

Alfabetização de Dados

Compreenda, organize e interprete os dados da sua organização

Aprendizado Supervisionado

Capítulo 11

Agenda

- Aprendizado Supervisionado
 - Classificação
 - Métodos Baseados em Distância
 - Métodos Probabilísticos
 - Métodos Baseados em Procura
 - Métodos Baseados em Otimização
 - Support Vector Machine
 - Deep Learning
 - Métodos Ensemble (mistos)
 - Regressão





Aprendizado Supervisionado

Classificação

Classificação

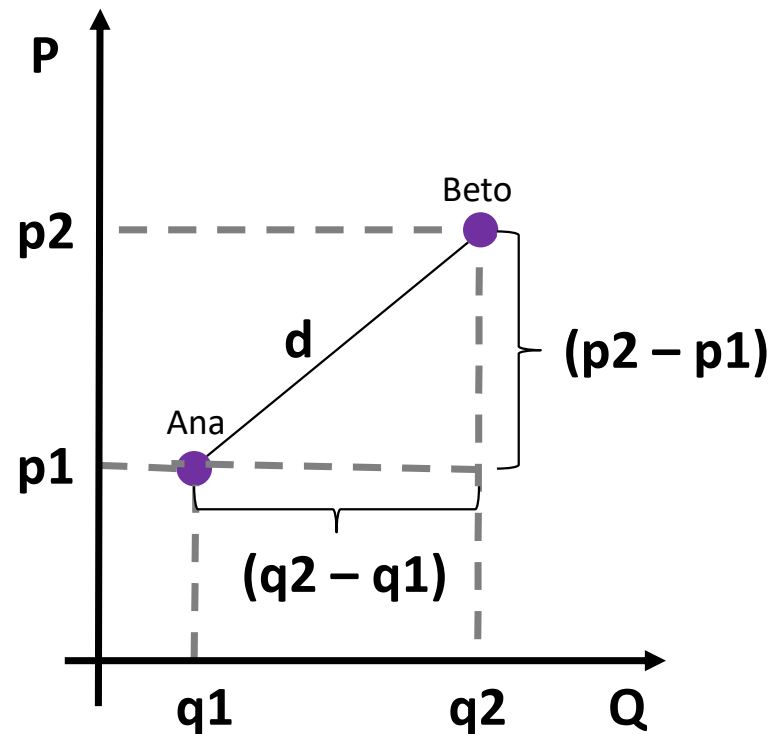
- ❑ Os principais métodos utilizados em classificação são:
 - ❑ Métodos baseados em distância.
 - ❑ Métodos probabilísticos.
 - ❑ Métodos baseados em procura.
 - ❑ Métodos baseados em otimização.
 - ❑ Métodos ensemble (mistos).

Métodos Baseados em Distância

- ❑ Os métodos de classificação baseados em distância, consideram a proximidade entre os itens de um conjunto de dados para a realização de previsões.
- ❑ O algoritmo K Nearest Neighbors (K-NN), que significa K-vizinhos mais próximos, parte da hipótese de que itens de dados similares tendem a estar concentrados na mesma região no espaço de dispersão dos dados.

