MAST90104: A First Course in Statistical Learning

Assignment 1

Late assignments will be penalised two marks for each day overdue.
Extensions should be supported by a medical certificate. Requests for extensions must come on or before the due date for the assignment.

Please submit a scanned or other electronic copy of your work via the Learning Management System in one file - see this link for instructions

The .pdf must have in **one file**:

- handwritten or typed answers to the questions
- handwritten or typed R code used to produce your answers
- graphics required to answer the questions

If you have more than one file submitted, *only the last LMS .pdf file with your name on it will be marked.*

This assignment is worth 5% of your total mark. You must fill in the online plagiarism declaration form prior to submitting your assignment

You may use R for this assignment, but for matrix calculations only (you may not use the `lm` function). If you do, include your R commands and output.

1. Suppose that A is a symmetric matrix with $A^k = A^{k+1}$ for some integer $k \geq 1$. Show that A is idempotent.

2. Let $A_1, A_2, \ldots, A_m$ be a set of symmetric $k \times k$ matrices. Suppose that there exists an orthogonal matrix P such that $P^T A_i P$ is diagonal for all i. Show that $A_i A_j = A_j A_i$ for every pair $i, j = 1, 2, \ldots, m$.

3. Show that for any random vector $\mathbf{y}$ and compatible matrix A, we have $\text{var } A\mathbf{y} = A(\text{var } \mathbf{y})A^T$.

4. Let $\mathbf{y}$ be a 3-dimensional multivariate normal random vector with mean and variance
$$\boldsymbol{\mu} = \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix}, \quad V = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$
Let
$$A = \frac{1}{10}\begin{bmatrix} 4 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 10 \end{bmatrix}.$$
   (a) Describe the distribution of $A\mathbf{y}$.
   (b) Find $E[\mathbf{y}^T A\mathbf{y}]$.
   (c) Describe the distribution of $\mathbf{y}^T A\mathbf{y}$.
   (d) Find all linear combinations of $\mathbf{y}$ elements which are independent of $\mathbf{y}^T A\mathbf{y}$.

5. The table below shows prices in US cents per pound received by fishermen and vessel owners for various species of fish and shellfish in 1970 and 1980. (Taken from Moore & McCabe, Introduction to the Practice of Statistics, 1989.)

Type of fish	Price (1970)	Price (1980)
Cod	13.1	27.3
Flounder	15.3	42.4
Haddock	25.8	38.7
Menhaden	1.8	4.5
Ocean perch	4.9	23.0
Salmon, chinook	55.4	166.3
Salmon, coho	39.3	109.7
Tuna, albacore	26.7	80.1
Clams, soft-shelled	47.5	150.7
Clams, blue hard-shelled	6.6	20.3
Lobsters, american	94.7	189.7
Oysters, eastern	61.1	131.3
Sea scallops	135.6	404.2
Shrimp	47.6	149.0

We will model this data using a linear model.

(a) The linear model is of the form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Write down the matrices and vectors involved in this equation.

(b) Using matrices, find the least squares estimators of the parameters.

(c) Calculate the sample variance s^2.

(d) A fisherman sold ocean trout for 18c/pound in 1970. Predict the price for ocean trout in 1980.

(e) Calculate the standardised residual for sea scallops.

(f) Calculate the Cook's distance for sea scallops.

(g) Does sea scallops fit the linear model? Justify your argument.