

Overview

Abel Soares Siqueira

14/08/2018

Machine Learning

Machine Learning - Parte II

- Começa a segunda parte - ML de verdade;
- Veremos os algoritmos agora;
- Três partes:
 - Aprendizagem Supervisionada - Regressão;
 - Aprendizagem Supervisionada - Classificação;
 - Aprendizagem Não-supervisionada - Agrupamento;
- Vamos tentar estudar três coisas:
 - Teoria e desenvolvimento do algoritmo;
 - Implementação e/ou ideias de como aplicar o algoritmo;
 - Uso prático computacional, com uma implementação já completa.
- Como já vimos, temos que separar os dados em treinamento e teste, a partir daqui já assumimos que isso é feito;

Classes de Problemas

Classes de Problemas

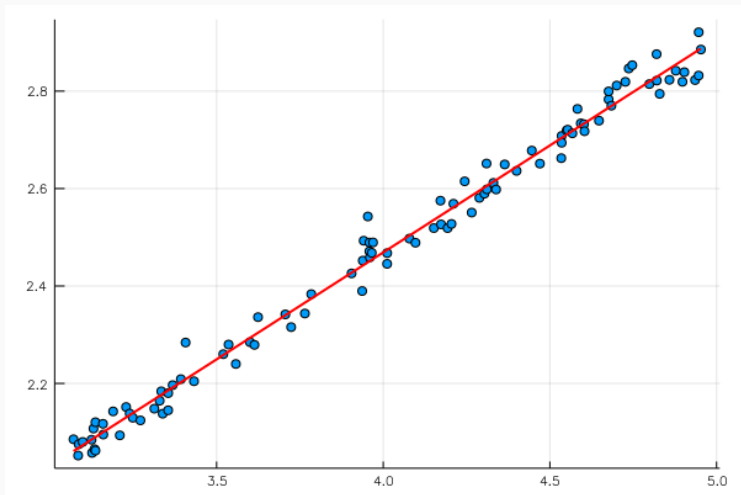
Aprendizagem supervisionada

- Temos um conjunto de dados $\{(x_i, y_i), i = 1, \dots, N\}$, $x_i \in X$ e $y_i \in Y$;
- Se os *targets* y_i forem valores numéricos contínuos, é um problema de regressão;
- Se os *targets* y_i forem discretos, ou categóricos, é um problema de classificação;

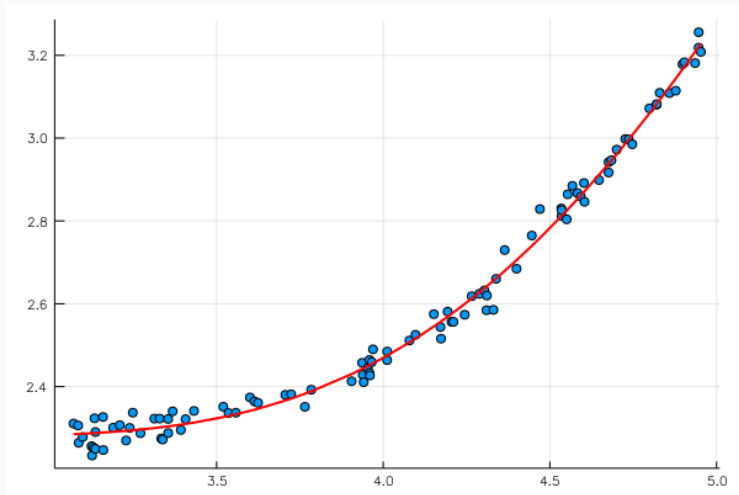
Aprendizagem Não-supervisionada

- Temos um conjunto de dados $x_i \in X$, sem target;
- Supõe-se haver uma classificação implícita desses dados, mas não temos exemplos;
- Na pior das hipóteses, não sabemos nem quantos grupos diferentes;