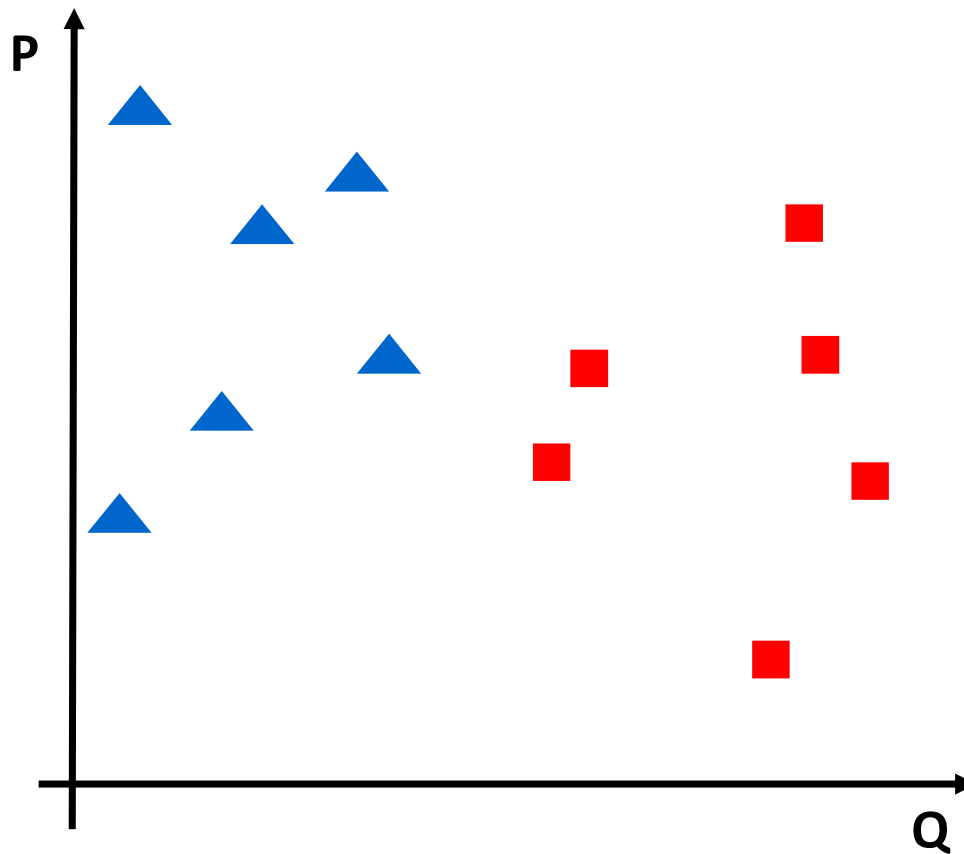
Distância Euclidiana

Nome	Idade (Q)	Salário (P)
Ana	34	12.000
Beto	42	18.000

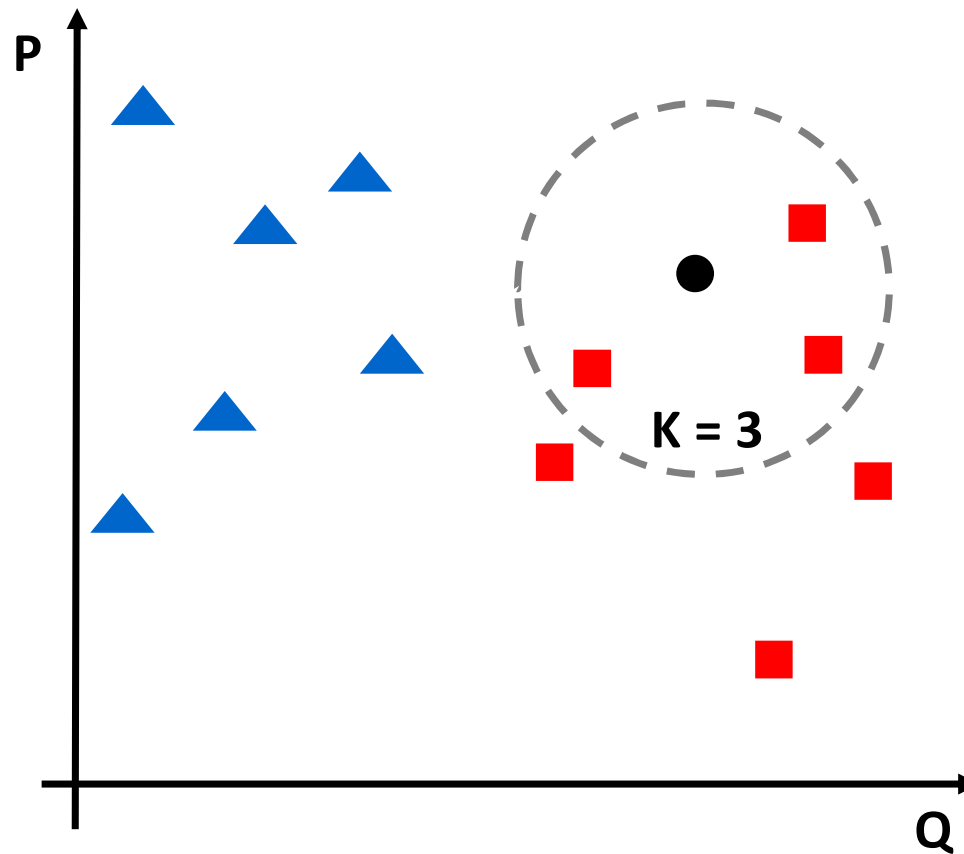


$$d = \sqrt{(q2 - q1)^2 + (p2 - p1)^2}$$

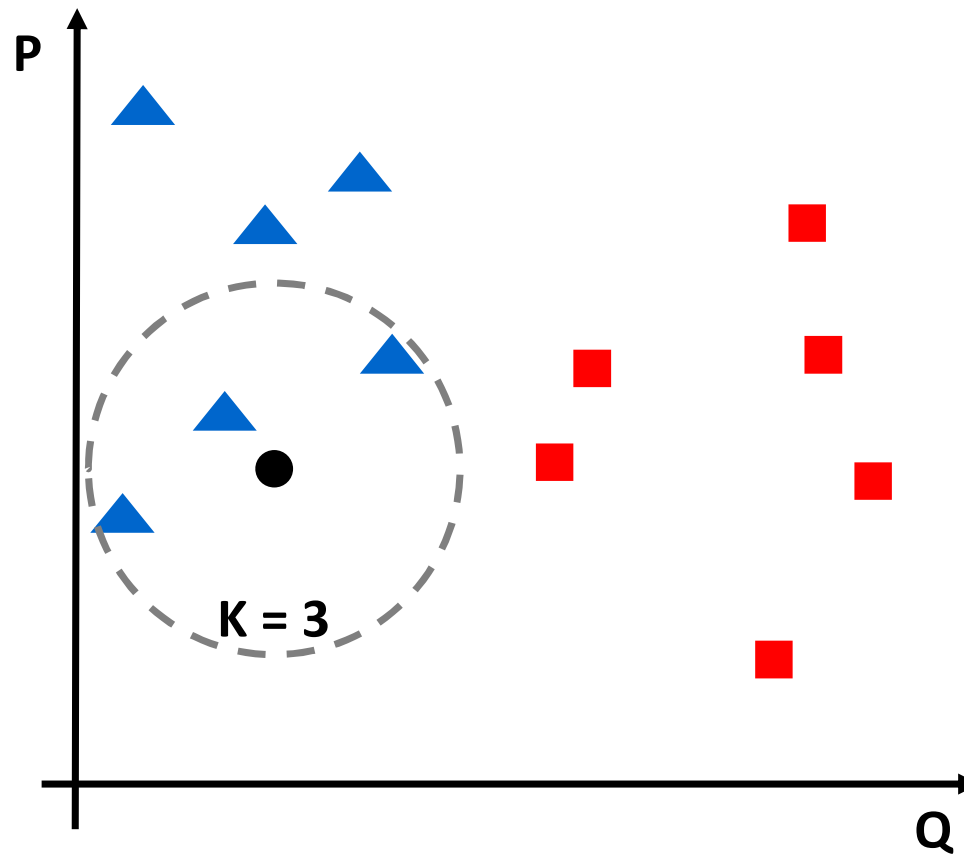
Distância Euclidiana



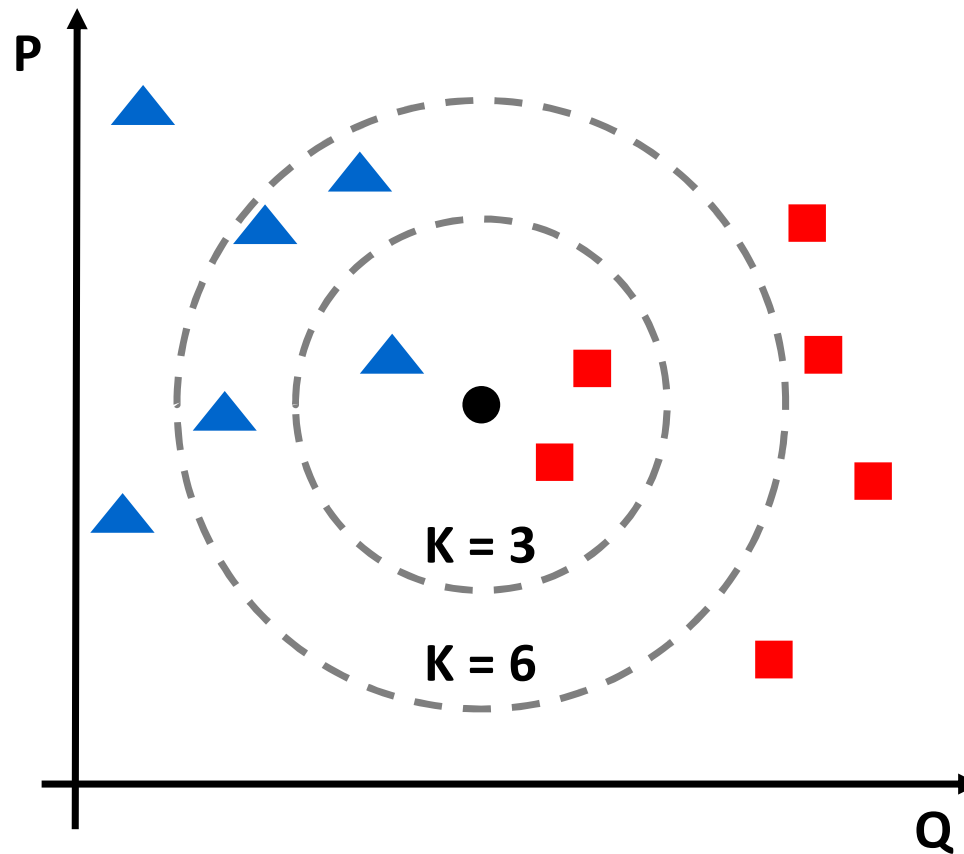
Distância Euclidiana



Distância Euclidiana



Distância Euclidiana



Métodos Probabilísticos

- ❑ Um dos métodos probabilísticos mais usuais para a classificação é fundamentado no Teorema de Bayes, denominado classificador Naive Bayes.
- ❑ O Teorema de Bayes é um modelo matemática utilizado para calcular a probabilidade de um evento ocorrer dado que outro evento já ocorreu, o que é denominado de probabilidade condicional.

Métodos Probabilísticos

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

- ❑ A e B são eventos observados.
- ❑ $P(A|B)$: representa a probabilidade de ocorrer o evento A dado que o evento B ocorreu.
- ❑ $P(B|A)$: representa a probabilidade de ocorrer o evento B dado que o evento A ocorreu.
- ❑ $P(A)$: representa a probabilidade de ocorrer o evento A.
- ❑ $P(B)$: representa a probabilidade de ocorrer o evento B.

Conjunto de Dados de Treinamento

Clima	Jogo
Sol	Não
Nublado	Sim
Chuva	Sim
Sol	Sim
Sol	Sim
Nublado	Sim
Chuva	Não
Chuva	Não
Sol	Sim
Chuva	Sim
Sol	Não
Nublado	Sim
Nublado	Sim
Chuva	Não



Tabela de Frequência

Clima	Não	Sim	Probabilidade	
Nublado	0	4	= 4/14	0,28
Sol	2	3	= 5/14	0,36
Chuva	3	2	= 5/14	0,36
Total	5	9		
	= 5/14	= 9/14		
	0,36	0,64		

Probabilidade

Exemplo

- ❑ Os jogadores irão praticar esporte se o tempo estiver ensolarado?
- ❑ $P(\text{Sim} \mid \text{Sol}) = ?$

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}$$



$$P(\text{Sim} \mid \text{Sol}) = \frac{P(\text{Sol} \mid \text{Sim}) P(\text{Sim})}{P(\text{Sol})}$$

Exemplo

$$P(\text{Sim} | \text{Sol}) = \frac{P(\text{Sol} | \text{Sim}) P(\text{Sim})}{P(\text{Sol})}$$

Tabela de Probabilidade

Clima	Não	Sim		
Nublado	0	4	= 4/14	0,28
Sol	2	3	= 5/14	0,36
Chuva	3	2	= 5/14	0,36
Total	5	9		
	= 5/14	= 9/14		
	0,36	0,64		

□ $P(\text{Sol} | \text{Sim}) = 3/9 = 0,33$

□ $P(\text{Sim}) = 9/14 = 0,64$

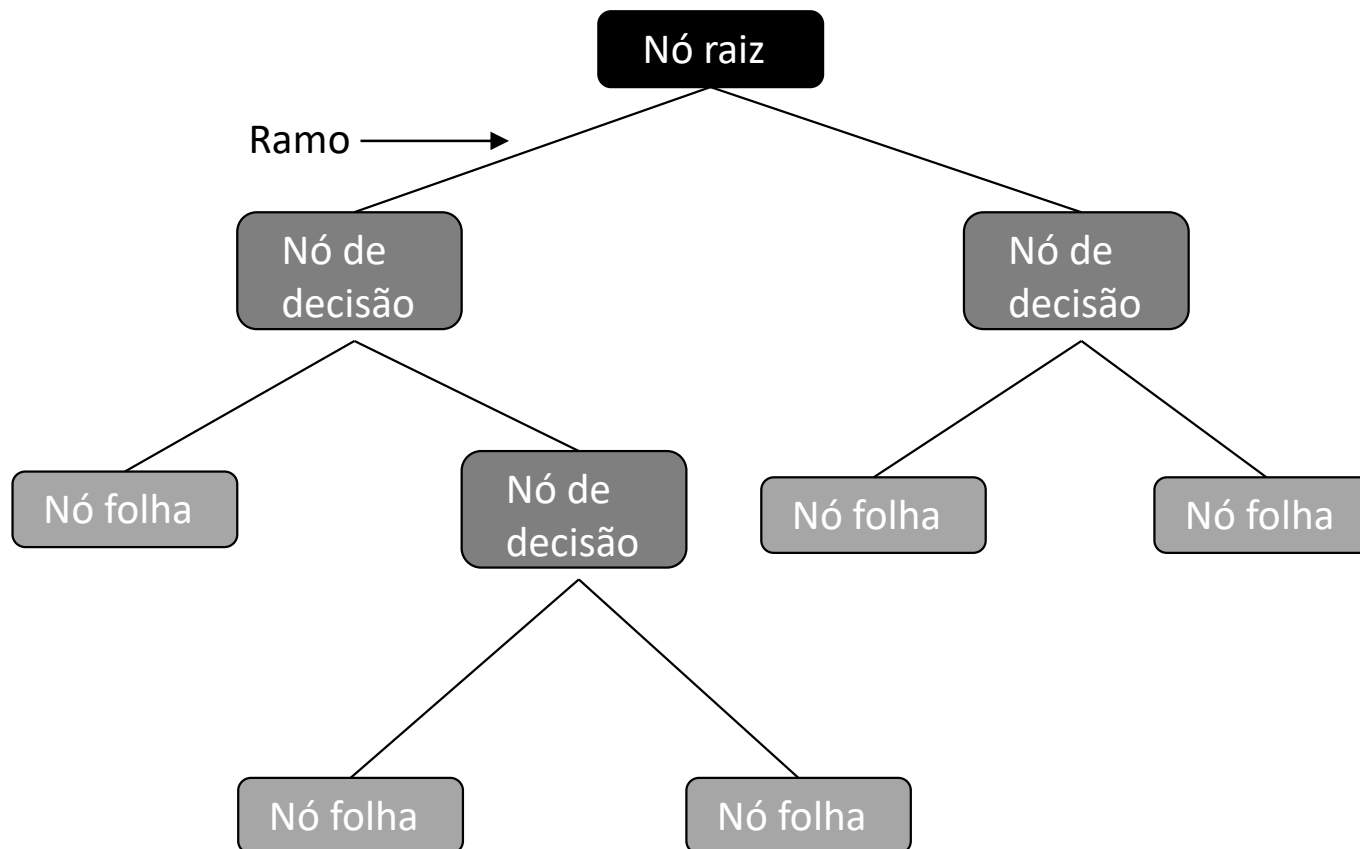
□ $P(\text{Sol}) = 5/14 = 0,36$

$$P(\text{Sim} | \text{Sol}) = (0,33 \times 0,64) / 0,36 = 0,58666 = 58,67\%$$

Métodos Baseados em Procura

- ❑ Árvore de decisão é o método baseado em procura mais usual em aprendizado de máquina supervisionado, pode ser utilizada em tarefas de classificação ou regressão.
- ❑ Uma árvore de decisão é composta por elementos chamados nós, dentre os quais um é denominado nó raiz.
- ❑ Esse nó origina uma relação de “paternidade” que gera uma estrutura hierárquica sobre os nós.
- ❑ Todos os nós que não possuem filhos são chamados nós terminais ou folhas.
- ❑ Os nós que contêm filhos são chamados nós internos ou de decisão.

