Tutorial - Week 2

Please read the related material and attempt these questions before attending your allocated tutorial. Solutions are released on Friday 4pm.

Question 1

Consider the matrix

$$\begin{pmatrix} -1 & 3 & -2 \\ 2 & 4 & 2 \\ 5 & 2 & 3 \end{pmatrix}.$$

(a) Calculate the matrix of deviations (residuals), given by

$$\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}'$$

where $\bar{\mathbf{x}}$ is the mean vector of observations and $\mathbf{1}$ is a vector of 1's. Is this matrix of full rank? Explain.

(b) Determine the sample covariance matrix $\mathbf{S}$ and calculate the sample generalised variance $|\mathbf{S}|$. Interpret the latter geometrically.

(c) Calculate the *sample total variance* which is given by the sum of the diagonal elements of the sample covariance matrix $\mathbf{S}$. The total variance is another measure of variance (i.e., a number to describe the covariance of the data).

Question 2

Consider the matrices

$$A = \begin{pmatrix} 4.000 & 4.001 \\ 4.001 & 4.002 \end{pmatrix}, \qquad B = \begin{pmatrix} 4 & 4.001 \\ 4.001 & 4.002001 \end{pmatrix}.$$

These matrices are identical except for a small difference in the (2,2) position. Moreover, the columns of $\mathbf{A}$ (and $\mathbf{B}$) are nearly linearly dependent. Show that

$$\mathbf{A}^{-1} = (-3)\mathbf{B}^{-1}.$$

Consequently, small changes – perhaps caused by rounding – can give substantially different inverses.

Question 3

In this question, we consider the Fashion MNIST dataset and use it as an example data set for calculating the sample total variation. Please refer to Workshop 1 for examples of how to load and manipulate the Fashion MNIST dataset. Details on the Fashion MNIST dataset is found here: https://github.com/zalandoresearch/fashion-mnist.

(a) Turn each (image) 28×28-sized observation of Fashion MNIST data set into a vector observation of size $p = 784$.

(b) Download the Fashion MNIST labels and use them to calculate the sample total variation of each clothing class.