

Conceitos e aplicações da aprendizagem de máquina

Eduardo Krempser
eduardo.krempser@fiocruz.br

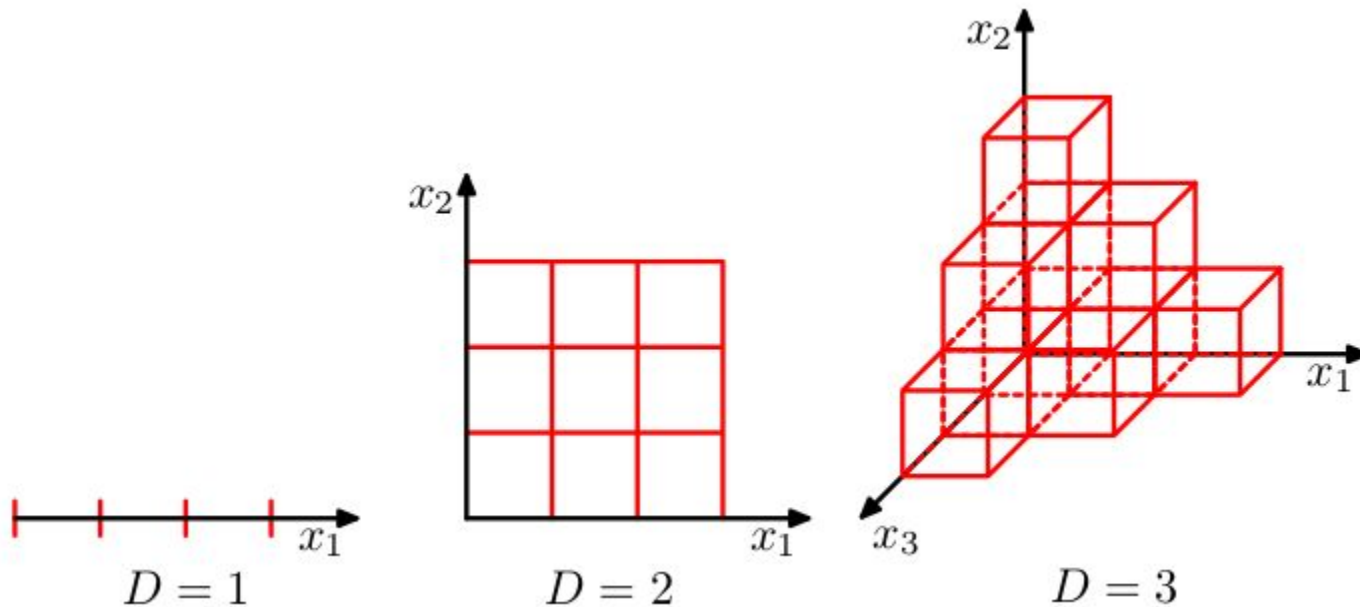
Matheus Müller
matheusp@posgrad.Incc.br

Fontes de dados

Características dos Dados

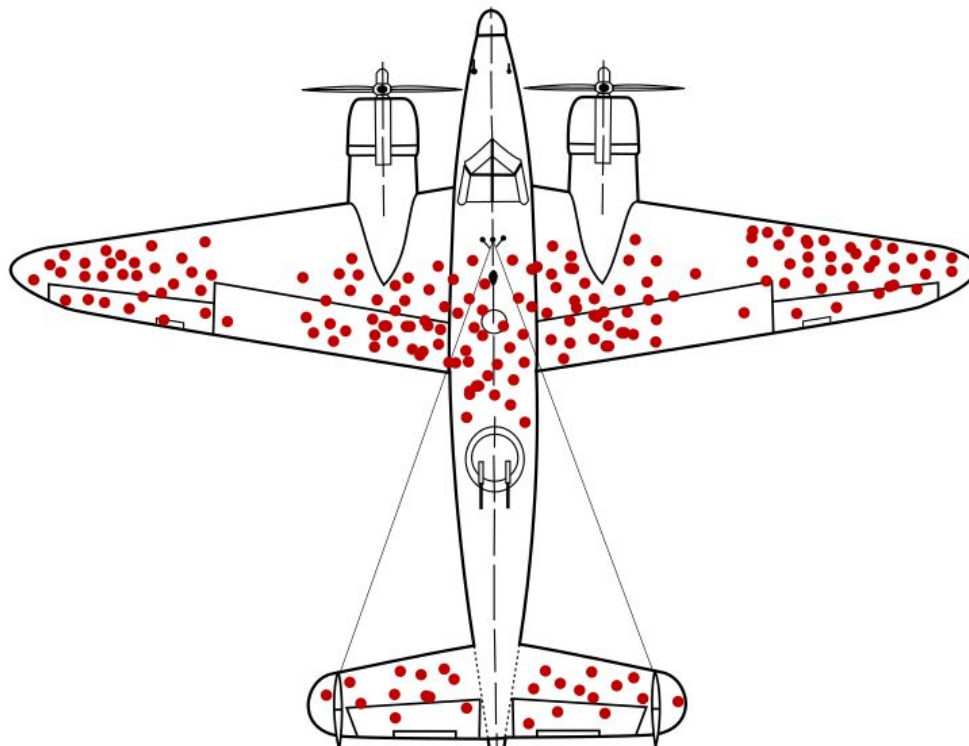
- Dimensionalidade (número de atributos)
 - Alta dimensionalidade geram uma série de desafios
- Esparsidade
 - Dados de presença, por exemplo
- Resolução
 - Padrões dependem da escala
- Tamanho
 - Tipos de análises podem depender do tamanho do conjunto de dados