Árvore de decisão



Exemplo

Nr	Aspecto	Humidade	Vento	Jogo
1	Sol	Elevada	Fraco	Não
2	Sol	Elevada	Forte	Não
3	Nublado	Elevada	Fraco	Sim
4	Chuva	Elevada	Fraco	Sim
5	Chuva	Normal	Fraco	Sim
6	Chuva	Normal	Forte	Não
7	Nublado	Normal	Fraco	Sim
8	Sol	Elevada	Fraco	Não
9	Sol	Normal	Fraco	Sim
10	Chuva	Normal	Forte	Não
11	Sol	Normal	Forte	Sim
12	Nublado	Elevada	Forte	Sim
13	Nublado	Normal	Fraco	Sim
14	Chuva	Elevada	Forte	Não

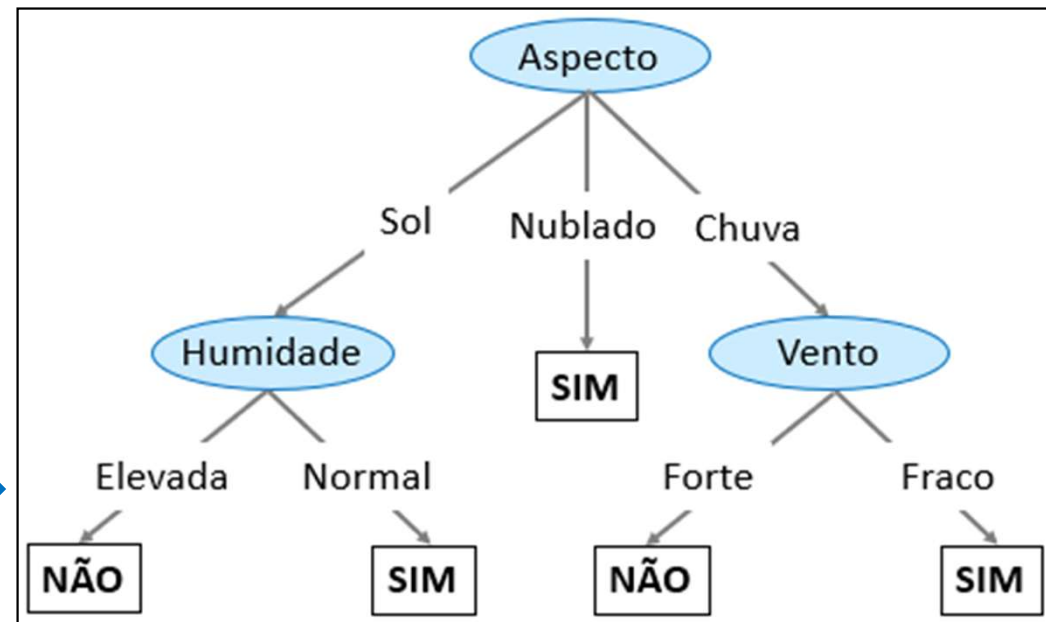
Aspecto	Humidade	Vento	Jogo
Sol	Elevada	Fraco	Não
Sol	Elevada	Forte	Não
Nublado	Elevada	Fraco	Sim
Chuva	Elevada	Fraco	Sim
Chuva	Normal	Fraco	Sim
Chuva	Normal	Forte	Não
Nublado	Normal	Fraco	Sim
Sol	Elevada	Fraco	Não
Sol	Normal	Fraco	Sim
Chuva	Normal	Forte	Não
Sol	Normal	Forte	Sim
Nublado	Elevada	Forte	Sim
Nublado	Normal	Fraco	Sim
Chuva	Elevada	Forte	Não

Dados

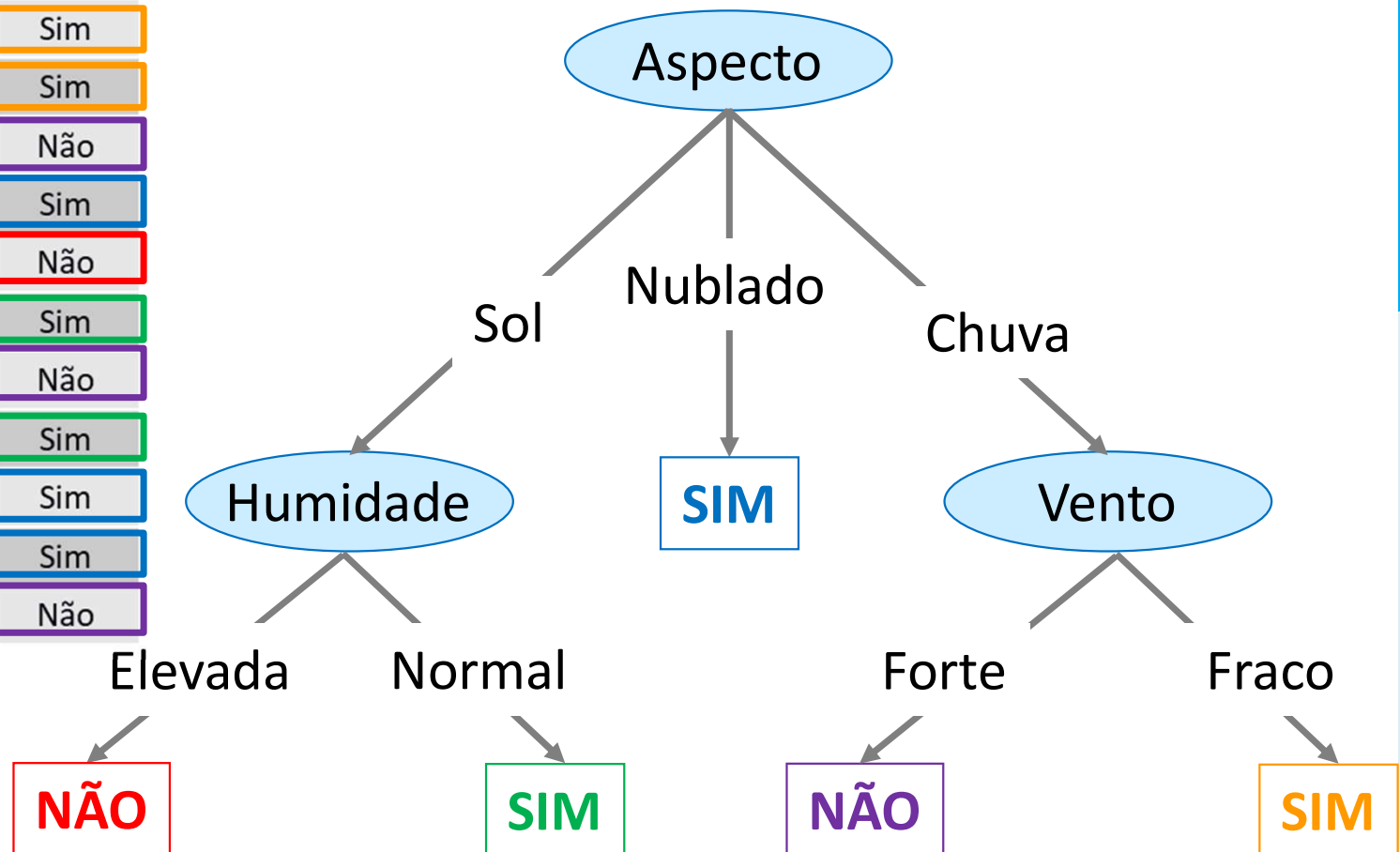
Respostas



Regras



Aspecto	Humidade	Vento	Jogo
Sol	Elevada	Fraco	Não
Sol	Elevada	Forte	Não
Nublado	Elevada	Fraco	Sim
Chuva	Elevada	Fraco	Sim
Chuva	Normal	Fraco	Sim
Chuva	Normal	Forte	Não
Nublado	Normal	Fraco	Sim
Sol	Elevada	Fraco	Não
Sol	Normal	Fraco	Sim
Chuva	Normal	Forte	Não
Sol	Normal	Forte	Sim
Nublado	Elevada	Forte	Sim
Nublado	Normal	Fraco	Sim
Chuva	Elevada	Forte	Não



Árvore de decisão

- ❑ Os nós folhas representam as duas classes desse conjunto de dados, ou seja, “há jogo” (sim) ou “não há jogo” (não).
- ❑ Essa árvore pode ser interpretada como um conjunto com as seguintes regras:
 - ❑ SE (Aspecto = Sol) E (Humidade = Elevada) ENTÃO Classe = Não.
 - ❑ SE (Aspecto = Sol) E (Humidade = Normal) ENTÃO Classe = Sim.
 - ❑ SE (Aspecto = Nublado) ENTÃO Classe = Sim.
 - ❑ SE (Aspecto = Chuva) E (Vento = Forte) ENTÃO Classe = Não.
 - ❑ SE (Aspecto = Chuva) E (Vento = Fraco) ENTÃO Classe = Sim.

Métodos Baseados em Otimização



- ❑ A aprendizagem supervisionada baseada em otimização visa encontrar o melhor conjunto de parâmetros para um modelo preditivo.
- ❑ Nesse contexto, “otimização” refere-se ao processo de ajuste dos parâmetros do modelo através da minimização ou maximização de uma função objetivo.
- ❑ Esse ajuste é realizado por meio do treinamento do modelo utilizando conjuntos de dados rotulados, ou seja, com as classes pré-definidas.

Métodos Baseados em Otimização



- ❑ Os algoritmos de aprendizagem supervisionada baseados em otimização incluem classificação por funções matemáticas, Regressão Logística, Support Vector Machine (SVM), Deep Learning, etc.
- ❑ Neste livro abordaremos apenas os algoritmos SVM e Deep Learning.

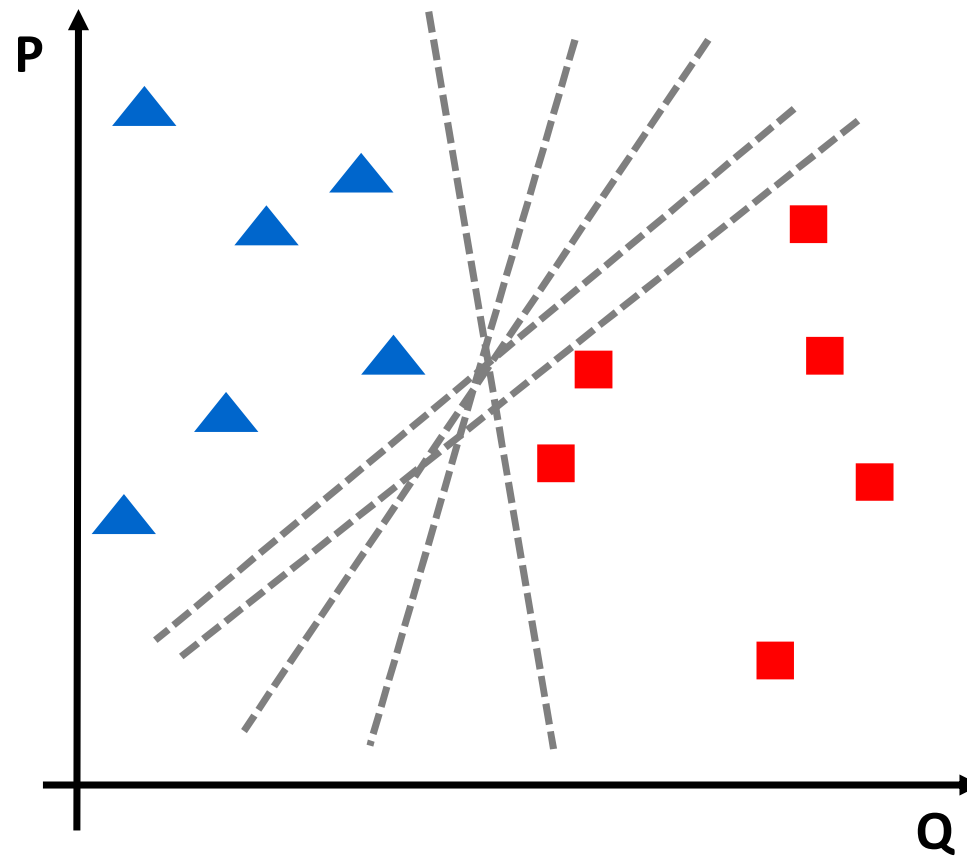
Support Vector Machines (SVM)

- ❑ Support Vector Machines (SVM) é um modelo de aprendizagem supervisionada que pode ser utilizado para classificação e análise de regressão, nesta seção será abordado apenas o seu uso para a classificação.

