desejado que os dados estejam em distribuição normal	É desejado eliminar a influência dos outliers
1	Normalizer	×	✓	×	×
2	MinMaxScaler	×	✓	×	×
3	StandardScaler	✓	✓	✓	×
4	RobustScaler				✓
5	QuantileTransformer				✓
6	PowerTransformer				

Separando Dados



Resampling

As principais formas resampling são:

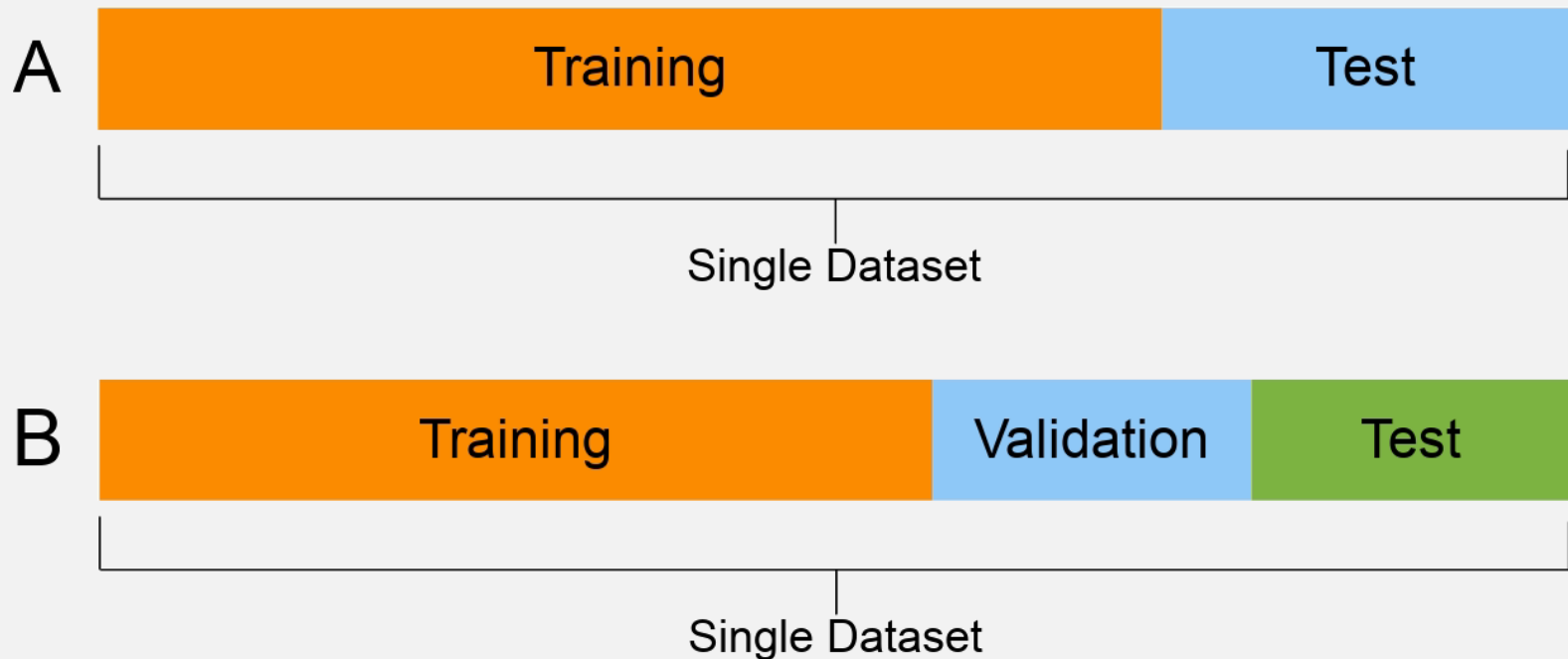
Hold-out validation

Bootstrap

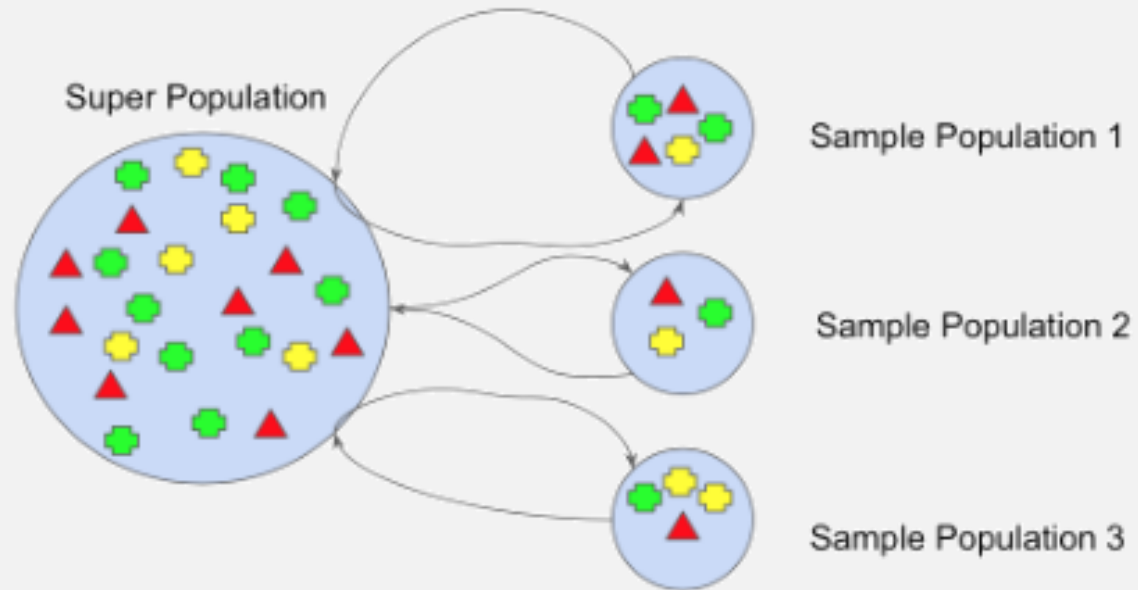
Cross validation (k-fold, Leave-one-out)

Hold-out

- Normalmente dividimos nossos dados em 2 ou 3 partições:




