

# MAST30025 Assignment 1 S1 2021 (1/3)

Michael Le (998211)

March 17, 2021

## Question 1 Solution:

Part a:

$$A^2 = A^3$$

Suppose  $A$  is a square matrix is (real and) symmetric then its eigenvalues are all real, and its eigenvectors are orthogonal.

### **Theorem 2.3**

Proof:

Take  $A$  to be a square matrix,  $n \times n$ . First we diagonalise  $A$ , i.e., find  $P$  such that.

$$= D = P^T A P$$
$$= \begin{bmatrix} \lambda_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \lambda_k \end{bmatrix}$$

where  $\lambda_1, \lambda_2, \dots, \lambda_k$  are the eigenvalues of  $A$ .

Since  $P$  is orthogonal both  $P$  and  $P^T$  are non-singular,

$$r(P^T A P) = r(P^T A) = r(A)$$

Because  $P^T A P$  is diagonal  $r(P^T A P)$  is the number of non zero eigenvalues of  $A$ .

But we wanted to prove **Theorem 2.2**

that  $A$  any symmetric matrix is idempotent. Which has eigenvalues of  $\lambda = 0$  or  $\lambda = 1$ .

The eigenvalues of idempotent matrices are always either

$$\lambda = 0 \text{ or } \lambda = 1.$$

$$A^2 = \lambda^2 x$$

Multiplying by  $A$ !!!

$$A^3 x = A^2 \lambda x = \lambda A^2 x = \lambda^3 x$$

$$(\lambda^3 - \lambda^2)x = 0$$

By definition,  $x \neq 0$ ,

$$\lambda^3 - \lambda^2 = 0$$

$$\lambda^2(\lambda - 1) = 0$$

Therefore there are two values with eigenvalues of 0 and one eigenvalue of 1! satisfies this theorem that A is idempotent!

Part b:

$$A = A^3$$

$$A^3 x = A \lambda x = \lambda A x = \lambda^3 x$$

Using the same theorem from the previous it has eigenvalues of 0,1 and -1. Since we care that A has to be positive semi-definite. Which has an eigenvalue of -1. Which does not satisfy Theorem 2.2! A is not idempotent!

### Question 2 Solution:

#### Theorem 2.4

There exists a matrix **P** which diagonalises  $A_1, \dots, A_m$ .

$$P^T A_i P = D_i$$

and

$$P^T A_j P = D_j$$

We take  $A_i$  and  $A_j$  to be  $k \times k$  matrices first we diagonalizes  $A_i, A_j$ , i.e. find P such that,

$$D_i = P^T A_i P = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \dots & \lambda_2 & \dots \\ 0 & \dots & \lambda_k \end{bmatrix}$$

for  $i = 1, \dots, k$

$$D_j = P^T A_j P = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \dots & \lambda_2 & \dots \\ 0 & \dots & \lambda_m \end{bmatrix}$$

for  $j = 1, \dots, m$

Proof:

$$A_i A_j = A_j A_i$$

$$P^T A_i A_j P = A_j A_i$$

$$P^T A_i A_j P = P^T A_i P P^T A_j P = D_i D_j$$

$$D_i D_j = P^T A_i A_j P = P^T A_i P P^T A_j P = P^T A_j P P^T A_i P = P^T A_j A_i P = D_j D_i$$

**Question 3 Solution:**

***Pre Proof Using Theorem 2.3***

For any matrix A

$$r(A) = r(A^T) = r(A^T A) = \text{tr}(A)$$

$$A = \begin{bmatrix} | & | & \dots & | & \dots & | \\ a_1 & a_2 & \dots & a_p & \dots & a_n \\ | & | & \dots & | & \dots & | \end{bmatrix}$$

Given A matrix with dimensions n x p with p independent columns.

Let  $x_1, x_2, x_3, \dots, x_k$  the basis for column space of A.