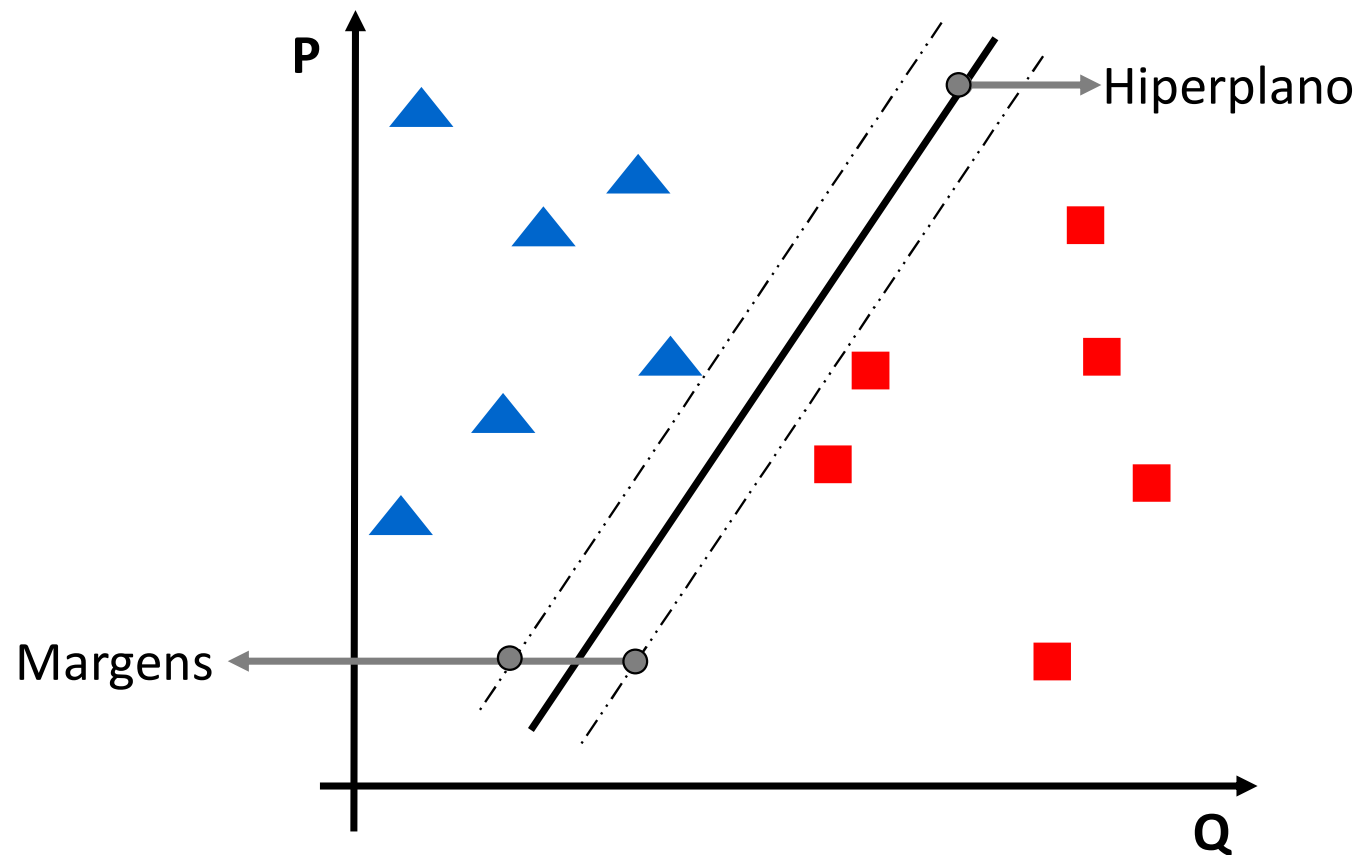
Support Vector Machines (SVM)

- ❑ Diferentemente do algoritmo K-NN, que é um método baseado em distância, o algoritmo SVM é embasado na definição de um hiperplano que separa as classes do conjunto de dados usado para o treinamento.
- ❑ Hiperplanos são limites de decisão que auxiliam a separação dos dados em classes.
- ❑ A dimensão do hiperplano depende do número de atributos do conjunto de dados.
- ❑ Se o número de atributos for 2, o hiperplano será apenas uma linha.
- ❑ Se o número de atributos for 3, o hiperplano se tornará um plano bidimensional e assim por diante.

Support Vector Machines (SVM)



Support Vector Machines (SVM)



Deep Learning

- ❑ Deep Learning pode ser traduzido para “Aprendizado Profundo” que, na verdade, é proporcionado por Redes Neurais Profundas resultantes da evolução das Redes Neurais Artificiais (RNA).
- ❑ RNA é uma técnica computacional representada por um modelo matemático inspirado na estrutura neuronal do ser humano que assimila conhecimento por meio da experiência (treino).
- ❑ Consiste em uma rede de neurônios artificiais interconectados, que processam dados através de conexões ponderadas.

Deep Learning

- ❑ Cada neurônio artificial em uma RNA recebe entradas, que são ponderadas para gerarem saídas geralmente transformadas por meio das funções de ativação.
- ❑ A rede neural pode ter várias camadas de neurônios, incluindo uma camada de entrada, uma ou mais camadas ocultas, que são as camadas intermediárias, e uma camada de saída.
- ❑ O processo de treinamento de uma rede neural geralmente envolve ajustar os pesos das conexões entre os neurônios para que a rede possa aprender a mapear corretamente as entradas para as saídas desejadas.

Deep Learning

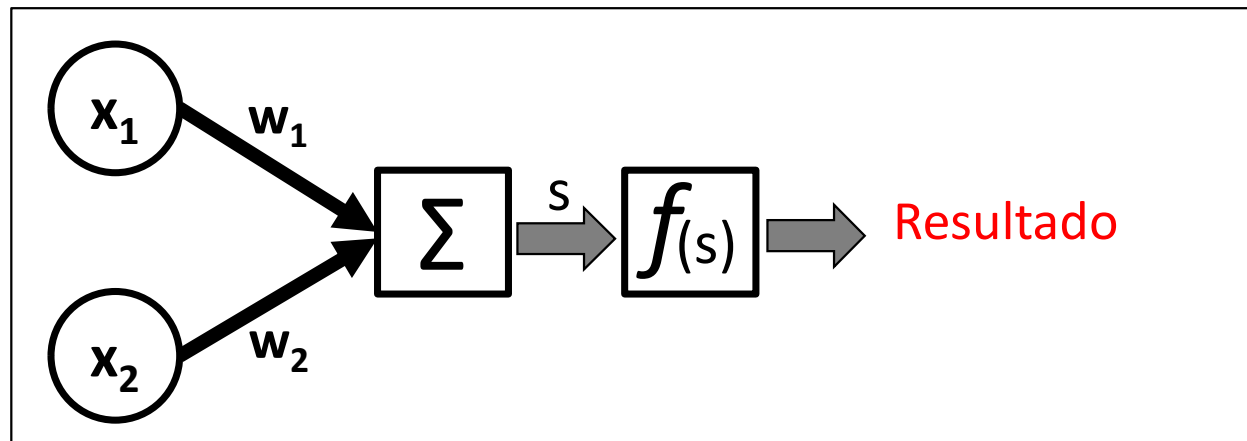
- ❑ As RNAs têm sido muito utilizadas para atividades de classificação, processamento de linguagem natural, reconhecimento de imagem, diagnósticos e predição de doenças, carros autônomos, detecção de fraudes financeiras e várias outras.
- ❑ Elas são especialmente poderosas em lidar com problemas complexos e não lineares, onde métodos tradicionais tendem a ser menos eficazes.

Modelo Perceptron

- ❑ Foi desenvolvido nas décadas de 1950 e 1960 pelo cientista Frank Rosenblatt, inspirado em trabalhos anteriores de Warren McCulloch e Walter Pitts.
- ❑ É uma rede neural de camada única capaz de atuar como um classificador linear binário.
- ❑ Para ser considerada uma RNA é necessário, no mínimo, um Perceptron de várias camadas.

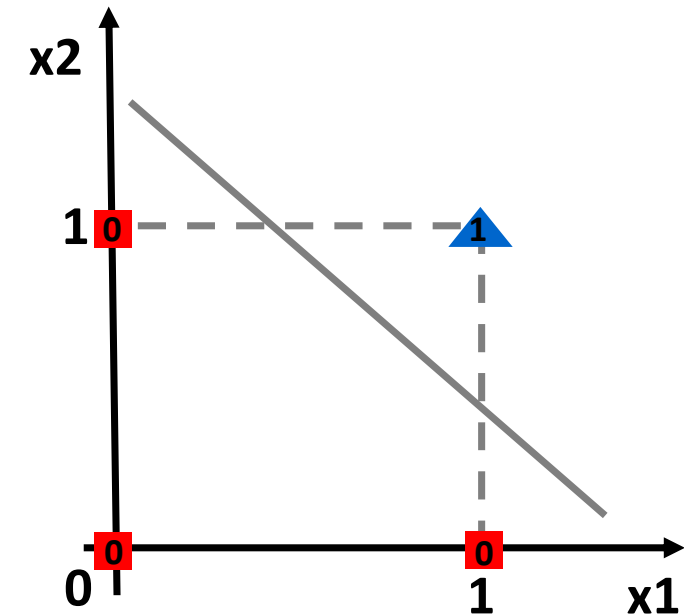
Modelo Perceptron

- ❑ Sinais são apresentados à entrada: x_1 e x_2 ;
- ❑ Cada sinal é multiplicado por um peso: w_1 e w_2 ;
- ❑ É feita a soma ponderada dos sinais: $s = (x_1 * w_1) + (x_2 * w_2)$;
- ❑ Este nível de atividade 's' passa pela função de ativação $f(s)$ para verificar se o somatório excedeu um certo limite (threshold) ou não.
- ❑ O processamento dessa verificação determinará um resultado de saída.



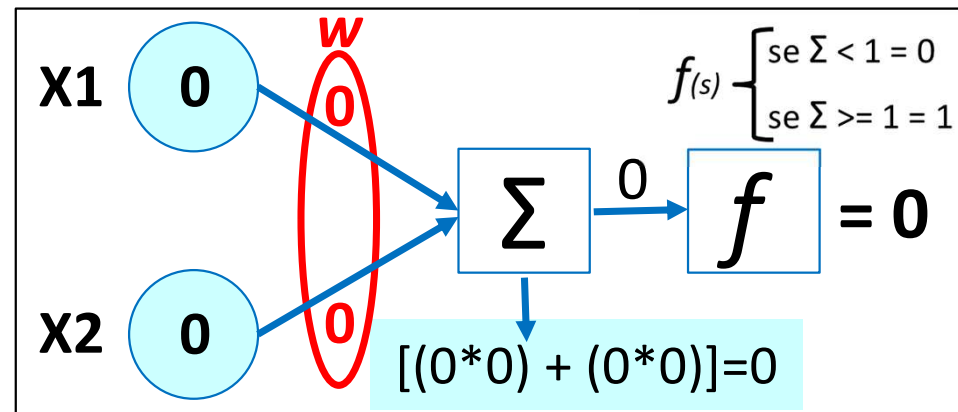
Modelo Perceptron

x1	x2	x1 AND x2
0	0	0
0	1	0
1	0	0
1	1	1

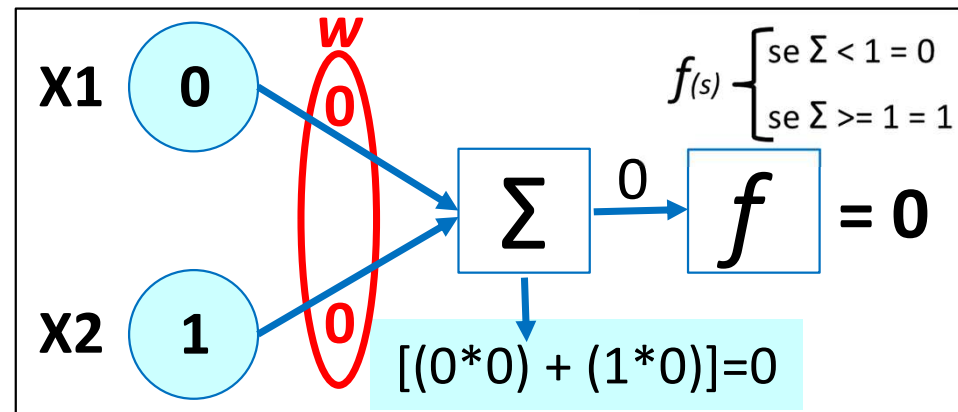


$$f(s) = \begin{cases} \text{se } \Sigma < 1 = 0 \\ \text{se } \Sigma \geq 1 = 1 \end{cases}$$