Botstrap



K-Fold

5-fold CV

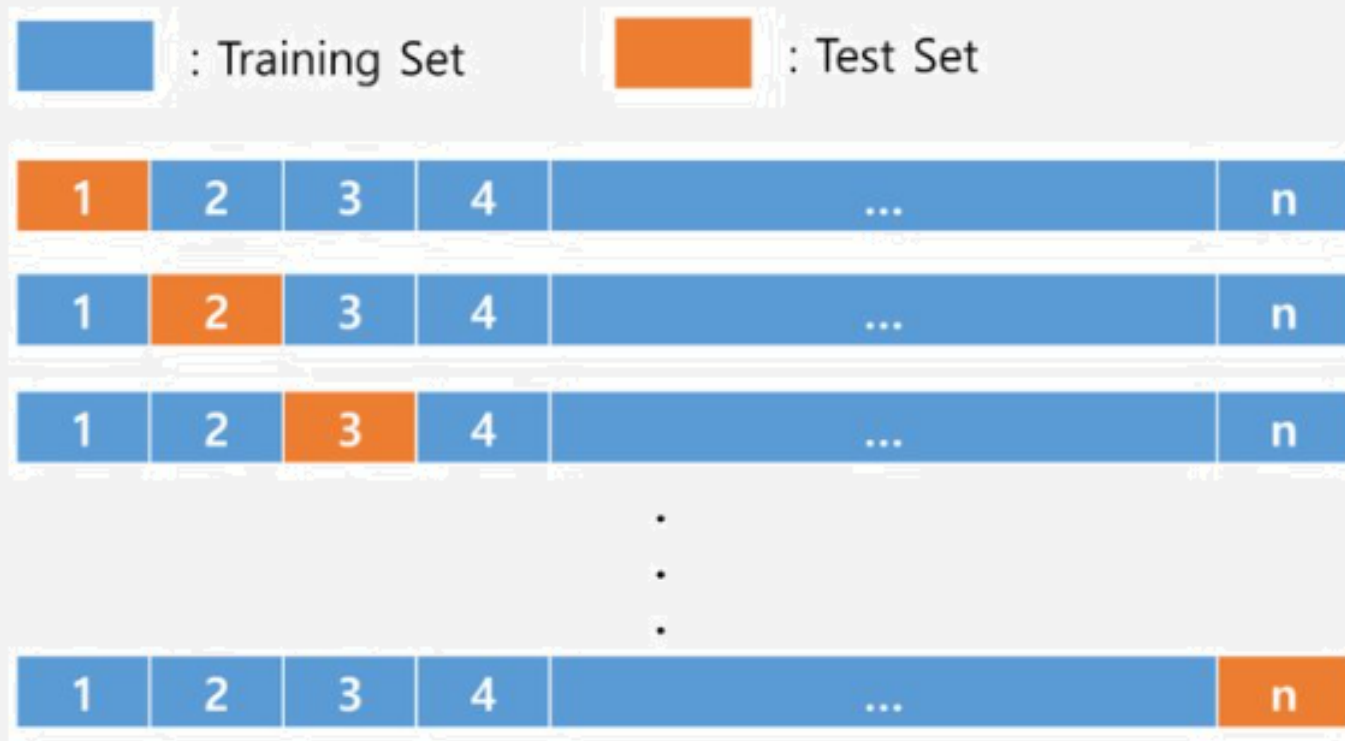
DATASET



A blue bracket above the table indicates that the five columns represent the entire dataset, which is partitioned into five different test sets across the five estimations.

Estimation 1	Test	Train	Train	Train	Train
Estimation 2	Train	Test	Train	Train	Train
Estimation 3	Train	Train	Test	Train	Train
Estimation 4	Train	Train	Train	Test	Train
Estimation 5	Train	Train	Train	Train	Test

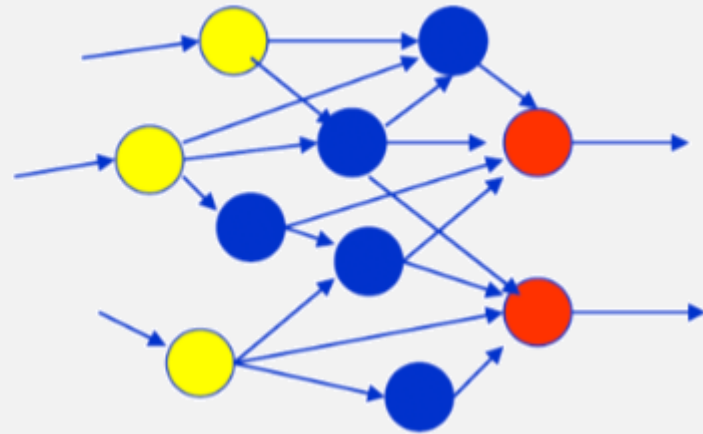
Leave-one-out



Redes Neurais

O que é uma rede neural?

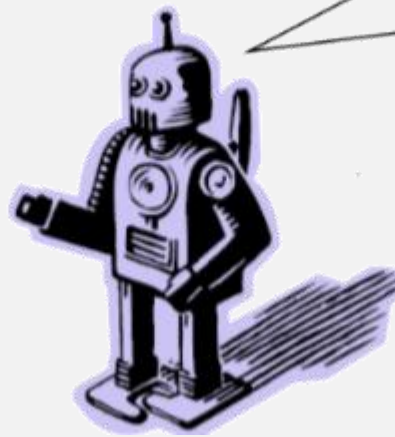
- Uma rede neural é um sistema de coordenação com neurônios como elementos básicos
- Um neurônio é uma célula cerebral em redes bio-neurais
- Um neurônio é uma unidade de processamento simples em redes neurais artificiais



Redes Neurais

Por que utilizar redes neurais?

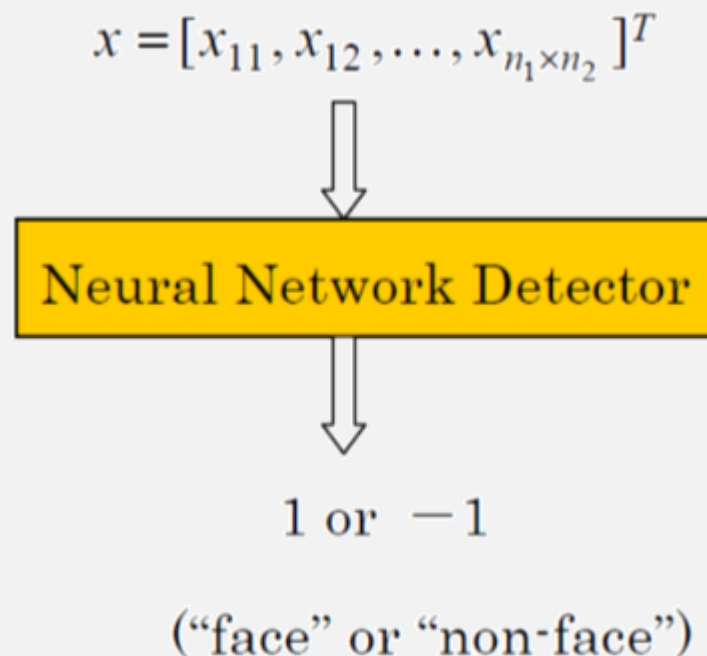
- Abrem uma maneira de resolver problemas sem fazer programas
- Podem aprender com a experiência e podem resolver diferentes tipos de problemas através da aprendizagem
- Podem aprender em tempo real e podem se adaptar aos ambientes em mudança de forma flexível



Você ainda está fazendo programas para aquele computador estúpido?

