

ISCTE – IUL
Licenciatura em Ciência de Dados
1.º Semestre 2024
Modelação Estocástica
Segundo Trabalho de Grupo

Pretende-se, com este trabalho, que utilize os métodos de validação cruzada (*cross validation*) para avaliar a qualidade da previsão de dois modelos (linear e polinomial), quando aplicados a um conjunto de dados.

O conjunto de dados deve ser simulado por cada grupo, utilizando a rotina:

```
set.seed(*****)

funcao_geradora_dos_dados <- function(n, ruido){
  x <- runif(n, -10, 10)
  y <- 2 + 1.5 * x - 0.7 * x^2 + 0.05 * x^3 + rnorm(n, mean = 0, sd = ruido)
  data.frame(x = x, y = y)
}

dados <- funcao_geradora_dos_dados(n = 1000, ruido = 1)

plot(dados$x, dados$y, main = "Dados gerados", xlab = "x", ylab = "y")
```

onde a *seed* a utilizar em `set.seed(*****)` deve ser constituída pelos primeiros 3 algarismos do número de aluno do elemento do grupo que surgiria primeiro caso os nomes fossem ordenados por ordem alfabética e pelos últimos 3 algarismos do número de aluno do outro elemento do grupo. Caso o grupo tenha 3 elementos, utilize o mesmo procedimento repartindo os 6 algarismos em blocos de 2.

Os dados gerados deverão ser semelhantes aos exibidos no gráfico abaixo:

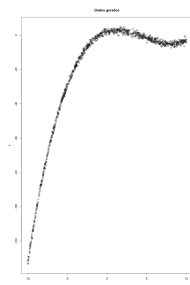


Figura 1: Exemplo de amostra gerada a partir da rotina descrita acima.

O relatório não pode exceder duas páginas A4 (incluindo gráficos) com letra 12pt e espaçamento entre linhas de 1.5. Deve conter:

1. Uma breve explicação dos métodos de validação cruzada.
2. Descrição dos modelos utilizados (pretende-se uma regressão linear simples e uma regressão polinomial de terceiro grau).
3. Utilização de diferentes números de *folds* e avaliação do Erro Quadrático Médio em termos do *variance-bias tradeoff*.
4. Análise crítica dos resultados obtidos.

Pode utilizar os *packages* `cvTools` ou `caret` para a sua análise. O relatório deve ser submetido eletronicamente até às 23h59 do dia 27 de Novembro de 2024.