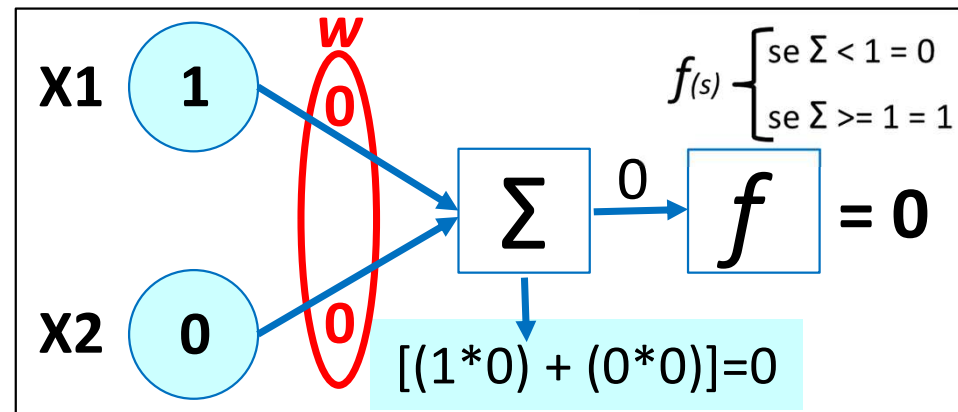
Modelo Perceptron



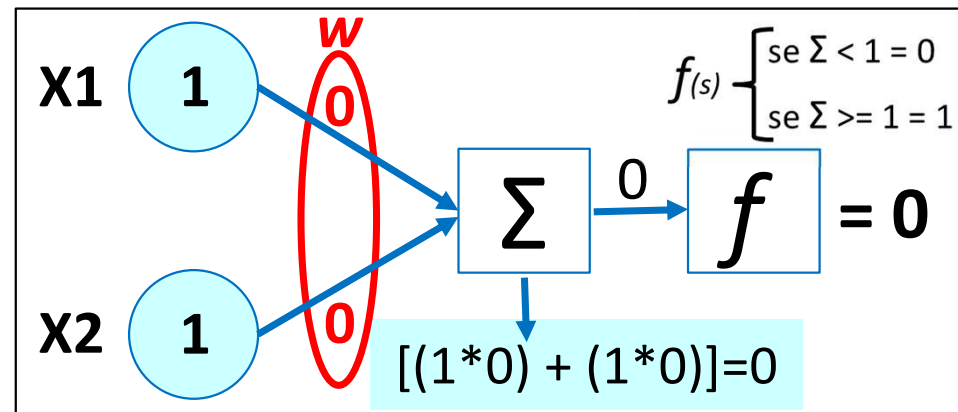
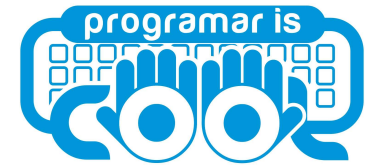
Modelo Perceptron



Modelo Perceptron



Modelo Perceptron

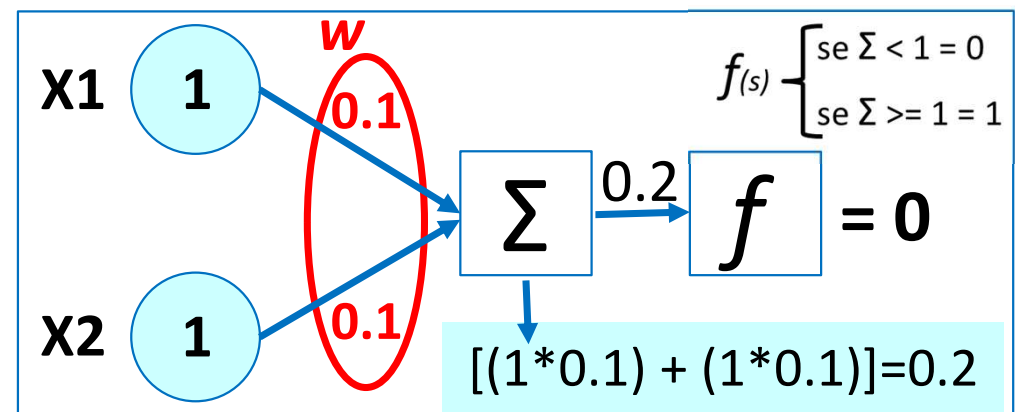
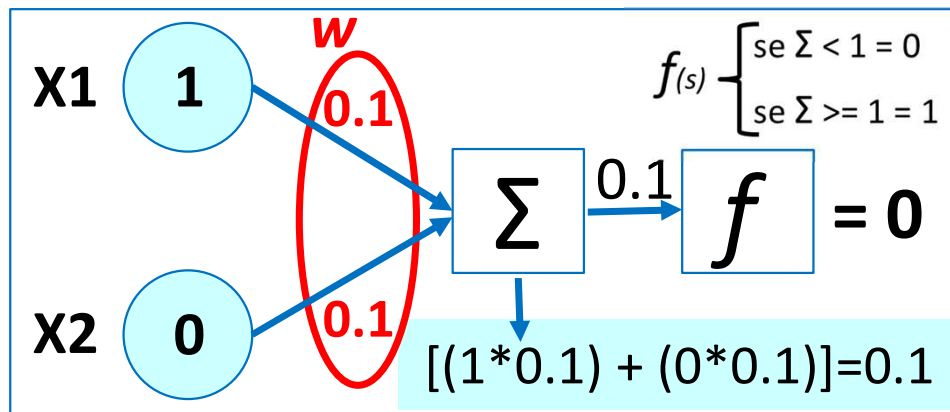
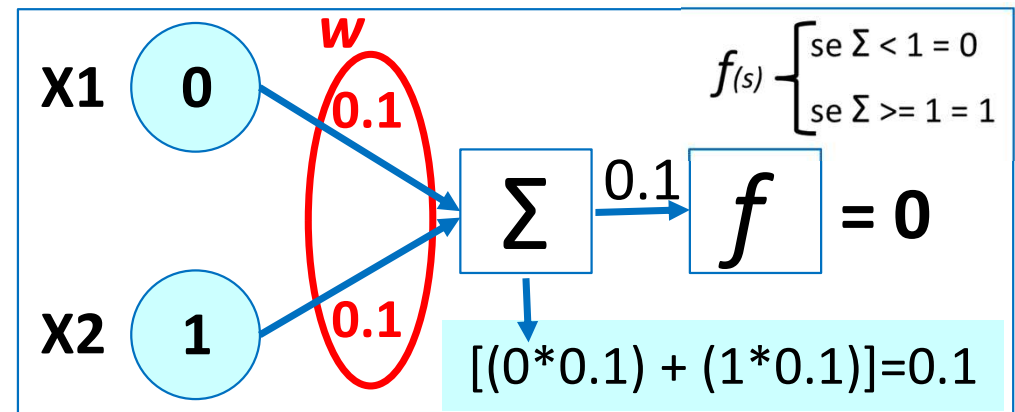
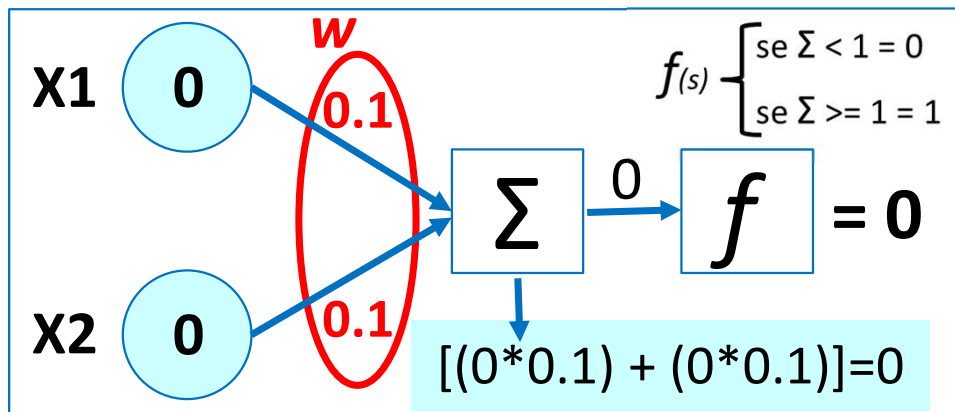


Modelo Perceptron

❑ Erro total:

- ❑ Para `0 AND 0` \Rightarrow erro = $0 - 0 = 0$.
- ❑ Para `0 AND 1` \Rightarrow erro = $0 - 0 = 0$.
- ❑ Para `1 AND 0` \Rightarrow erro = $0 - 0 = 0$.
- ❑ Para `1 AND 1` \Rightarrow erro = $1 - 0 = 1$.

