



REDES NEURAIS COM TENSORFLOW

DIEGO RODRIGUES DSC
INFNET

Agenda

Parte 1 : CRISP

- Metodologia CRISP
- Aplicação aos Dados Iris

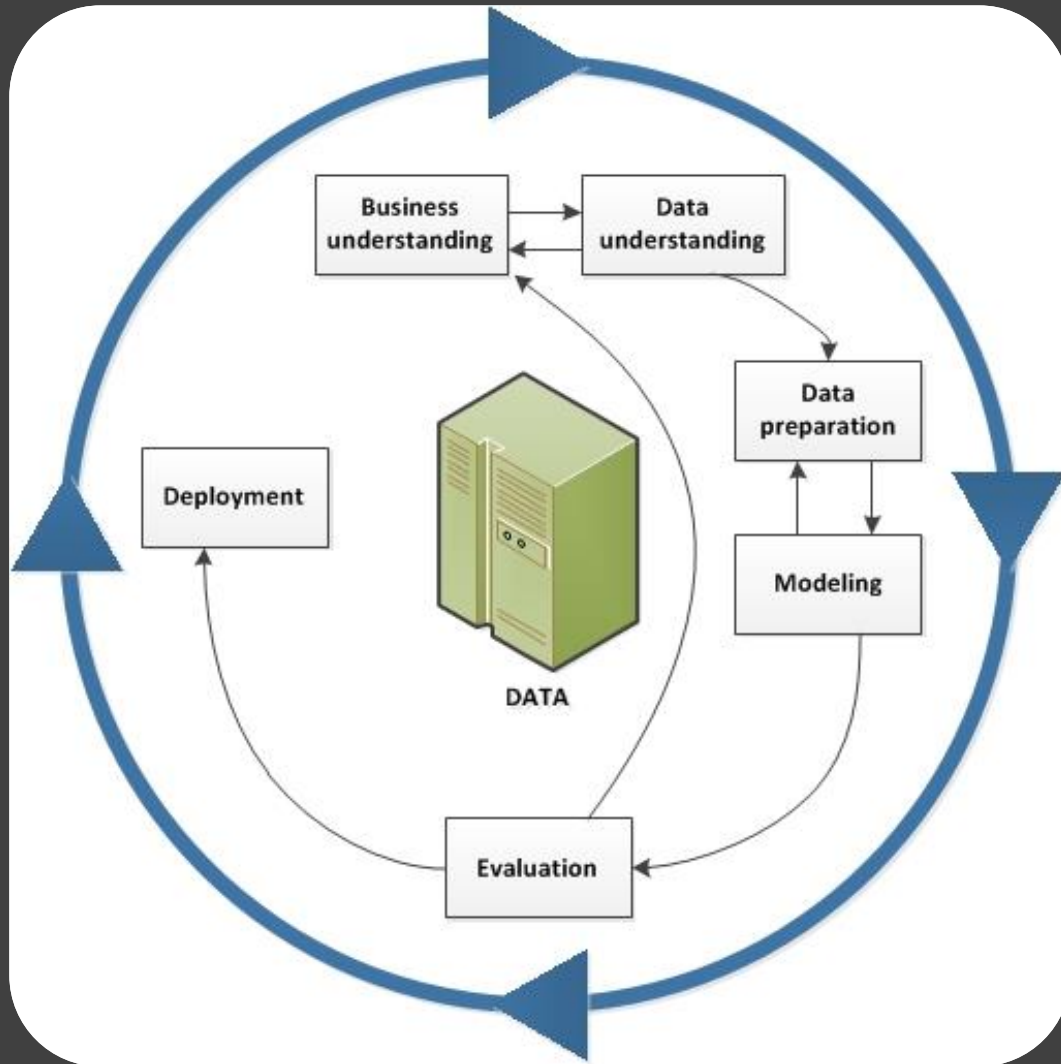
Parte 2 : Prática

- Notebook: Classificador Iris

Parte 3 : Trabalhos

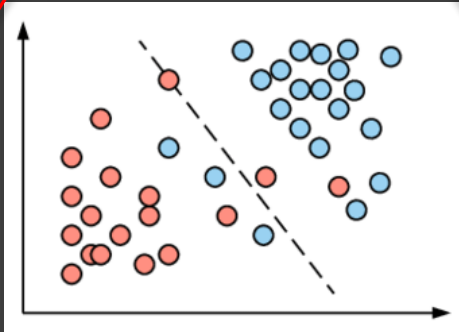
- Escopo & Evolução



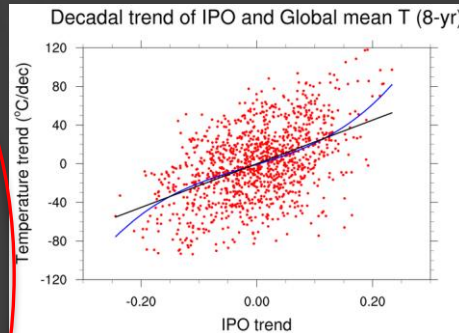


Cross
Industry
Process for
Data Mining
(CRISP-DM)

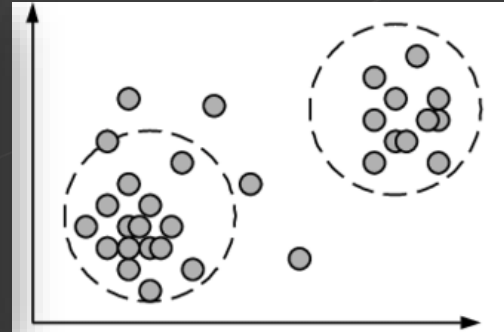
Paradigmas de Modelagem Estatística



