

Curso de Especialização em Aprendizagem de Máquina em Inteligência Artificial

Disciplina: Aprendizagem de Máquina

AULA 02

Prof. Gustavo Gattass Ayub

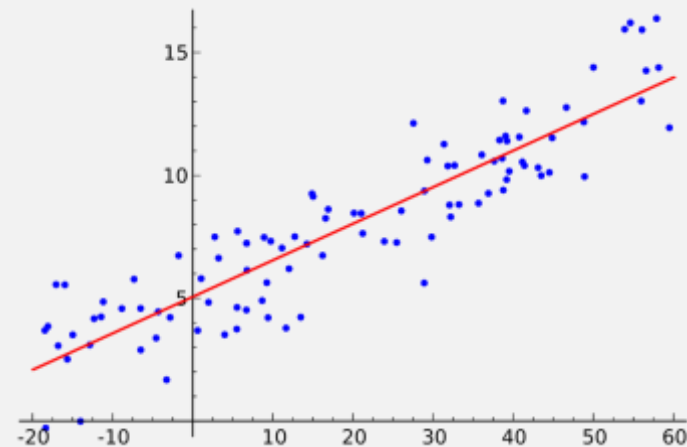




Revisão

■ Regressão Linear

- A regressão linear simples é um método de análise de regressão onde temos uma única relação linear entre uma variável independente X e uma variável dependente Y.
- $Y = a + bX$
- A função custo nos auxilia a determinar os valores ótimos para a e b de modo a obtermos o melhor ajuste (ou “fit”) para a reta de ajuste de modo a estimar Y em função de X.



$$\text{minimize } \frac{1}{n} \sum_{i=1}^n (\text{pred}_i - y_i)^2$$

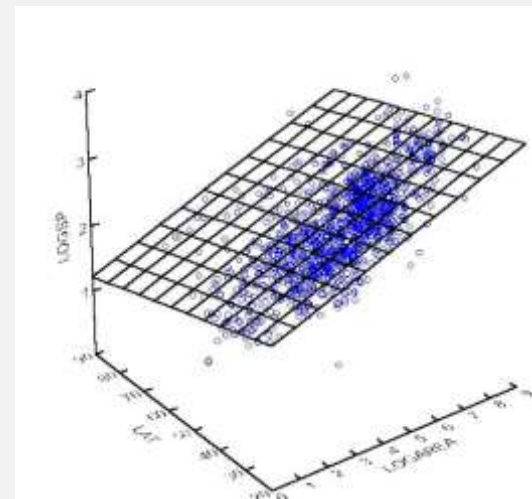
■ Regressão Linear Múltipla

The **multiple linear regression** model has the form

$$y_i = b_0 + \sum_{j=1}^p b_j x_{ij} + e_i$$

for $i \in \{1, \dots, n\}$ where

- $y_i \in \mathbb{R}$ is the real-valued **response** for the i -th observation
- $b_0 \in \mathbb{R}$ is the regression **intercept**
- $b_j \in \mathbb{R}$ is the j -th predictor's regression **slope**
- $x_{ij} \in \mathbb{R}$ is the j -th **predictor** for the i -th observation
- $e_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ is a Gaussian **error term**



Fonte: <http://users.stat.umn.edu/~helwig/notes/mvlr-Notes.pdf>

■ Regressão Linear Múltipla (Cont.)

The model is **multiple** because we have $p > 1$ predictors.

- If $p = 1$, we have a **simple** linear regression model

The model is **linear** because y_i is a linear function of the parameters (b_0, b_1, \dots, b_p are the parameters).

The model is a **regression** model because we are modeling a response variable (Y) as a function of predictor variables (X_1, \dots, X_p).

Fonte: <http://users.stat.umn.edu/~helwig/notes/mvlr-Notes.pdf>

The fundamental assumptions of the MLR model are:

- ➊ Relationship between X_j and Y is **linear** (given other predictors)
- ➋ x_{ij} and y_i are **observed random variables** (known constants)

Regressão Linear Múltipla (Cont.)

Matrix form writes MLR model for all n points simultaneously

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{pmatrix}$$

The **ordinary least squares** (OLS) problem is

$$\min_{\mathbf{b} \in \mathbb{R}^{p+1}} \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 = \min_{\mathbf{b} \in \mathbb{R}^{p+1}} \sum_{i=1}^n \left(y_i - b_0 - \sum_{j=1}^p b_j x_{ij} \right)^2$$

where $\|\cdot\|$ denotes the Frobenius norm.

