

Machine Learning



Regressão, *k-Nearest Neighbours* (kNN) e Naive Bayes

Bloco 1

Lucas Claudino

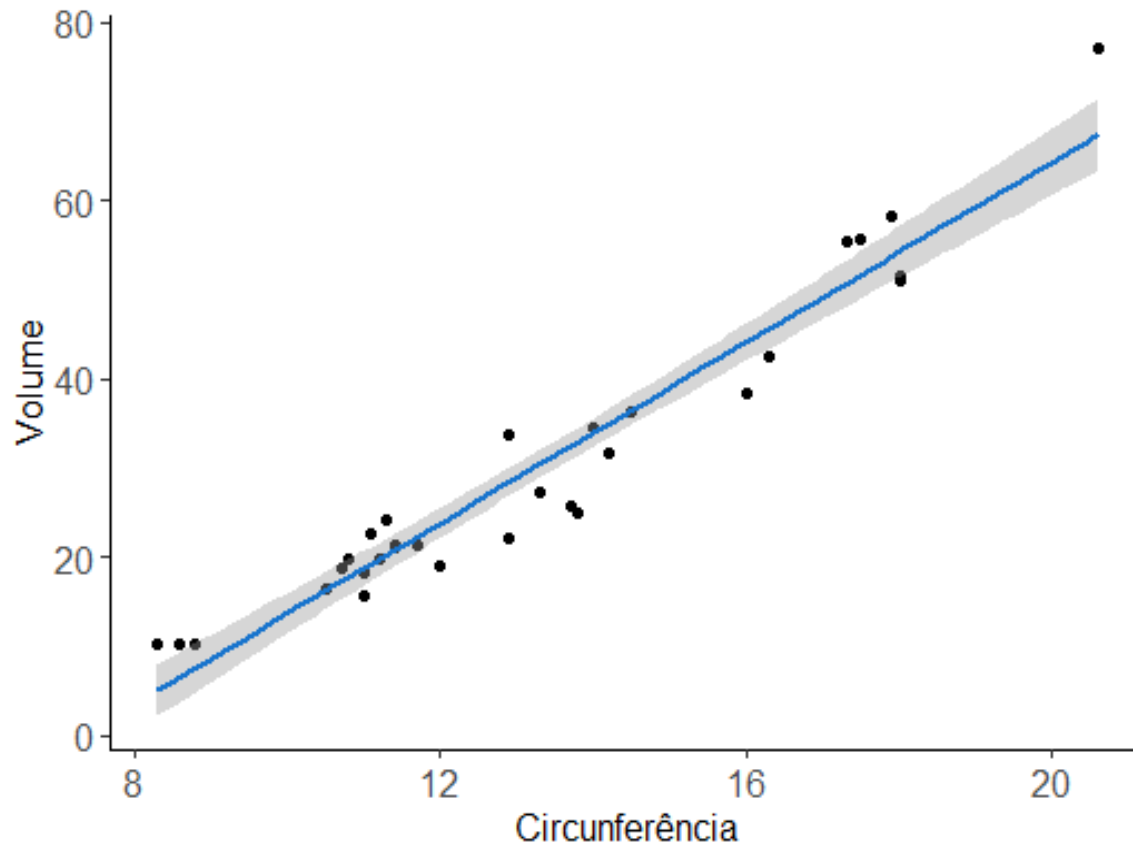


► Regressão

- **Objetivo:** criar uma função matemática a partir dos dados.
- Estimar $\hat{f}(x_i) \approx f(x_i)$.
- A função tem que satisfazer $[\hat{f}(x_i) - f(x_i)] < \delta$.