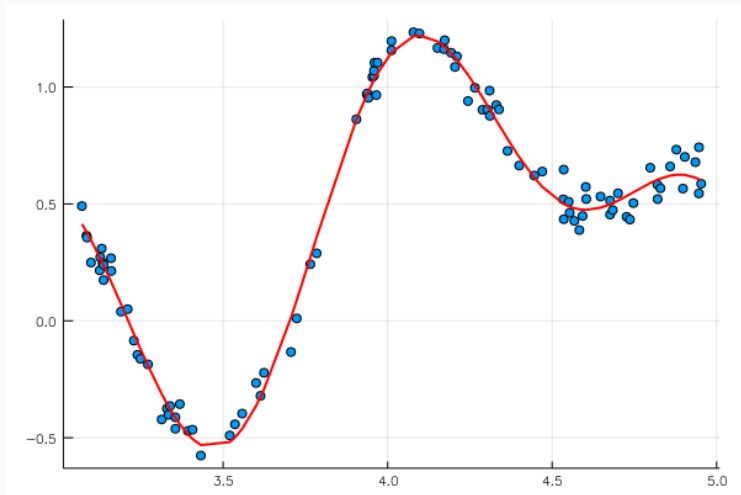
Aprendizagem supervisionada - Regressão



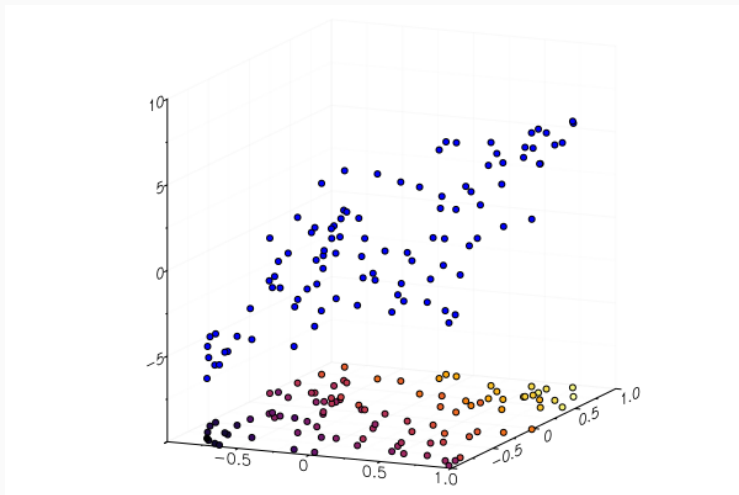
Aprendizagem supervisionada - Regressão



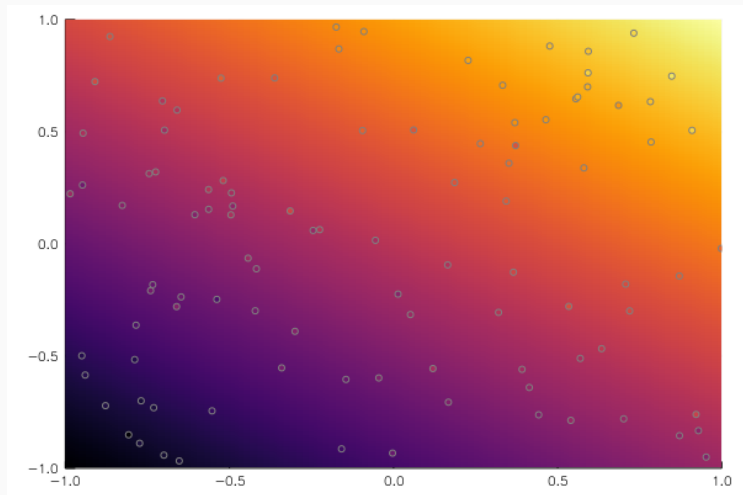
Aprendizagem supervisionada - Regressão



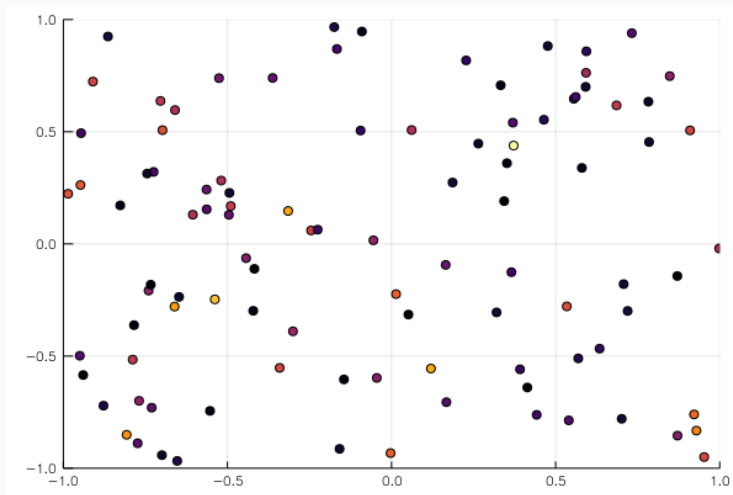
Aprendizagem supervisionada - Regressão



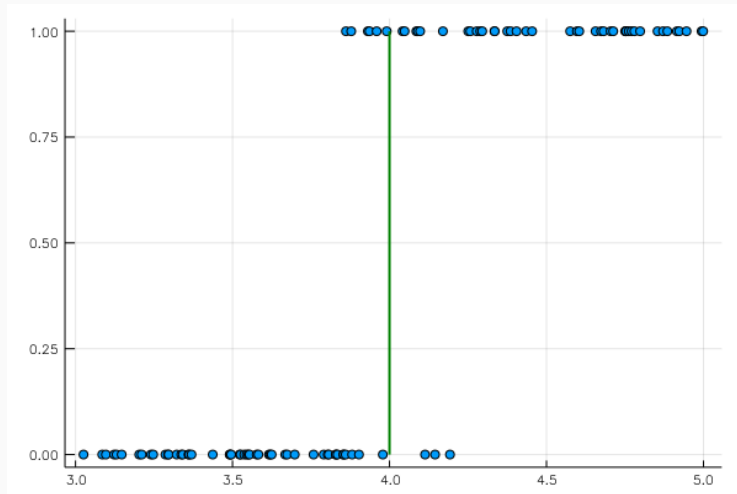
Aprendizagem supervisionada - Regressão



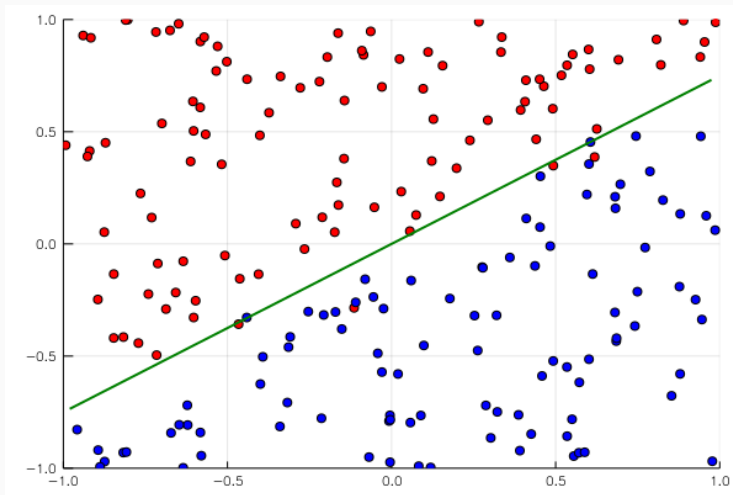
Aprendizagem supervisionada - Regressão



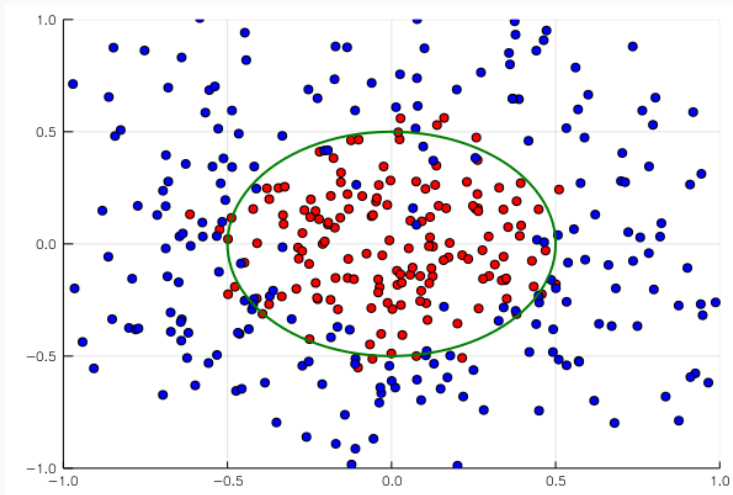
Aprendizagem supervisionada - Classificação



