

Problem n.2

The file `fish.txt` contains the coordinates (x, y) of 250 fishing locations in the Mediterranean sea. The fishing areas have been classified in two types $\{L, H\}$ based on the abundance of fish in the area (L = low, H = high).

- a) Build a classifier for the type of the fishing area based on its spatial coordinates. Report the model for the data, the estimates of its parameters (means and covariances), the priors within the groups and verify the model assumptions. Report the plot of the classification regions.
- b) Analyse the possible weaknesses of the model. Estimate the AER of the classifier through leave-one-out cross-validation.
- c) Build a k -nearest neighbor classifier for the type of the fishing area based on its spatial coordinates. Choose parameter k in the range $[10, 30]$ as to optimize the misclassification error, assessed via leave-one-out cross-validation (set the random seed equal to 19 prior to perform cross-validation). Report the error rate associated with the optimal classifier, and the plot of the classification regions.
- d) Using the best classifier among those at points (a) and (c), how would you classify a new fishing area located at position (10.8, 39.4)?

Upload your results here:

<https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9AS0Gwf2lHjvGX24HiqFVUQkJKQVNVrzhYVOE2N0J0TFFEN0xRWE5aRi4u>