# MAST90104: A First Course in Statistical Learning

## Assignment 1, 2024 Solution

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)

   Note:

   (⇒)For any $n$ by $m$ matrix, we may write:

   $$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_m \end{bmatrix},$$

   where each $\mathbf{x}_j$ is a column vector of size $n$ for each $j = 1, \dots, m$. Then, we have

   $$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T\mathbf{x}_1 & \mathbf{x}_1^T\mathbf{x}_2 & \dots & \mathbf{x}_1^T\mathbf{x}_m \\ \mathbf{x}_2^T\mathbf{x}_1 & \mathbf{x}_2^T\mathbf{x}_2 & \dots & \mathbf{x}_2^T\mathbf{x}_m \\ & \vdots & \vdots & \\ \mathbf{x}_m^T\mathbf{x}_1 & \mathbf{x}_m^T\mathbf{x}_2 & \dots & \mathbf{x}_m^T\mathbf{x}_m \end{bmatrix}.(1m)$$

   Since $\mathbf{X}^T\mathbf{X} = \mathbf{0}$, we have $\mathbf{x}_j^T\mathbf{x}_j = 0$ (1m), for each $j = 1, \dots, m$. By (***) below, $\mathbf{x}_j = \mathbf{0}$ for all $j = 1, \dots, m$ (1m).

   (***) $\mathbf{b}^T\mathbf{b} = \mathbf{0}$ if and only if $\mathbf{b} = \mathbf{0}$.
   Now, let $f(\mathbf{b}) = \mathbf{b}^\top\mathbf{b}$. Then, vector calculus yields $\nabla_\mathbf{b} f = 2\mathbf{b}^T$ and hence the only critical point occurs at $\mathbf{b} = \mathbf{0}$. Also, have $\partial^2 f/\partial\mathbf{b}\partial\mathbf{b}^T = 2\mathbf{I}$ for all $\mathbf{b} \in \mathbb{R}^n$. Hence, $f$ is convex everywhere and attains its global minimum at $\mathbf{b} = \mathbf{0}$ and consequently $\mathbf{b} = \mathbf{0}$ is the only solution to the equation $\mathbf{b}^T\mathbf{b} = \mathbf{0}$. The converse direction of the proof is trivial

   Note: Acceptable to quote result that $\mathbf{b}^\top\mathbf{b} = \mathbf{0}$ if and only if $\mathbf{b}$ is a zero column vector without proof.

   (⇐) The proof in this direction is obvious. (1m)

2. (2 pt) $\nu$ is an eigenvalue of $\mathbf{X} \Rightarrow |\mathbf{X} - \nu\mathbf{I}| = 0 \Rightarrow |(\mathbf{X} - \nu\mathbf{I})^T| = 0$ (1m) $\Rightarrow |\mathbf{X}^T - \nu\mathbf{I}| = 0 \Rightarrow \nu$ is an eigenvalue of $\mathbf{X}^T$ (1m).

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

$$A\mu = \begin{bmatrix} 16 \\ -7 \\ 7 \end{bmatrix}, \quad AVA^T = \begin{bmatrix} 54 & -10 & 39 \\ -10 & 7 & 10 \\ 39 & -10 & 40 \end{bmatrix}.$$

(b) Find $E[\mathbf{y}^T A\mathbf{y}]$. (2pt)
**Solution** $E[\mathbf{y}^T A\mathbf{y}] = tr(AV) + \mu^T A\mu = 95$

(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 non-central chi-square distribution.