“Maldição” da Dimensionalidade



Pattern Recognition and Machine Learning. Bishop, C. M.; 2006

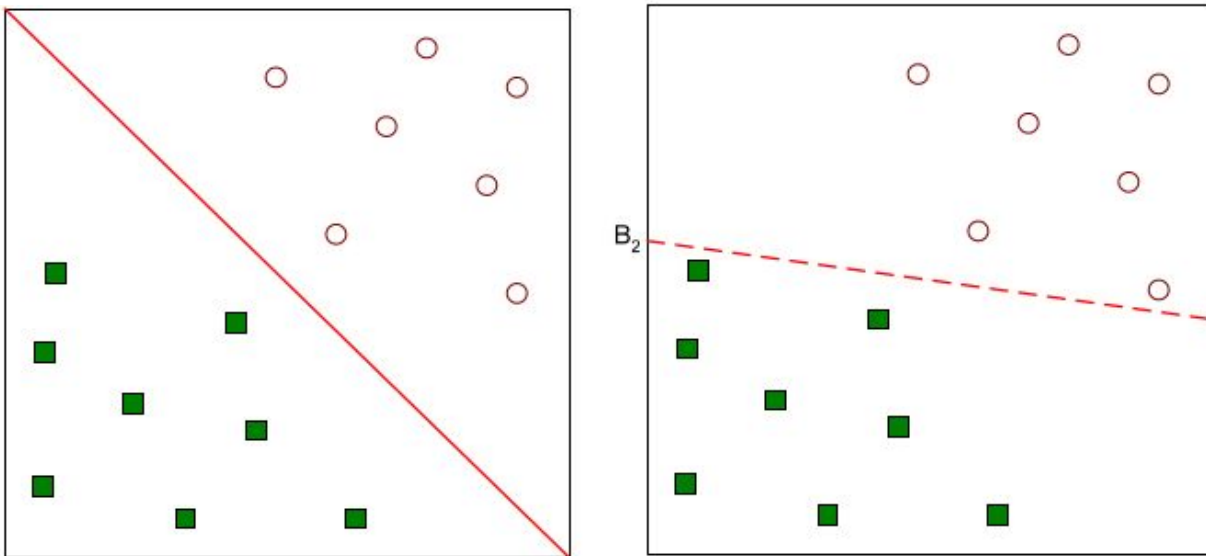
Viés do Sobrevivente



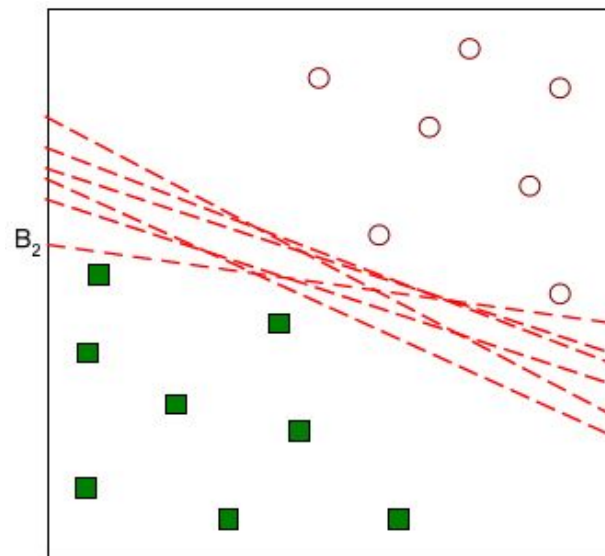
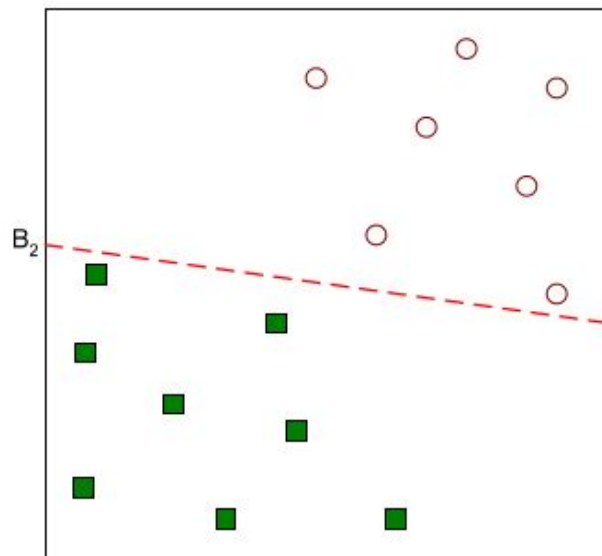
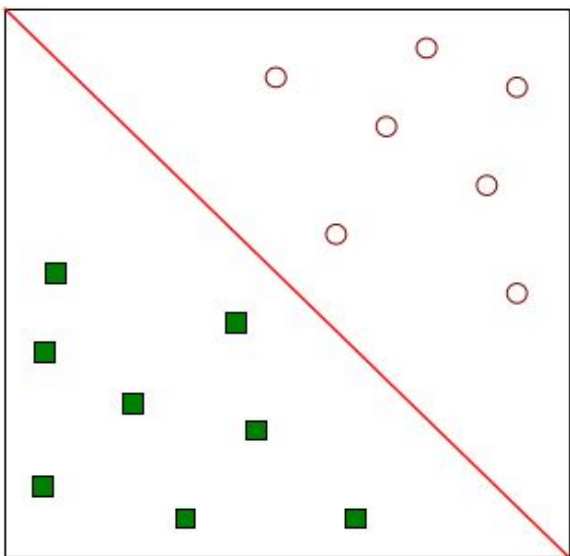
<https://pt.wikipedia.org/wiki/Ficheiro:Survivorship-bias.svg>