Definition of basis every column vector of A is a linear combination of the column vectors of x.

$$a_1 = b_1 x_1 + b_2 x_2 + \dots + b_k x_k$$

Definition of linear combination

where b is scalar

$$B = \begin{bmatrix} - & - & b_1 & - & - \\ - & - & b_2 & - & - \\ & & | & & \\ - & - & b_p & - & - \end{bmatrix}$$

$$\begin{bmatrix} | & | & \dots & | & \dots & | \\ a_1 & a_2 & \dots & a_p & \dots & a_n \\ | & | & \dots & | & \dots & | \end{bmatrix} = \begin{bmatrix} | & | & \dots & | & \dots & | \\ x_1 & x_2 & \dots & x_p & \dots & x_n \\ | & | & \dots & | & \dots & | \end{bmatrix} \begin{bmatrix} - & - & b_1 & - & - \\ - & - & b_2 & - & - \\ & & | & & \\ - & - & b_p & - & - \end{bmatrix}$$

$$A = XB$$

$$A^T = (XB)^T = B^T X^T$$

$$r(A) \leq r(A^T) \text{ or } r(A) \geq r(A^T) \text{ to satisfy!}$$

$$r(A) = r(A^T) = r(A^T A) = \text{tr}(A) = p$$

Since P is orthogonal both P and  $P^T$  are non-singular. Therefore we need to sum up the diagonal elements

$$r(A) = r(P^T A P) = \text{tr}(P^T A P) = \text{tr}(P P^T A) = \text{tr}(A) = p$$

Because  $D = P^T A P$  is diagonal  $r(P^T A P)$  is the number of nonzero values of A!

But A is idempotent so it takes eigenvalues between 0 or 1. To Prove Theorem 2.7! We need only the identity matrix to allow  $A^T A$  to be positive definite!

***Using Theorem 2.7***

**Proof ( $\Leftarrow$ ) :**

We want  $A^T A$  to be symmetric

and have all the eigenvalues to be strictly positive to prove  $A^T A$  is a positive definite matrix!

we know  $r(A^T A) = p$  is a  $p \times p$  matrix so it has to be a full rank matrix, p!

Let,  $\lambda_1, \lambda_2, \dots, \lambda_p > 0$  be the eigenvalues of  $A^T A$  for every  $x$  and for each eigenvalue has to have a value of 1.

for  $z = P^T x = (z_1, \dots, z_p)^T$

$$x^T (A^T A) x = x^T P D P^T x = z^T D z = \sum_{i=1}^p z_i^2 \lambda_i$$

since  $\lambda_i = 1$ !

$$= \sum_{i=1}^p z_i^2$$

$> 0$

Thus  $A^T A$  is positive definite as required!

**Proof ( $\Rightarrow$ ) :**

Suppose  $A^T A$  is positive definite let  $x_i$  be its normalised  $i$ -th eigenvector then,

$$x_i^T (A^T A) x_i = \lambda_i x_i^T x_i = \lambda_i$$

From theorem 2.3 we want  $A^T A$  to be symmetric and idempotent. We want the eigenvalues to be 0 or 1. This case all of the eigenvalues must equal to 1.

$\lambda_i = 1 > 0$

So, the eigenvalues of  $A^T A$  are strictly positive as required!!

#### Question 4 Solution:

Part a:

Given information:

Let,

$x_1, x_2, x_3 \sim (N(\mu, \sigma^2))$  be a sequence of independent normal random variables,

$$\bar{x} = \frac{x_1 + x_2 + x_3}{3}$$

$$\mathbf{x}^T = (x_1, x_2, x_3)^T$$

Supposed to be  $x^T$  as noted!

$$\mathbf{y} = (x_1 - \bar{x}, x_2 - \bar{x}, x_3 - \bar{x})^T$$

To solve A from:

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

$$\begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

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

Where A is symmetric and idempotent!