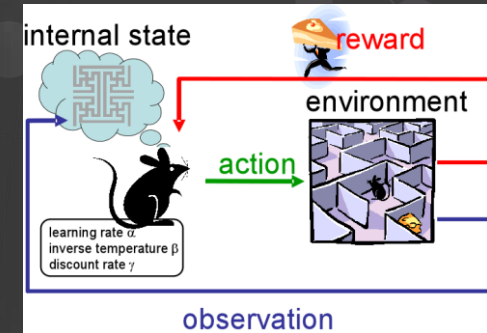
SUPERVISIONADO
– CLASSIFICAÇÃO



SUPERVISIONADO
– REGRESSÃO



NÃO
SUPERVISIONADO



APRENDIZADO
POR REFORÇO

Problema de Negócio

Classificação

Um bebê consegue separar e ordenar blocos com diferentes tamanhos, formas e cores. Ele também consegue **identificar os tipos diferentes de objetos**.

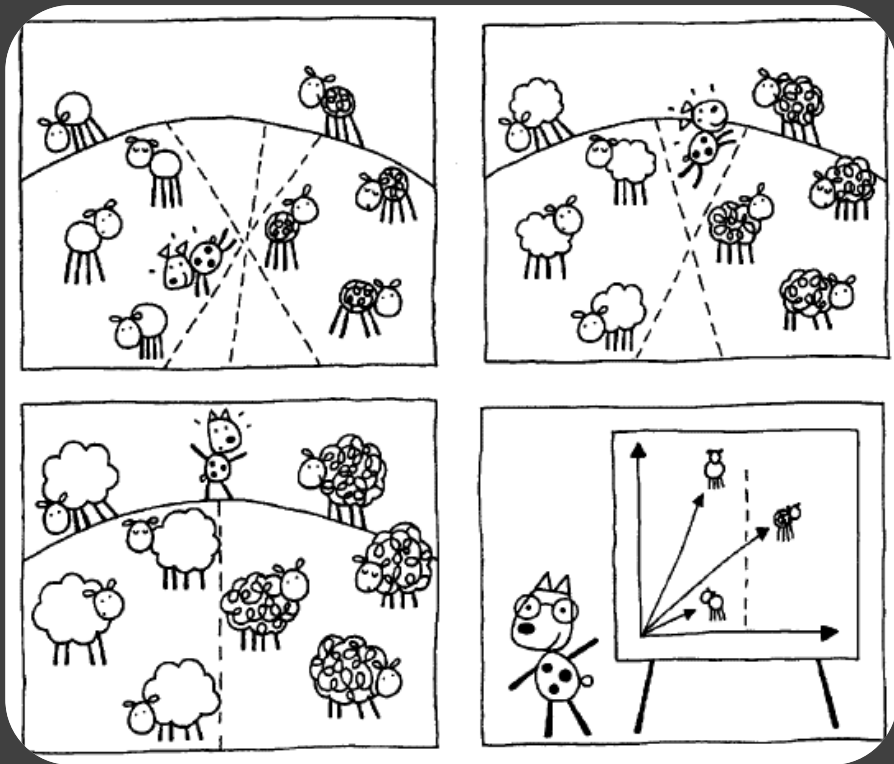
Os diferentes tipos de objetos são chamados de **classes**. As características dos objetos são chamadas de **variáveis** ou **atributos**.