Máquina de Vetores Suporte - SVM

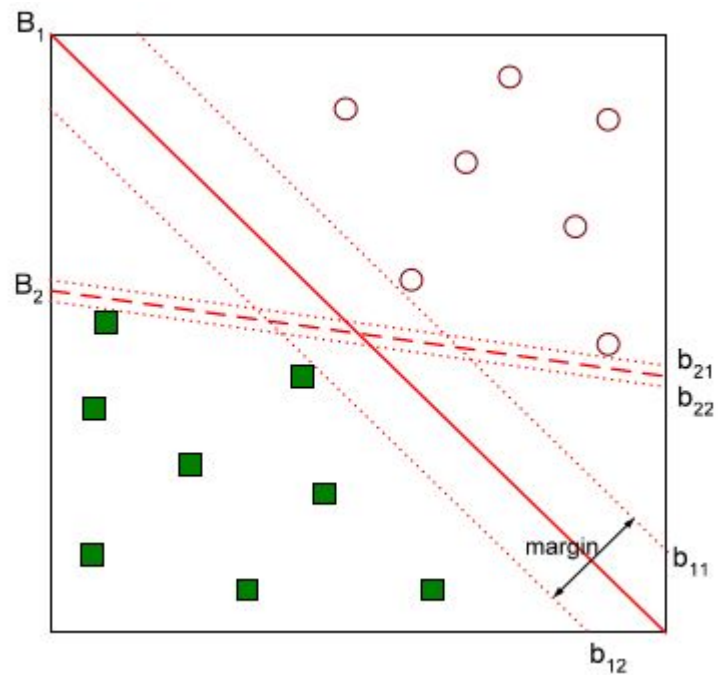
SVM



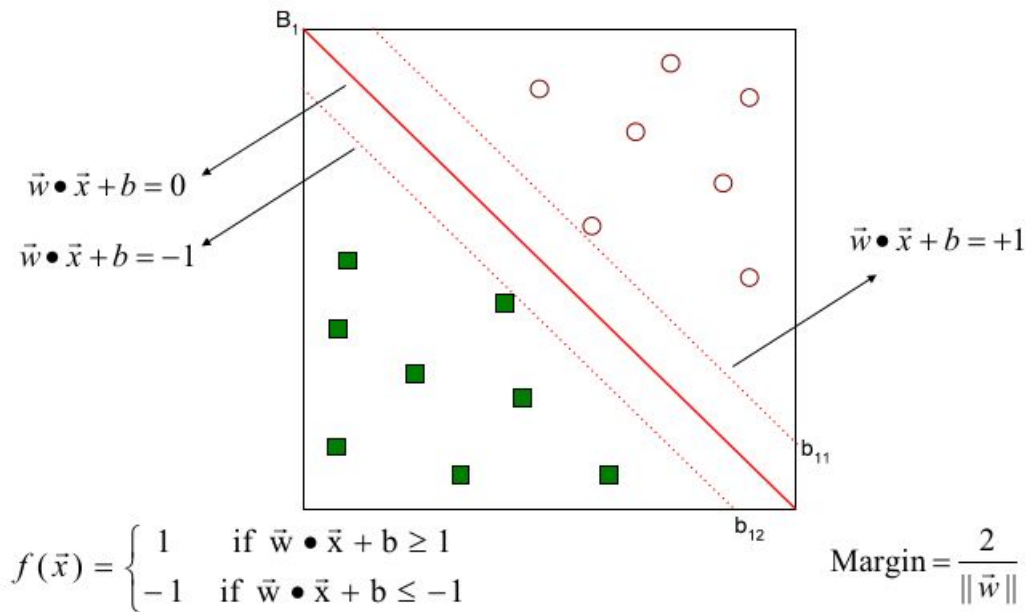
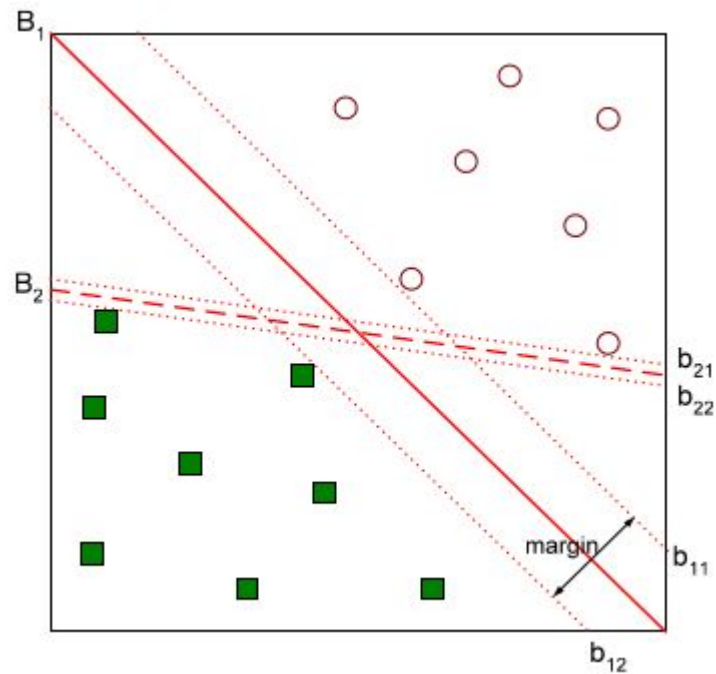
SVM