Part b: Finding the rank of A

Proof: that there is a linear combination for any columns?

```
%% {r}
A = matrix(c(2,-1,-1,-1,2,-1,-1,-1,2)/3,3,3)
%%
```

```
      [,1]      [,2]      [,3]
[1,] 0.6666667 -0.3333333 -0.3333333
[2,] -0.3333333 0.6666667 -0.3333333
[3,] -0.3333333 -0.3333333 0.6666667
```

```
# Finding rank of A
```

```
```{r}
rankMatrix(A)[1]
```
```

```
[1] 2
```

Each column all added up together gives us 0!.

$$x_1 + x_2 + x_3 = 0$$

Can be written as,

$$x_1 = -x_2 - x_3$$

That are linearly dependant and similar for  $x_2$  and  $x_3$

$$\text{Hence } r(A) = 2$$

Part c: Computing  $E[y^T y]$

Finding  $E[\mathbf{y}^T \mathbf{y}]$

$$= E\left(\frac{2x_1 - x_2 - x_3}{3}, \frac{-x_1 + 2x_2 - x_3}{3}, \frac{-x_1 - x_2 + 2x_3}{3}\right) \begin{bmatrix} \frac{2x_1 - x_2 - x_3}{3} \\ \frac{-x_1 + 2x_2 - x_3}{3} \\ \frac{-x_1 - x_2 + 2x_3}{3} \end{bmatrix}$$

$$= E[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2]$$

$$= E[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2]$$

$$\begin{aligned}
&= E[\sum_{i=1}^3 (x_i - \bar{x})^2] \\
&= E[\sum_{i=1}^3 x_i^2 - 2x_i\bar{x} + \bar{x}^2] \\
&= E[\sum_{i=1}^3 x_i^2 - n\bar{x}^2]
\end{aligned}$$

Since we have 3 x's that are random variables!!

$$\begin{aligned}
&= E[\sum_{i=1}^3 x_i^2 - 3\bar{x}^2] \\
&= E[(\sum_{i=1}^3 x_i^2) - 3\bar{x}^2]
\end{aligned}$$

since  $x_1, x_2$  and  $x_3$  are identical independent distributions!!

$$= (3-1)\sigma^2 = 2\sigma^2$$

Assuming that the sample variance is unbiased! and we can imply  $\lambda =$   
 0! Following similarly to Theorem 3.2. for the Non-central distribution!

Alternative method:

Theorem 3.5:

$$E[y^T A y] = \text{tr}(A V) + \mu^T A \mu$$

since  $A = I$ ,

$$= \text{tr}(V) + \mu^T \mu$$

$$V = \text{var}y = \text{var}Ax = A \text{var}x A^T$$

since  $A$  is symmetric and idempotent!!

$$\text{var}(x_i) = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

$$V = \frac{1}{3} \begin{bmatrix} 2\sigma^2 & -1 & -1 \\ -1 & 2\sigma^2 & -1 \\ -1 & -1 & 2\sigma^2 \end{bmatrix}$$

$$\mu = E[y] = E[Ax] = AE[x]$$

$$= \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} \mu \\ \mu \\ \mu \end{bmatrix} = 0$$

$$E[y^T y] = \text{tr}\left(\frac{1}{3} \begin{bmatrix} 2\sigma^2 & -1 & -1 \\ -1 & 2\sigma^2 & -1 \\ -1 & -1 & 2\sigma^2 \end{bmatrix}\right) + 0$$

$$= 2\sigma^2$$

Part d:

Using Theorem 3.5:

Proof:

Assuming that  $A$  is idempotent and has rank  $k$ . Because it is symmetric, it can be diagonalised. Let the (orthogonal) diagonalising matrix be  $P$ .

$$D = P^T A P = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \dots & \lambda_2 & \dots \\ 0 & \dots & \lambda_k \end{bmatrix}$$

since A is symmetric and idempotent, all eigenvalues are either 0 or 1. We know from definition:

$$tr(A) = r(A) = k$$

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

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

from Part 4b, we find out the rank and trace of matrix A we found in Part 4a. Is also is the same number of degrees of freedom for the chi squared distribution.

$$tr(A) = r(A) = 2$$

Therefore, A must have two eigenvalues of 1 and one eigenvalue of 0.

