

10ª Lista de exercícios – Teoria dos Modelos Lineares

Nas listas 8 e 9, foi utilizada a base de dados "boston_corrected.txt", na qual estão disponíveis 13 potenciais variáveis preditoras. Nesta lista (10), desejamos investigar a capacidade preditiva do estimador *LASSO* (Least Absolute Shrinkage Selection Operator). Novamente, o foco do ajuste não será a relevância estatística. Entretanto, o estimador *LASSO* permite que alguns coeficientes sejam estimados como zero. Ou seja, permite a seleção automática de variáveis. O objetivo do exercício é encontrar o modelo que apresente o maior valor do coeficiente de determinação preditivo (R^2_{pred}). O estimador *LASSO* pode ser escrito na forma:

$$\tilde{\beta} = \arg \min_{\beta} (\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta)$$

$$\text{Sujeito a: } \sum_{i=1}^p |\beta_i| \leq d$$

Faça o ajuste do estimador *LASSO* de forma a maximizar o coeficiente de determinação preditivo. Determine o valor ótimo de d . Faça uma comparação entre os coeficientes estimados com o método de mínimos quadrados (opção padrão), os coeficientes estimados utilizando o método de *ridge regression* (lista passada) e o método *LASSO*.

Utilizando o pacote **glmnet**, faça também o ajuste de um quarto modelo. Faça a comparação com o resultado obtido pelo método **glmnet**.

Na sua opinião, quais são as vantagens e desvantagens de utilizar o estimador *LASSO*?

Elabore o seu relatório utilizando o **Rmarkdown** e envie o documento em formato **PDF** para avaliação pelo sistema minha.ufmg. O documento deve conter, no máximo, 6 páginas.