COLORADOSCHOOLOFMINES

MATH437/537 **Fall, 2022**

Homework 1. Due September 9 (submit through CANVAS)

1. Consider the lizard data in Example 1.6 in the textbook (Wichern and Johnson). The data file can also be downloaded from the CANVAS page. (In Chapters 1-3 of the book by Zelterman, you can see examples of how to use R code to answer the questions.)

   (a) Make boxplots for SVL and HLS next to each other. Is there evidence that the median SVL is different from the median HLS?

   (b) Make a matrix plot with scatterplots for each pair of variables (m, SVL, HLS) with a lowess curve fit in each (use the function 'pairs' in R). Is there evidence of nonlinear correlations?

   (c) What is a bivariate boxplot?

   (d) Make a bivariate boxplot for the variables SVL and HLS (R function 'bvbox' in library 'MVA').

   (e) Determine the 3×3 sample covariance matrix and the 3×3 sample correlation matrix of the data.

   (f) Compute the mean and median of each variable.

   (g) Compute the standard deviation and MAD (median of the absolute deviations from the median) of each variable.

2. Draw an iid sample of size 500 from the Gaussian mixture

$$0.99\, N(0,1) + 0.01\, N(3,16).$$

   (a) Make a boxplot of the sample.

   (b) Make a histogram of the sample.

   You may use the following R-code to generate a sample from the Gaussian mixture:

```
N = 500
components = sample(1:2,prob=c(.99,.01),size=N,replace=TRUE)
mus = c(0,3)
sds = sqrt(c(1,4))
samples = rnorm(n=N,mean=mus[components],sd=sds[components])
plot(samples)
```

3. Use $n = 500$ equally spaced points $x_1, \ldots, x_n$ in $[0, 2\pi]$ to simulate observations

$$y_{i,j} = \sin(x_j) + \epsilon_{i,j} \qquad i = 1, \ldots 500,\ j = 1, \ldots, 500$$

where the random variables $\epsilon_{i,j}$ are iid from the Gaussian mixture

$$0.99\, N(0, 0.1^2) + 0.01\, N(0, 200^2).$$

This leads to a 500×500 data matrix $\boldsymbol{Y}$. (You may modify the R-function in question (2) to generate the noise from the Gaussian mixture.)

(a) Let $\boldsymbol{m}$ and $\boldsymbol{s}$ be the column means and standard deviations. Make a plot that shows $\boldsymbol{m}$, $\boldsymbol{m} + \boldsymbol{s}$ and $\boldsymbol{m} - \boldsymbol{s}$ all vs x on the same plot.

(b) Repeat (a) but using the median and mad instead of the mean and standard deviation

(c) Comment on the results.

4. Let $\boldsymbol{S}$ be the matrix

$$\boldsymbol{S} = \begin{pmatrix} 1.1 & 2.1 & 3.1 & 4.1 \\ 2.1 & 4.1 & 6.1 & 8.1 \\ 3.1 & 6.1 & 9.1 & 12.1 \end{pmatrix}.$$

(a) Determine a SVD of $\boldsymbol{S}$ (you may use code), write the orthogonal matrices and the matrix of singular values.

(b) Find the best rank-one approximation of $\boldsymbol{S}$.