Data Scientist da Nova Geração

Então, um classificador é um modelo **treinado para discriminar objetos** pertencentes a duas ou mais classes, baseado em seus atributos.

Representação



Exercício (1): qual representação o cachorro deve escolher para diferenciar ovelhas pretas, cinzas e brancas?

Exercício (2): qual seria uma boa representação para diferenciar alunos e alunas do curso de redes neurais?

Modelagem

1) Aprendizado Supervisionado

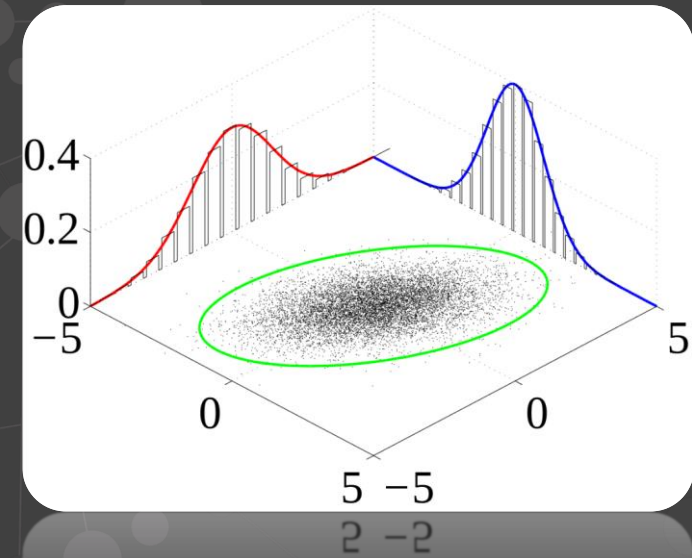
Tarefas de classificação e regressão pertencem a esta categoria. O treinamento consiste em **encontrar parâmetros** para o modelo que **minimiza uma função de risco/erro** para uma amostra de treinamento, baseado na diferença entre os **valores previstos e reais**, para cada observação.

Modelos Densos

Algoritmos que estimam a **função densidade de probabilidade** dos dados, ou aproximações locais, para determinar a classe de observações fora da amostra de treino.

- 1) Classificador Bayesiano & Naive
- 2) K-Vizinhos mais próximos.
- 3) Neurônio RBF
- 4) Mapa Auto Organizável / Rede Art

Algoritmos baseados em densidade dependem da **DENSIDADE (!!!)**. Consequentemente, se beneficiam de um **conjunto grande de observações e de baixa esparsidade do espaço de atributos**. O Classificador Bayesiano é considerado o classificador “ótimo”, mas é raramente utilizado, dada a dificuldade de estimar a função densidade de probabilidade dos dados. É normalmente utilizado como benchmark para comparação teórica entre os algoritmos de classificação.