Redes Neurais

- Uma RN é representada por muitos parâmetros chamados pesos
- O processo para encontrar os pesos é chamado aprendizagem ou treinamento
- Geralmente é muito difícil afinar os pesos de uma RN porque o número de pesos pode ser extremamente grande



Redes Neurais

- Em geral, a aproximação de função é encontrar uma função desconhecida de dados observados
- A aproximação de função é um problema comum para muitas aplicações:
 - Controle de sistemas, design e identificação
 - Restauração de sinal ou imagem e reconstrução
 - Mesmo o projeto de redes neurais é um caso especial de aproximação de função

Redes Neurais

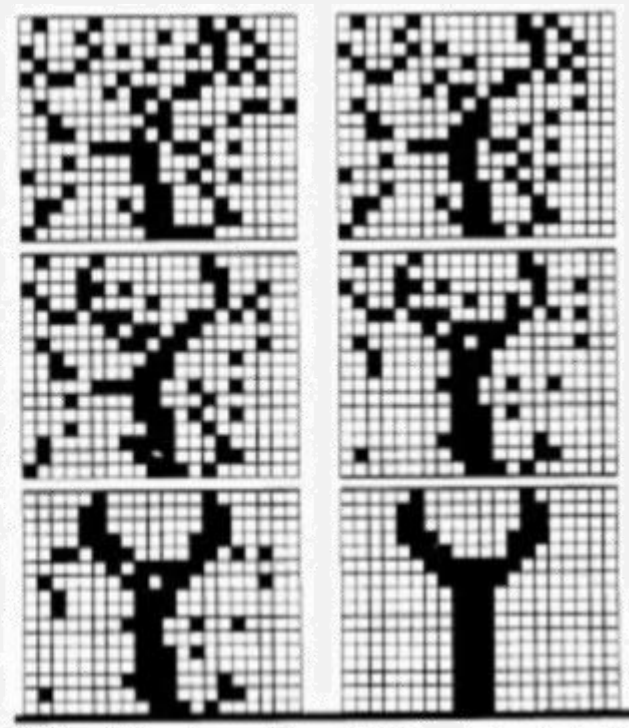
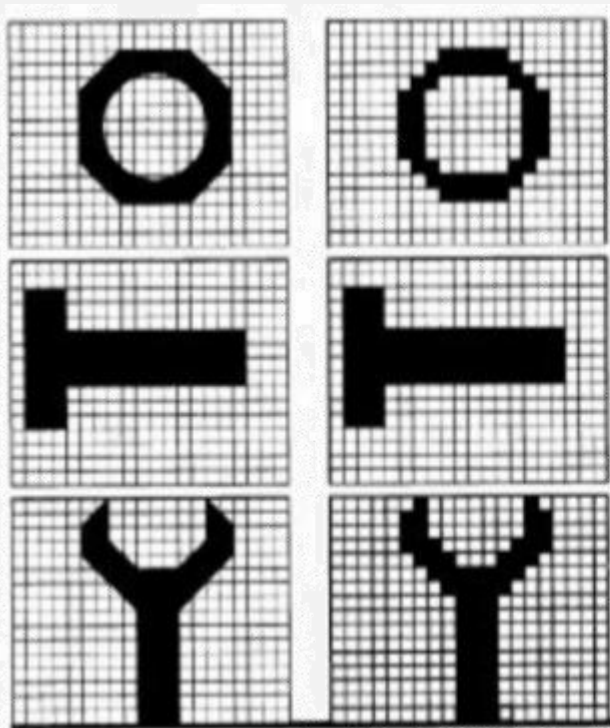
Memória associativa

- A recuperação de dados baseada em conteúdo é geralmente chamada de memória associativa
- O conteúdo pode ser ligeiramente diferente daqueles armazenados na memória
- Podemos reconhecer o padrão de entrada, encontrar o mais semelhante na memória e externar esse padrão



Redes Neurais

- Exemplo de memória associativa



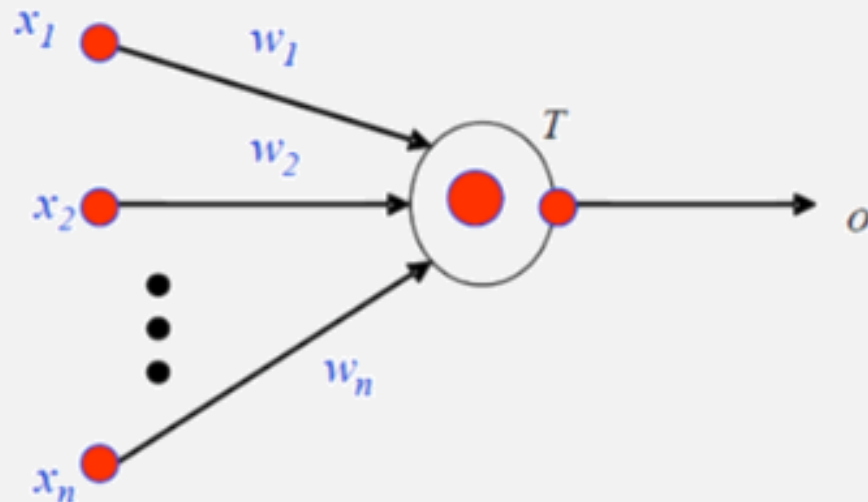
Redes Neurais

Neurônio

- Os parâmetros usados para dimensionar as entradas são chamados de pesos (w_i)
- A entrada efetiva é a soma ponderada das entradas ($w_i x_i$)
- O parâmetro para medir o nível de comutação é o limiar ou viés (bias) (T)
- A função para produzir a saída final é chamada a função de ativação ($f(u)$)

$$o = f\left(\sum_{i=1}^n w_i x_i - T\right)$$

$$f(u) = \begin{cases} 1 & \text{if } u \geq 0 \\ 0 & \text{else} \end{cases}$$



Redes Neurais

Funções de Ativação

- Em geral, existem muitos tipos diferentes de funções de ativação
- Uma das “responsáveis” pela não linearidade do modelo
- Pode afetar bastante o comportamento do modelo
- As funções típicas usadas em um neurônio artificial são funções sigmoide, função de base radial, funções senoidais, etc...



