MAST90104: A First Course in Statistical Learning

Assignment 1, 2023

Due: 11:59 pm Sunday August 13. Please submit a scanned or electronic copy of your work via the Learning Management System. Late submissions will have their score deducted (10% for every 12 hrs late)

This assignment is worth 5% of your total mark.

You may use R for this assignment, but only for question 4. If you do, include your R commands and output in your answer.

1. (3pt) Let X be an $n \times m$ matrix and Y be an $m \times n$ matrix. Prove that $tr(XY) = tr(YX)$.
2. (3 pt) Show that for any random vector $\mathbf{y}$ and compatible matrix A, we have var $A\mathbf{y} = A(\text{var } \mathbf{y})A^T$.
3. (6 pt) Let $\mathbf{y}$ be a 3-dimensional multivariate normal random vector with mean and variance

$$\boldsymbol{\mu} = \begin{bmatrix} -4 \\ 2 \\ 1 \end{bmatrix}, \quad V = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 4 \end{bmatrix}.$$

Let

$$A = \begin{bmatrix} 1 & -3 & 2 \\ -3 & 2 & -1 \\ 2 & -1 & 3 \end{bmatrix}.$$

   (a) Describe the distribution of $A\mathbf{y}$.

   (b) Find $E[\mathbf{y}^T A\mathbf{y}]$.

   (c) Does $\mathbf{y}^T A\mathbf{y}$ have a (noncentral) chi-square distribution? Explain your answer.

4. (8pt) A researcher is interested in predicting price of houses and have collected data from several suburbs of a big city. The following table contains measurements from 10 houses in the collected data. The 3 variables are price (in \$10,000s), distance from the suburb to the city's employment centres and pupil-teacher ratio in the area.

Price	37.0	32.5	48.5	13.9	14.4	24.1	22.2	21.7	19.2	21.4
Distance	6.48	2.85	5.12	4.45	1.44	6.27	7.04	5.45	2.75	1.95
Ratio	15.2	17.8	14.7	21.0	21.2	14.7	16.6	15.2	17.8	20.2

Using this small dataset, we will build a linear model to predict house prices based on distance to CBD and pupil-teacher ratio.

   (a) Plot the price of houses against distance to CBD and pupil-teacher ratio (*Hint: You need to produce 2 plots. You can use the function* `plot()` *in R*)

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

   (c) Is this model a full rank model? Explain your answer.

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

   (e) Estimate the price of a house in a suburb that is 3 kilometres from the city's employment centres. The pupil-teacher ratio in this area is 17.5.