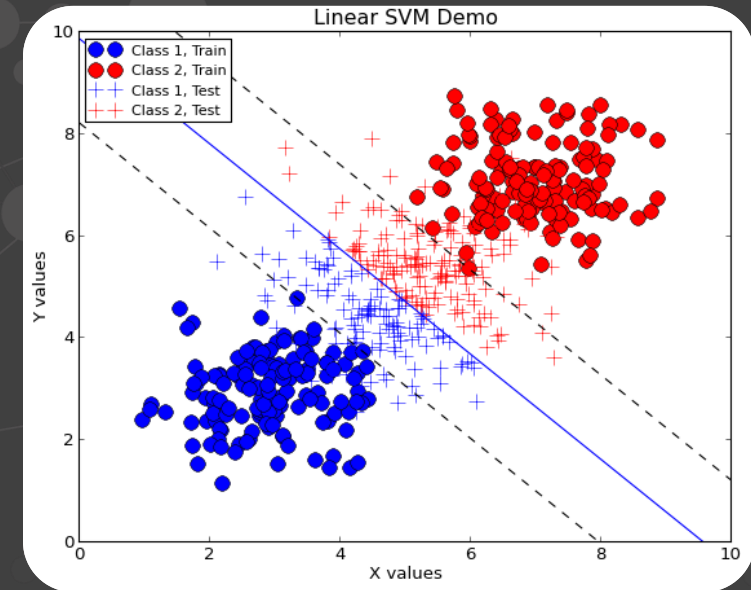


Modelos Funcionais

Algoritmos que dependem da **estimação dos parâmetros de uma função** que é utilizada como **superfície de separação** entre as classes.

- 1) Funções Polinomiais
- 2) Regressão Logística
- 3) Máquina de Vetores Suporte
- 4) Neurônio Sigmoidal / Tangente Hiperbólica
- 5) Árvores de Decisão

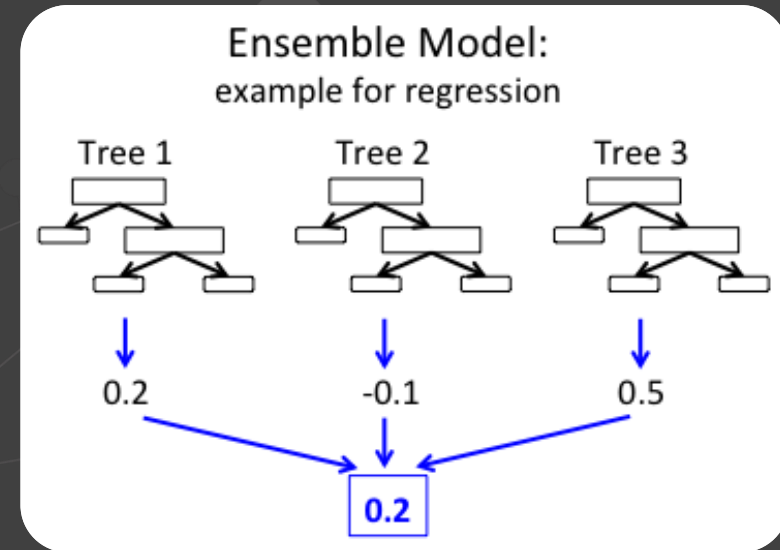
Algoritmos baseados em funções são **mais simples**, usualmente tem um **número menor de parâmetros** e não dependem em armazenar muitos dados para manter uma “memória”, como por exemplo K-vizinhos mais próximos.