Fonte: <http://users.stat.umn.edu/~helwig/notes/mvlr-Notes.pdf>

■ 0 coeficiente R²

The **coefficient of multiple determination** is defined as

$$\begin{aligned} R^2 &= \frac{SSR}{SST} \\ &= 1 - \frac{SSE}{SST} \end{aligned}$$

and gives the amount of variation in y_i that is explained by the linear relationships with x_{i1}, \dots, x_{ip} .

When interpreting R^2 values, note that. . .

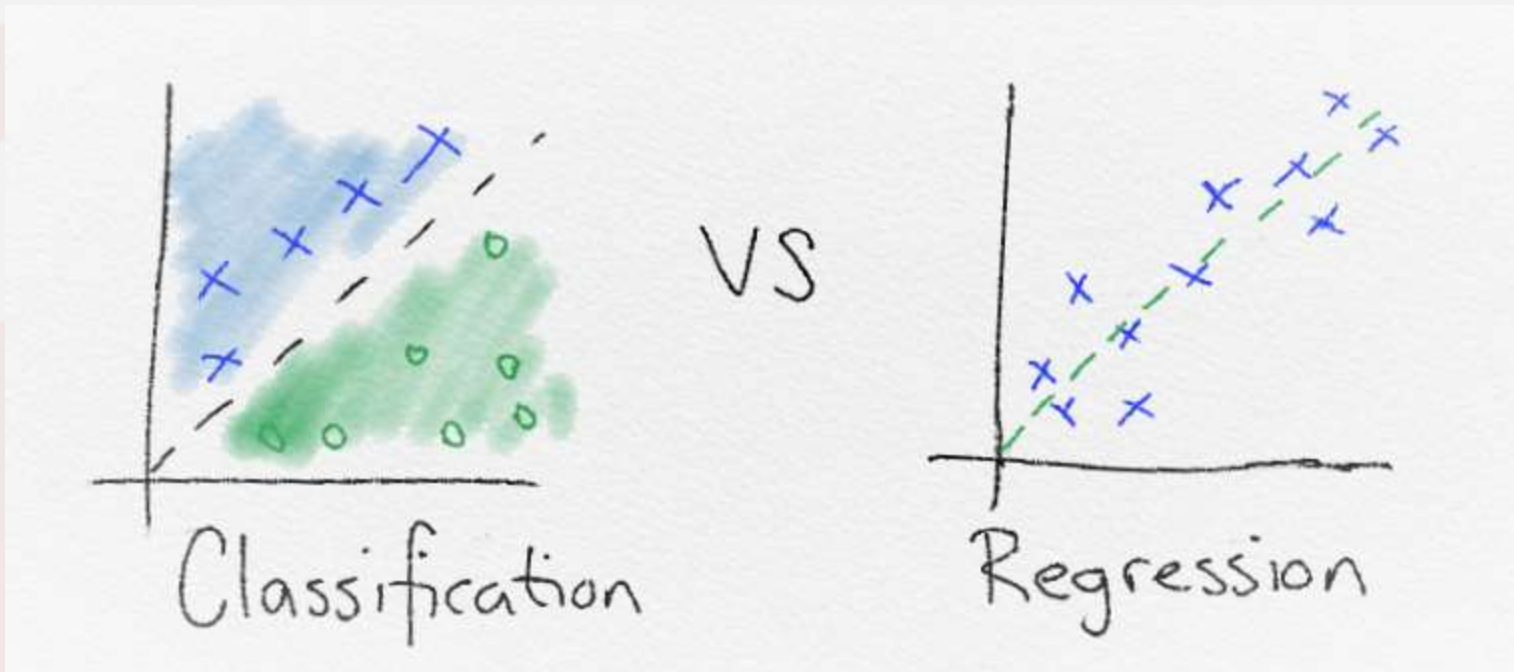
- $0 \leq R^2 \leq 1$
- Large R^2 values do not necessarily imply a good model

