Project #2

Milica Cudina

2024-03-15

Problem #1 (55 points)

The `iris` data set is built-in in `R`. Start by studying the documentation of the data set, i.e., by entering `?iris` in the console. To familiarize yourselves with the architecture of an iris flower, go to:

US Forest Service

Your next step is exploratory data analysis.

(10 points) Which plot would you use to display pairwise associations between different measurements? How do you make sure that the different species are color-coded? Display the plot and write a few sentences about your conclusions.

Principal Component Analysis (PCA)

(20 **points)** Perform the PCA on the explanatory components of the above data, provide the report, and the relevant plots.

Principal Components Regression (PCR)

Your next task is to predict `Sepal.Length` from the other variables in the `iris` dataset.

(15 points) Run the PCR, provide an explanation for the output, and display the relevant plots (both validation and prediction).

(10 points) Split your dataset into training (4/5 of the data) and testing (1/5 of the data). Provide the mean squared error and an appropriate plot.

Problem #2 (20+5+10+10=45 points)

Solve **Problem 3.7.15** (page 128) from the textbook.

Hint: The command `lapply` could be useful.