Assignment 1 Solution

Linear Statistical Models (University of Melbourne)

MAST30025: Linear Statistical Models

Assignment 1 Solutions

Total marks: 36

1. Prove that if a symmetric matrix A has eigenvalues which are all either 0 or 1, it is idempotent.

   Solution [4 marks]: Diagonalise A and write it as

   $$A = PDP^T.$$

   Then

   $$\begin{aligned} A^2 &= PDP^TPDP^T \\ &= PD^2P^T \\ &= PDP^T = A \end{aligned}$$

   since D has only 0 or 1 on the diagonal and hence $D^2 = D$.

2. Prove (without using Theorem 2.5) that if A and B are symmetric matrices, $A + B$ is idempotent and $AB = BA = 0$, then both A and B are idempotent. (*Hint*: Use Theorem 2.4. Then derive two relations between the diagonalisations of A and B.)

   Solution [5 marks]: By Theorem 2.4, there exists a matrix P which diagonalises both A and B:

   $$\begin{aligned} P^TAP &= D_1 \\ P^TBP &= D_2 \\ P^T(A+B)P &= D_1 + D_2. \end{aligned}$$

   Since $A + B$ is idempotent, $D_1 + D_2$ has only 1s and 0s on the diagonal. Also

   $$\begin{aligned} D_1D_2 &= P^TAPP^TBP \\ &= P^TABP = 0. \end{aligned}$$

   Now consider the ith diagonal elements of D_1 and D_2. Their product must be 0, so one of them is 0. Their sum is either 0 or 1, so the other must be 0 or 1. Hence all diagonal elements of D_1 and D_2 are either 0 or 1, and therefore A and B are idempotent.

3. Let $\mathbf{y}$ be a 3-dimensional multivariate normal random vector with mean and variance

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

   Let

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

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

   Solution [3 marks]: $\mathbf{y} \sim MVN(A\boldsymbol{\mu}, AVA^T)$.

```
> mu <- c(1,-1,0)
> V <- matrix(c(2,0,1,0,1,0,1,0,2),3,3)
> A <- matrix(c(2,0,-1,0,3,0,-1,0,2),3,3)/3
> A%*%mu
           [,1]
[1,]  0.6666667
[2,] -1.0000000
[3,] -0.3333333
```

```
> A%*%V%*%t(A)
           [,1] [,2]       [,3]
[1,]  0.6666667    0 -0.3333333
[2,]  0.0000000    1  0.0000000
[3,] -0.3333333    0  0.6666667
```

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

Solution [2 marks]:

```
> sum(diag(A%*%V)) + t(mu)%*%A%*%mu
         [,1]
[1,] 4.666667
```

(c) Describe the distribution of $\mathbf{y}^T A\mathbf{y}$.

Solution [3 marks]: $AV = I_3$, which is idempotent. Hence $\mathbf{y}^T A\mathbf{y}$ has a non-central χ^2 distribution with 3 degrees of freedom and noncentrality parameter:

```
> t(mu)%*%A%*%mu/2
          [,1]
[1,] 0.8333333
```

(d) Find a matrix B such that $\mathbf{y}^T B\mathbf{y}$ is independent of $\mathbf{y}^T A\mathbf{y}$.

Solution [2 marks]: We require a matrix B such that $AVB = 0$. But $AV = I$, and so the only possible choice is $B = 0$.

4. Let $\mathbf{y} \sim MVN(\boldsymbol{\mu}, V)$ be a $n \times 1$ random vector and suppose V is nonsingular. Find A and $\mathbf{b}$ such that $A\mathbf{y} + \mathbf{b}$ is an n-length vector of independent standard normals.

Solution [4 marks]: We know that $A\mathbf{y} + \mathbf{b} \sim MVN(A\boldsymbol{\mu} + \mathbf{b}, AVA^T)$. Therefore we can choose

$$\begin{aligned} A &= V^{-1/2} \\ \mathbf{b} &= -V^{-1/2}\boldsymbol{\mu}. \end{aligned}$$

5. A study is conducted to determine if (and how) the fuel mileage of a car is dependent on its weight, and the speed at which it is driven. A linear model is assumed, and the following data is obtained:

Weight (tons)	Speed (km/hr)	Mileage (km/litre)
1.35	50	8.5
1.33	55	8
2	60	7.5
1.4	52	10
1.43	47	11
1.2	45	15
1.3	49	13.5
1.28	63	14

(a) Write down the linear model as a matrix equation, writing out the matrices in full.

Solution [2 marks]: $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where

$$\mathbf{y} = \begin{bmatrix} 8.5 \\ 8 \\ 7.5 \\ 10 \\ 11 \\ 15 \\ 13.5 \\ 14 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 1.35 & 50 \\ 1 & 1.33 & 55 \\ 1 & 2 & 60 \\ 1 & 1.4 & 52 \\ 1 & 1.43 & 47 \\ 1 & 1.2 & 45 \\ 1 & 1.3 & 49 \\ 1 & 1.28 & 63 \end{bmatrix}$$

and $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ are obvious.

(b) Calculate the least squares estimator of the parameters.

Solution [2 marks]:

```
> n <- 8
> p <- 3
> y <- c(8.5, 8, 7.5, 10, 11, 15, 13.5, 14)
> X <- t(matrix(c(1,1.35,50,1,1.33,55,1,2,60,1,1.4,52,
+                         1,1.43,47,1,1.2,45,1,1.3,49,1,1.28,63),p,n))
> (b <- solve(t(X)%*%X,t(X)%*%y))
```

```
             [,1]
[1,] 21.048202099
[2,] -7.418663570
[3,]  0.006819703
```

(c) Calculate the residual sum of squares SS_{Res} and sample variance s^2.

Solution [2 marks]:

```
> e <- y - X%*%b
> (SSRes <- sum(e^2))
```

```
[1] 36.43937
```

```
> (s2 <- SSRes/(n-p))
```

```
[1] 7.287874
```

(d) Predict (using a point estimate) the average fuel mileage of a car which weighs 1.8 tons and is driven at 59 km/hr.

Solution [2 marks]:

```
> c(1,1.8,59)%*%b
```

```
        [,1]
[1,] 8.09697
```

6. Let A be a symmetric and idempotent matrix with entries a_{ij}. Prove that $0 \leq a_{ii} \leq 1$. Use this to derive limits on the leverage of a point in the full rank model.

Solution [5 marks]:

$$\begin{aligned} a_{ii} &= [A^2]_{ii} \\ &= \sum_j a_{ij} a_{ji} \\ &= \sum_j a_{ij}^2. \end{aligned}$$

Since this is a sum of squares, the lower bound follows. We also have $a_{ii} \geq a_{ii}^2$ and since a_{ii} is non-negative, the upper bound follows. Since the hat matrix is symmetric and idempotent, the leverage of a point must lie in [0,1].