MAST90104: A First Course in Statistical Learning

Assignment 1, 2023 Solution

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 $tr(XY) = tr(YX)$.

   Solution Let $C = XY$, easy to see that C is a $n \times n$ matrix, and the diagonal elements of C are
   $$c_{ii} = \sum_{j=1}^{m} x_{ij} y_{ji}.$$
   Therefore, the trace of C is
   $$tr(C) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij} y_{ji}$$
   Similarly, $D = YX$ is a $m \times m$ matrix with diagonal elements
   $$d_{jj} = \sum_{i=1}^{n} y_{ji} x_{ij}$$
   and trace of D is
   $$\begin{aligned} tr(D) =& \sum_{j=1}^{m} \sum_{i=1}^{n} y_{ji} x_{ij} \\ =& \sum_{j=1}^{m} \sum_{i=1}^{n} x_{ij} y_{ji} \\ =& \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij} y_{ji} \\ =& tr(C) \end{aligned}$$
   This proves that $tr(C) = tr(D)$.

2. (3 pt) Show that for any random vector $\mathbf{y}$ and compatible matrix A, we have $\text{var}\ A\mathbf{y} = A(\text{var}\ \mathbf{y})A^T$.

   Solution:

   Note that if y is a random vector of length n, then $\text{var}\ \mathbf{y}$ is an $n \times n$ matrix with entries $[\text{var}\ \mathbf{y}]_{ij} = \text{cov}(y_i, y_j)$. Then $Q = A(\text{var}\ \mathbf{y})A^T$ has elements $Q_{ij} = \sum_{h=1}^{n} \sum_{k=1}^{n} a_{ik} a_{jh} \text{cov}(y_k, y_h)$

   Assume that A is an $m \times n$ matrix, $\mathbf{z} = A\mathbf{y}$ is a random vector of length m with element $z_i = \sum_{k=1}^{n} a_{ik} y_k$.

   $V = \text{var}\ \mathbf{z}$ has elements $V_{ij} = \text{cov}(z_i, z_j) = \sum_{h=1}^{n} \sum_{k=1}^{n} a_{ik} a_{jh} \text{cov}(y_k, y_h)$.

   Alternatively, let $\boldsymbol{\mu} = E[\mathbf{y}]$. From the definition,
   $$\begin{aligned} \text{var}\ A\mathbf{y} &= E[(A\mathbf{y} - A\boldsymbol{\mu})(A\mathbf{y} - A\boldsymbol{\mu})^T] \\ &= E[A(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T A^T] \\ &= A\, E[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T]\, A^T \\ &= A(\text{var}\ \mathbf{y})A^T. \end{aligned}$$

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 $A\mathbf{y}$. (2pt)

Solution $A\mathbf{y} \sim MVN(A\mu, AVA^T)$, where

$$A\mu = \begin{bmatrix} -8 \\ 15 \\ -7 \end{bmatrix}, \quad AVA^T = \begin{bmatrix} 19 & -11 & 18 \\ -11 & 23 & -20 \\ 18 & -20 & 40 \end{bmatrix}.$$

(b) Find $E[\mathbf{y}^T A\mathbf{y}]$. (2pt)

Solution $E[\mathbf{y}^T A\mathbf{y}] = tr(AV) + \mu^T A\mu = 66$

(c) Does $\mathbf{y}^T A\mathbf{y}$ have a (noncentral) chi-square distribution? Explain your answer. (2pt)

Easy to check that AV is not idempotent. So $\mathbf{y}^T A\mathbf{y}$ does not follow a chi-square distribution.

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*) (2pt)

Solution Refer to Figure 1.

```
plot(ratio,price,xlab = 'pupil-teacher ratio',ylab = 'price',main = 'plot 1')
plot(dis,price,xlab = 'distance',ylab = 'price', main = 'plot 2')
```

(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. (1pt)

Solution

$$\mathbf{y} = \begin{bmatrix} 37.0 \\ 32.5 \\ 48.5 \\ 13.9 \\ 14.4 \\ 24.1 \\ 22.2 \\ 21.7 \\ 19.2 \\ 21.4 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 6.48 & 15.2 \\ 1 & 2.85 & 17.8 \\ 1 & 5.12 & 14.7 \\ 1 & 4.45 & 21.0 \\ 1 & 1.44 & 21.2 \\ 1 & 6.27 & 14.7 \\ 1 & 7.04 & 16.6 \\ 1 & 5.45 & 15.2 \\ 1 & 2.75 & 17.8 \\ 1 & 1.95 & 20.2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \end{bmatrix}$$

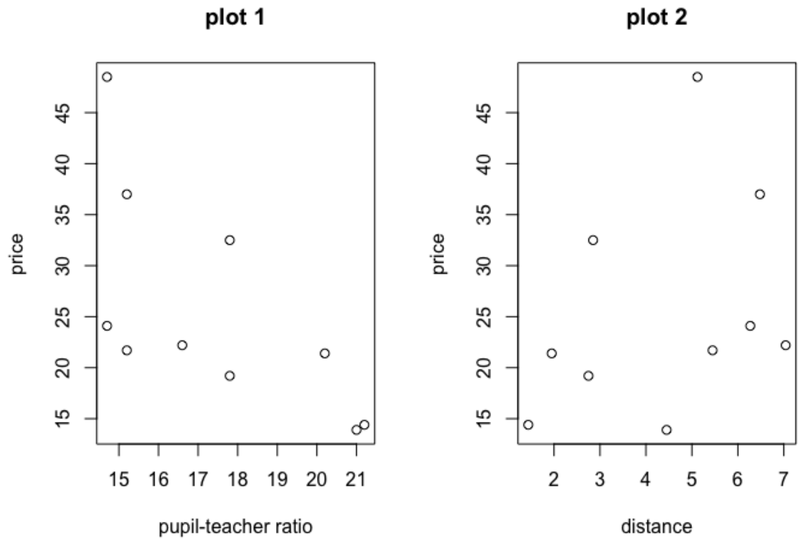

Figure 1: Plots of price vs distance and price vs pupil-teacher ratio.

(c) Is this model a full rank model? Explain your answer. (1pt)

Solution Yes, easy to check that X is of full rank

There are several ways to check, for example using R to check rank, or using $r(X) = r(X^T X) = 3$

(d) Using matrices, find the least squares estimators of the parameters. (2pt)

Solution

$\mathbf{b} = (X^T X)^{-1} X^T \mathbf{y}$

```
> price = c(37.0 ,32.5, 48.5, 13.9, 14.4, 24.1, 22.2, 21.7, 19.2, 21.4)
> dis = c(6.48, 2.85, 5.12, 4.45, 1.44, 6.27, 7.04, 5.45, 2.75, 1.95 )
> ratio = c( 15.2 ,17.8, 14.7, 21.0, 21.2, 14.7, 16.6, 15.2, 17.8, 20.2)
> X = cbind(rep(1,10),dis,ratio)
> (betahat= solve(t(X)%*%X,t(X)%*%price))
           [,1]
      102.058797
dis    -1.937970
ratio  -3.903698
```

(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 the area is 17.5. (2pt)

Solution We can estimate the price of the new observation by

$$\hat{y}^* = b_0 + b_1 x_1^* + b_2 x_2^* = (\boldsymbol{x}^*)^\top \boldsymbol{b},$$

where $\boldsymbol{x}^* = \begin{bmatrix} 1 \\ 3 \\ 17.5 \end{bmatrix}$.

The predicted price is

```
> (c(1,3,17.5)%*%betahat)
         [,1]
[1,] 27.93018
```

Or $\$279,300$.