

Exercise 9

Problem 1: Linear Discriminant Analysis

Install the package **MASS** and use the function **lda()** to perform Fisher's linear discriminant analysis to the Fisher's iris data set. The data set can be accessed with the command: **data(iris)**. It contains measurements of sepal width, sepal length, petal width and petal length. In your analysis, consider the two species that are the most difficult to separate: "versicolor" and "virginica", i.e. leave the first 50 observations out.

- Use **lda()** to find the vector a . Verify that a is equal to the eigenvector of $W^{-1}B$ that corresponds to the largest eigenvalue.
- Suppose we have a new flower with the following measurements: (sepal length, sepal width, petal length, petal width) = (6,3,4,1). In which group will this flower be classified?
- Use the leave-one-out method to determine the missclassification rate.

Problem 2: Fisher's Linear Discriminant Function

Show that the solution for the problem:

$$\max_a \left\{ \frac{a^\top B a}{a^\top W a} \right\},$$

is obtained by setting a equal to the eigenvector of $W^{-1}B$ that corresponds to the largest eigenvalue.

Homework Assignment 9: Linear Discriminant Analysis

Perform the Fisher's linear discriminant analysis to the data set ALCOHOL.txt. Suppose that we have a new drink with the following measurements: MEOH=500, ACET=400, BU1=3, MEPR=30, ACAL=20, LNPRO1=10. Answer the following questions in your report and provide the requested figure for part (b).

- Give the vector a , such that $\|a\|_2 = 1$.
- Visualize the original data using a pairwise scatterplot. Add the new drink to the plot. Furthermore, use different colors for KIRSCH, MIRAB, POIRE and the new drink.
- In which group will the new drink be classified to?
- Provide the corresponding R-code for parts (a),(b) and (c).