Fonte: <http://users.stat.umn.edu/~helwig/notes/mvlr-Notes.pdf>

Exercício Prático



■ Regressão vs Classificação



Fonte: <https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0>



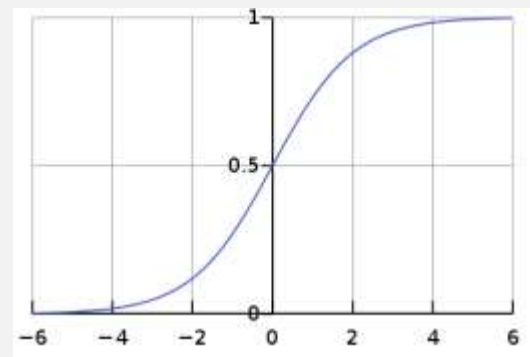
Regressão Logística

■ Regressão Logística

- Algoritmo de Classificação Binária
- Utilizado para prever a probabilidade de uma variável dependente categórica 1 (UM, Sim, Positivo, etc) ou 0 (ZERO, Não, Negativo, etc).
- Em outras palavras a regressão logística prevê $P(Y=1)$ como uma função de X .

The logistic regression can be understood simply as finding the β parameters that best fit:

$$y = \begin{cases} 1 & \beta_0 + \beta_1 x + \varepsilon > 0 \\ 0 & \text{else} \end{cases}$$

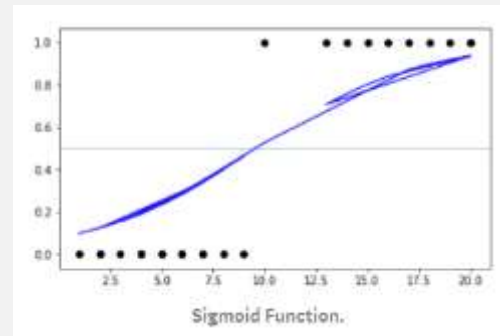


■ Regressão Logística (Premissas)

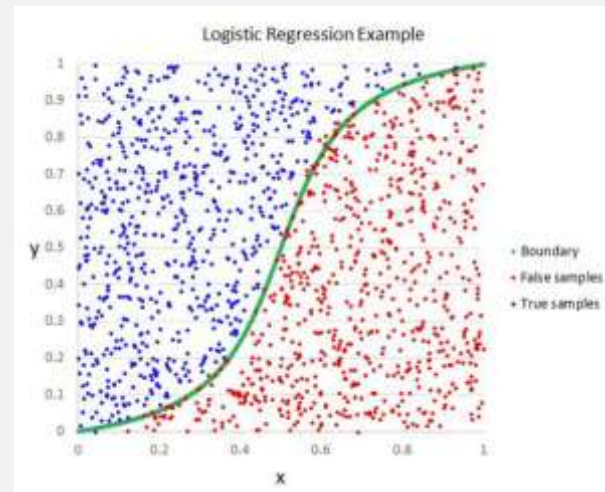
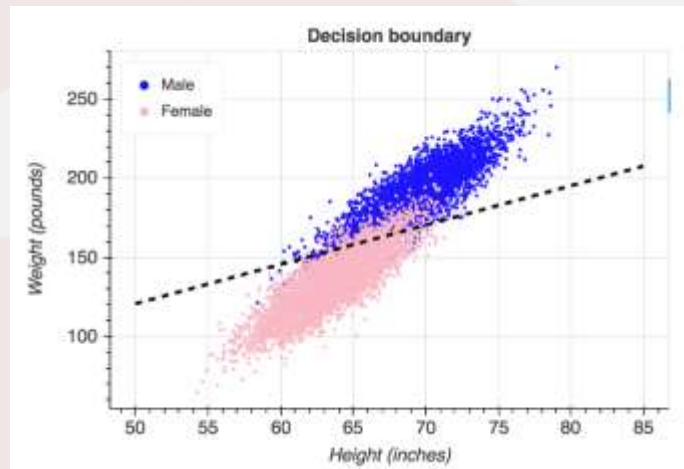
- A variável dependente Y deve ser binária
- O valor objetivo é 1 (UM, Sim, Sucesso, etc)
- O vetor X deve conter variáveis independentes com ZERO ou baixa co-linearidade
- Utilizamos a função logit como função de ativação para projetar a variável dependente a partir de uma regressão linear múltipla