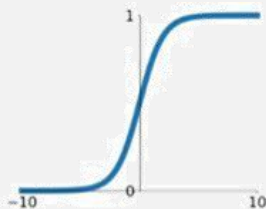
Redes Neurais

Funções de Ativação

- Geralmente utilizado 3 funções:

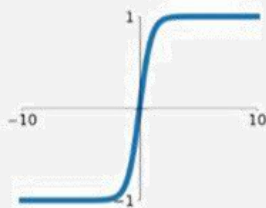
Sigmoid

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



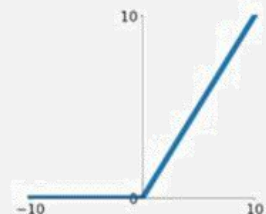
tanh

$$\tanh(x)$$



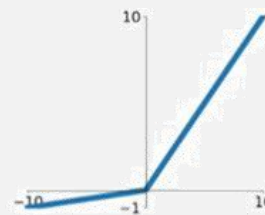
ReLU

$$\max(0, x)$$



Leaky ReLU

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

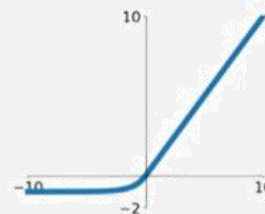


Maxout

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

ELU

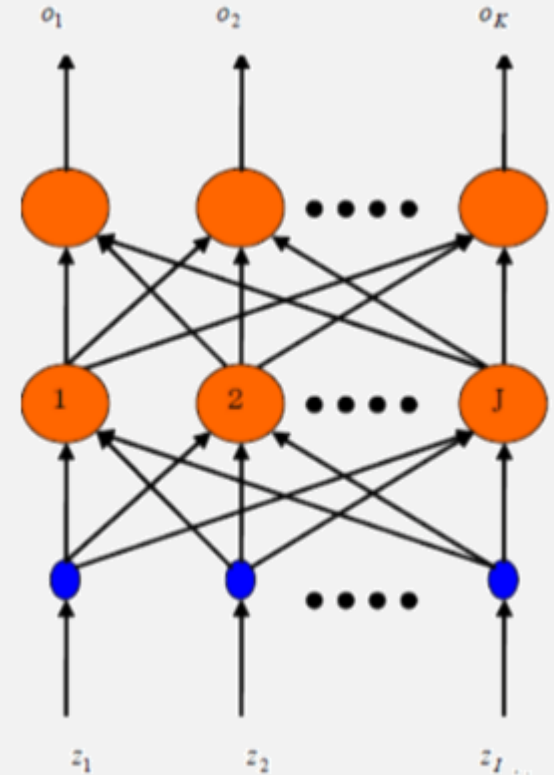
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Redes Neurais

Camadas Escondidas

- A utilização de um grande número de camadas escondidas não é recomendado.
- Cada vez que o erro médio durante o treinamento é utilizado para atualizar os pesos das sinapses da camada imediatamente anterior, ele se torna menos útil ou preciso.
- Para grande maioria dos problemas utiliza-se apenas uma camada escondida.



Redes Neurais

Número de Neurônios na Camadas Escondidas

- O número de neurônios nas camadas escondidas, geralmente é definido EMPIRICAMENTE.
- Deve-se ter cuidado para não utilizar nem unidades demais, o que pode levar a rede a memorizar os dados de treinamento (overfitting)
- Nem um número muito pequeno, que pode forçar a rede a gastar tempo em excesso tentando encontrar uma representação ótima.

Redes Neurais

Número de Neurônios na Camadas Escondidas

- Existem várias propostas de como determinar a quantidade adequada de neurônios:
- Média aritmética ou ainda como sendo a média geométrica entre tamanho da entrada e da saída da rede.
- Utilizar um número de sinapses dez vezes menor que o número de exemplos disponíveis para treinamento. Se o número de exemplos for muito maior que o número de sinapses, overfitting é improvável, no entanto pode ocorrer underfitting (a rede não converge durante o seu treinamento)

