Support Vector Machines

Mineração de Dados

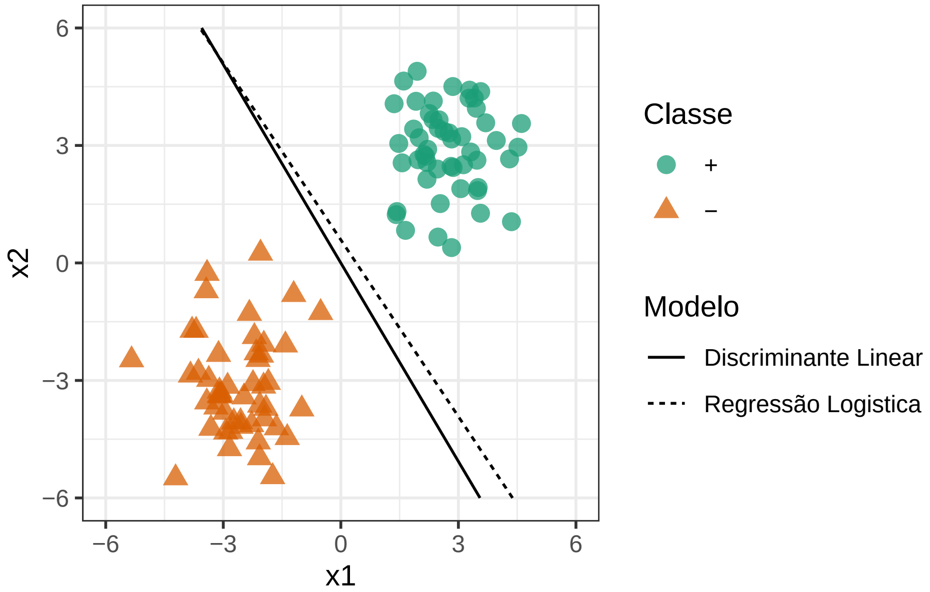
Ronaldo C. Prati[1]

[1]Universidade Federal do ABC (UFABC), ronaldo.prati@ufabc.edu.br

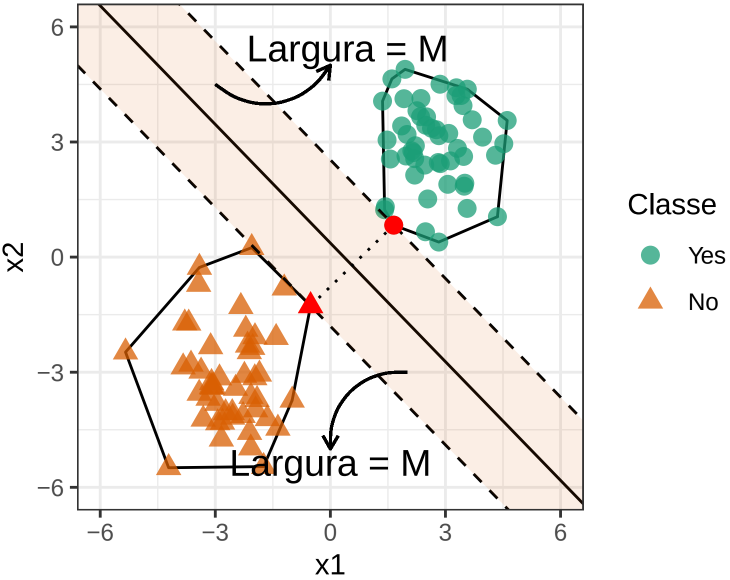
SVM's

- Máquinas de vetores de suporte são uma outra família de algoritmos de aprendizado de máquina supervisionado
  - Na forma mais simples, é um classificador linear
  - Busca maximizar a margem de separação entre as classes
  - O uso de funções *kernel* pode extender para classificadores não lineares

Classificadores Lineares



SVM



Função de custo (Regressão Logística)

Seja $z = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_m x_m = b_0 + \mathbf{b}^\intercal \mathbf{x}$ a reta que define a fronteira de decisão da regressão logística

A função de custo é:

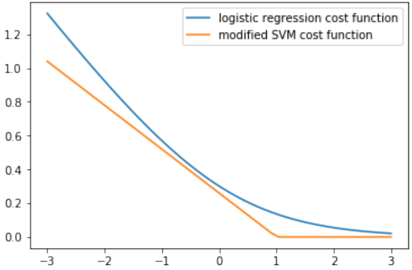
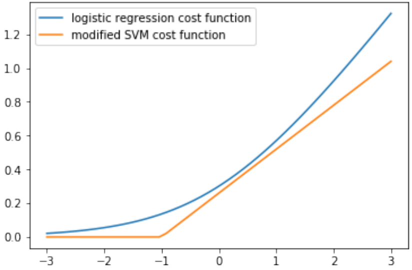
$$Custo(y, \hat{y}) = \begin{cases} -\log\left(\frac{1}{1+e^z}\right), \text{se } y = 1 \\ -\log\left(1 - \frac{1}{1+e^z}\right), \text{se } y = 0 \end{cases}$$

Função de custo (SVM)

Nas SMVs, função de custo é:

$$Custo(y, \hat{y}) = \begin{cases} \max(0, 1 - z), \text{se } y = 1 \\ \max(0, 1 + z), \text{se } y = 0 \end{cases}$$

Comparando os custos

$y = 1$	$y = 0$
	

Comparando os custos

- Na regressão logística queremos:
  - se $y = 1$, $\hat{y} \approx 1$ e $z > 0$
  - se $y = 0$, $\hat{y} \approx 0$ e $z < 0$
- SVM também é conhecida como **classificador de margem larga**, queremos:
  - se $y = 1$, $z \geq 1$ (não apenas > 0)
  - se $y = 0$, $z \leq -1$ (não apenas < 0)

Minimizando o custo

Para encontrar o hiperplano, queremos que

$$\sum_{i=1}^{n} y \times custo(y, \hat{y}) + (1 - y) \times custo(y, \hat{y})$$

seja próximo a zero. Isso é equivalente a

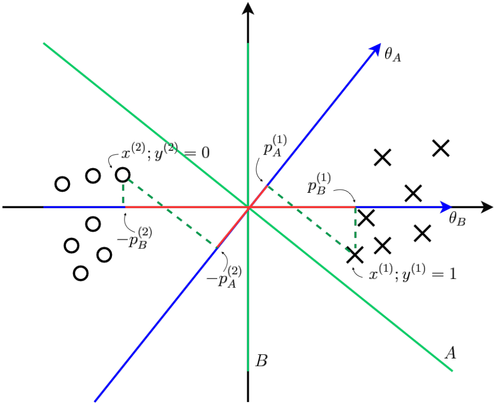
$$\min_{\mathbf{b}} \sum_{j=1}^{m} b_j$$

sujeito a:

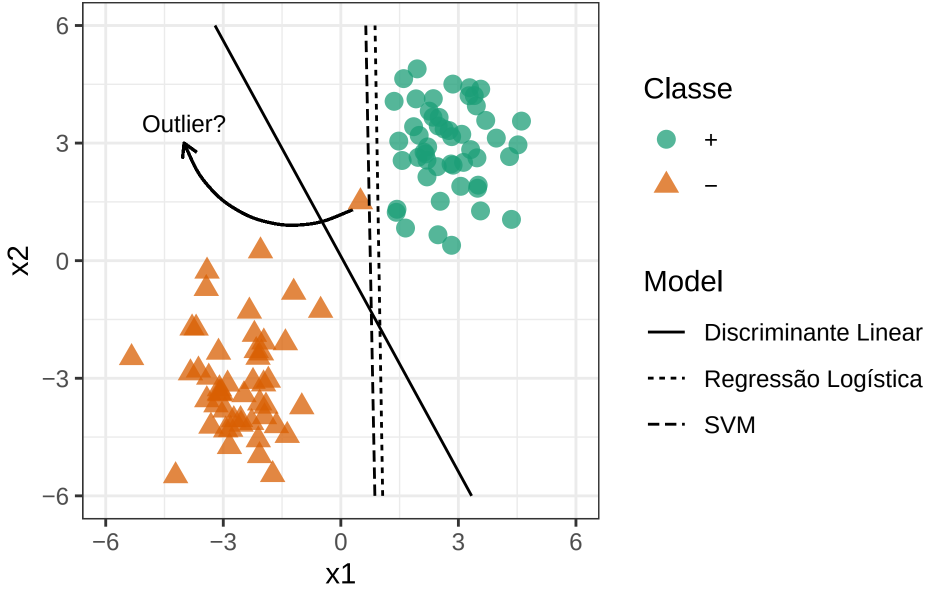
$$\mathbf{b}^\mathsf{T}\mathbf{x} \geq 1 \text{ se } y = 1, \mathbf{b}^\mathsf{T}\mathbf{x} \leq 1 \text{ se } y = 0$$