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X.$$

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}.$$



Exemplos



■ Vantagens e Pontos de Atenção

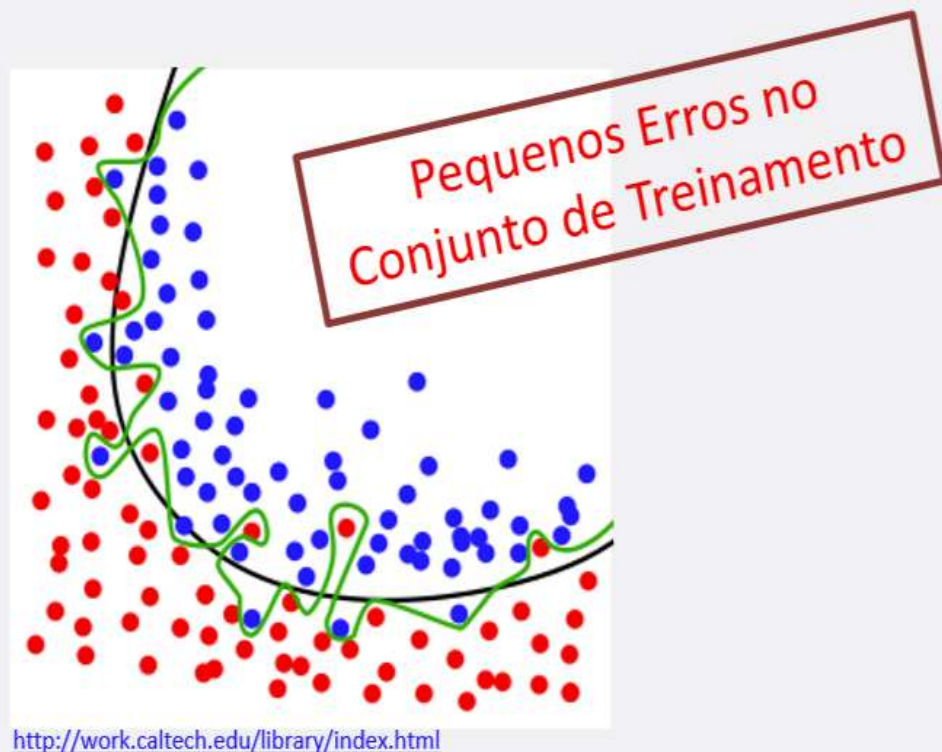
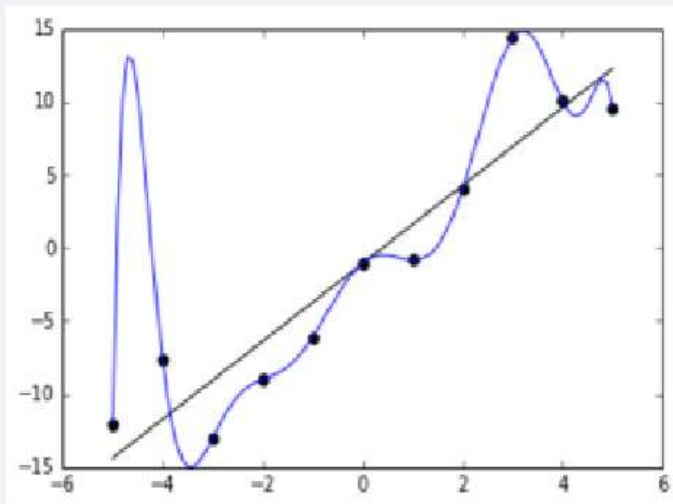
- Método simples e eficiente
- Pouca variância
- Além de classificar o método também fornece a probabilidade de ocorrência de um determinado valor
- Aplicável a classificação binária
- As features (variáveis) devem ser independentes e lineares em relação ao

Exercício Prático



Overfitting (ou sobreajuste)

