Homework 6

STAT 448 - Advanced Data Analysis

Due: Monday, December 4, 2017 at 11:00 pm

The data sets are contained in **HW6Data.sas** in the compass2g course space. For discriminant analysis, assume proportional priors.

1. (3 parts) For the Quitting Smoking Experiment data set described on data description pdf file:

    (a) Perform a discriminant analysis for quitting procedure (the **proc** variable in the data set) as a function of the 4 rating variables. Comment on what the MANOVA tests tell us about ability to discriminate quitting procedures based on these variables, and comment on the cross-validation error results and how well the discrimination matches the quitting procedures. Be sure to note quitting procedures that are often misclassified as other quitting procedures.

    (b) Perform a complete linkage cluster analysis based on the 4 rating variables. How many clusters would you choose based on the dendrogram and any relevant diagnostics?

    (c) Obtain 3 clusters and compare the frequencies for each quitting procedure within the clusters. Comment on how the clusters do or dont match the procedures, and which procedures are most prominent in each cluster. What does this tell us about similarities and differences of ratings for different quitting procedures?

2. (3 parts) For the Egyptian Skulls data set described on data description pdf file:

    (a) Perform a discriminant analysis for **epoch** as a function of the 4 skull measurements. Comment on what the MANOVA tests tell us about ability to discriminate **epoch**s based on these variables, and comment on the cross-validation error results and how well the discrimination matches the epochs. Be sure to note **epoch**s that are often misclassified as other **epoch**s.

    (b) Perform a complete linkage cluster analysis based on the 4 measurements. How many clusters would you choose based on the dendrogram and any relevant diagnostics?

(c) Obtain 5 clusters and compare the frequencies for **epoch** within the clusters. Comment on how the clusters do or dont match the **epoch**s. What does this tell us about similarities and differences of skull measurements for different **epoch**s?

3. Repeat Exercise 2 using only the measurements chosen in a stepwise discrimination as predictors for the classification. Comment on how the quality of the discrimination and clustering results have changed (e.g. how has the classification of **epoch**s changed and how has the clustering of **epoch**s changed?).