

# CD - 2024-2 Avaliação 2a

Nome e Sobrenome: \_\_\_\_\_

Disciplina: Ciência dos Dados

Semestre: 2024.2

Professor: Dr. Diego Pinheiro

**Observações: As respostas deverão ser escritas à CANETA. NÃO SERÃO ACEITAS QUESTÕES APENAS COM RESPOSTAS );**

## Informações preliminares

As questões a seguir são relacionadas a um modelo de classificação para prever se pessoas que são fumantes ( $x_1 = 1$ ) ou não fumantes ( $x_1 = 0$ ) vão apresentar ( $y = 1$ ) ou não ( $y = 0$ ) câncer de pulmão conforme os dados a seguir.

$$\begin{bmatrix} x_1 & | & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ y & | & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

## Questão 1 (0,5 pontos)

A regressão logística utiliza a **função logit** que é a transformação das probabilidades  $Pr(y = 1|x_1)$  para a escala **log-odds**.

$$\text{logit}(x_1) = \ln \left[ \frac{Pr(y = 1|x_1)}{1 - Pr(y = 1|x_1)} \right] = w_0 + w_1 x_1$$

Qual a expressão da probabilidade  $Pr(y = 1|x_1)$  de uma pessoa apresentar câncer de pulmão ?

$Pr(y = 1 x_1)$
-----------------

## Questão 2 (0,5 pontos)

Considerando um modelo inicial  $\mathbf{w}^{[0]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ , qual a probabilidade  $Pr(y = 1|x_1)$  de câncer de pulmão (**com duas casas decimais**) para uma pessoa

fumante  $Pr(y = 1|x_1 = 1)$  e para uma pessoa não fumante  $Pr(y = 1|x_1 = 0)$  ?

$Pr(y = 1 x_1 = 0)$	
$Pr(y = 1 x_1 = 1)$	

### Questão 3 (0,5 pontos)

Qual o **F<sub>1</sub>score** do modelo  $\mathbf{w}^{[0]}$  (com duas casas decimais)?

<b>F<sub>1</sub>score</b>	
---------------------------	--

### Questão 4 (0,5 pontos)

A partir do modelo  $\mathbf{w}^{[0]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$  e considerando  $\lambda = 0.1$ ,  $\left( \frac{\partial E(\mathbf{w})}{\partial \mathbf{w}^T} \right) \Big|_{\mathbf{w}=\mathbf{w}^{[0]}} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$ ,  $\left( \frac{\partial^2 E(\mathbf{w})}{\partial \mathbf{w} \partial \mathbf{w}^T} \right) \Big|_{\mathbf{w}=\mathbf{w}^{[0]}}^{-1} = \begin{bmatrix} 0.67 & -0.67 \\ -0.67 & 1.33 \end{bmatrix}$ , qual seria um melhor modelo  $\mathbf{w}^{[1]}$  ?

$\mathbf{w}^{[1]}$	
--------------------	--

### Questão 5 (0,5 pontos)

Considerando o modelo  $\mathbf{w}^{[1]}$  melhorado, qual a probabilidade  $Pr(y = 1|x_1)$  de câncer de pulmão (**com duas casas decimais**) para uma pessoa fumante  $Pr(y = 1|x_1 = 1)$  e para uma pessoaal não fumante  $Pr(y = 1|x_1 = 0)$  ?

$Pr(y = 1 x_1 = 0)$	
$Pr(y = 1 x_1 = 1)$	

### Questão 6 (0,5 pontos)

Qual o **F<sub>1</sub>score** do modelo  $\mathbf{w}^{[1]}$  (com duas casas decimais)?

<b>F<sub>1</sub>score</b>	
---------------------------	--