Problem n.2

Accurate and granular wind forecasting is essential for wind-based sports involving the use of a sail such as sailing, windsurfing or paragliding. Accurate wind estimation requires the exact knowledge of air pressure at any location of the globe which is unfortunately not possible since air pressure is only measured at specific locations by weather stations. This motivates the use of statistical techniques to infer air pressure at any location from the locations where its value is known.

For the sake of simplicity we are only interested in knowing if the air pressure at a certain location is high or low, with respect to the average atmospheric pressure value ≈ 1013 hPa.

The file `pressure.txt` contains the coordinates (x, y) of 200 weather stations in Europe along with a binary variable $\{L, H\}$ indicating the pressure category ($L =$ low, $H =$ high) at the station location.

a) Build a classifier for the pressure level 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 pressure level 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), which pressure level would you predict at position (10.8,39.4)?

Upload your results here:
https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9AS0GWf2lHjvGX24Hiq
FVUNTBSMDFPVkVUU1pMWUNOVzlWNVRWNUpMSC4u