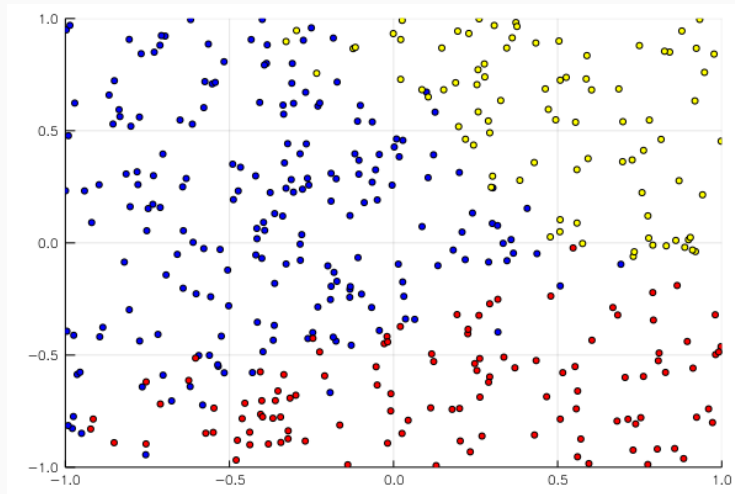
Aprendizagem supervisionada - Classificação



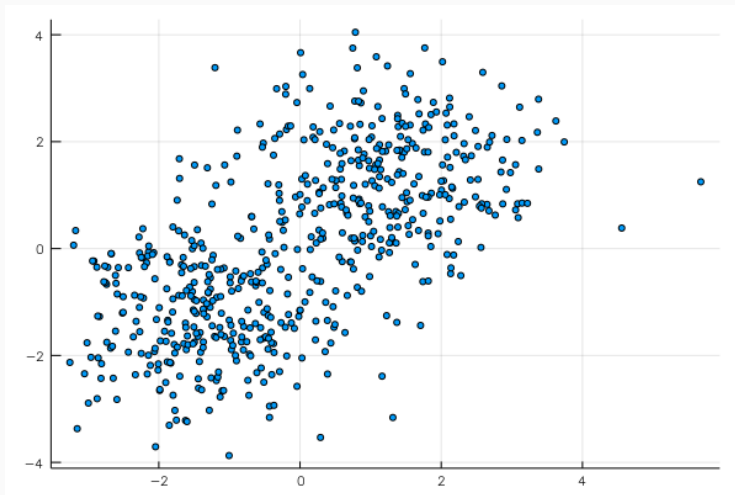
Aprendizagem supervisionada - Classificação



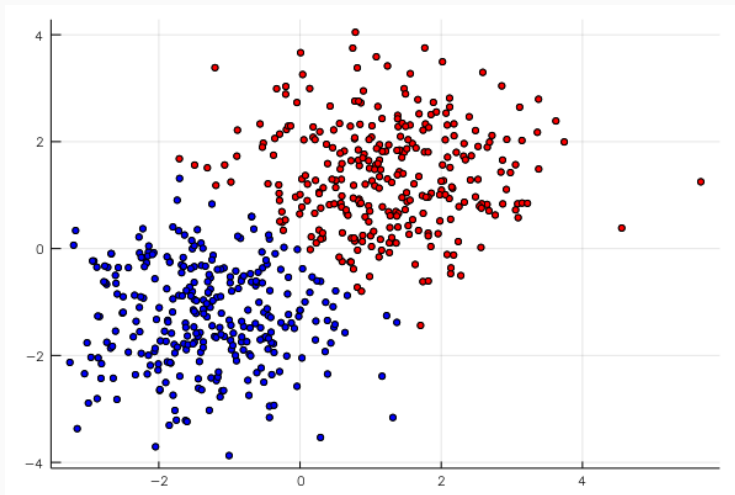
Aprendizagem supervisionada - Classificação



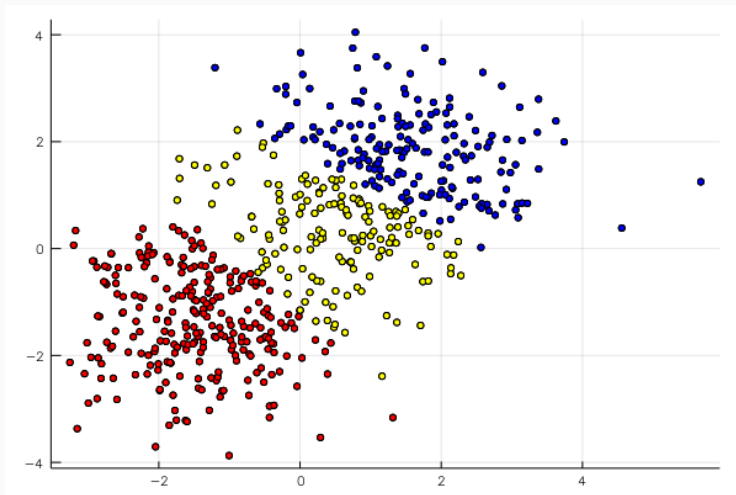
Aprendizagem não-supervisionada - Agrupamento



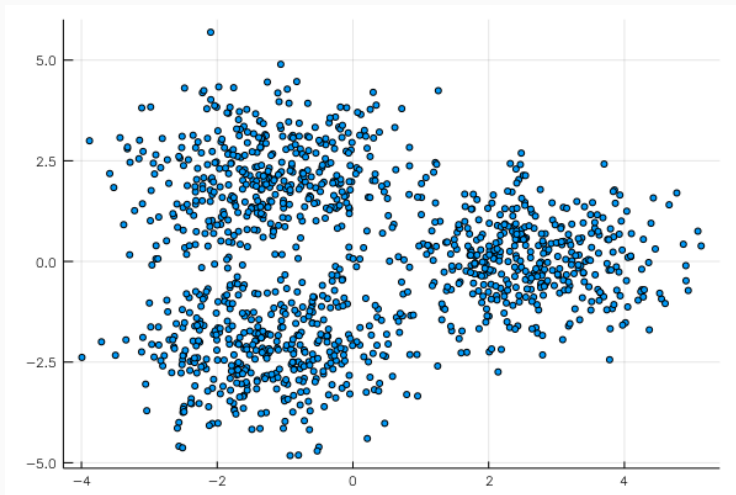
Aprendizagem não-supervisionada - Agrupamento



