

## 2a Prova de Modelos de Regressão Paramétricos e Não-Paramétricos

### (Implementação computacional)

Em geral, os parâmetros de regressão para uma variável aleatória  $Y$  que pertence à família exponencial podem ser estimados utilizando o algoritmo de *Newton-Raphson*, descrito na seção 4.3 (Algoritmo de estimação para a família exponencial, supondo ligação canônica). O algoritmo assume uma forma iterativa onde:

$$\beta^{(k+1)} = \left( \mathbf{X}^T \mathbf{W}^{(k)} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{W}^{(k)} \mathbf{Z}^{(k)}$$

Mesmo na forma iterativa, é possível inserir restrições nos parâmetros do modelo, como é o caso de restrições do tipo  $l_2$  (Ridge-Regression), onde aplica-se a restrição  $\sum_{i=1}^p \beta_i^2 \leq c$ . Nesta situação, caso o modelo seja de regressão linear múltipla, a solução é dada na forma:

$$\tilde{\beta} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}^*)^{-1} \mathbf{X}^T \mathbf{Y}$$

Uma possível escolha para o parâmetro  $\lambda$  pode ser obtido a partir da matriz  $H$ . Caso o modelo de regressão linear múltiplo (Gaussiano) seja utilizado, é possível utilizar o coeficiente de determinação preditiva  $R_{pred}^2$  ():

$$\hat{\lambda} = \arg \min_{\lambda} \sum_{i=1}^n \left( \frac{r_{i(\lambda)}}{1 - h_{ii(\lambda)}} \right)^2$$

Na família exponencial, deseja-se verificar se é possível o cálculo da matriz  $H$  é o uso do  $R_{pred}^2$  como critério de seleção do parâmetro  $\lambda$ . Neste caso, pode-se utilizar a pseudo-variável  $Y$ :  $Z^{(k)}$  (após a convergência do algoritmo). Alguns autores indicam o uso dos resíduos de *deviance* para o cálculo de uma estatística PRESS (*Predicted Error Sum of Squares*).

Portanto, esta avaliação tem como objetivos:

- Implementação do algoritmo de Newton Raphson para um problema de classificação de padrões (resposta binária).
- Comparação do resultado obtido no item (a) com a função `glm()`
- Implementação do algoritmo de Newton Raphson para o problema de classificação incorporando a restrição  $\sum_{i=1}^p \beta_i^2 \leq c$ .
- Estimar o parâmetro  $\lambda$  utilizando os conceitos de matrix  $H$  e dos pseudo-resíduos do modelo de regressão generalizada (ex. *deviance residual*).

Sugere-se que seja entregue um relatório em formato Rmd (RMarkdown), compilado para um arquivo final em **pdf** contendo códigos, resultados de execução e comentários. O relatório deve incluir uma descrição sucinta da base de dados utilizada.