

## RP: Exercícios

Entrega: 16/03

**Questão 1.** O arquivo <https://www.dropbox.com/s/d3epin1n6e4821i/dados-ex1-1.csv> possui amostras de treinamento para um problema de classificação binária com  $\mathcal{X} = \mathbb{R}^2$  e  $\mathcal{Y} = \{0, 1\}$ . Os dados estão na forma de um arquivo `.csv` com três colunas:

1. **Coluna 1:** atributo numérico 1 ( $x_1$ );
2. **Coluna 2:** atributo numérico 2 ( $x_2$ );
3. **Coluna 3:** classe ( $y$ ).

Considerando o classificador  $k$ -NN, plote as superfícies de decisão no intervalo  $[-1, 1]^2$  para  $k = 1, 3, 9, 27$ . Qual a sua conclusão sobre o efeito do valor de  $k$  para o problema em questão?

**Questão 2.** Considere um problema de regressão com  $\mathcal{X} = \mathbb{R}$  e  $\mathcal{Y} = \mathbb{R}$ . Gere 3 conjuntos de dados com  $N = 4, 10, 100$  amostras da função alvo  $f(x) = 2x^2 \sin(15x)$  no intervalo  $x \in [-1, 1]$ . Em seguida, gere um conjunto independente de teste (entradas e saídas) com 1000 amostras no mesmo intervalo. Implemente o modelo  $k$ -NN para regressão conforme visto em sala, e para cada um dos 3 conjuntos, faça:

- Calcule o erro quadrático médio  $\mathcal{L}$  para predições no conjunto de treinamento e no conjunto de teste para  $k = 1, 3$  e  $5$ ;

$$\mathcal{L} = \frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2$$

- Plote a função aproximada pelo  $k$ -NN com  $k = 3$  para entradas no intervalo  $[-1, 1]$ .

**Sobre a entrega.** Um relatório deve ser enviado até a data especificada no início deste documento via *google classroom*. Preferencialmente, o relatório deve consistir de *um único python notebook* com as respostas, gráficos, comentários e códigos.

Opcionalmente, você pode enviar um arquivo `.zip` com os códigos e um único arquivo `.pdf` com as respostas e gráficos. Nesse caso, é altamente recomendada a escrita em *Latex*.

Cada relatório pode ser desenvolvido por até 2 pessoas.

Bons estudos!