MAST90104: A First Course in Statistical Learning

Assignment 1, 2024

Due: 6:00 pm Friday August 16. 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. (4pt) Let $\mathbf{X}$ be an $n \times m$ real-valued matrix. Prove that $\mathbf{X}^T\mathbf{X} = \mathbf{0}$ if and only if $\mathbf{X} = \mathbf{0}$. *Hint: To prove an "if and only if" result, say "statement A if and only if statement B", first assume that statement A is true and prove that statement B is true. Next, assume that statement B is true and prove that statement A is true.*

2. (2 pt) Let $\mathbf{X}$ be a square matrix and ν an eigenvalue of $\mathbf{X}$. Show that ν is an eigenvalue of $\mathbf{X}^T$. *Hint: For any real-valued matrix* $\mathbf{A}$*, we have* $|\mathbf{A}| = |\mathbf{A}^T|$.

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

$$A = \begin{bmatrix} 4 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & 3 \end{bmatrix}.$$

   (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	37.3	32.1	47.5	14.2	14.0	23.7	22.6	21.7	19.5	22.0
Distance	6.40	2.80	5.15	4.40	1.50	6.30	7.10	5.55	2.85	2.00
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.* Be sure to provide your codes in your submission)

   (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) A house is located in a suburb that is 3 kilometres from the city's employment centres. Based on your estimated linear regression fit, what is the largest pupil-teacher ratio such that the estimated price is at least \$250,000? Round your final answer to 2 decimal places.