→ Modelos Melhores



*All models are wrong,
but some are useful.*

George Box

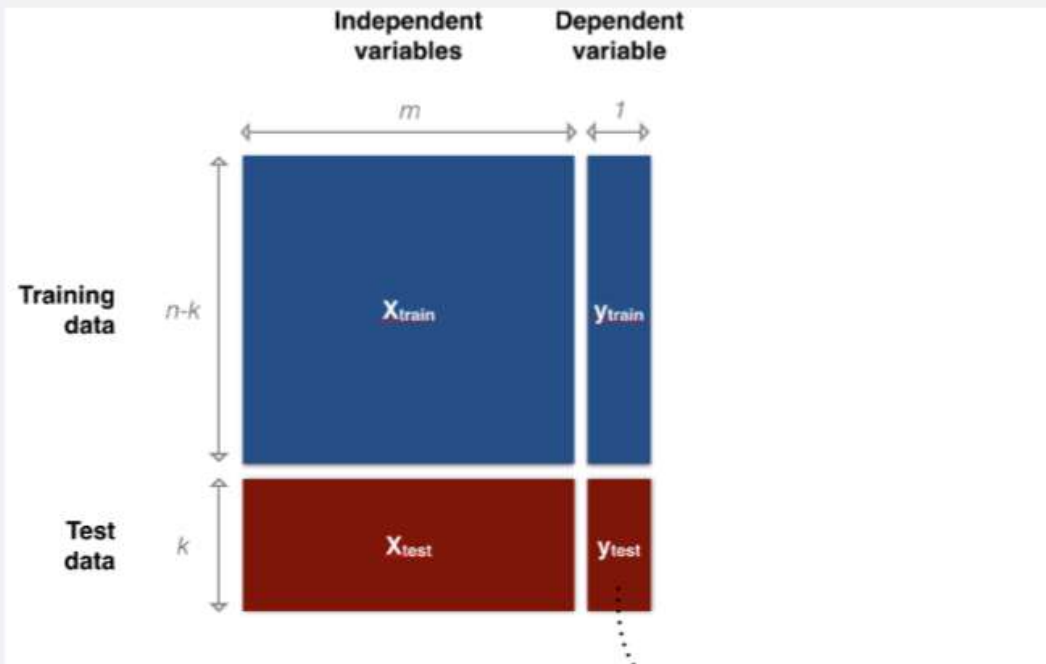
Um bom modelo



A função g produzida pelo modelo é suficientemente próxima de f ?

$$g(.) \sim f(.)$$

Conjunto de treinamento e de testes



A função g produzida pelo modelo é suficientemente próxima de f ?

$$g(.) \sim f(.)$$

Erro($g(.)$, $f(.)$)

<http://work.caltech.edu/library/index.html>

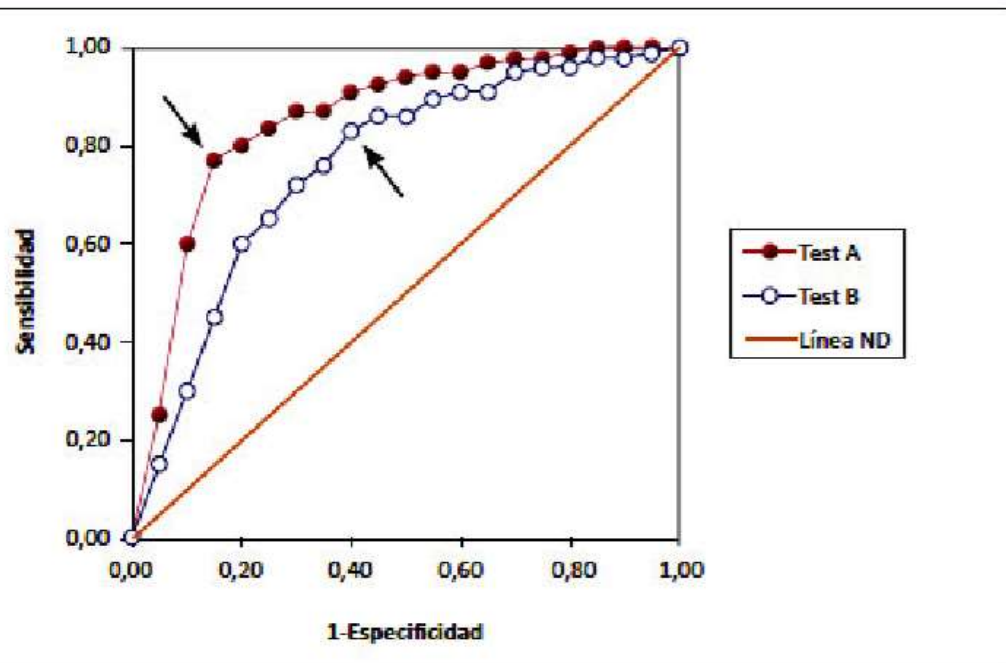
Matriz de Confusão

- **Acurácia** $(VP + VN) / (Total)$
- Sensibilidade $VP / (VP + FN)$
- Especificidade $VN / (VN + FP)$