Modelos de Comitê

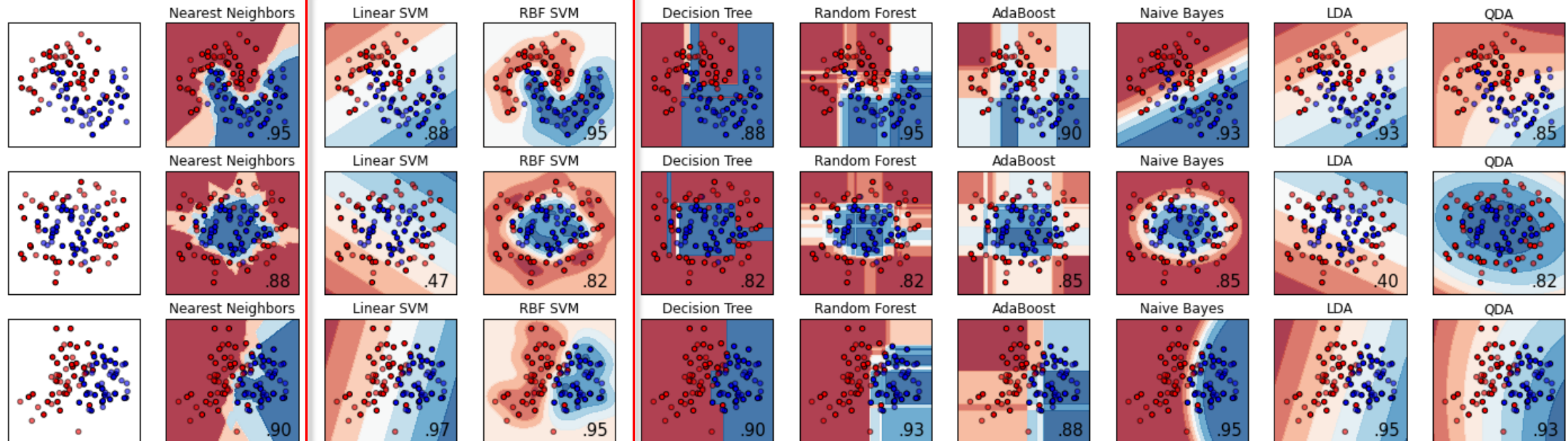
Algoritmos que **combinam modelos simples**,
usualmente através de **votação ou ponderação**,
para atingir maiores taxas de classificação.

- 1) Random Forest
- 2) Boosting
- 3) Redes Neurais Profundas

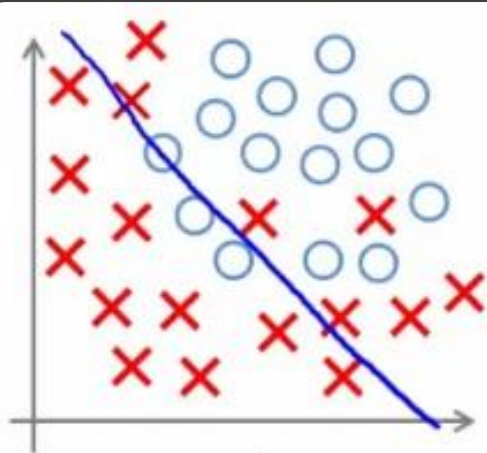


Boa **capacidade de generalização** gerado através de **arranjos complexos** de múltiplos modelos simples de machine learning.

Comparação dos Classificadores

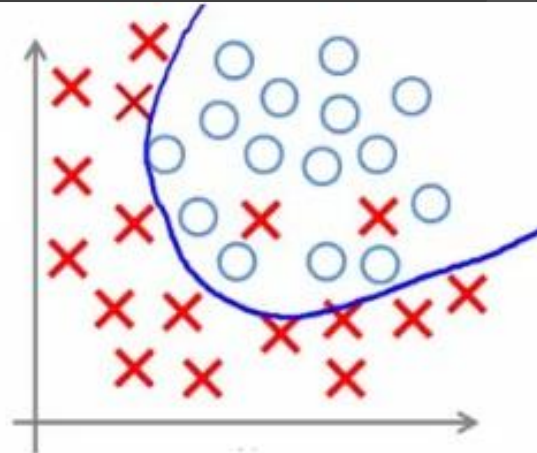


Validação

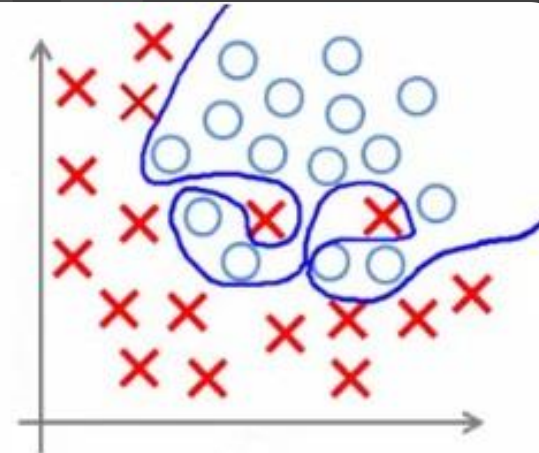


Under-fitting

(too simple to
explain the
variance)



Appropriate-fitting



Over-fitting

(forcefitting -- too
good to be true)