Modelo Perceptron

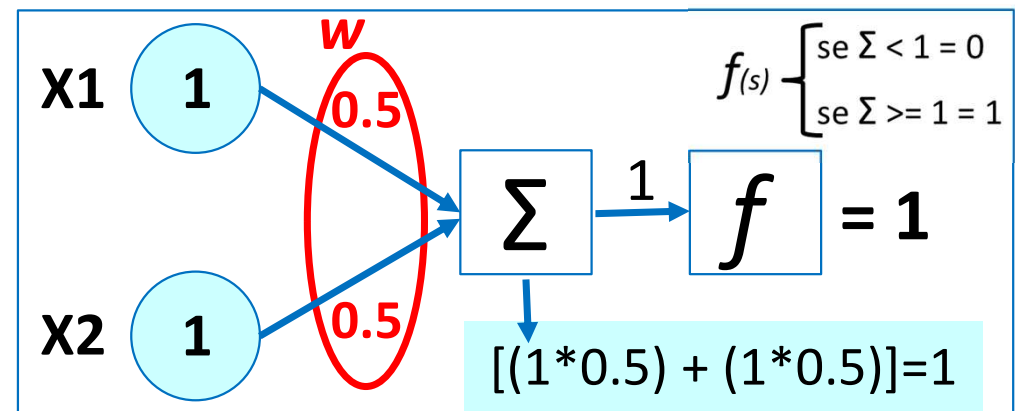
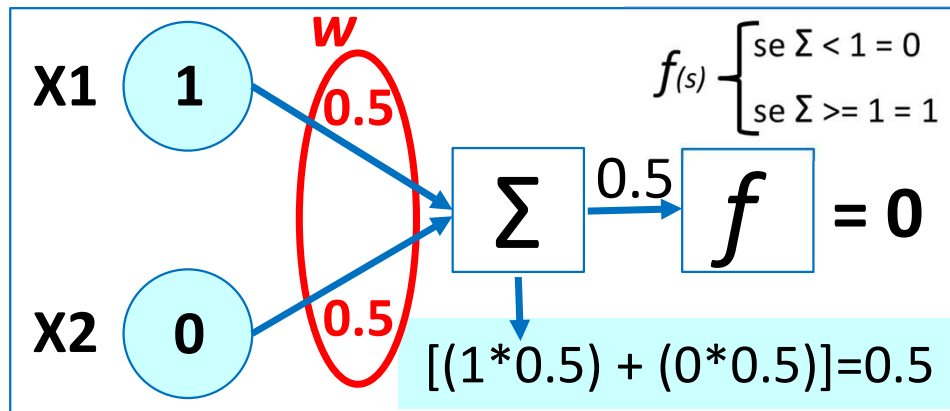
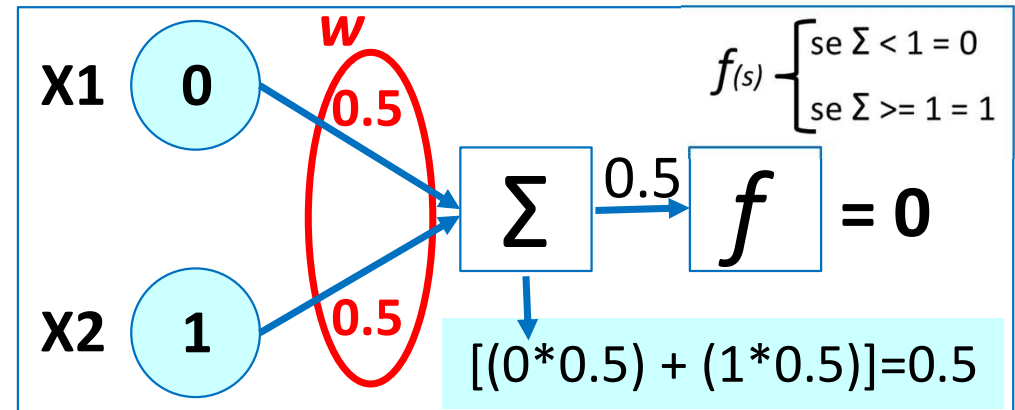
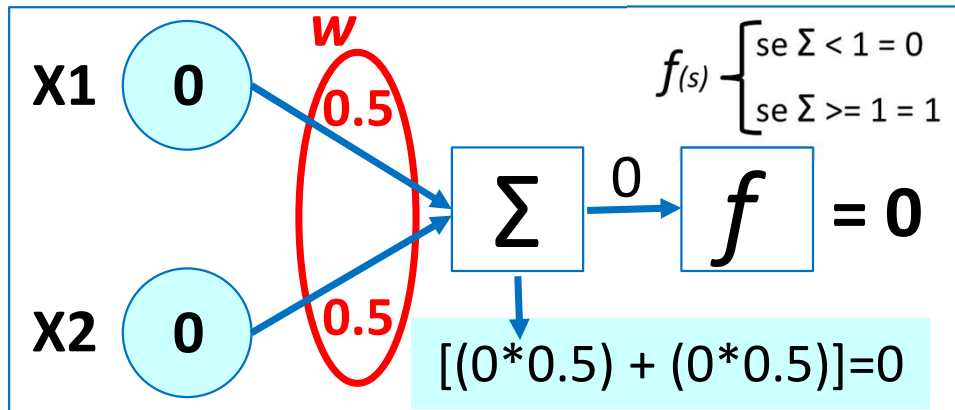


Modelo Perceptron

❑ Erro total:

- ❑ Para `0 AND 0` \Rightarrow erro = $0 - 0 = 0$.
- ❑ Para `0 AND 1` \Rightarrow erro = $0 - 0 = 0$.
- ❑ Para `1 AND 0` \Rightarrow erro = $0 - 0 = 0$.
- ❑ Para `1 AND 1` \Rightarrow erro = $1 - 0 = 1$.

Modelo Perceptron



Modelo Perceptron

❑ Erro total:

❑ Para `0 AND 0` \Rightarrow erro = $0 - 0 = 0$.

❑ Para `0 AND 1` \Rightarrow erro = $0 - 0 = 0$.

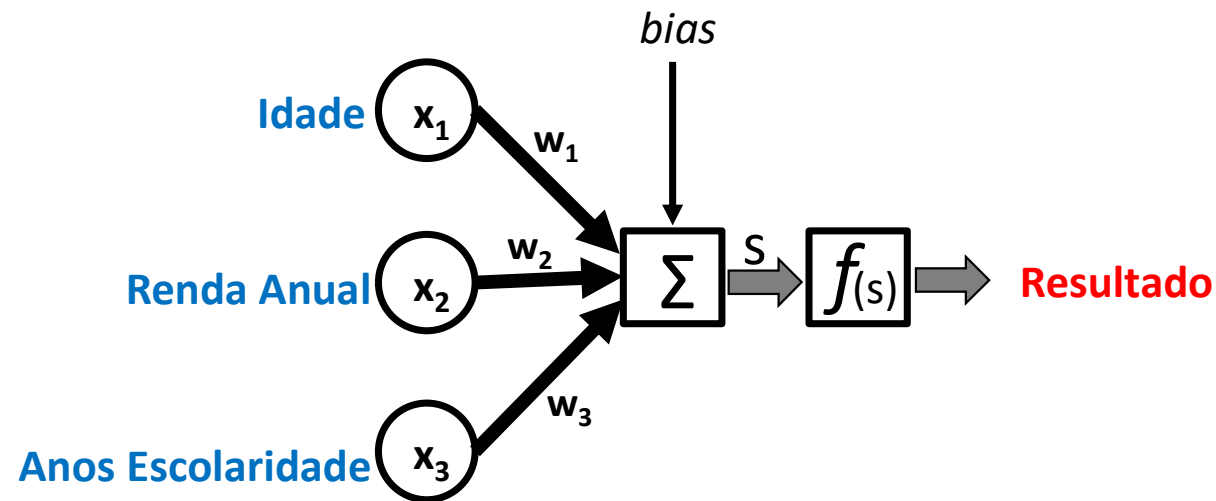
❑ Para `1 AND 0` \Rightarrow erro = $0 - 0 = 0$.

❑ Para `1 AND 1` \Rightarrow erro = $1 - 1 = 0$.

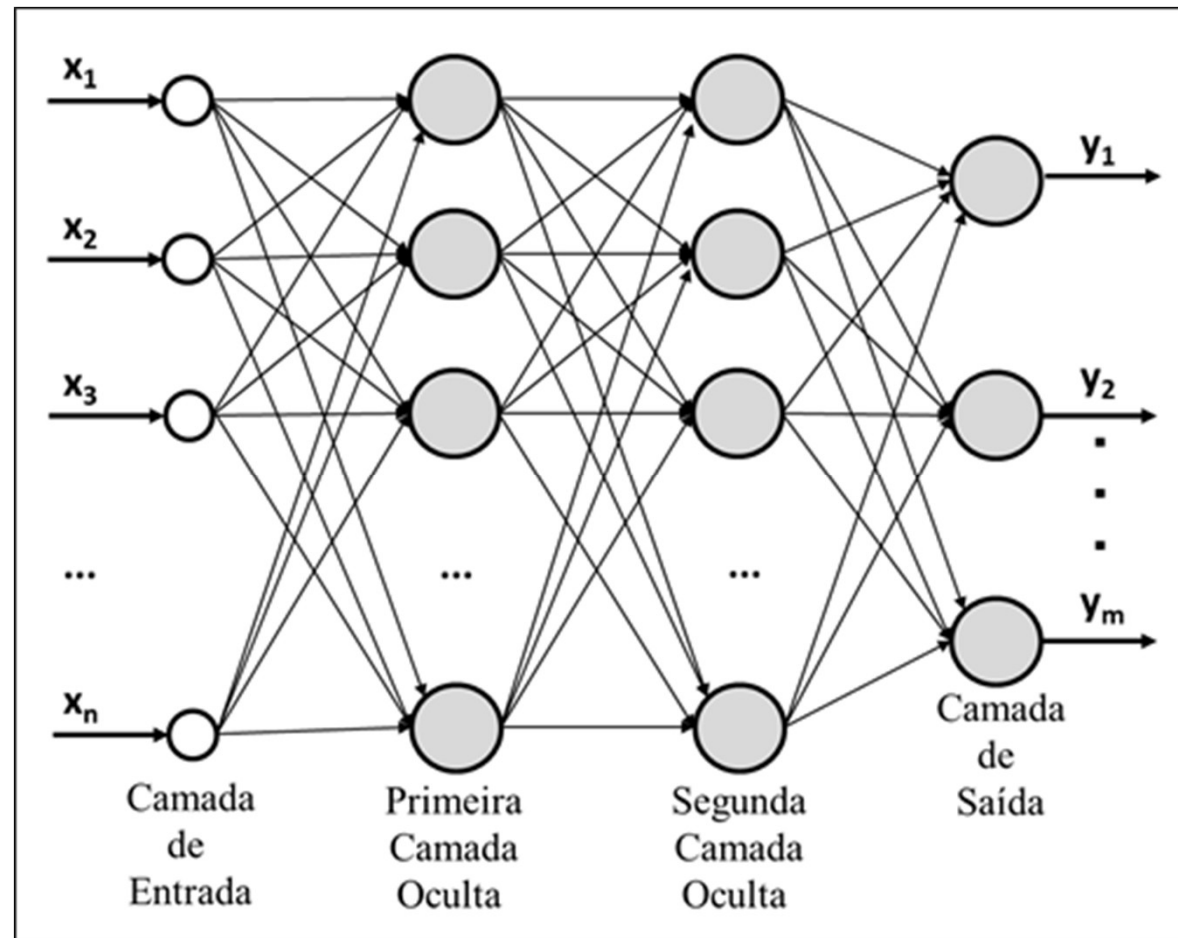
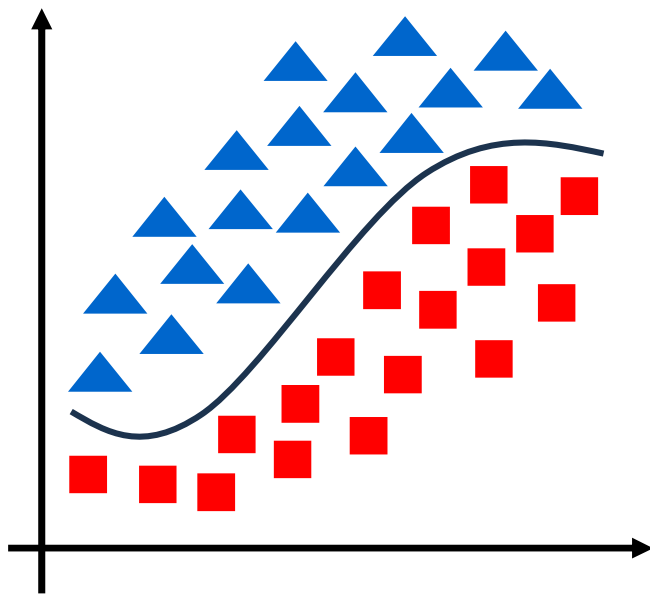
Conjunto
de
Treino

Idade	Renda Anual	Anos Escolaridade	Risco
35	100.000	10	Alto
45	150.00	15	Baixo
50	250.000	18	Baixo
38	180.000	8	Baixo
24	75.000	16	Alto
...
54	175.000	12	?