Using Theorem 3.5 and Corollary 3.7:

with our non central parameter  $\lambda$ !

$$\begin{aligned} \lambda &= \frac{1}{2} \mu^T A \mu \\ &= \frac{1}{2} \begin{bmatrix} \mu \\ \mu \\ \mu \end{bmatrix} \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} \mu & \mu & \mu \end{bmatrix} \\ &= 0 \end{aligned}$$

$\iff$  : if and only if

$$E[y] = E \begin{bmatrix} x_1 - \mu \\ x_2 - \mu \\ x_3 - \mu \end{bmatrix}$$

Since  $x_1, x_2$  and  $x_3$  is identically independently distributed! and taking the expectation of the expectation is the expectation itself!

$$E[y] = E \begin{bmatrix} \mu - \mu \\ \mu - \mu \\ \mu - \mu \end{bmatrix} = 0$$

NOTE:  $\mu = \bar{x}$

In which case,

$$\frac{y^T y}{\sigma^2}$$

is just the sum of two independent standard normal's. This is just an ordinary (central) chi squared distribution  $\chi^2_2$ .  
with expectation of 2 and variance of 4.



# MAST30025 Assignment 1 S1 2021 (3/3)

Michael Le (998211)

March 17, 2021

## Question 5 Solution:

Part a: Computing  $y, X, \beta$  and  $\epsilon$

$$y = \begin{bmatrix} 27.3 \\ 42.7 \\ 38.7 \\ 4.5 \\ 23.0 \\ 166.3 \\ 109.7 \\ 80.1 \\ 150.7 \\ 20.3 \\ 189.7 \\ 131.3 \\ 404.2 \\ 149 \end{bmatrix} \quad X = \begin{bmatrix} 1 & 13.1 \\ 1 & 15.3 \\ 1 & 25.8 \\ 1 & 1.8 \\ 1 & 4.9 \\ 1 & 55.4 \\ 1 & 39.3 \\ 1 & 26.7 \\ 1 & 47.5 \\ 1 & 6.6 \\ 1 & 94.7 \\ 1 & 61.1 \\ 1 & 135.6 \\ 1 & 47.6 \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \end{bmatrix}$$

$y = X\beta + \epsilon$  becomes,

$$\begin{bmatrix} 27.3 \\ 42.7 \\ 38.7 \\ 4.5 \\ 23.0 \\ 166.3 \\ 109.7 \\ 80.1 \\ 150.7 \\ 20.3 \\ 189.7 \\ 131.3 \\ 404.2 \\ 149 \end{bmatrix} = \begin{bmatrix} 1 & 13.1 \\ 1 & 15.3 \\ 1 & 25.8 \\ 1 & 1.8 \\ 1 & 4.9 \\ 1 & 55.4 \\ 1 & 39.3 \\ 1 & 26.7 \\ 1 & 47.5 \\ 1 & 6.6 \\ 1 & 94.7 \\ 1 & 61.1 \\ 1 & 135.6 \\ 1 & 47.6 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \end{bmatrix}$$

Part b: Solving the least squares estimator

```
```{r}
b = solve(t(X)%*%X, t(X)%*%y)
b
```
```

```
      [,1]
[1,] -1.233836
[2,]  2.701553
```

Part c:

```
```{r}
e = y - X%*%b
e #Residual errors
```
```

```
          [,1]
[1,] -6.8565106
[2,]  2.3000724
[3,] -29.7662361
[4,]  0.8710405
[5,] 10.9962256
[6,] 17.8677893
[7,]  4.7627957
[8,]  9.2023660
[9,] 23.6100596
[10,]  3.7035852
[11,] -64.9032511
[12,] -32.5310639
[13,]  39.1032233
[14,] 21.6399042
```

```
```{r}
n = 14 #sample size
p = 2 #number of parameters
SSRes = sum(e^2)
ssquared = SSRes/(n-p)
ssquared

...