Identificando os meta- parâmetros

LEAVE ONE OUT

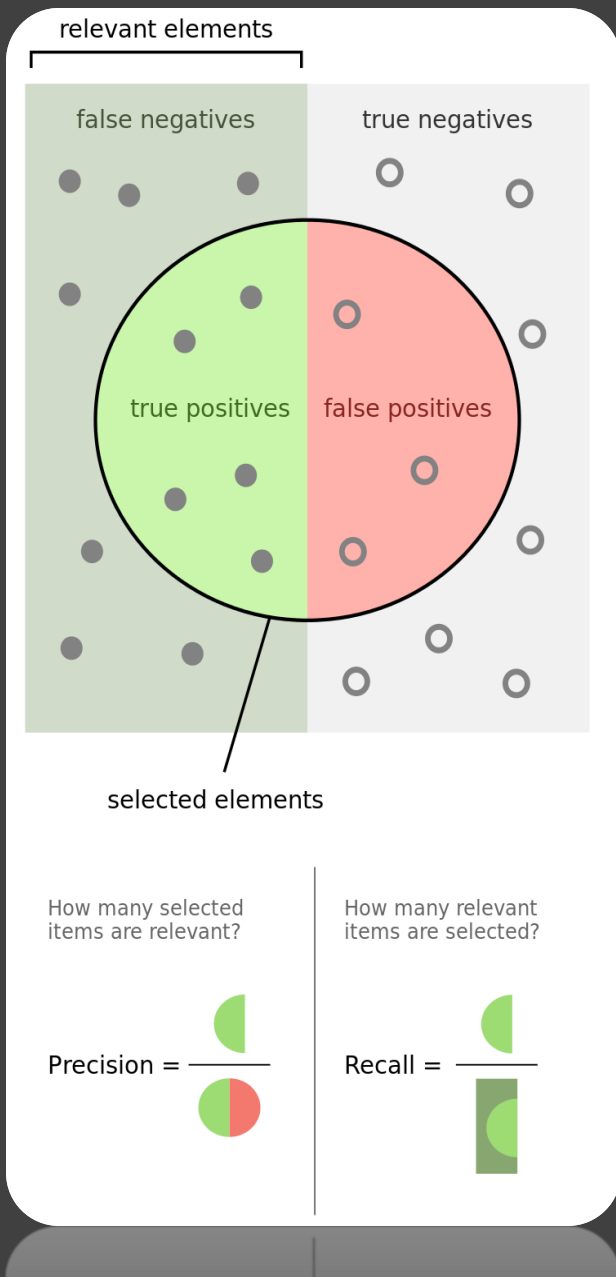
- Uma única observação é deixada de fora a cada treinamento. N treinamentos são realizados para calcular a estatística de erro.

K FOLDS

- Amostra é dividida em K conjuntos. K treinamentos são realizados, mantendo um conjunto como fora-da-amostra.

BOOTSTRAPPING

- O algoritmo itera, amostrando aleatoriamente M observações, para a quantidade Q desejada de treinamentos.



Figuras de Mérito

Acurácia

- $(TP+TN)/(P+N)$

Taxa de Erro

- $1 - \text{Acurácia}$

Sensibilidade (Recall)

