Practical Sheet 2: k-Nearest Neighbours

APL 405 - 2023W (Machine Learning for Mechanics)

[20 marks] Predicting beam state with k-NN: Consider a synthetic binary classification problem (two classes) with a given training dataset having $N = 10$ (statistically independent) observations of 2-dimensional input variables $\mathbf{x} = [x_1 \ \ x_2]^T$ and one categorical output y, representing either *Damaged* or *Undamaged* state of linearly behaving elastic beam. x_1 represents the Young's modulus and x_2 represents the Poisson's ratio.

i	x_1	x_2	y
1	175.25×10^6	0.12	Damaged
2	206×10^6	0.10	Undamaged
3	165×10^6	0.11	Damaged
4	175.25×10^6	0.11	Undamaged
5	175.10×10^6	0.09	Undamaged
6	195.75×10^6	0.10	Damaged

1. Normalize the inputs (using min-max scaler) to have them in a similar range.
2. Write a function to calculate the Euclidean distance between two vectors. Using a test input $\mathbf{x}^* = [195.75 \times 10^6, \quad 0.11]^T$, compute the Euclidean distance between each training data point $\mathbf{x}^{(i)}$ and the test data point. Print the result in tabular fashion.
3. Output the k-NN prediction $\hat{y}(\mathbf{x}^*)$ for case (a) $k = 1$ (one neighbour), and case (b) $k = 3$ (three neighbours).
4. Repeat step (2) for the test input $\mathbf{x}^* = [185.50 \times 10^6, \ 0.11]^T$. What beam state prediction do you get for the two cases?
5. Repeat step (2) for the test input $\mathbf{x}^* = [190.63 \times 10^6, \ 0.11]^T$. What beam state prediction do you get for the two cases?
6. Plot the decision boundaries of the two k-NN classifiers.