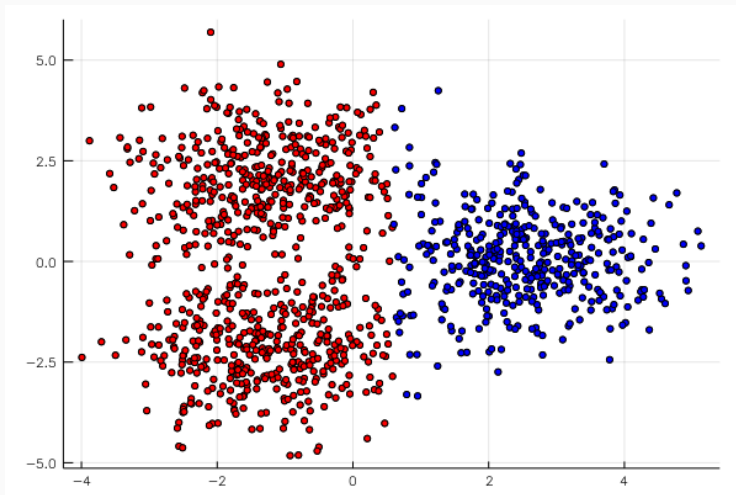
Aprendizagem não-supervisionada - Agrupamento



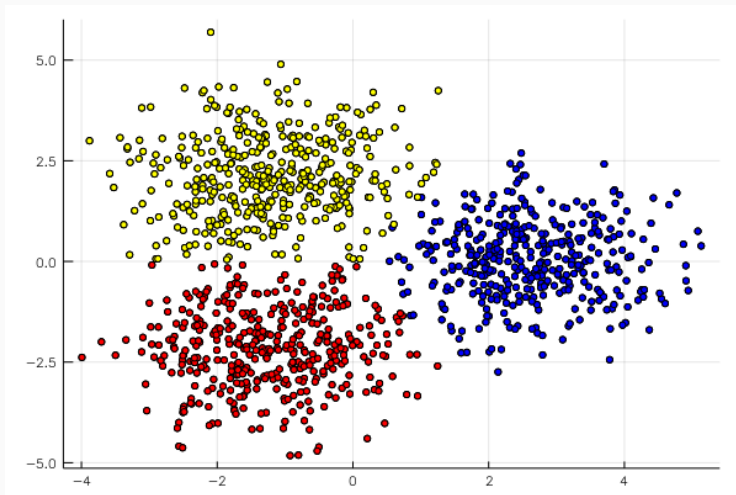
Aprendizagem não-supervisionada - Agrupamento



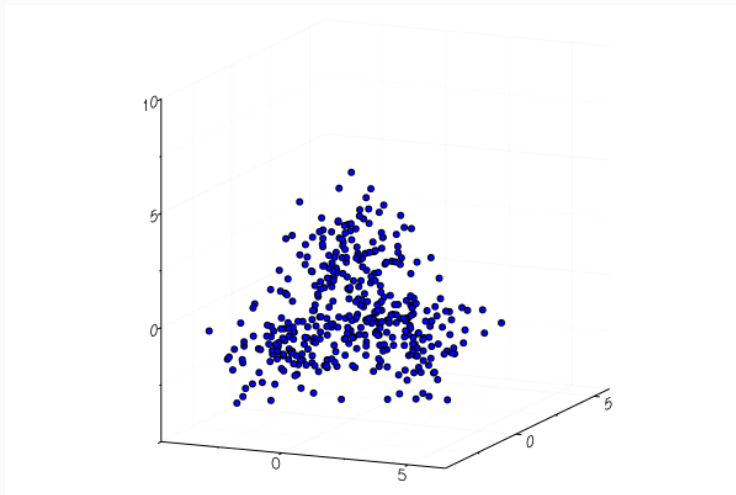
Aprendizagem não-supervisionada - Agrupamento



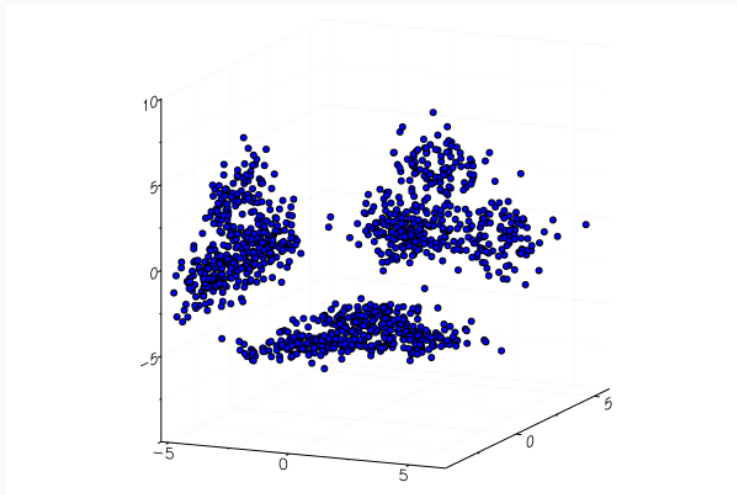
Aprendizagem não-supervisionada - Agrupamento



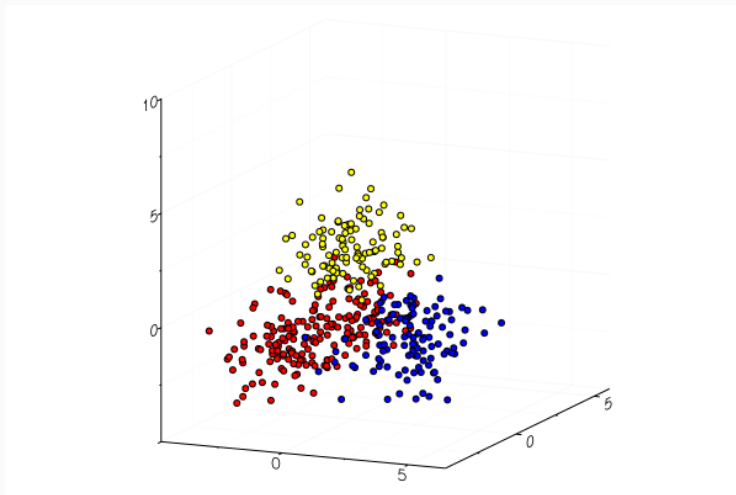
Aprendizagem não-supervisionada - Agrupamento