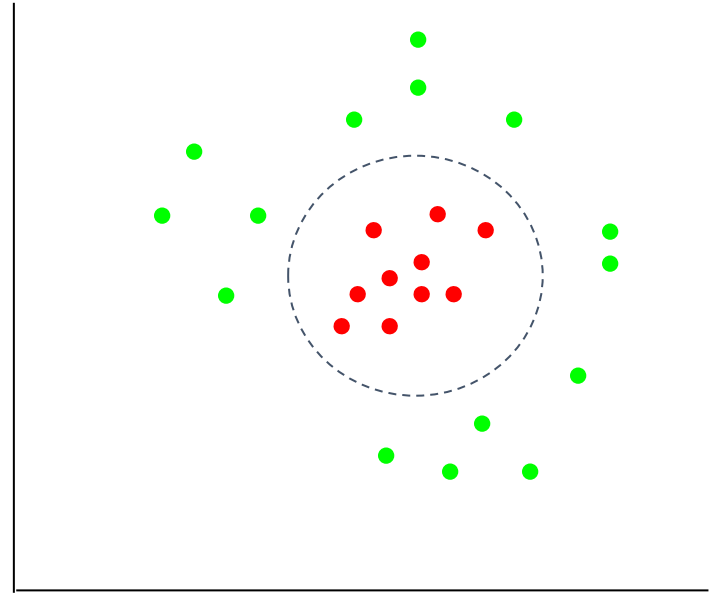
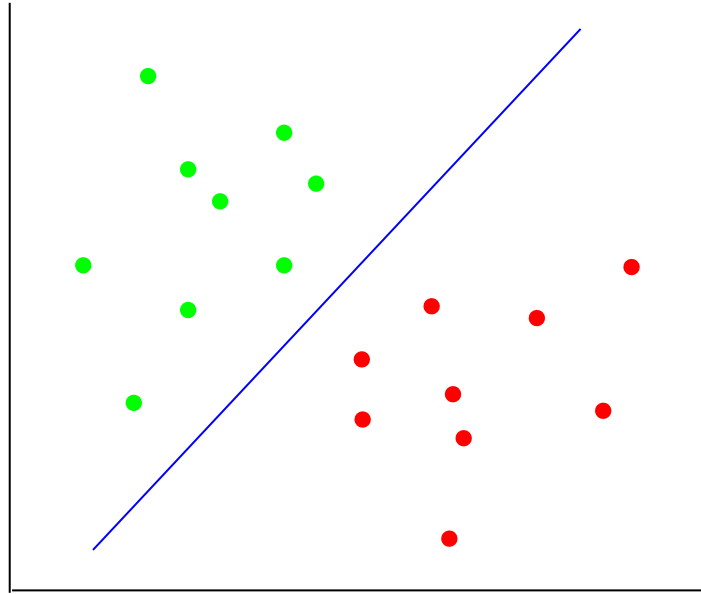
SVM