► Regressão linear e intervalo de confiança

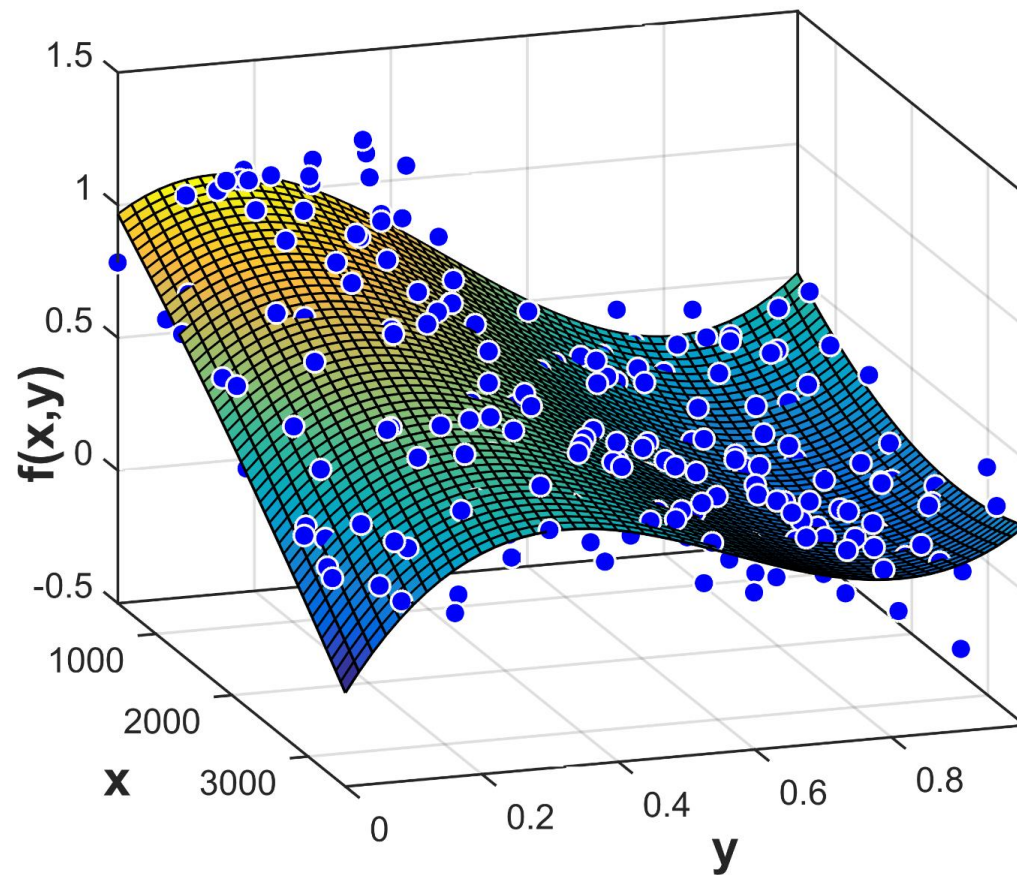
Figura 1 – Regressão linear com as observações e o intervalo de confiança



Fonte: elaborada pelo autor.

► Regressão multivariável

Figura 2 – Regressão multivariável



Fonte: elaborada pelo autor.