← Predição



Rede Neural Profunda

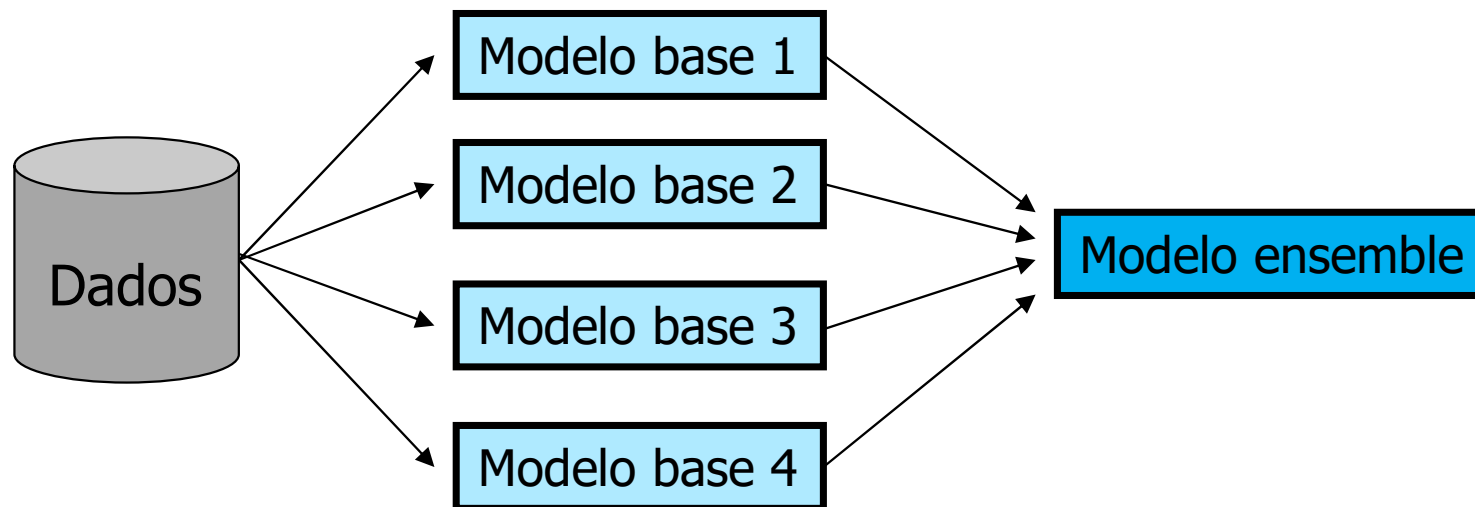


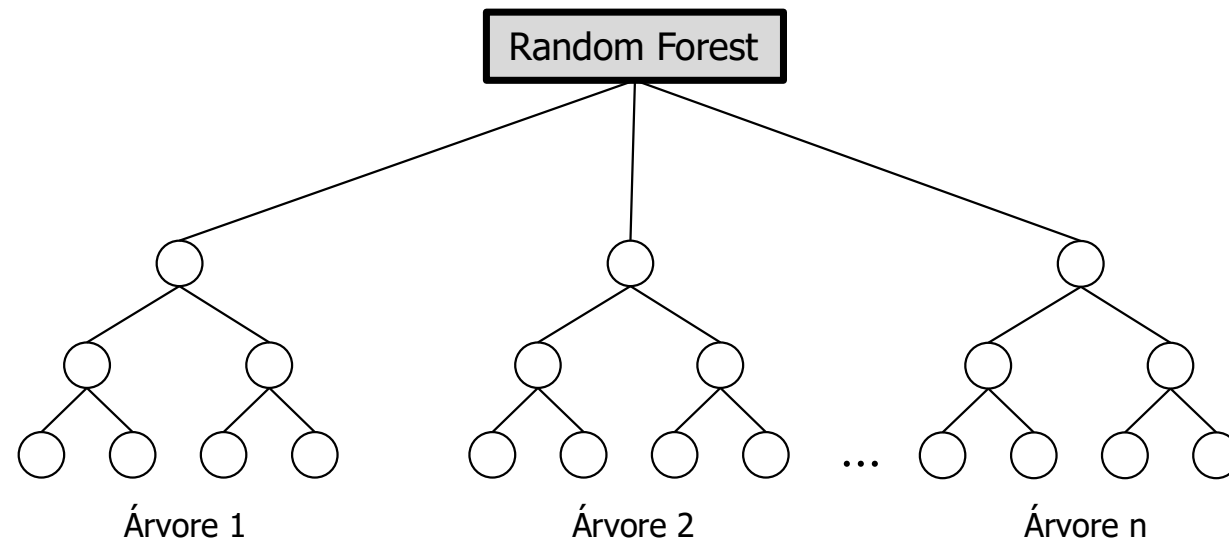
Métodos Ensemble (mistos)

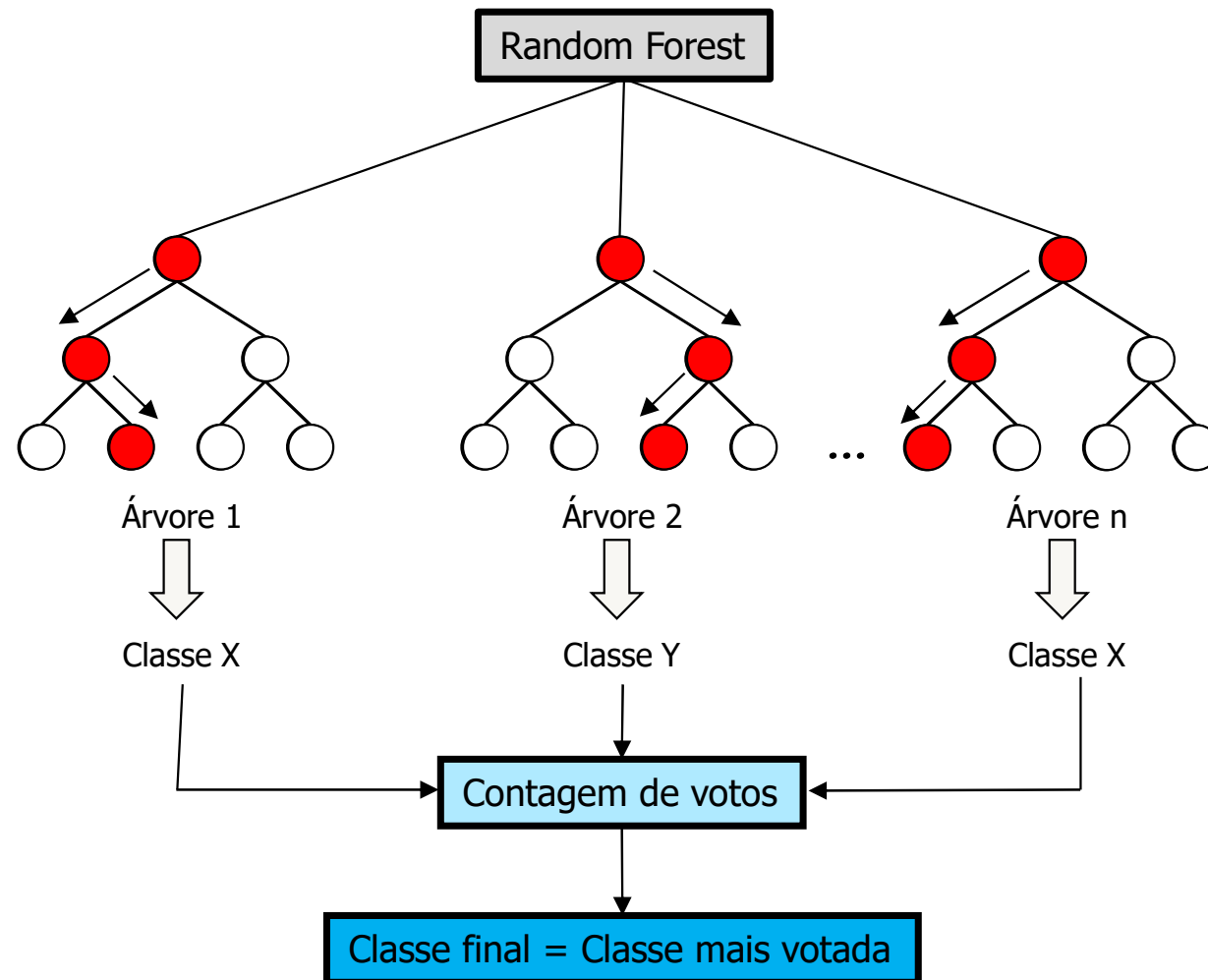
- ❑ São técnicas de aprendizado de máquina que combinam os resultados de um conjunto de modelos para obter um modelo preditivo mais eficiente.
- ❑ Esse conjunto pode ser composto por algoritmos distintos ou remodelações do mesmo algoritmo.
- ❑ O nome desses métodos tem origem na palavra francesa “ensemble” cujo significado pode ser conjunto ou conjuntamente, entre outros.
- ❑ Assim, o resultado da predição é obtido por meio da combinação de vários métodos conjuntos que utilizam o aprendizado coletivo de vários modelos.

Métodos Ensemble (mistos)

- Um classificador ensemble consiste em um conjunto de classificadores treinados individualmente, denominados classificadores base, cujos resultados são combinados de alguma forma.







Regressão Linear

Regressão Linear

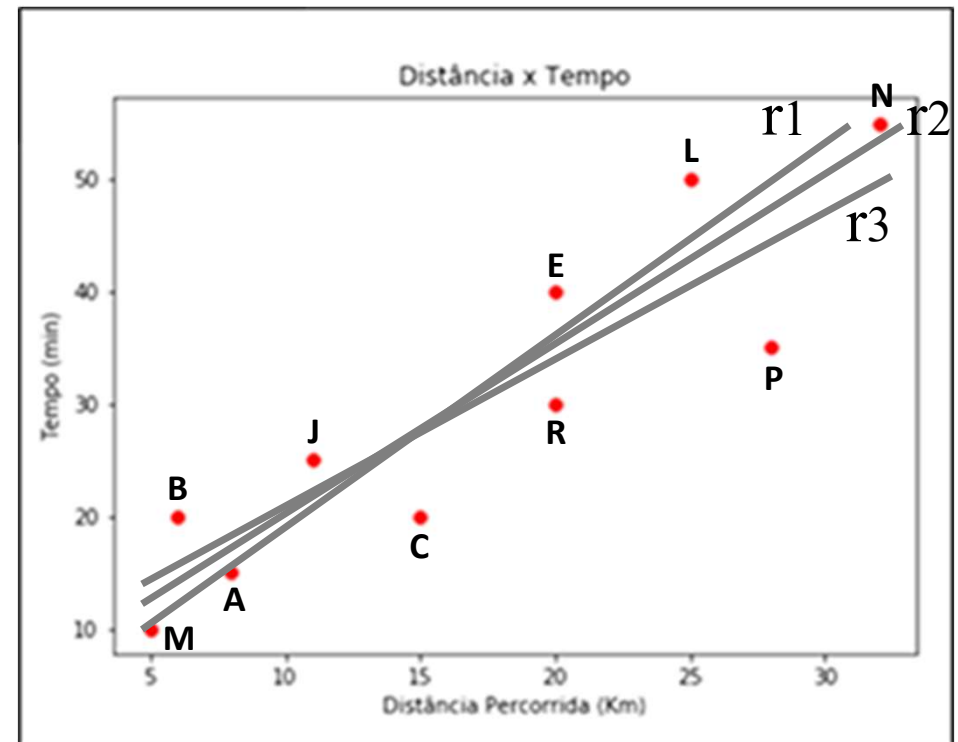
- ❑ A análise de regressão é um método de aprendizado supervisionado utilizado para calcular a relação entre duas ou mais variáveis, sendo uma delas dependente e as outras independentes.
- ❑ Com esse método, é possível criar um modelo que represente a relação entre as variáveis selecionadas e prever (estimar) valores com base no modelo.
- ❑ A regressão linear simples é um caso especial da regressão linear em que há apenas uma variável independente que influencia a variável dependente.