Maximização da margem

- **b** é ortogonal a **bx** (geometria analítica)
- é possível mostrar que $\min_{\mathbf{b}} \sum_{j=1}^{m} b_j$ corresponde a maximizar a margem



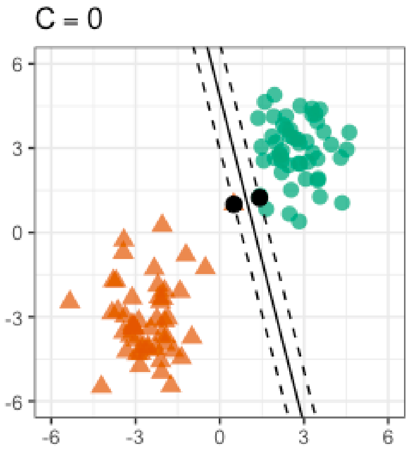
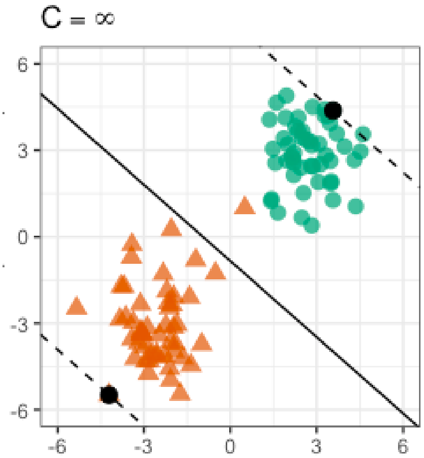
Outliers


Outlier?
x2
x1
6
3
0
−3
−6
−6
−3
0
3
6
Classe
+
−
Model
Discriminante Linear
Regressão Logística
SVM

C

- Para lidar com possível outliers, podemos usar o conceito de margem "*soft*"
- De maneira simplista, corresponde a admitir uma pequena violação na margem, de maneira que alguns exemplos possam estar dentro dela
- É um hiperparâmetro do algoritmo

C

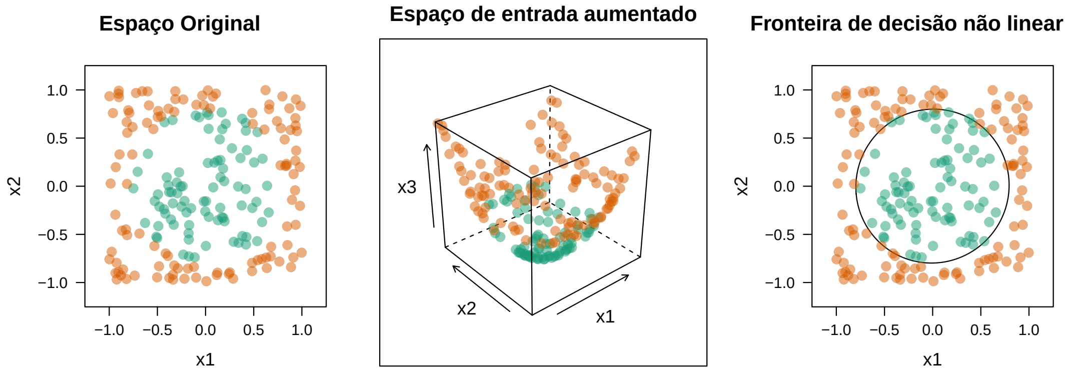
$C = 0$	$C = \infty$
	

Kernel

- O **kernel** é útil quando queremos computar uma fronteira de decisão não linear
- A ideia consiste em calcular um novo espaço de atributos
- Kernels são uma outra maneira de fazer isso

Kernel

- O *kernel gaussiano* usa uma função de similarided baseada na função gaussiana.



SVM com kernel gaussiano

- Parâmetro C da SVM
  - C grande leva a uma hipótese com baixo bias/alta variância: overfitting
  - C pequeno leva a uma hipótese com alto bias/baixa variância: underfitting
- Parâmetro σ^2 do Kernel Gaussiano
  - σ^2 grande - os atributos variam mais suavemente - alto bias, baixa variância
  - σ^2 pequeno - os atributos variam mais abruptamente - baixo bias, alta variância

Outros Kernels

- Linear (não usar kernel)
- Polinomial (parecido com a ideia de usar polinômios de maior grau)
- String
- X^2
- Intersecção de histograma

Cada um tem um conjunto própio de parâmetros

Fronteira de decisão