- $TP/(TP+FN)$

Especificidade

- $TN/(TN+FP)$

Precisão

- $TP/(TP+FP)$

Produto Sp

- $\text{SQRT}[\text{SQRT}(R1 * R2) * (R1 + R2)/2]$

Classificação Iris

Problema de Negócio

Características das flores

Largura & comprimento da pétala

Largura & comprimento da sépala



Iris Setosa



Iris Versicolor



Iris Virginica

Iris Setosa

Iris Versicolor

Iris Virginica

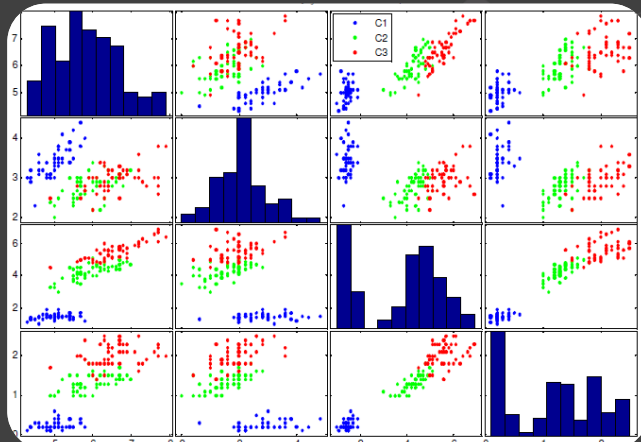
Representação



Características das flores

Largura & comprimento da pétala

Largura & comprimento da sépala

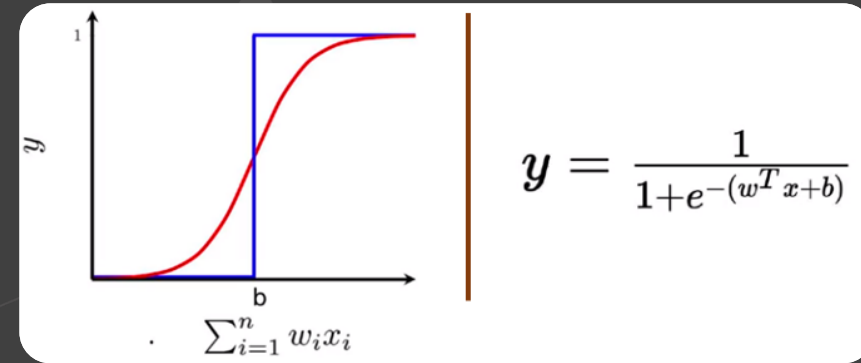


Espaço de
atributos com
4 dimensões!

Modelagem

Rede Neural Feed Forward

- Representação: 2 atributos
- Meta-parâmetros: 1 neurônio Sigmoidal
- Treinamento: base de treino completa.
 - Base de Validação Percentual
 - Critério de Parada: Número de épocas
- Algoritmo RMSProp
 - Batelada Completa
 - Taxa de Aprendizado
- Figura de Mérito: Precisão



$$\Delta w_{ij} = (\eta * \frac{\partial E}{\partial w_{ij}})$$

weight increment learning rate weight gradient

$$\Delta w_{ij} = (\eta * \frac{\partial E}{\partial w_{ij}}) + (\gamma * \Delta w_{ij}^{t-1})$$

momentum factor weight increment, previous iteration



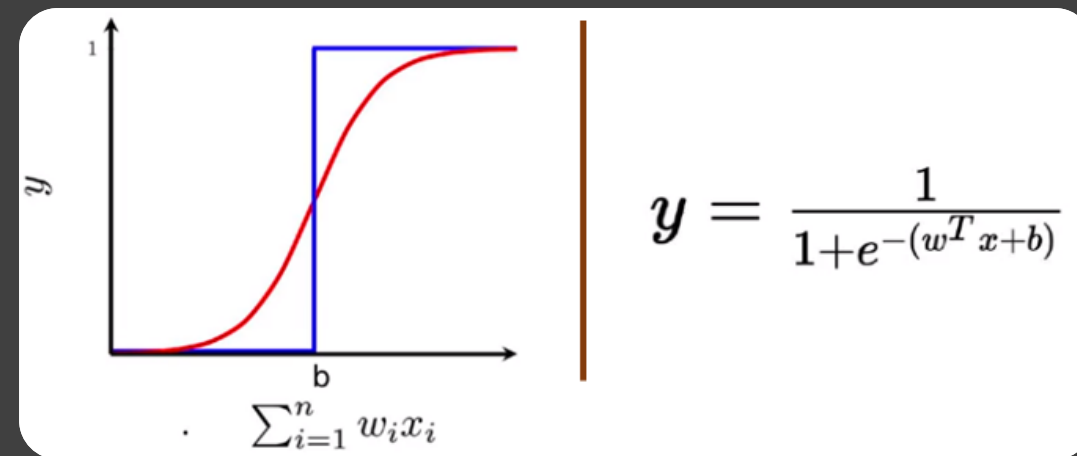
Iris Setosa



Iris Versicolor



Iris Virginica



Classificador Iris