MA 751
Data Assignment 1
Due Tues. Mar. 1

Analyze the following quadratic discriminant analysis problem. Consider the vowel data at the text web site https://web.stanford.edu/~hastie/ElemStatLearn/data.html . There are 11 vowels to distinguish, based on 10 features for each training/test sample of a vowel. Please include all graphs and diagrams, and the code/commands that you have used, and turn in this assignment (in hardcopy) in class.

(a) Write a program to classify these vowels using linear discriminant analysis (LDA), which assumes that each class has the same covariance. This involves determining different centroids μ_k as well as a single common (pooled) covariance for the classes. You can use your own code or functions, e.g. in R, Python or Matlab.

(b) Write a program to classify vowels in this way using quadratic discriminant analysis, which fits a different Gaussian model for each of the 11 classes. Note this involves determining both a different centroid μ_k and a different covariance Σ_k for each class.

(c) Test your two models on the test data and determine the misclassification rate. Explain the models and procedures, illustrating your discussion where needed through qualitiative diagrams for visualization. Comment on the differences between these models.

(d) For LDA, try to increase your input space beyond the 10 features above by including quadratic and interaction terms (e.g. x_i^2, $x_i x_j$) as well (see the discussion at the end of section 4.1). How does this affect the results?