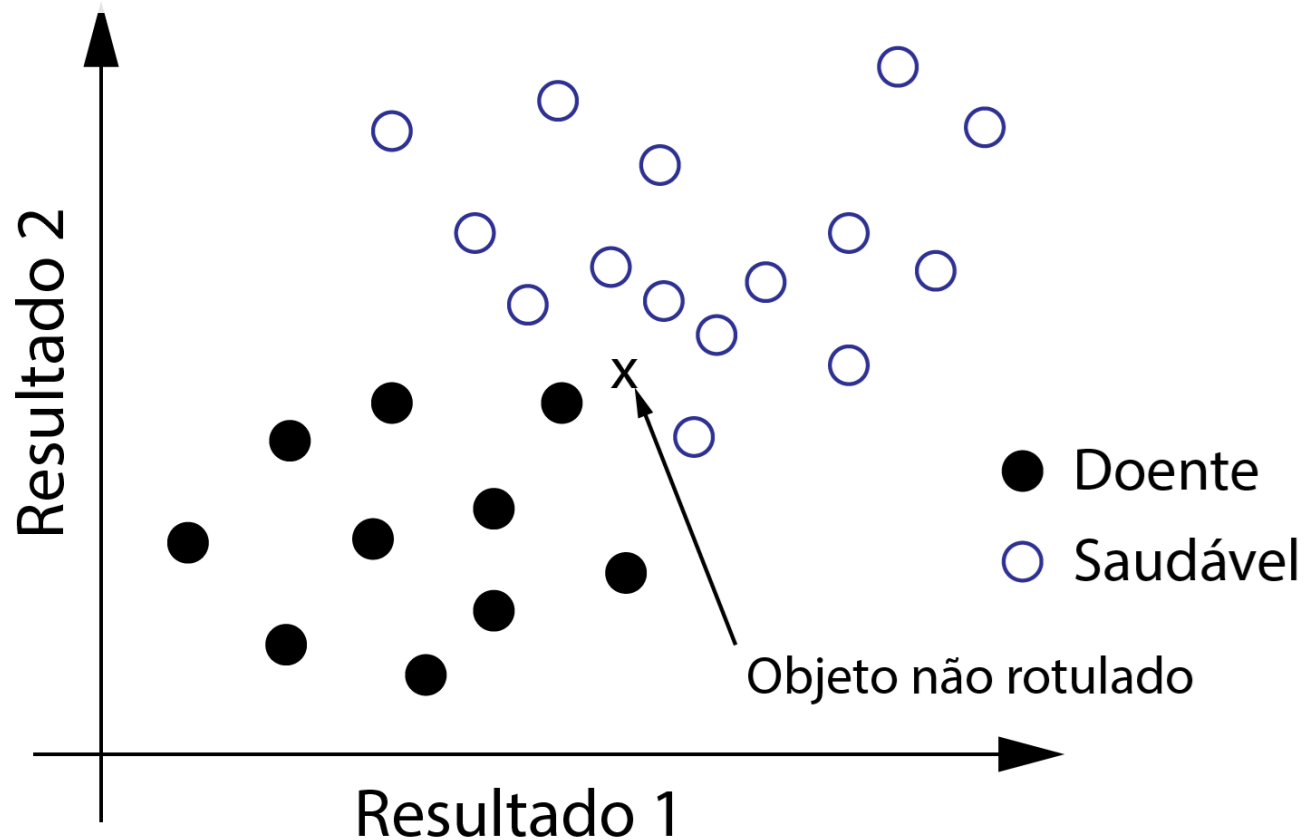
► kNN

- AM supervisionado baseado em distância.
- Treinamento: memoriza o conjunto -> “*lazy*”.
- Descobre o rótulo através da distância entre os objetos:

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^d (x_i^l - x_j^l)^2}$$

► Regressão multivariável

Figura 3 – Classificação 1-NN



Fator k pode gerar classificações diferentes!

Fonte: elaborada pelo autor.

► *Naive Bayes*



- Fácil e rápido para prever o conjunto de dados.
- Bom para variáveis independentes.



- Ruim para variáveis dependentes.

► *Naive Bayes*

Verossimilhança
Prob. do preditor
dada a classe.

Probabilidade original.

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

Probabilidade posterior.


Evidência.

Preditor da probabilidade.



► *Naive Bayes*

Aplicações:

- Previsão em tempo real.
 - Previsões multiclasse.
 - Classificação de textos.
 - Filtragem de spam.
 - Sistema de recomendação.
- 