Redes Neurais

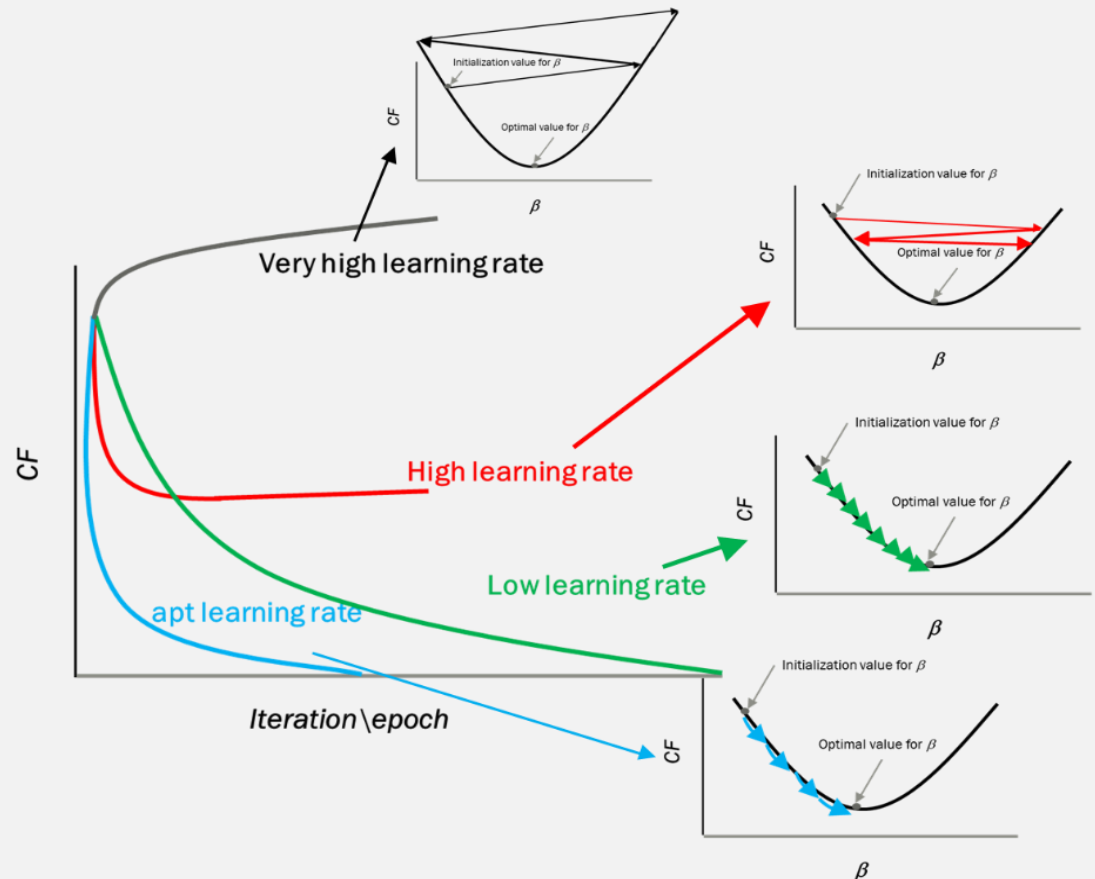
Taxa de Aprendizado

- O parâmetro taxa de aprendizado tem grande influência durante o processo de treinamento da rede neural
- Uma taxa de aprendizado muito baixa torna o aprendizado da rede muito lento, ao passo que uma taxa de aprendizado muito alta provoca oscilações no treinamento e impede a convergência do processo de aprendizado
- Geralmente seu valor varia de 0.1 a 1.0
- Alguns softwares disponíveis no mercado possuem este parâmetro adaptativo, por isso a escolha de um valor inicial não constitui um grande problema.

Redes Neurais

Taxa de Aprendizado

- Controla o quanto o modelo muda em resposta ao erro estimado de treinamento toda vez que os pesos são atualizados.



Redes Neurais

Momentum

- A inclusão do termo momentum tem por objetivo aumentar a velocidade de treinamento da rede neural e reduzir o perigo de instabilidade
- Este termo pode ou não ser utilizado durante o treinamento e seu valor varia de 0.0 (não utilização) a 1.0

Redes Neurais

Por ciclo (batch ou epoch)

- Na abordagem por ciclo os pesos são atualizados após todos os exemplos de treinamento terem sido apresentados
- Esta técnica é geralmente mais ESTÁVEL e o treinamento é menos influenciado pela ordem de apresentação dos padrões, mas ela pode ser lenta se o conjunto de treinamento for grande e redundante.

Redes Neurais

Critérios de Parada do Treinamento

- Existem vários métodos para a determinação do momento em que o treinamento de uma rede neural deve ser encerrado. Os critérios de parada mais utilizados são:
 - Número de ciclos
 - Erro
 - Combinação dos métodos (Ciclos + Erro)
 - Validação (Best Model)

Redes Neurais

Critérios de Parada do Treinamento

- Número de ciclos: Um número excessivo de ciclos pode levar a rede à perda do poder de generalização (overfitting). Por outro lado, com um pequeno número de ciclos a rede pode não chegar ao seu melhor desempenho (underfitting).
- Erro: Consiste em encerrar o treinamento após o erro médio quadrático ficar abaixo de um valor pré-definido. Vale lembrar que um erro médio quadrático muito pequeno não implica necessariamente numa boa generalização (bom desempenho com dados não vistos anteriormente). Este valor depende muito do problema.

Redes Neurais

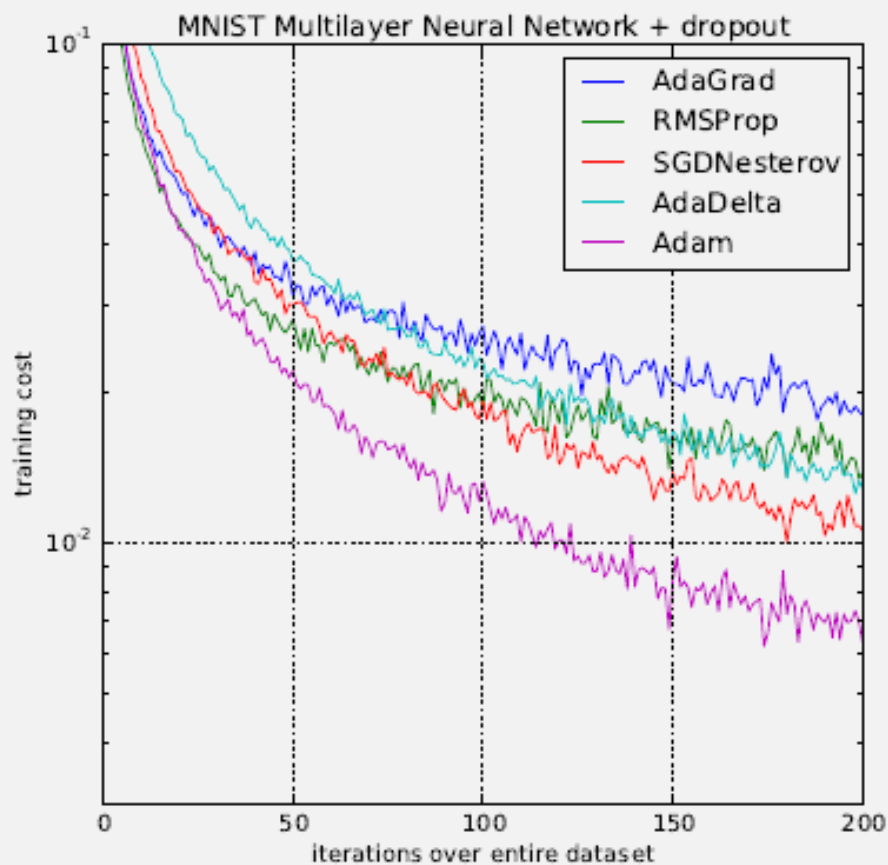
Critérios de Parada do Treinamento

- Validação (Best Model): Na técnica de parada pela validação, o treinamento é interrompido a cada x ciclos e é realizada uma estimação de erro da rede sobre o conjunto de dados de teste. A partir do momento em que o erro medido no conjunto de teste apresentar crescimento, o treinamento é encerrado. O que se deseja com esta técnica é descobrir o momento exato em que a rede começa a perder generalização