		Valor Verdadeiro (confirmado por análise)	
		positivos	negativos
Valor Previsto (predito pelo teste)	positivos	VP Verdadeiro Positivo	FP Falso Positivo
	negativos	FN Falso Negativo	VN Verdadeiro Negativo

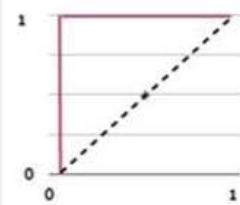
SILVA, Leandro Nunes De Castro; FERRAR, Daniel Gomes. Introdução À Mineração de Dados

Curva ROC

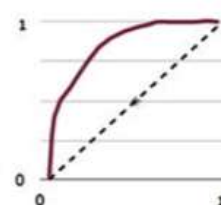


**Diagnóstico
Perfeito**

AC=1,0

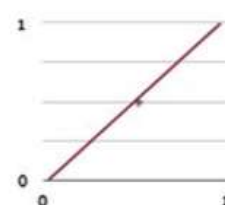


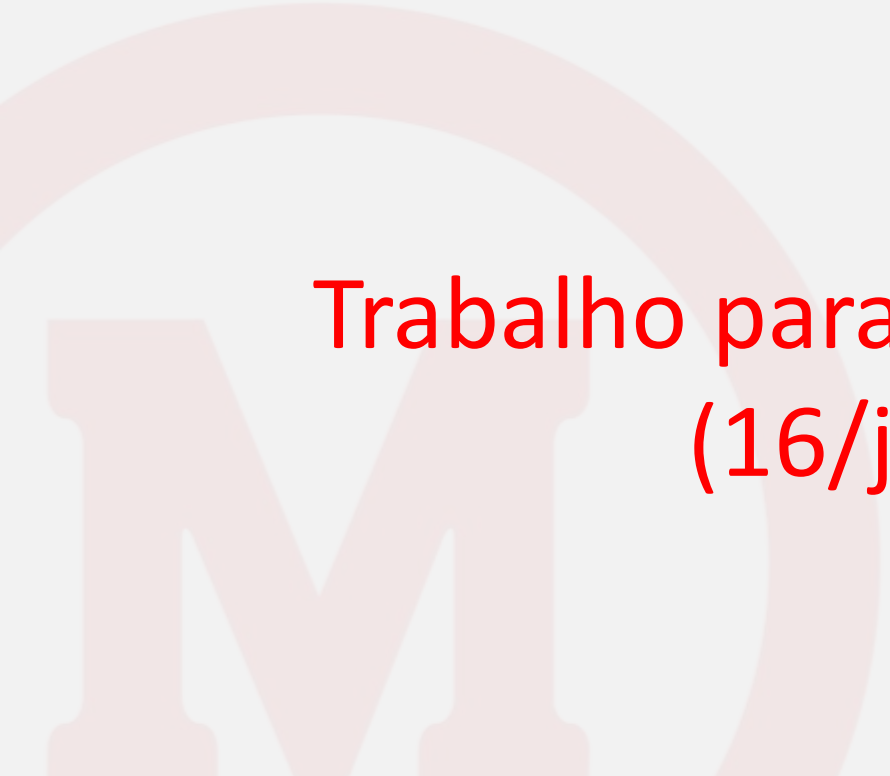
0,8



**Sem
Diagnóstico**

0,5





Trabalho para a próxima aula (16/jun/20)

■ Para a próxima aula (16/jun/20)

Exercício Complementar - **RECOMENDADO** (não vale nota)

- [Introduction to Logistic Regression](#)
- [Building a Logistic Regression in Python, Step by Step](#)

Exercício de Aprofundamento (vale nota)

- Vamos utilizar o mesmo dataset de vinhos tintos para criar um classificador. O dataset categoriza os vinhos em 6 classes de qualidade (3-8). Para criar um classificador binário, você deve considerar que as notas (3-6) indicam “Baixa Qualidade” e (7-8) indicam “Alta Qualidade”
- Você deve particionar o dataset em dois conjuntos: treinamento e validação usando a proporção 80-20. Tome o cuidado de manter um bom balanço em termos de exemplos nos dois conjuntos.
- Você deve treinar o classificador utilizando o algoritmo da Regressão Logística. Você também deve produzir uma matriz de confusão aplicando esse classificador ao conjunto de validação.
- Dataset: arquivo .csv “Wine Dataset” (no moodle – Aula 02)



Até a próxima aula

MUITO OBRIGADO!