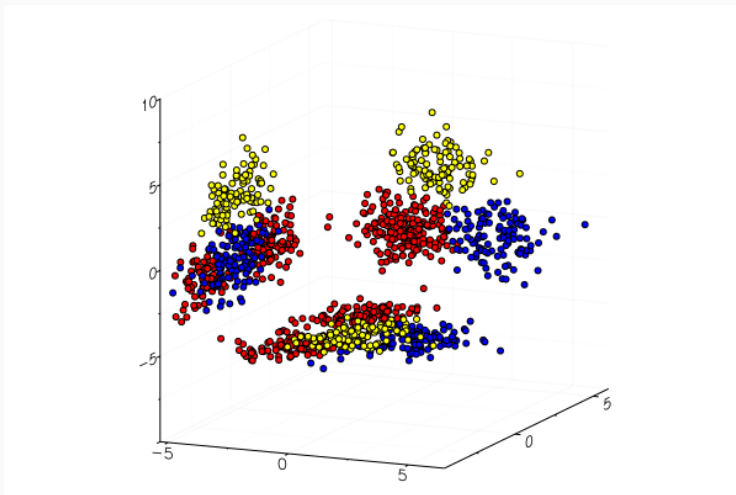
Aprendizagem não-supervisionada - Agrupamento



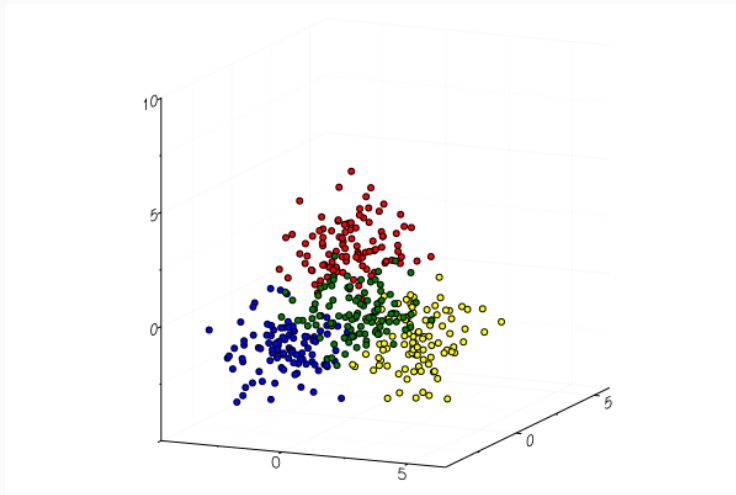
Aprendizagem não-supervisionada - Agrupamento



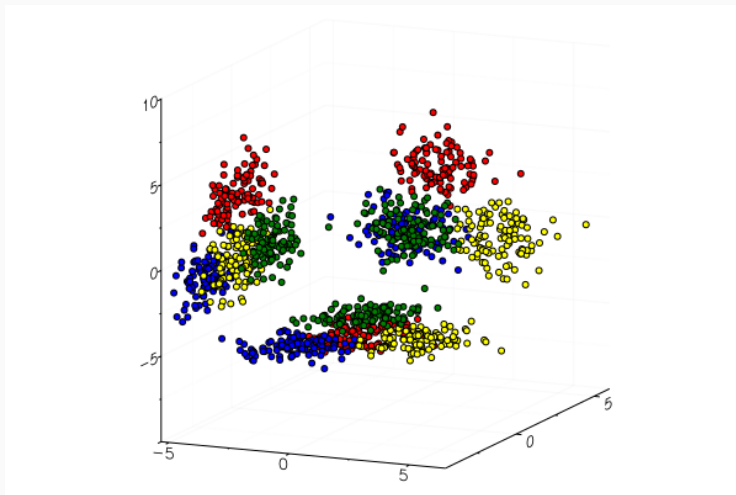
Aprendizagem não-supervisionada - Agrupamento



Aprendizagem não-supervisionada - Agrupamento



Aprendizagem não-supervisionada - Agrupamento



Algoritmos

Regressão

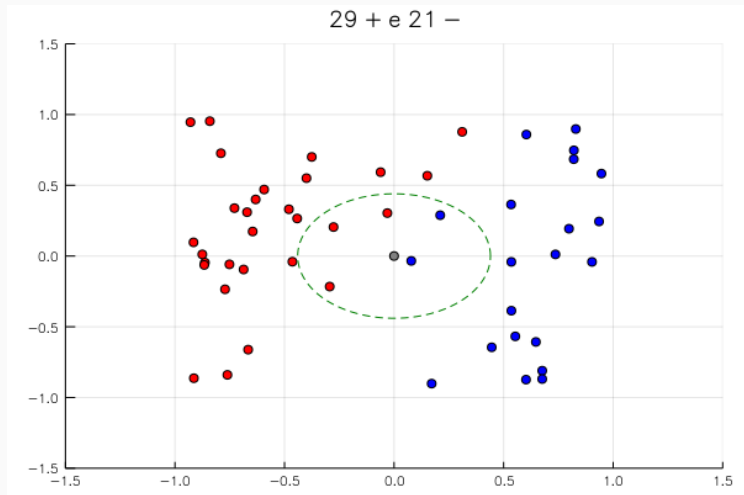
- Queremos encontrar um modelo $h_{\theta}(x)$ tal que $h_{\theta}(x_i)$ aproxime bem y_i ;
- Utilizamos uma função de perda, $\ell(h_{\theta}(x_i), y_i)$ para medir o erro;
- Minimizamos a média das perdas