Regressão Linear Simples

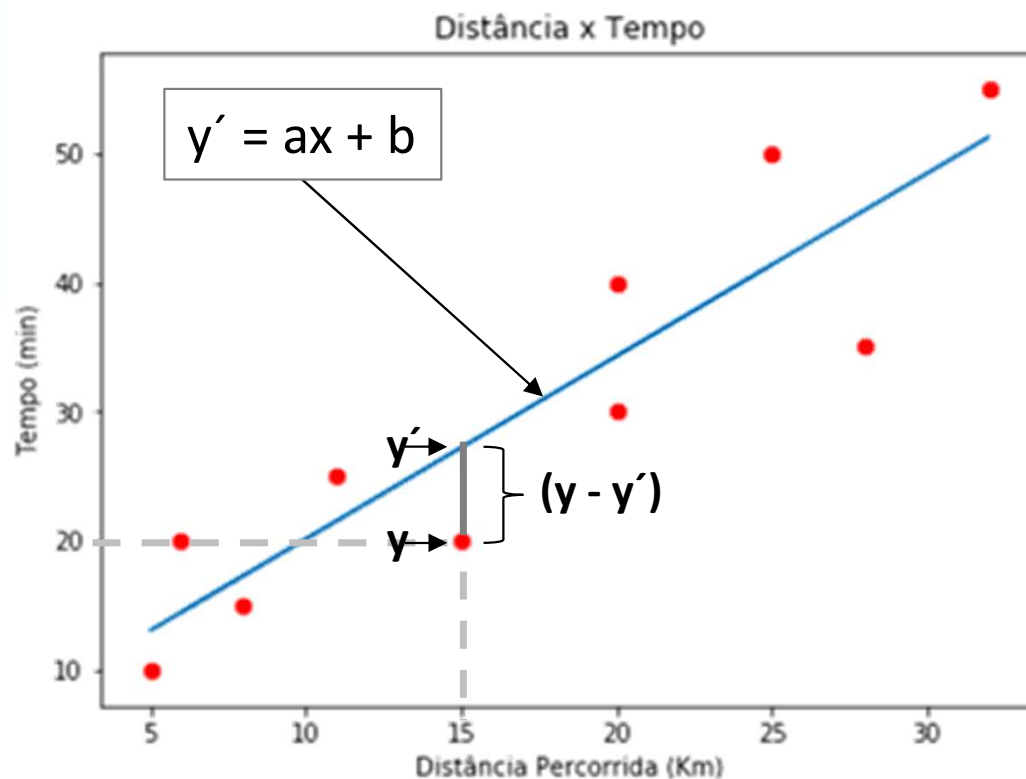
- ❑ A equação da regressão linear simples consiste em uma equação de primeiro grau, que pode ser retratada por uma reta em um espaço bidimensional.
- ❑ De forma simplificada, uma equação de primeiro grau pode ser caracterizada da seguinte forma: $y = ax + b$.
- ❑ Onde:
 - ❑ `y` = variável dependente;
 - ❑ `x` = variável independente;
 - ❑ `a` = o coeficiente de inclinação da reta;
 - ❑ `b` = intercepto no eixo y, ou seja, o valor de y quando x é igual a zero.

Regressão Linear Simples

Estudante	Distância (Km)	Tempo (min)
Ana	8	15
Beto	6	20
Caio	15	20
Eduardo	20	40
Lia	25	50
Júnior	11	25
Miguel	5	10
Nádia	32	55
Paula	28	35
Rita	20	30



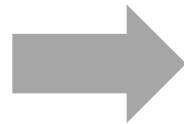
Regressão Linear Simples



- ❑ $\text{erro} = y - y'$;
- ❑ $y' = ax + b$;
- ❑ substituindo y' por $(ax + b)$;
- ❑ $(y - y') = (y - (ax + b))$;
- ❑ $(y - (ax + b)) = (y - ax - b)$;
- ❑ $\text{erro} = (y - ax - b)$.

Método dos Mínimos Quadrados

$$\sum_{i=1}^n (y_i - y'_i)^2$$



$$\sum_{i=1}^n (y_i - ax_i - b)^2$$

$$a = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b = \bar{y} - a\bar{x}$$

\bar{y} = média dos valores de y.
 \bar{x} = média dos valores de x.

Regressão Linear Simples

Estudante	Distância (Km)	Tempo (min)
Ana	8	15
Beto	6	20
Caio	15	20
Eduardo	20	40
Lia	25	50
Júnior	11	25
Miguel	5	10
Nádia	32	55
Paula	28	35
Rita	20	30

1	2	3	4	5	6	7
Estudante	Distância (x)	Tempo (y)	(x - media_x)	(x - x_media)^2	(y - media_y)	(x - media_x) * (y - media_y)
Ana	8	15	-9	81	-15	135
Beto	6	20	-11	121	-10	110
Caio	15	20	-2	4	-10	20
Eduardo	20	40	3	9	10	30
Lia	25	50	8	64	20	160
Júnior	11	25	-6	36	-5	30
Miguel	5	10	-12	144	-20	240
Nádia	32	55	15	225	25	375
Paula	28	35	11	121	5	55
Rita	20	30	3	9	0	0
Média →	17	30	Somatório →	814	Somatório →	1.155

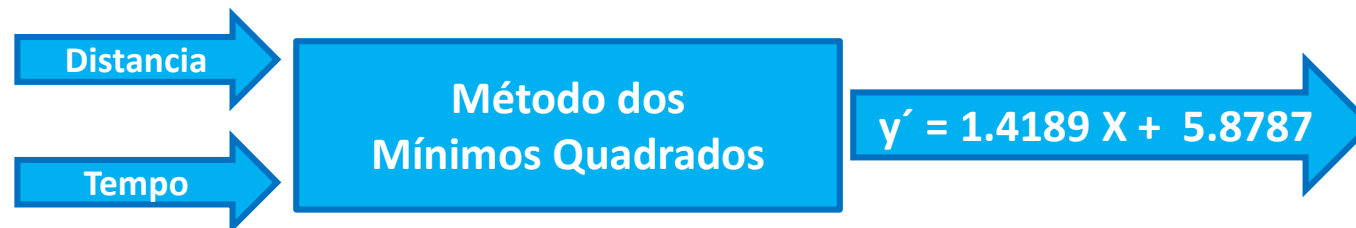
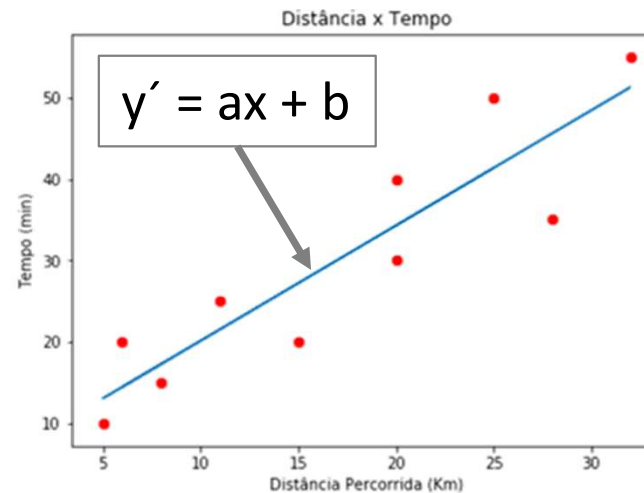
$$a = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b = \bar{y} - a\bar{x}$$

- $a = 1.155 / 814 = 1,4189$
- $b = 30 - (1,4189 \times 17) = 5,8787$
- $y' = 1,4189x + 5,8787$

Regressão Linear Simples

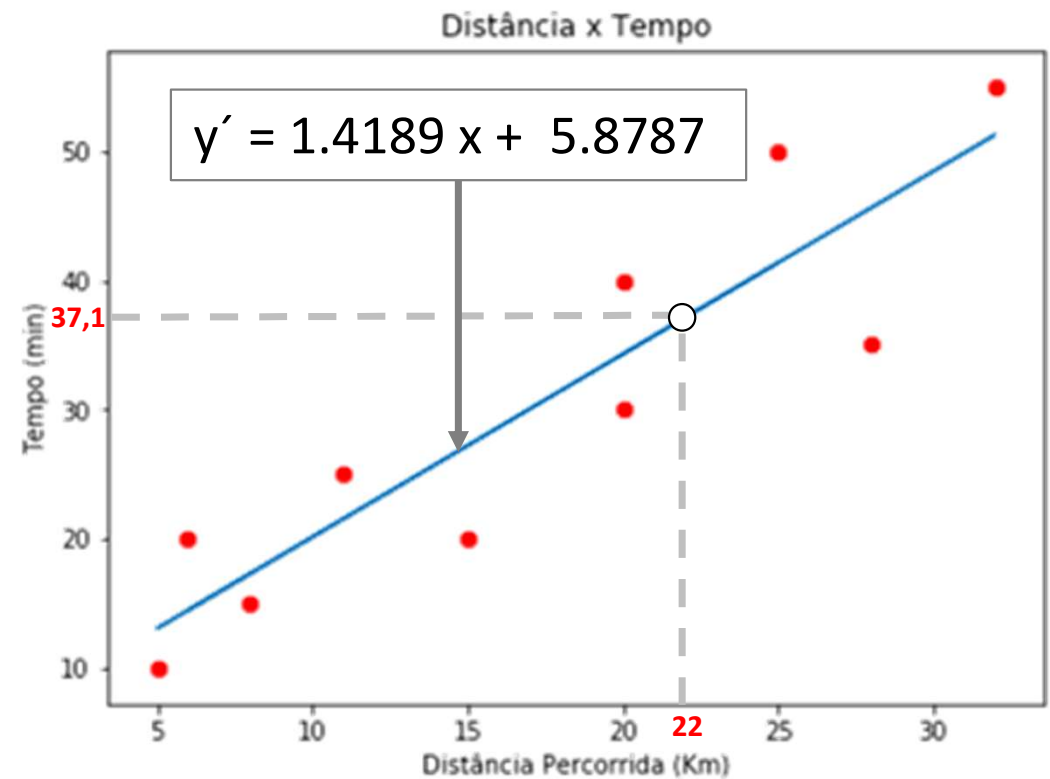
Estudante	Tempo	Distância
Gabriela	15	8
Dalila	20	6
Gustavo	20	15
Letícia	40	20
Luiz	50	25
Leonor	25	11
Ana	10	5
Antônio	55	32
Júlia	35	28
Mariana	30	20



Regressão Linear Simples

- Qual será o tempo necessário para um estudante percorrer 22 Km?

$$y' = 1,4189 (22) + 5,8787 = 37,1$$





Fim