Redes Neurais

Otimizador

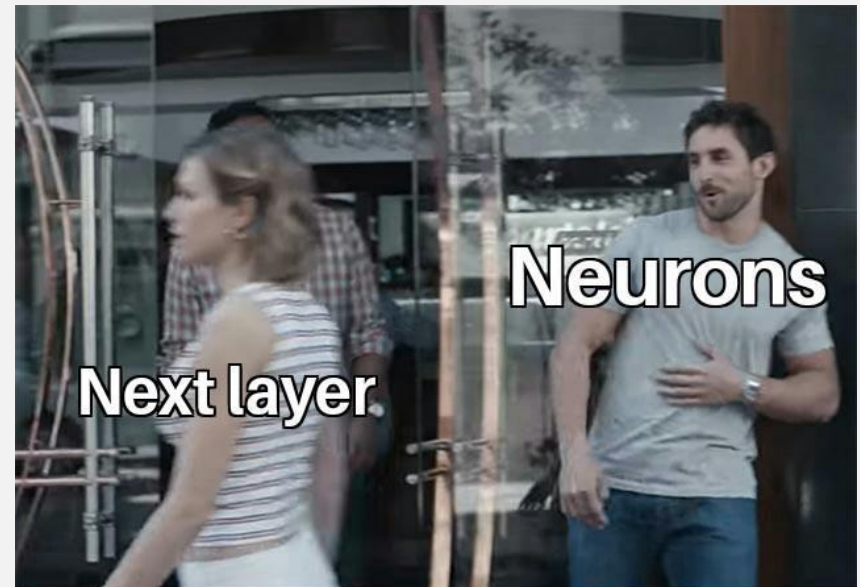
- É o algoritmo que busca minimizar o erro.
- Ou seja, é no otimizador que a taxa de aprendizagem atua.



Redes Neurais

Regularização Dropout

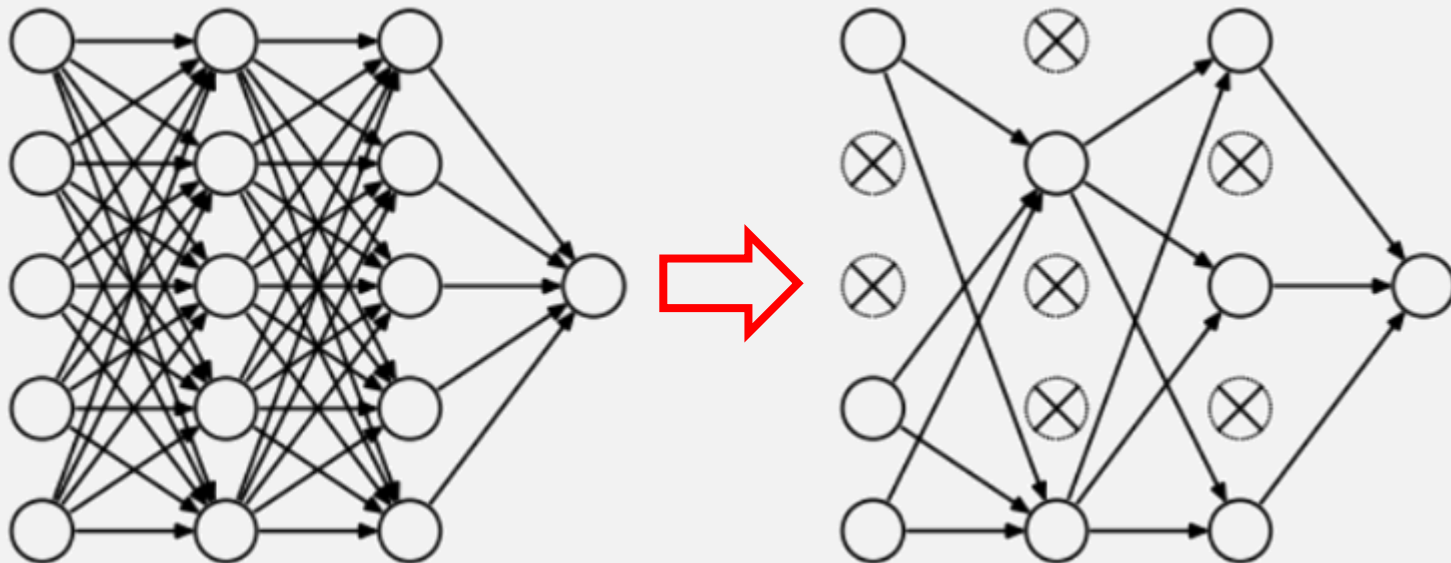
- Aleatoriamente, neurônios são “apagados” durante o treinamento.



Redes Neurais

Regularização Dropout

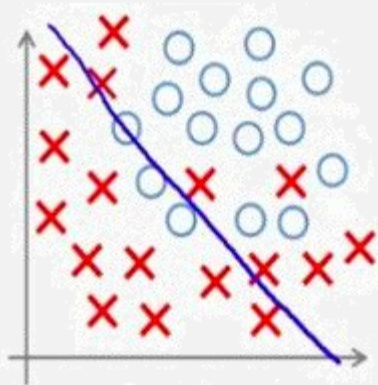
- Aleatoriamente, neurônios são “apagados” durante o treinamento.
- Com uma certa probabilidade, um neurônio pode ter todos os pesos de suas conexões zeradas
- “Poda” uma rede neural, diminuindo sua complexidade desnecessária



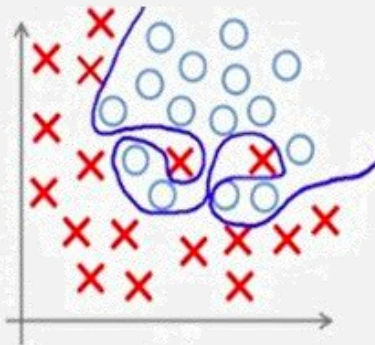
Redes Neurais

Overfit e Underfit

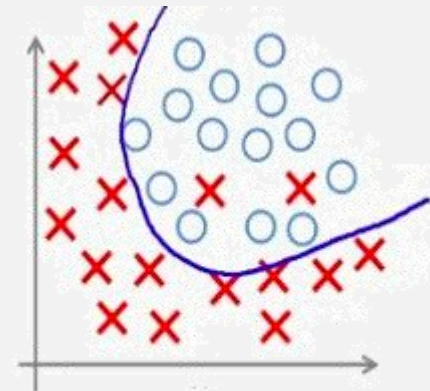
- Deve ser avaliado em amostras que não foram usadas para construir ou ajustar o modelo, de modo que forneçam um resultado imparcial de eficácia.