[1] 777.1528
```

Part d:

```
``{r}  
c(1,28)*%b  
``
```

```
      [,1]  
[1,] 74.40965
```

Part e:

```

```{r}
a = solve(t(X)%*%X)
a
```

```

```

           [,1]      [,2]
[1,]  0.163081936 -2.230009e-03
[2,] -0.002230009  5.425812e-05

```

```

```{r}
H = X%*%a%*%t(X)
H
```

```

```
```{r}
z = e/sqrt(ssquared * (1 - diag(H)))
z[13]
```
```

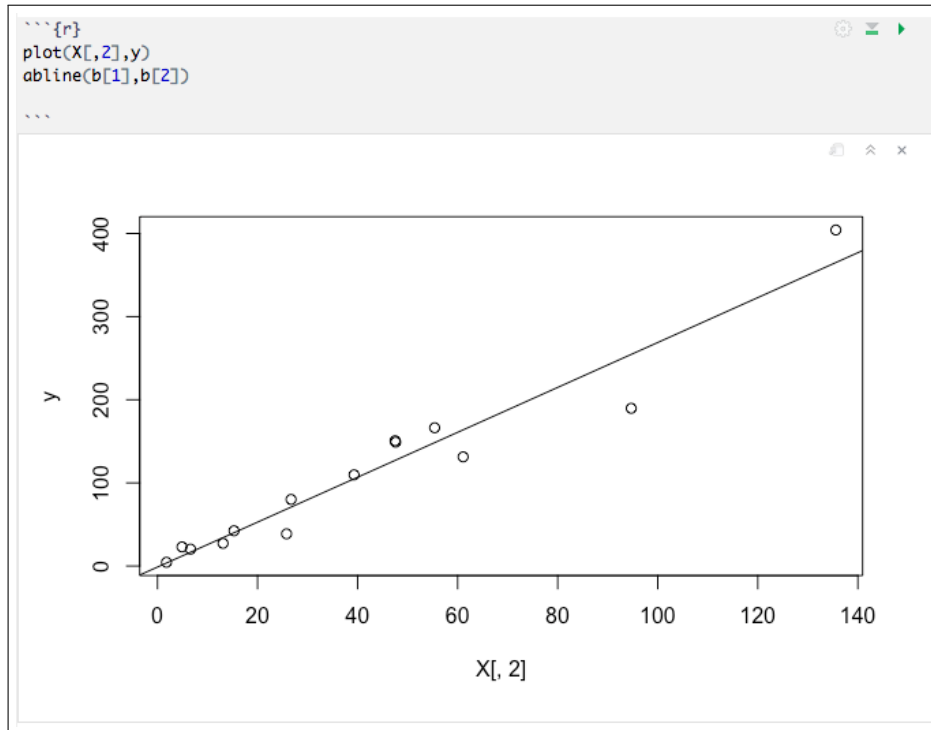
```
[1] 2.104999
```

Part f:

```
```{r}
k = 1
D = z^2 * (diag(H)/(1-diag(H))) * 1/(k+1)
D[13]
```
```

```
[1] 2.774008
```

Part g:



Full explanation: The Cook's distance certainly indicates it should be of some concern; however looking at the plot, it seems that the fit is actually okay. There is considerable evidence for heteroskedasticity — the variance increases with  $x$  (the design variable). Sea scallops has (by far) the largest  $x$  and so may be prone to a larger variance than the remaining points. The high Cook's distance therefore comes primarily from a very high leverage, rather than a bad fit to the model.

**END OF ASSIGNMENT!!**