Problemas e aplicações do *Machine Learning*

Bloco 2

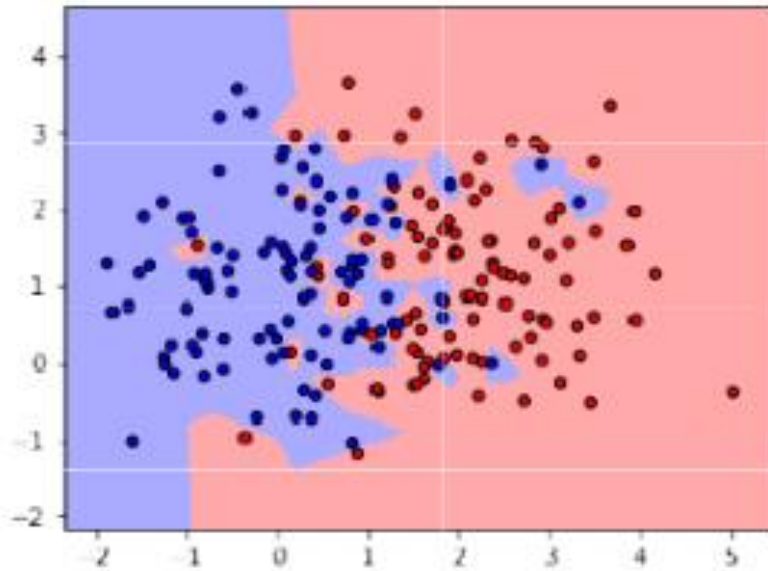
Lucas Claudino



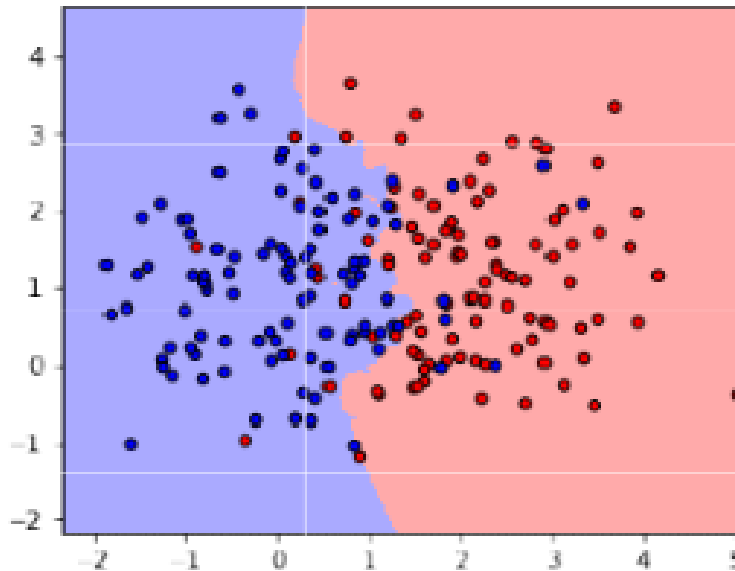
► kNN – influência de k

Figura 4 – Exemplo de influência do parâmetro k em algoritmos kNN

$k=1$



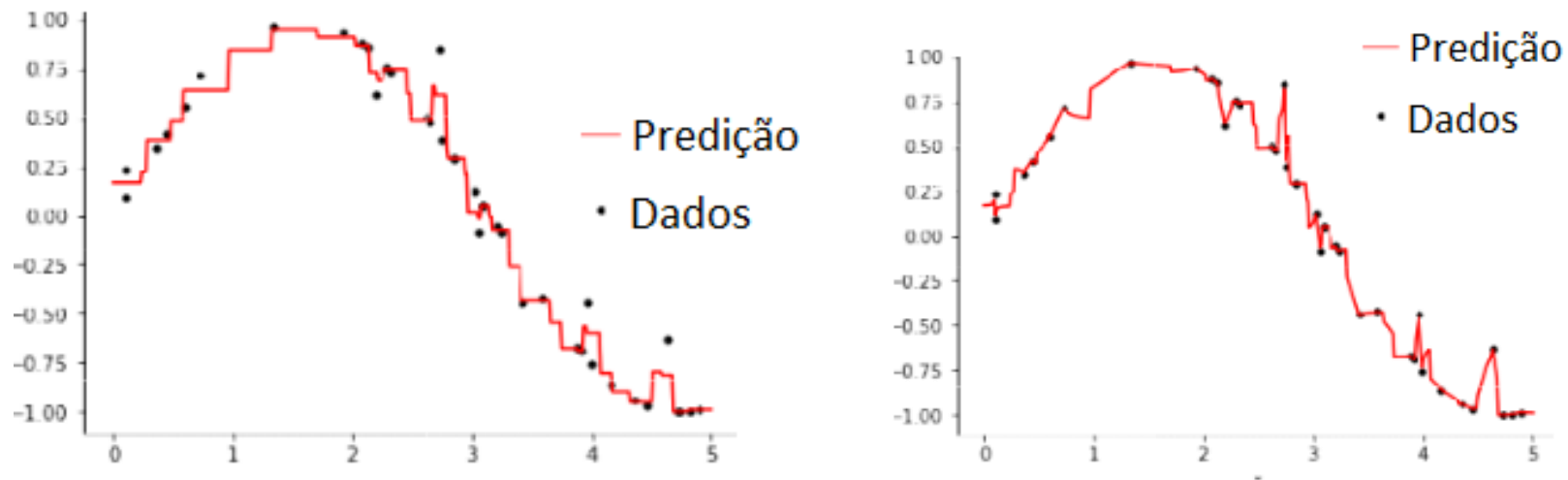
$k=5$



Fonte: Boccato e Attux (2018).