$$L(\theta) = \sum_{i=1}^N \ell(h_{\theta}(x_i), y_i);$$

- Usualmente o modelo é linear em θ , mas não necessariamente em x ;
- Exemplo: $h_{\theta}(x) = \theta_0 + \theta_1 x$, $\ell(\hat{y}, y) = \frac{1}{2}(y - \hat{y})^2$,

$$L(\theta) = \frac{1}{2} \sum_{i=1}^m [y_i - h_{\theta}(x_i)]^2.$$

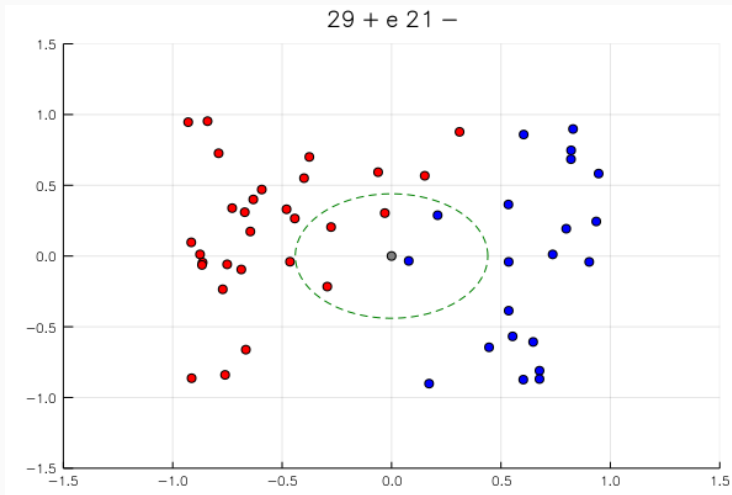
Classificação - Naive Bayes



Classificação - Naive Bayes

- Duas classes: + e -;
- Novo ponto X , e uma região de avaliação local;
- Quero $P(+|X)$ prob. de ser +;
- Teorema de Bayes: $P(+|X) = \frac{P(X|+)P(+)}{P(X)}$;
- $P(+)$ é calculado usando todos os pontos;
- $P(X|+)$ é a probabilidade de estar na região e dado a classe +;

Classificação - Naive Bayes



- $P(+) = 29/50;$
- $P(-) = 21/50;$
- $P(X|+) = 3/29;$
- $P(X|-) = 2/21;$

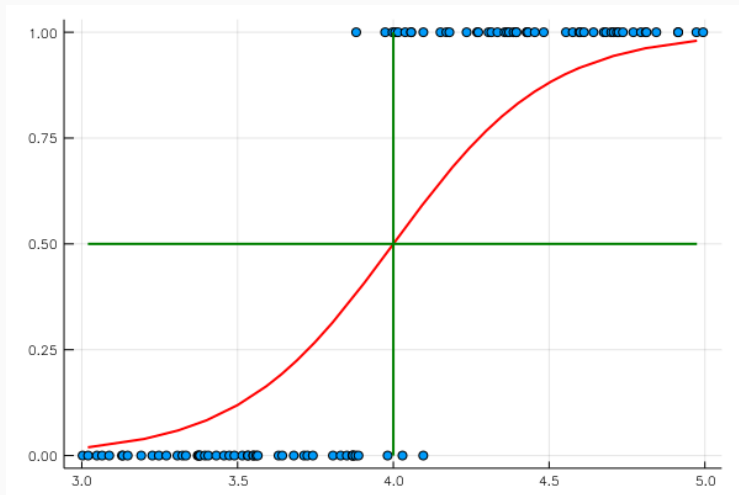
Classificação - Naive Bayes

- $P(+|X) = \frac{\frac{29}{50} \frac{3}{29}}{\frac{P(X)}{P(X)}} = \frac{3}{50P(x)};$
- $P(-|X) = \frac{\frac{21}{50} \frac{2}{21}}{\frac{P(X)}{P(X)}} = \frac{2}{50P(x)};$
- $P(+|X) > P(-|X)$ então X será da classe $+$.

Classificação - Logística

- $y_i \in \{0, 1\}$;
- Hipótese: y é Bernoulli, $P(y = 1) = p = 1 - P(y = 0)$; logo, $P(y = k) = p^k(1 - p)^{1-k}$, $k \in \{0, 1\}$;
- Fazemos um modelo $h_\theta(x) = \sigma(\theta^T x)$ onde $\sigma(t) = \frac{1}{1 + e^{-t}}$;
- Nosso classificador é
$$\begin{cases} 1, & h_\theta(x) > \frac{1}{2}, \\ 0, & h_\theta(x) < \frac{1}{2}. \end{cases}$$
- Nossa perda: $\ell(h_\theta(x), y) = -\log \left[h_\theta(x)^y (1 - h_\theta(x))^{1-y} \right]$;

Classificação - Logística

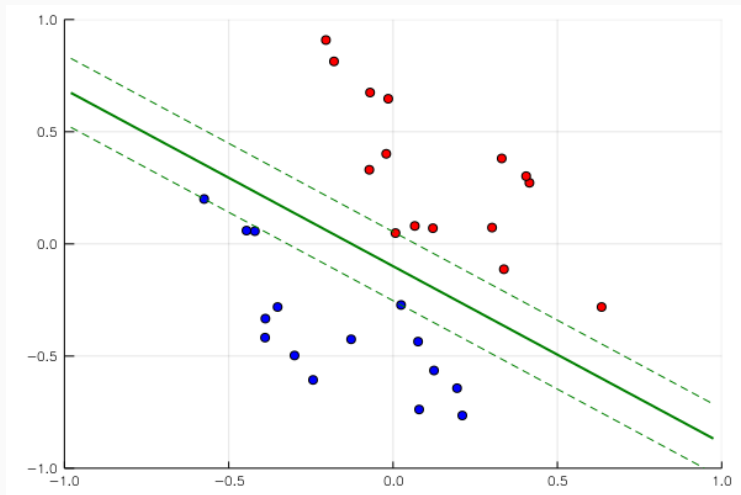


Classificação - Máquina de Vetor Suporte

- $y_i \in \{-1, 1\}$;
- $b + w^T x = 0$ define um hiperplano;
- Queremos $y_i(b + w^T x_i) \geq 1$ para cada i ;
- A margem entre $b + w^T x = 1$ e $b + w^T x = -1$ é $\frac{2}{\|w\|}$;
- Para maximizar essa margem podemos minimizar $\|w\|$;
- Otimização:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 \\ & y_i(b + w^T x_i) \geq 1, \quad \forall i = 1, \dots, m. \end{aligned}$$