4. (8pt)

(a) 1m for each plot
**Solution** Refer to Figure 1.

```
> price = c(37.3, 32.1, 47.5, 14.2,  14.0, 23.7, 22.6, 21.7, 19.5, 22.0)
> dis = c(6.40, 2.80, 5.15, 4.40, 1.50, 6.30, 7.10, 5.55, 2.85, 2.00 )
> ratio = c( 15.2 ,17.8, 14.7, 21.0, 21.2, 14.7, 16.6, 15.2, 17.8, 20.2)
> par(mfrow=c(1,2))
> plot(ratio,price,xlab = ‘pupil-teacher ratio’,ylab = ‘price’,main = ‘plot 1’)
> plot(dis,price,xlab = ‘distance’,ylab = ‘price’, main = ‘plot 2’)
```

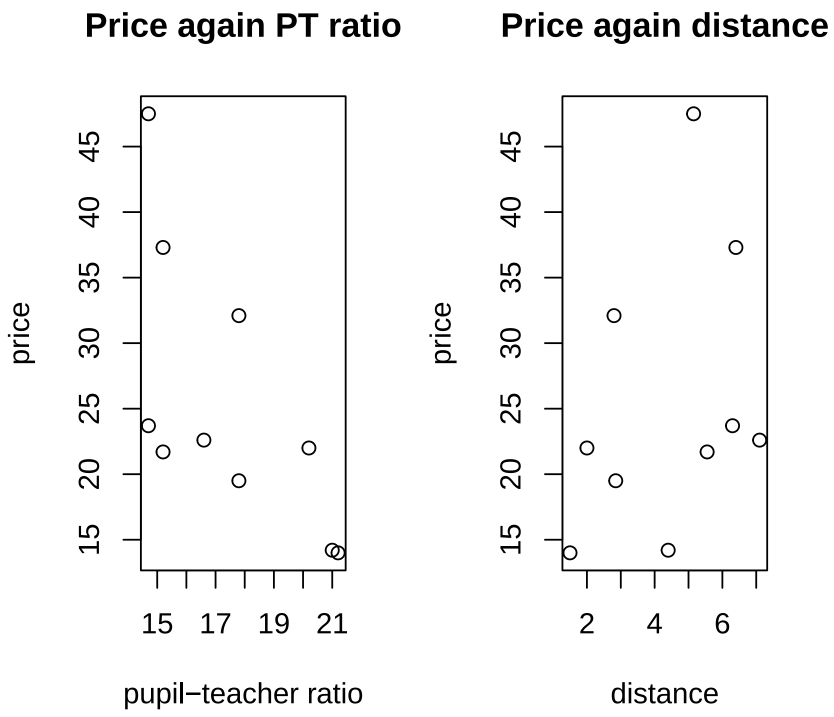


Figure 1: Plots of price vs distance and price vs pupil-teacher ratio.

(b) 1m if all vectors and matrices are correct.

**Solution**

$$\mathbf{y} = \begin{bmatrix} 37.3 \\ 32.1 \\ 47.5 \\ 14.2 \\ 14.0 \\ 23.7 \\ 22.6 \\ 21.7 \\ 19.5 \\ 22.0 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 6.40 & 15.2 \\ 1 & 2.80 & 17.8 \\ 1 & 5.15 & 14.7 \\ 1 & 4.40 & 21.0 \\ 1 & 1.50 & 21.0 \\ 1 & 6.30 & 14.7 \\ 1 & 7.10 & 16.6 \\ 1 & 5.55 & 15.2 \\ 1 & 2.85 & 17.8 \\ 1 & 2.00 & 20.2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \end{bmatrix}$$

(c) 1m for any valid explanation or R output

**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) **Solution** 1m for formula of least squares estimator. 1m for correct R output.

$\widehat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y}$

```
> X = cbind(1,dis,ratio)
> (betahat= solve(t(X)%*%X,t(X)%*%price))
           [,1]
      101.651873
dis    -1.986060
ratio  -3.867161
```

(e) **Solution** 1m for setting up equations correctly (see first line below). 1m for correct answer.

$$\begin{aligned} &\begin{pmatrix} 1 \\ 3 \\ x_2^\star \end{pmatrix}^T \widehat{\boldsymbol{\beta}} \geq 25.0 \\ &\Leftrightarrow 101.651873 - 1.986060 \times 3 - 3.867161 x_2^\star \geq 25.0 \\ &\Leftrightarrow x_2^\star \leq 18.28051 \end{aligned}$$

Therefore, the largest pupil-teacher ratio is 18.28.