Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
September 9th, 2022

Problem n.2

The file `fossil.txt` contains the locations (latitute and longitude) of 182 fossil records of two types: perisphinctes and nerinea.

a) Build a classifier to identify the regions associated to the fossils of the two types. Report the model for the data, the estimates of its parameters (means and covariances) and verify the model assumptions. Report the plot of the classification regions.

b) Compute the APER. Estimate the AER of the classifier via cross validation.

c) Build a k-nearest neighbor classifier for the type of the fossil based on its spatial coordinates. Choose parameter k in the range $[5, 20]$ as to optimize the misclassification error, assessed via leave-one-out cross-validation (set the random seed equal to 9 prior to perform cross-validation). Report the cross validation error rate associated with the optimal classifier, and the plot of the classification regions.

d) Using the best classifier, how would you classify a new fossil that it is found at lon = -159.5 and lat = 21.9 ?

Upload your results here:
`https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9AS0GWf2lHjvGX24HiqFVUNUU3Mjk5RTZPSjFUVEFIQ1IzV1BBUUlNRy4u`