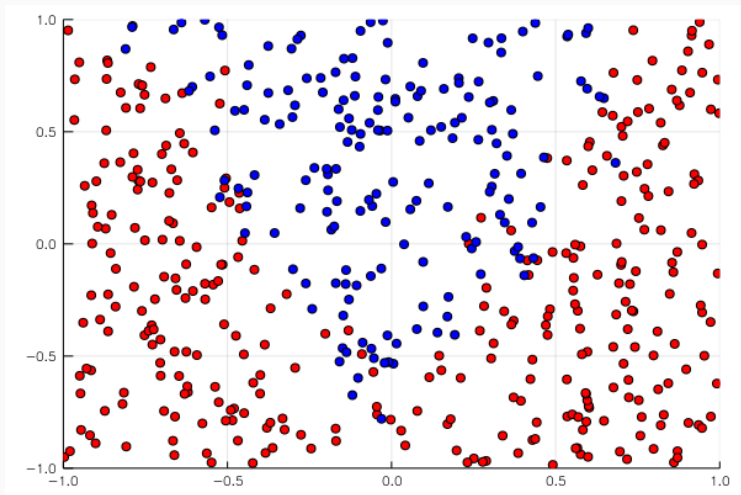
Classificação - Máquina de Vetor Suporte



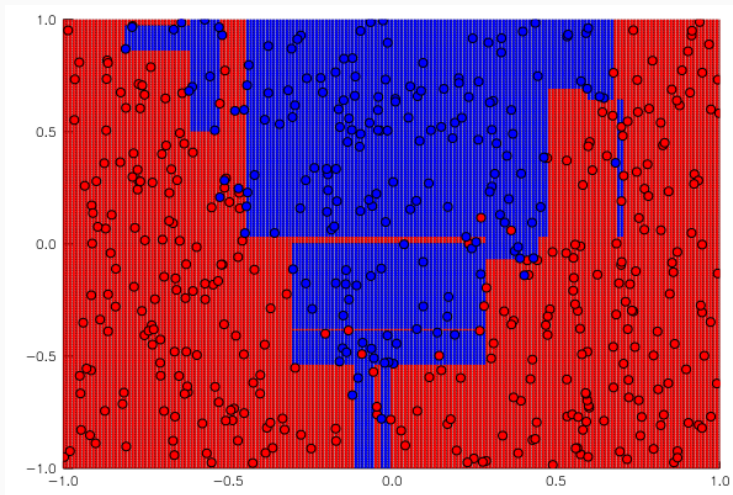
Classificação - Árvore de decisão

- Uma árvore de decisão é uma sequência de escolhas;
- A cada passo de escolha separamos uma característica em duas partes;
- A árvore essencialmente divide o espaço em retângulos;

Classificação - Árvore de decisão



Classificação - Árvore de decisão



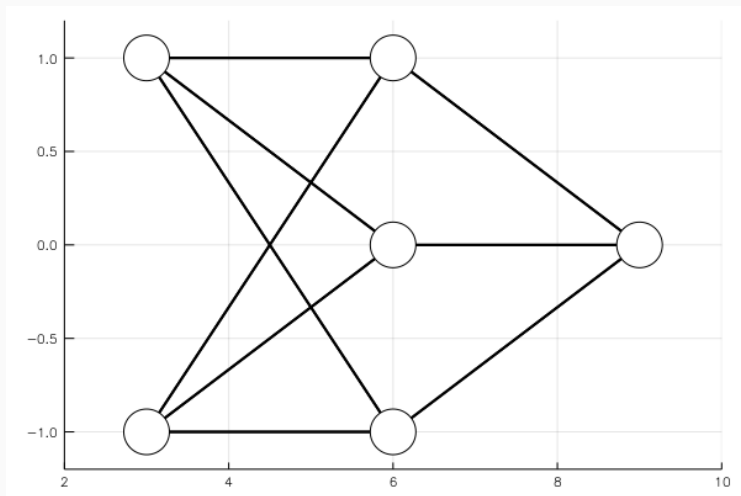
Supervisionado - Redes Neurais

- A rede é feita de camadas, a primeira sendo a entrada x_i e a última a saída y_i ;
- As outras camadas são ditas escondidas;
- Entre cada camada existem ligações de cada nó, significando multiplicação por um parâmetro;
- Em cada nó também podemos aplicar uma função de ativação;
- Existem discussões, mas muito de RN pode ser visto simplesmente como um modelo:

$$h_{\theta}(x) = \sigma_2(\Theta_2 \sigma_1(\Theta_1 x)),$$

onde Θ_j são matrizes de parâmetros θ e σ_1 são as funções de ativação dessa camada;

Supervisionado - Redes Neurais



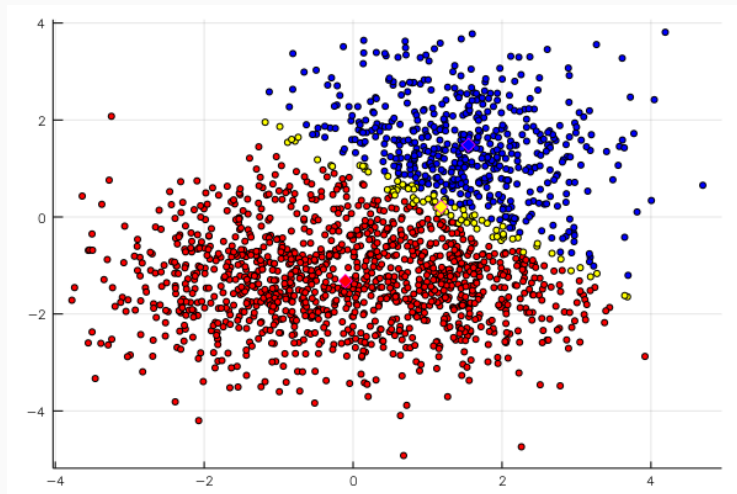
Supervisionado - Redes Neurais

- Primeira: x_1, x_2 ;
- Interna: $z_1 = \sigma_{11}(\theta_1 x_1 + \theta_2 x_2)$ $z_2 = \sigma_{12}(\theta_3 x_1 + \theta_4 x_2)$ $z_3 = \sigma_{13}(\theta_5 x_1 + \theta_6 x_2)$;
- Final: $\hat{y} = \sigma_2(\theta_7 z_1 + \theta_8 z_2 + \theta_9 z_3)$;
- Num modelo para duas variáveis de entrada e uma de saída usamos 9 variáveis (ignoramos o viés);