► kNN para regressão

Figura 5 – Algoritmo kNN utilizado para regressão multivariável



Fonte: Boccato e Attux (2018).



► kNN – Escolha do valor de k

- *K-Fold Cross Validation* (KOHAVI, 1995).
- *Leave-One-Out Cross Validation* (MOLINARO, 2005).

PÓS-GRADUAÇÃO

Teoria em prática

Bloco 3

Lucas Claudino



► Teoria em prática

- Problema Iris.

Figura 6 – Plantas setosa, versicolor e virgínica



Fonte: <https://mc.ai/visualization-and-understanding-iris-dataset/>. Acesso em: 14 ago. 2019.



► Teoria em prática

- Vantagens do kNN:
 - Fácil implementação.
 - Pacote para uso de kNN em R.
 - Treinamento: vai memorizar a distribuição.
 - Banco com 150 objetos (50 de cada classe).
 - Pode-se variar a proporção de treinamento e teste.

► Teoria em prática

Figura 7 – Exemplo de implementação de algoritmo para resolução do problema Iris

```
1 require("class") # carrega pacote
2 require("datasets")
3 data("iris") # carrega Iris
4 #set.seed(99) # caso queira sempre o mesmo resultado
5 rnum<- sample(rep(1:150)) # num aleat. de 1 a 150
6 iris<- iris[rnum,] #embaralha o "iris"
7 normalize <- function(x){
8   return ((x-min(x))/(max(x)-min(x)))
9 }
```

Fonte: elaborada pelo autor.

► Teoria em prática

Figura 8 – Exemplo de implementação de algoritmo para resolução do problema Iris

```
10 iris.new<- as.data.frame(lapply(iris[,c(1,2,3,4)],normalize))(2)
11 head(iris.new)
12 iris.train<- iris.new[1:120,]
13 iris.train.target<- iris[1:120,5]
14 iris.test<- iris.new[121:150,]
15 iris.test.target<- iris[121:150,5]
16 summary(iris.new)
17 modell<- knn(train=iris.train, test=iris.test,
18             cl=iris.train.target, k=26) #modell
19
20 table(iris.test.target, modell)
```

Fonte: elaborada pelo autor.

PÓS-GRADUAÇÃO

Dica do professor

Bloco 4

Lucas Claudino





► Dica do professor

- Faça sempre a comparação entre os métodos!
- RAY, Sunil. **Commonly used Machine Learning Algorithms (with Python and R Codes)**. 2017. Disponível em: <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>. Acesso em: 16 ago. 2019.



► Referências

MOLINARO, A. M.; SIMON, R.; PFEIFFER, R. M.. Prediction error estimation: a comparison of resampling methods. **Bioinformatics**, [s.l.], v. 21, n. 15, p. 3301-3307, maio/2005. Oxford University Press (OUP).

BOCCATO, Levy; ATTUX, Romis. **K-Nearest Neighbors**. 2018. Disponível em: http://www.dca.fee.unicamp.br/~lboccato/topico_5_k_nearest_neighbor.pdf. Acesso em: 15 ago. 2019.

FACELI, K. *et al.* **Inteligência artificial**: uma abordagem de aprendizado de máquina. São Paulo: LTC Editora, 2011.



► Referências

KOHAVER, Ron. A Study of Cross Validation and Bootstrap for Accuracy Estimation and Model Selection. **International Joint Conference On Artificial Intelligence (IJCAI)**. San Mateo, p. 1137-1143. ago. 1995.

RAY, Sunil. **Commonly used Machine Learning Algorithms (with Python and R Codes)**. 2017. Disponível em:

<<https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>>. Acesso em: 16 ago. 2019.

SMOLA, Alex; VISHWANATHAN, S.V.N. **Introduction to Machine Learning**. New York, NY: Cambridge University Press, 2008. 226 p.