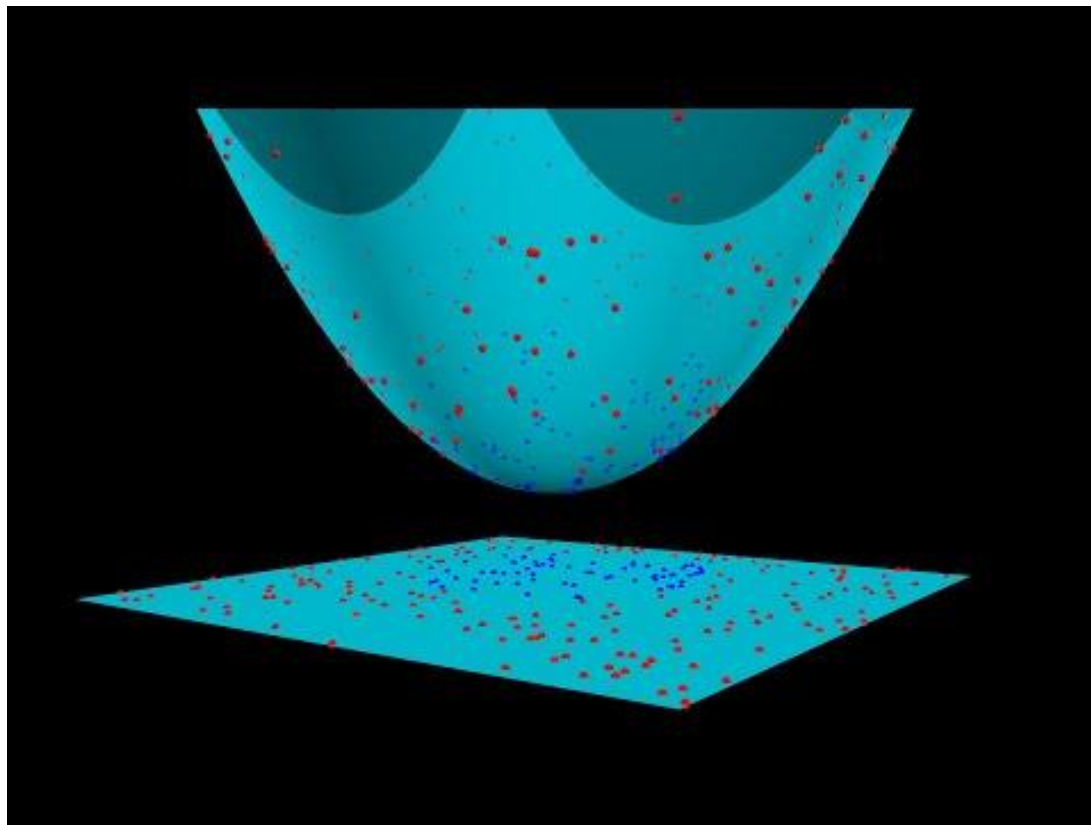
SVM



“Truque” do Kernel

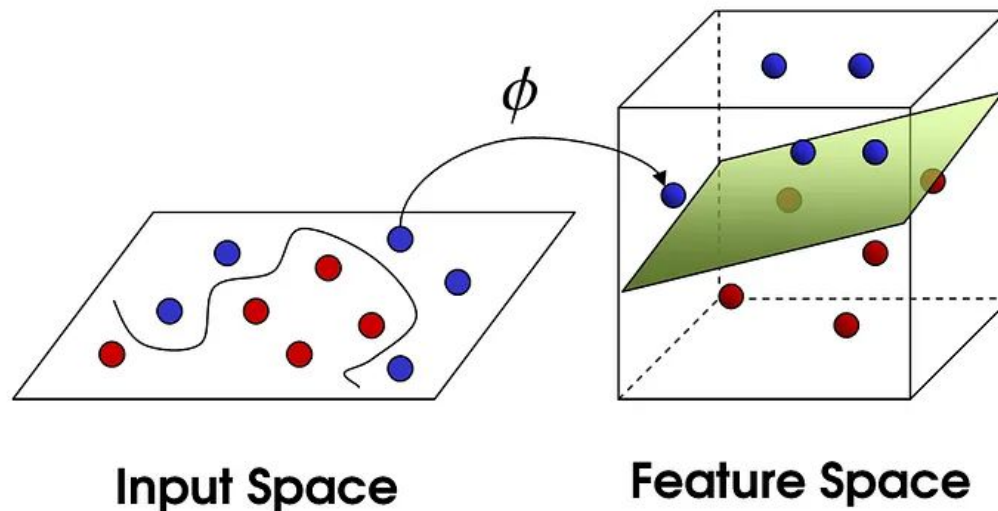


“Truque” do Kernel



“Truque” do Kernel

- The Kernel Trick in Support Vector Classification:
 - <https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f>



“Truque” do Kernel

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ x_{31} & x_{32} & \cdots & x_{3n} \\ x_{41} & x_{42} & \cdots & x_{4n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \vdots \\ y_m \end{bmatrix}$$

$\mathbf{X} \quad \theta \quad = \quad \mathbf{y}$

“Truque” do Kernel

$$\begin{aligned}\mathbf{X}\theta &= \mathbf{y} \\ \mathbf{X}^\top \mathbf{X}\theta &= \mathbf{X}^\top \mathbf{y} \\ \theta &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}\end{aligned}$$

“Truque” do Kernel

$$h(\mathbf{x}) = \sum_{i=0}^n \theta_i x_i = \theta^\top \mathbf{x}.$$

$$h(\mathbf{x}) = \sum_{i=0}^n \theta_i \phi(x_i) = \theta^\top \phi.$$

Kernel Gaussiano

$$k(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x} - \mathbf{y}\|^2}, \gamma > 0$$

Métodos Combinados (Ensemble)

