Multivariate Analysis: Homework 3

Zhaoxia Yu

2023-05-24

Due on Tuesday, May 30 2023. R Code should be included as appendices.

- Problem 1: Find an example that has three classes and choose two quantitative variables/features to analyze. Do NOT use any data set we have analyzed in class. Randomly select 10 observations for testing and the rest for training.

Method 1

a. Compute the pooled sample covariance, denoted as S_{pooled}.
b. For each data point in the testing data set, calculate its statistical distance to each of the three groups. Based on this, classify each data point into one of the three groups.
c. Visualize the classification results in the two-dimensional space of the two chosen features, with the x-axis representing the first feature and the y-axis representing the second feature.

Method 2

d. Compute the between-group covariance matrix, denoted as $\mathbf{B}$, and the within-group covariance matrix, denoted as $\mathbf{W}$.
e. Find the two linear discriminants by using the eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$. Visualize the directions of the two linear discriminants.
f. Project the training data onto the two linear discriminants and plot the projected data.
g. Transform the testing data to the same two-dimensional space using the linear discriminants. Calculate the distance between each data point and each of the three groups. Classify each data point into one of the three groups based on the minimum distance (Euclidean distance).

Compare Method 1 and Method 2

h. Do the two methods give you the same predictions?

- Problem 2: Read the following article. Then summarize what you learned from it by writing a paragraph with 100-300 words. Please comment on what extension did the paper make, and why it is useful extension.

"Penalized classification using Fisher's linear discriminant". Link of pdf: https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1467-9868.2011.00783.x