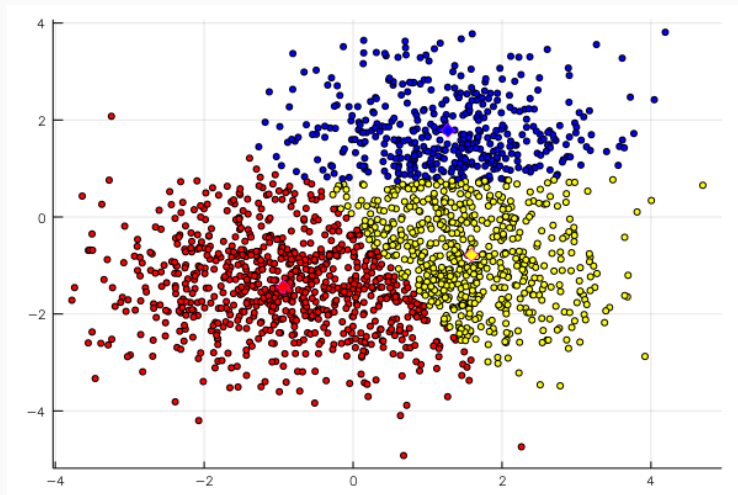
Agrupamento - K-Médias

- N pontos são escolhidos como centros;
- Cada ponto é classificado pela proximidade a um centro;
- Os centros são atualizados como o ponto médio de cada região;

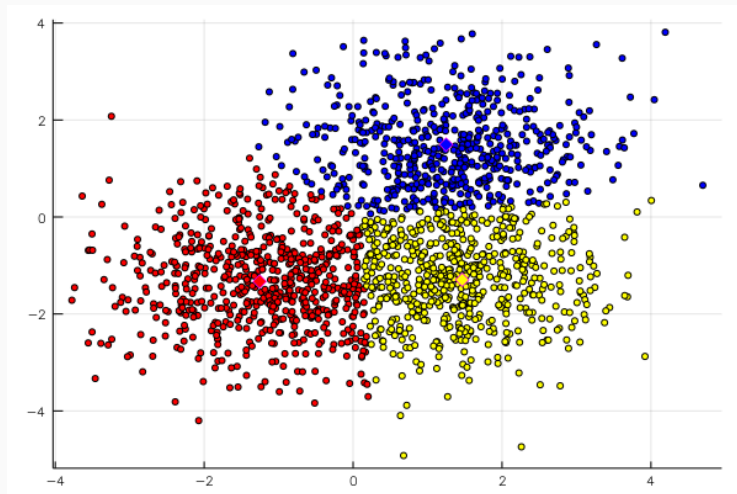
Agrupamento - K-Médias



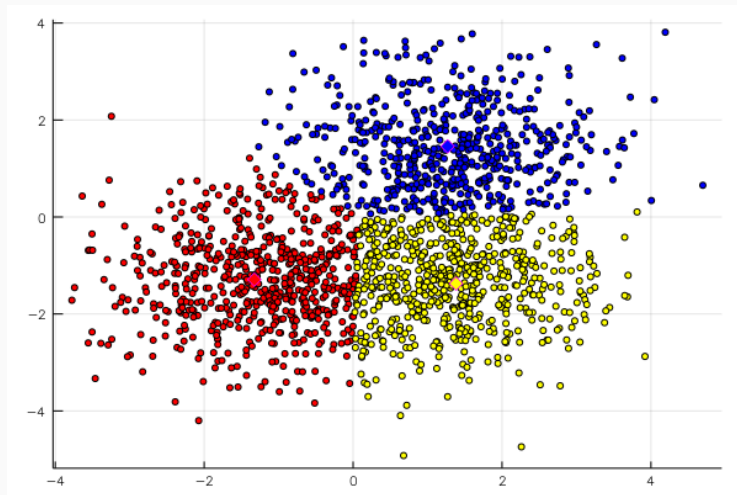
Agrupamento - K-Médias



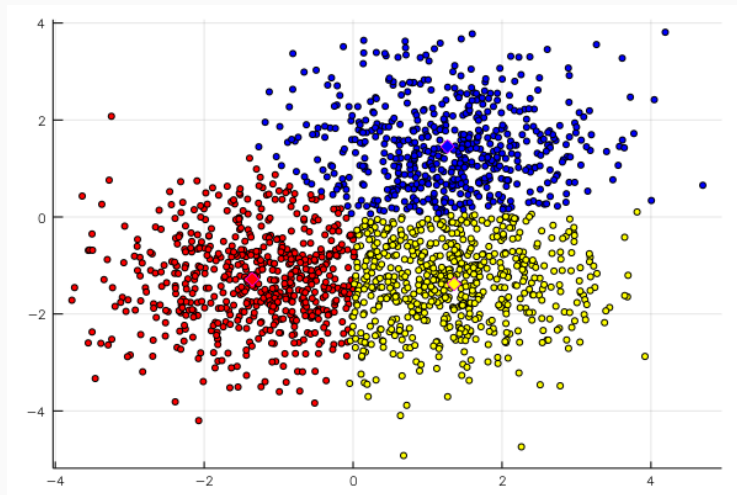
Agrupamento - K-Médias



Agrupamento - K-Médias



Agrupamento - K-Médias



Lista de Algoritmos

Regressão	Classificação	Agrupamento
Reg. Mul.	Logística	Hierárquicos
	Naive Bayes	K-médias
Redes Neurais	Redes Neurais	K-medóides
SVMr	SVMc	DBSCAN
Árvores Reg.	Árvores Class.	Red. Dim.
KNN	KNN	Anal. Fatorial
	A. Discrimante	ACP

FIM