

FA084 - Pre Aula 04

Use o script `pre_aula04_base.R`.

Não serão aceitos envios que não usem o script.

Entregue um arquivo `pre_aula_04_RA_primeironome.zip` contendo:

- Um arquivo `pre_aula_04_RA_primeironome.pdf` com as respostas
- Um arquivo `pre_aula_04_RA_primeironome.R` (originado do script `pre_aula04_base.R`)

Exercício 01

Use `set.seed(seu_ra)` e faça o sample de 6 números de 1 a 10, salve na variável `d`.

Utilizando a equação 4.2 (pg 132) (função logística) crie uma funcao que calcula os valores da saída da funcao logística para cada valor de um vetor de entrada. Faça os gráficos de $p(x)$, com x sendo um vetor de -10 a 10 , utilizando:

- a. $B0_a = (d[3] + d[4])/10$, $B1_a = (d[5] + d[6])/10$
- b. $B0_b = (d[3] + d[4])/10$, $B1_b = - (d[5] + d[6])/10$
- c. $B0_c = (d[3] + d[4] + d[1] + d[2])/10$, $B1_c = (d[5] + d[6])/10$

Descreva os efeitos dos valores de $B0$ e $B1$ no comportamento do gráfico utilizando os resultados dos itens a, b e c.

Recomendações:

- Use `plot(x,y,type='l')` para plotar o item a. Em seguida, use `lines(x,y)` para colocar os dados gerados em b e c. Use os valores do vetor x no eixo x e da saída da funcao criada ($p(x)$) no eixo y.

Exercício 02

Faça download do conjunto de dados disponibilizado. Esse conjunto de dados é referente a dados meteorológicos e foi processado para a tarefa de predizer se no dia seguinte, ocorrerá chuva.

Os atributos no conjunto são Chuva, T_{max} e T_{min} para temperaturas máxima e mínima respectivamente. Eles são apresentados com 3 tipos de defasagem: T_{max1} corresponde à temperatura máxima do dia anterior, T_{max2} para dois dias antes, e T_{max3} para 3 dias antes. A mesma lógica se aplica para T_{min} .

- a. Crie três atributos novos: `AmpTerm1`, `AmpTerm2`, `AmpTerm3`
- b. Divida o conjunto em treino e teste utilizando `set.seed(seu_ra)` antes de `sample()`. **Não é necessário normalizar os atributos.**
- c. Treine 3 modelos:
 - 1. Um usando dados de 1 dia antes para prever ocorrência de chuva
 - 2. Usando dados dos 2 últimos dias para prever ocorrência de chuva
 - 3. Usando dados dos 3 últimos dias para prever ocorrência de chuva
- d. Crie a matriz de confusao e calcule as acurácias de cada modelo.

Responda:

1. Para cada modelo, qual o coeficiente com maior valor absoluto?

2. Considerando as matrizes de confusão, o que são falsos positivos e falsos negativos nesse caso? Discuta (brevemente) sobre como podemos ponderar o custo de cada um em uma situação onde esse modelo é usado para decidir se é feita ou não a irrigação no dia em questão.
3. Compare a performance dos modelos utilizando a acurácia. Qual modelo apresenta melhores resultados? Justifique.

Recomendações:

- Recomendamos criar os 3 modelos sem utilizar loop
- A seção 4.6.1 do livro-texto tem detalhes de uma modelagem com regressão logística
- As páginas 157 e 158 são muito importantes para a avaliação de modelos
- Para treinar uma regressão logística, é necessário utilizar `family=binomial` como argumento na função `glm()`
- Para fazer a predição, é necessário utilizar `type='response'` para obter as probabilidades de ser `TRUE`.
- Para transformar o vetor de probabilidades (saída do `predict()`), em um vetor lógico (com `TRUE` e `FALSE`), basta usar:
 - `vetor_logico = vetor_de_probabilidades > 0.5` (porque 0.5?)