STAT 503 Homework 5

Due Sunday, April 10, 2022
Submit online on Canvas by 11:59pm

Part 1. Let the population covariance matrix be given by

$$\Sigma = \begin{bmatrix} 9 & -3 \\ -3 & 4 \end{bmatrix}$$

(a) Find the *second* population principal component direction based on Σ, and the population variance along this direction.

(b) Find the projection of the data points contained in the following data matrix

$$X = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}$$

onto the second population principal component direction you found in (a).

Part 2. The auto data set (`auto-mpg.txt` on Canvas or `Auto` in the ISLR package) concerns city-cycle fuel consumption in miles per gallon (mpg) and other attributes collected for 398 vehicle instances. The variables are: mpg, cylinders, displacement, horsepower, weight, acceleration, model year, origin and car name.

1. Perform PCA on this dataset. In particular:
   (a) Describe the data and present some initial pictorial and numerical summaries, such as scatterplots, histograms, etc.
   (b) Consider which variables should or should not be included in PCA. Compare PCA on covariances and correlations and choose one to proceed with for the subsequent questions.
   (c) Comment on the percentage of variance explained and number of principal components to retain. Include a scree plot.
   (d) Comment on variable loadings and their potential interpretations.
   (e) Make a plot of the data projected on the first two PCs. Comment on any interesting features, including potential outliers, if any.
   (f) Make a PCA biplot and comment on any interesting features.
2. Construct a dissimilarity measure that will include at least one of the categorical variables you had to remove for PCA. Explain the rationale for your dissimilarity. Perform MDS using this measure and plot the resulting projection into 2-d. Comment on similarities/differences with the projection onto the first two PCs.
3. Pick one non-linear dimensionality reduction method, apply to the data, and plot the resulting projection into 2-d. Comment on similarities/differences with the plots you obtained for PCA and MDS.