Ensemble

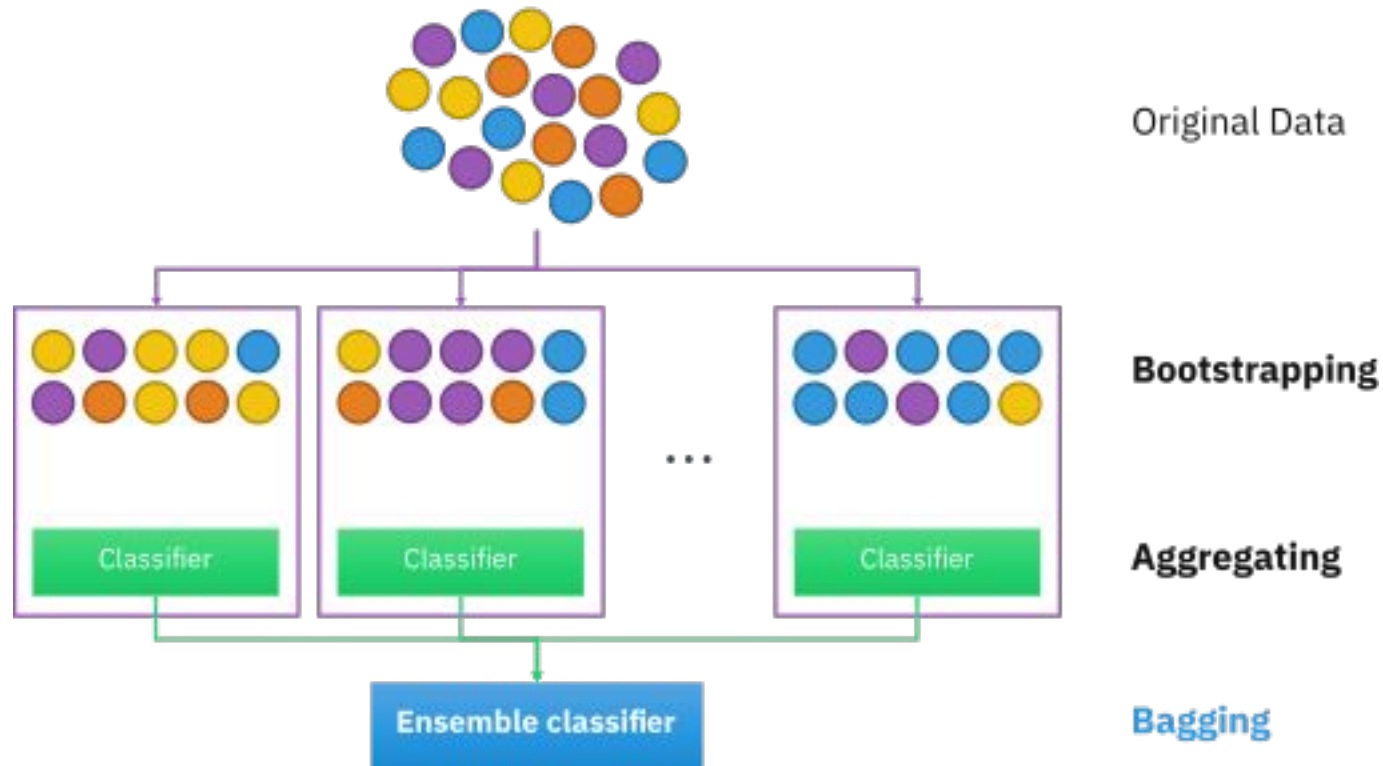
- Bagging
- Boosting
- Stacking

Ensemble

- **Bagging**
- **Boosting**
- **Stacking**

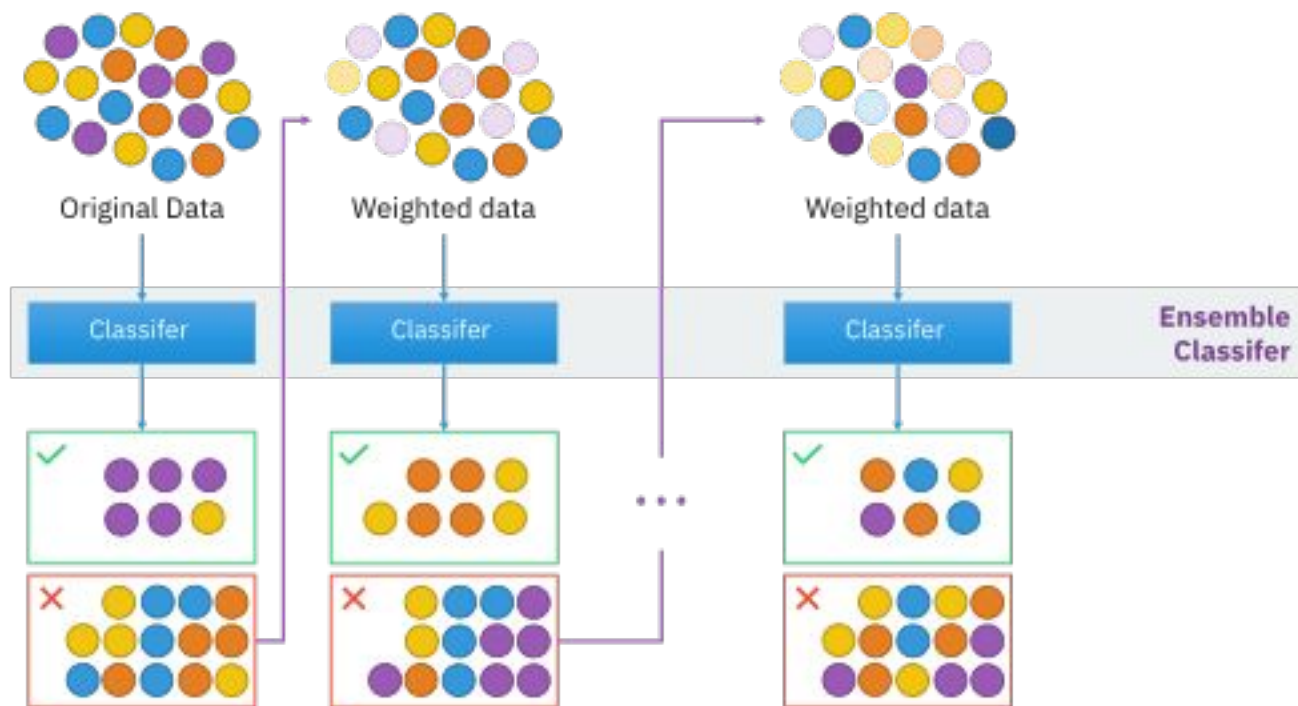
Ensemble

- Bagging