Under-fitting



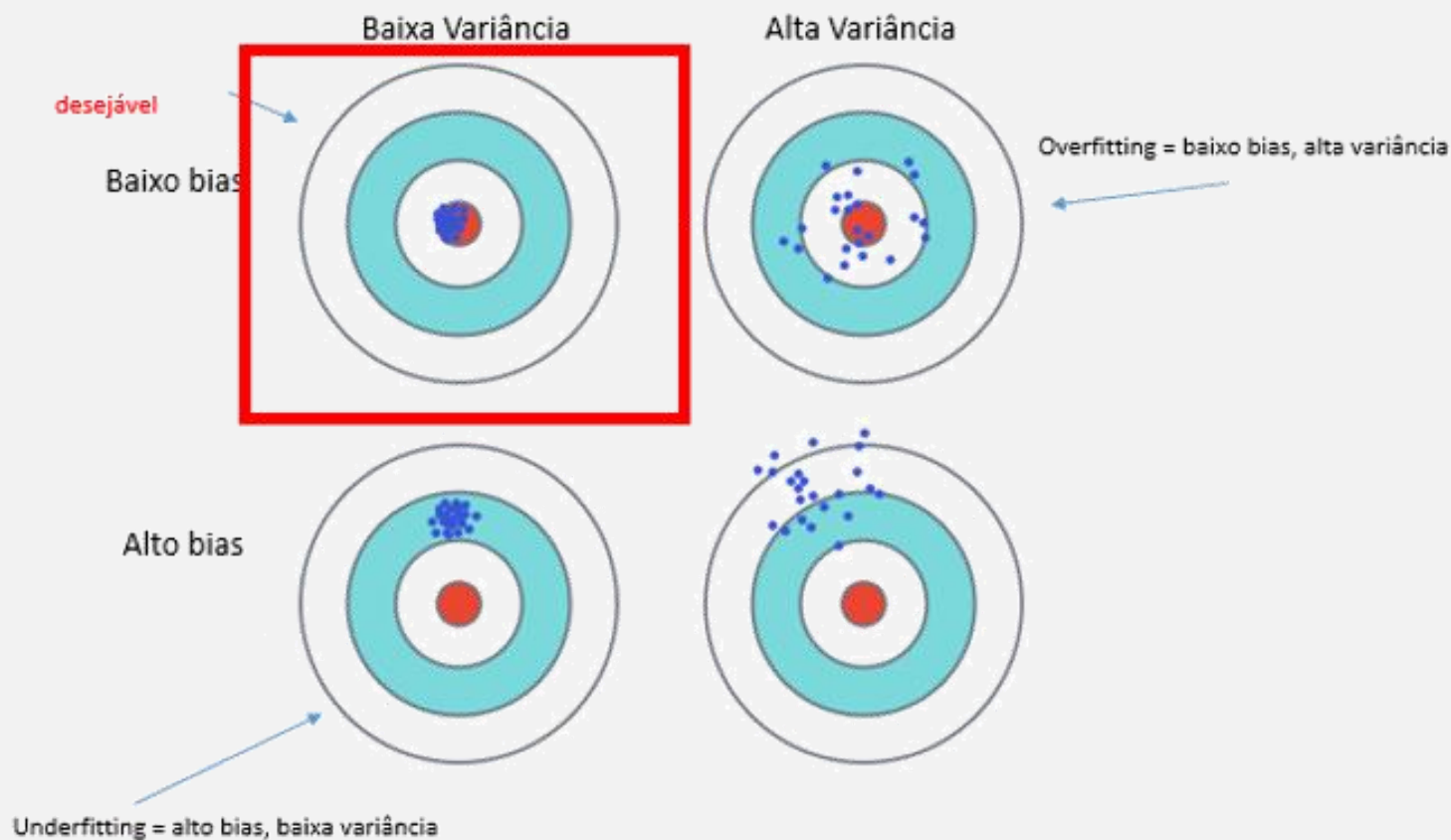
Over-fitting



Appropriate-fitting

Redes Neurais

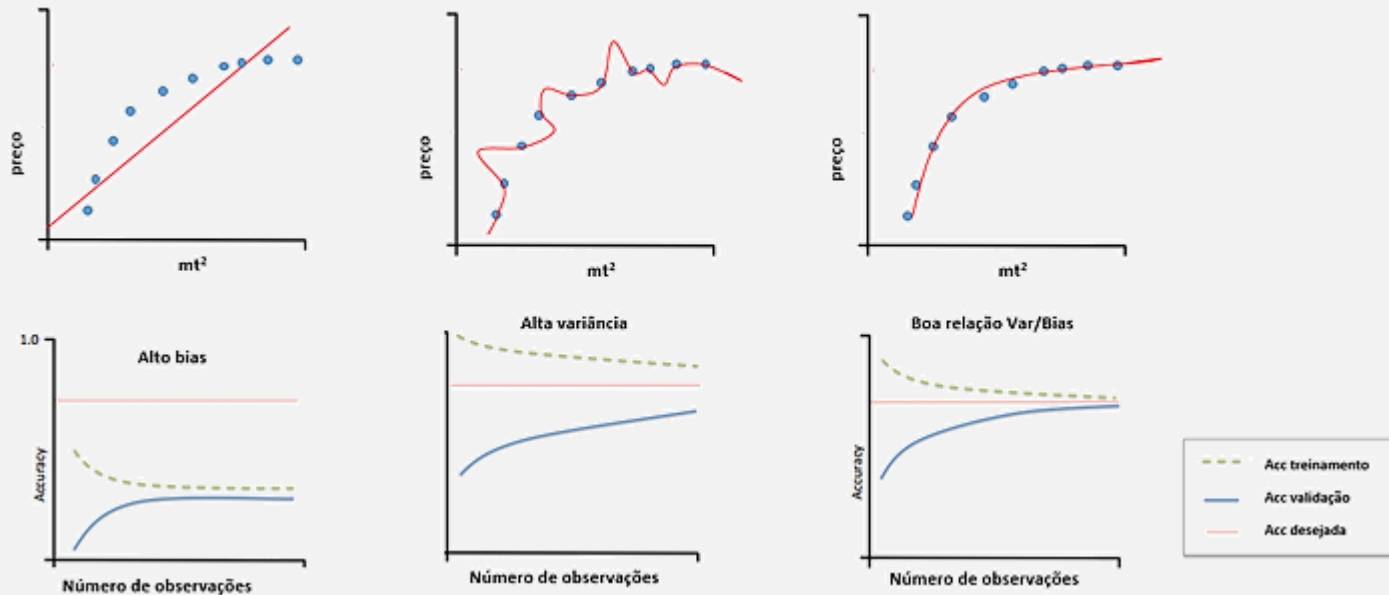
Overfit, Underfit



Redes Neurais

Overfit, Underfit

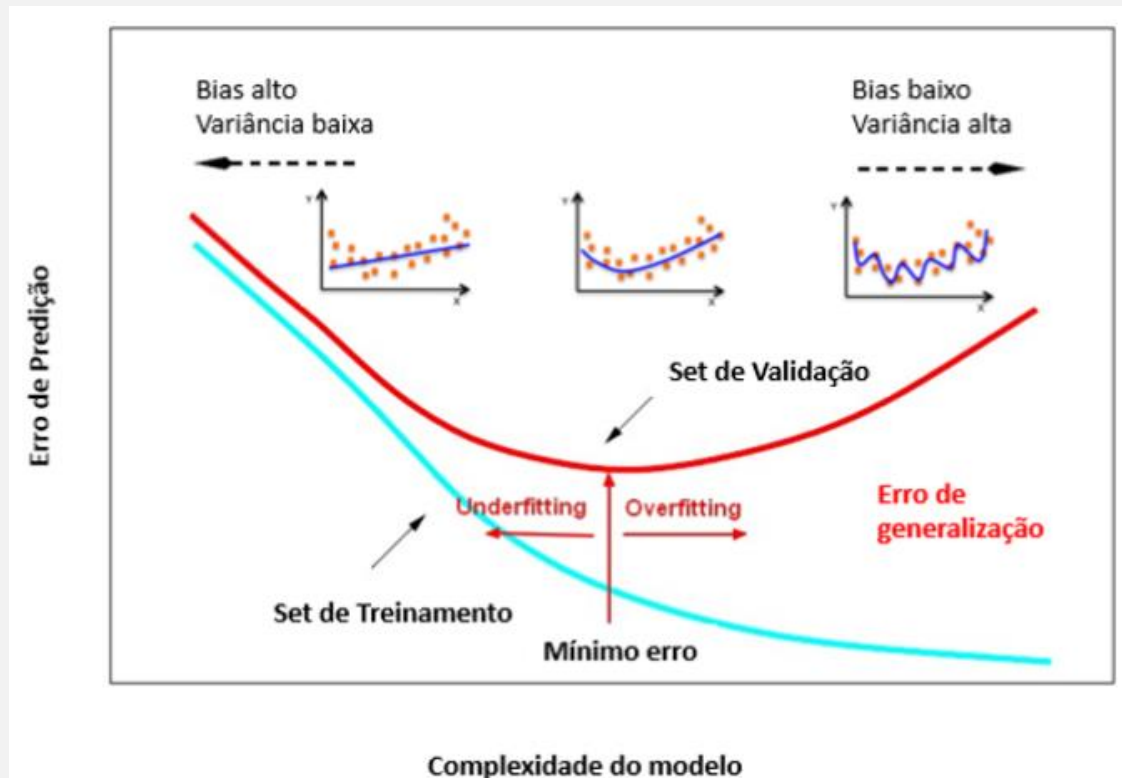
- Uma curva de aprendizagem é essencialmente um gráfico da precisão da predição (accuracy) versus o número de observações.



Redes Neurais

Overfit, Underfit

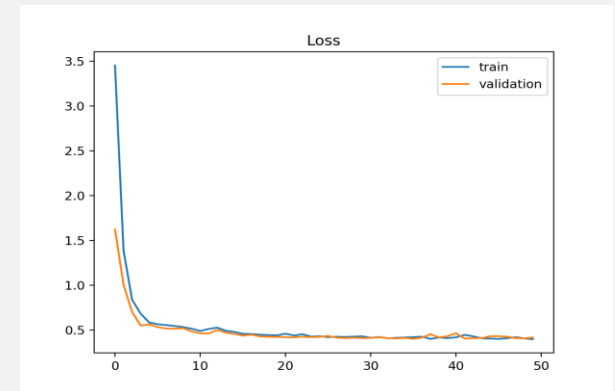
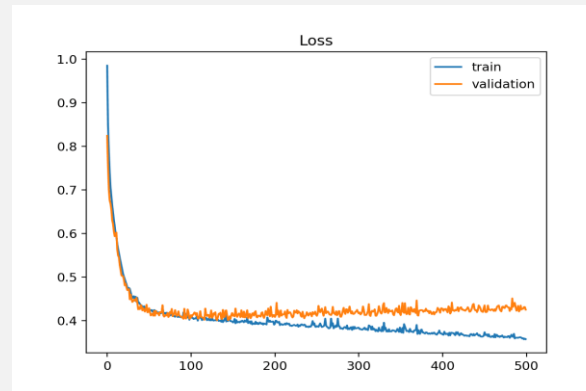
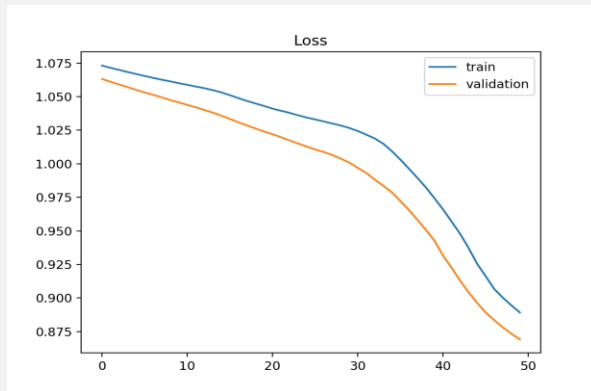
- Representação das regiões de overfitting e underfitting que relaciona a complexidade do modelo com o erro de predição.



Redes Neurais

Overfit, Underfit

- Essa divisão gera “curvas” de loss
- Essas curvas indicam como a rede neural está aprendendo, se está ocorrendo underfit ou overfit



Resumo

- Escalonamento auxilia na resolução de problemas de escala (grande variações)
- Redes Neurais podem aprender em tempo real e podem se adaptar aos ambientes em mudança de forma flexível
- A técnica de divisão da base de dados varia conforme a base de dados
- A escolha adequada dos parâmetros pode interferir diretamente no resultado final (overfitting e underfitting).

Trabalho Final (parte 2)

Implementação MultiLayer Perceptron.

Data de entrega do relatório final: 07/12

Dúvidas e envio do trabalho:

e-mail: bernarddss62gmail.com