PLOC – Trabalho 1

Ano letivo: 2021/22 2 de outubro de 2022

Data de entrega - 23h59, 14 de outubro de 2022

Nos ficheiros `trabalho1_2022_tabela.txt` e `trabalho1_2022_tabela.txt` encontramos duas colunas de valores. Em ambos os casos, a primeira coluna contém 50 números igualmente espaçados entre -3 e 3, enquanto que a segunda regista o valor que um certo polinómio toma nos valores da primeira. Infelizmente, em ambos os casos nos esquecemos de que polinómio avaliámos. Apenas sabemos que o grau é menor ou igual a 10. Mais ainda, as medições guardadas na segunda coluna contêm erros: todas elas têm ruído normalmente distribuído e numa certa percentagem de entradas há um erro por excesso adicional. O nosso objetivo é tentar recuperar o polinómio.

1. Usando o vosso software/linguagem preferido, criem uma função que dada uma lista x de valores, uma lista y de medições e um grau d, encontre o polinómio p de grau menor ou igual a d que minimize:

   (a) $\sum_i (p(x_i) - y_i)^2$;

   (b) $\sum_i |p(x_i) - y_i|$.

2. Aplica essas funções aos dados nos ficheiros:

   (a) Qual das normas parece dar melhores resultados?

   (b) Apresenta uma estimativa de qual o polinómio usado para criar cada uma das tarefas. Justifica a escolha do grau da aproximação.

NOTA: Se escrevermos um polinómio de grau d como $p(x) = a_0 + a_1 x + a_2 x^2 + ... + a_d x^d$, então se chamarmos a ao vector $(a_0, a_1, ..., a_d)^t$ dos coeficientes, temos que $p(x) = \langle a, (1, x, x^2, ..., x^d)\rangle$. Dados n valores diferentes para x, podemos criar a matriz:

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^d \end{pmatrix}$$

Então $Xa = (p(x_1), p(x_2), ..., p(x_n))$. Podemos então reescrever os problemas da pergunta 1 como encontrar o vector a que minimiza $\|Xa - y\|_2$ e $\|Xa - y\|_1$ respectivamente. O primeiro é simplesmente resolver um sistema no sentido dos mínimos quadrados, o segundo um problema de programação linear.