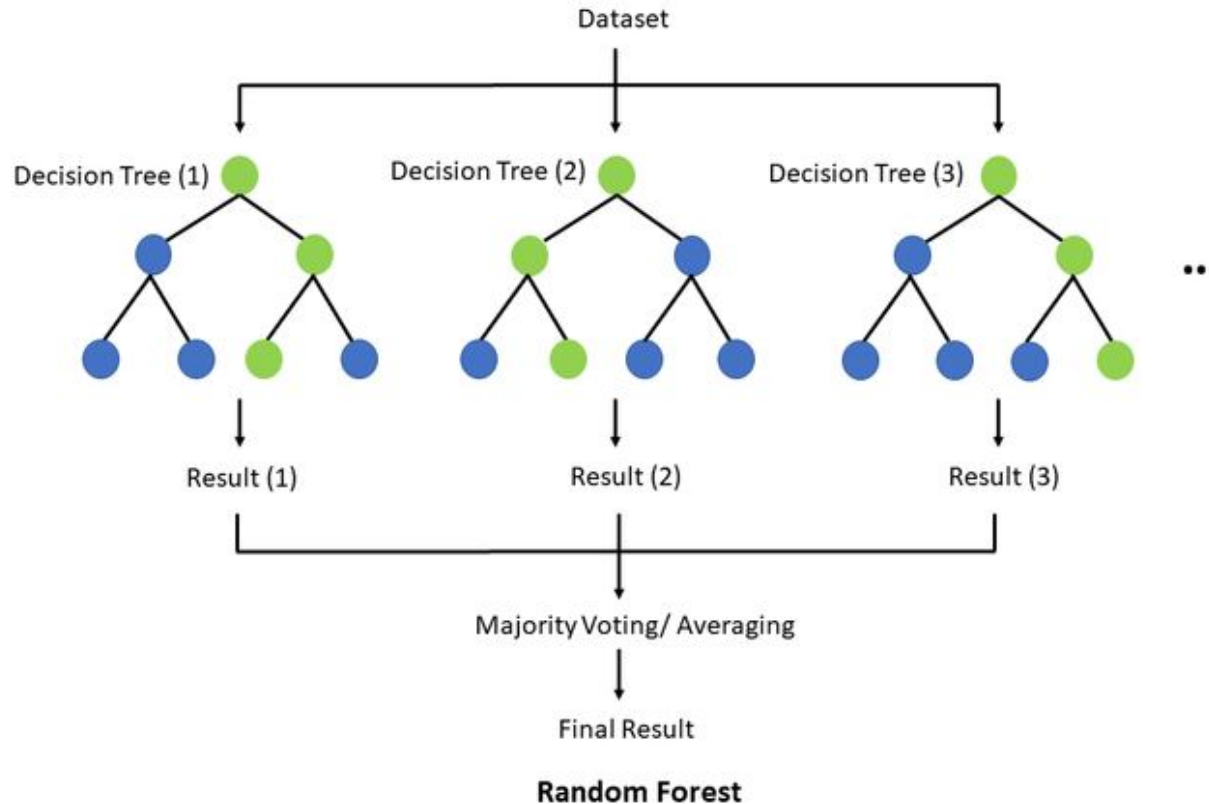
Ensemble

- Boosting



Random Forest

Random Forest



Conceitos e aplicações da aprendizagem de máquina

Eduardo Krempser
eduardo.krempser@fiocruz.br

Matheus Müller
matheusp@posgrad.Incc.br