STA 521 HW 2

William Tirone

1

a)

TRUE. It is unbiased, since $E(\hat{\beta}) - \beta = 0$

b)

c)

TRUE. JUSTIFY

d)

TRUE.
It is a projection matrix, and thus idempotent. $H = H^T$ so it is symmetric, and it is PSD since every $\lambda \geq 0$.

e)

FALSE. $tr(I - H) = tr(I) - tr(H) = n - p$ since the trace of an idempotent matrix equals it's rank, and H has rank = p.

f)

FALSE. We saw an example in class of outliers that had low leverage scores.

g)

answer

h)

TRUE. Per wikipedia: https://en.wikipedia.org/wiki/Ggplot2#:~:text=Created%20by%20Hadley%20Wickh

2

a)

projection $= [x_1, 0, 0]^T$

b)

projection $= [x_1, x_2, 0]^T$ - just did this visually.

c)

$$\text{part A:} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\text{part B:} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

d)

```
# proj matrices
part_A = matrix(c(1,0,0,0,0,0,0,0,0),3,3)
part_B = matrix(c(1,0,0,0,1,0,0,0,0),3,3,byrow=TRUE)

# making random unif vectors
set.seed(345)
v1 = matrix(runif(3,0,1))
v2 = matrix(runif(3,0,1))

proj1 = part_A %*% v1
proj2 = part_B %*% v2
```

```
proj1
```

```
          [,1]
[1,] 0.2162537
[2,] 0.0000000
[3,] 0.0000000
```

```
proj2
```

```
          [,1]
[1,] 0.6557397
[2,] 0.4358664
[3,] 0.0000000
```

e)

$$\frac{a^T x}{a^T a} a$$

and the matrix is:

$$P = \frac{aa^T}{a^T a}$$

f)

```
proj = function(x,a) {
  proj_matrix = (a %*% t(a)) / as.double((t(a) %*% a))
  x_onto_a = proj_matrix %*% x
  x_onto_a
}

proj(
  x=matrix(c(3,2,-1)),
  a=matrix(c(1,0,1))
)
```

```
     [,1]
[1,]    1
[2,]    0
[3,]    1
```

g)

$$P = A(A^TA)^{-1}A^T$$

Where the columns of A are a_1, a_2.

h)

As hinted by the problem, we can use Gram-Schmidt to find an orthonormal basis that spans the same subspace given by a_1, a_2. With A = QR, and Q being the orthonormal basis, we can construct a projection matrix:

$$P = Q(Q^TQ)^{-1}Q^T$$

i)

```
A = matrix(c(1,0,1,1,-1,0),3,2)
qr_decomp = qr(A)
Q = qr.Q(qr_decomp)

x = matrix(c(3,2,-1))

projection_matrix = Q %*% solve(t(Q) %*% Q) %*% t(Q)
projection_matrix %*% x
```

```
             [,1]
[1,] 1.000000e+00
[2,] 1.665335e-16
[3,] 1.000000e+00
```

j)

I believe this is the same as my answer in h) (and looking ahead to part k) we can construct a projection matrix onto the k-dimensional subspace spanned by $a_1, ..., a_k$ by computing the QR decomposition, taking the Q, and constructing:

$$P = Q(Q^TQ)^{-1}Q^T$$

l)

The answer below looks the same as part i).

```
A %*% solve(t(A)%*% A) %*% t(A) %*% x
```

```
              [,1]
[1,]  1.000000e+00
[2,] -5.551115e-17
[3,]  1.000000e+00
```

m)

$$P = Q(Q^TQ)^{-1}Q^T = Q(I)Q^T = QQ^T = I$$

3

a)

Joint Distributions:

$$p(X = x, Z = 1; \theta) = \pi_1 N(x; \mu_1, 1)p(X = x, Z = 2; \theta) = \pi_2 N(x; \mu_2, 1)$$

Marginal likelihood:

$$p(X = x; \theta) = \sum_{j=1}^{2} p_\theta(X = x|Z = j)p_\theta(Z = j)$$

log-likelihood:

$$l(X = x; \theta) = \sum_{i=1}^{N} logp_\theta(X = x_i)$$

w/ n i.i.d. samples:

$$l(X = x; \theta) = -\frac{n}{2}ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2$$

b)

c)

d)

4

1.

We know the distribution of $\hat{\beta}$ must be normal because we assumed the errors are normally distributed. (Why?) Then it is sufficient to find the mean and variance of $\hat{\beta}$

$$\begin{aligned}
E(\hat{\beta}) &= E[(X^TX)^{-1}X^Ty] \\
&= (X^TX)^{-1}X^TE[x\beta^* + \epsilon] \rightarrow \text{since X is constant} \\
&= (X^TX)^{-1}X^TXE(\beta^*) \\
&= \beta^*
\end{aligned}$$

(Note for self: below mirrors form of $Var(X) = E(X - E(X))^2$

$$\begin{aligned}
Var(B^*) &= E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))^T] \\
&= E[(\hat{\beta} - (X^TX)^{-1}X^T\epsilon)(\hat{\beta} - (X^TX)^{-1}X^T\epsilon)^T] \\
&\text{next line follows since} E(\hat{\beta}) = \beta \text{and} X^TX \text{is symmetric} \\
&= E[(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1}] \\
&= (X^TX)^{-1}X^TE(\epsilon\epsilon^T)X(X^TX)^{-1} \\
&= \sigma^2(X^TX)^{-1}
\end{aligned}$$

Noting that $E(\epsilon\epsilon^T)$ is the covariance matrix of ϵ with mean $= 0$.
Thus we have the mean and the variance of $\hat{\beta}$ and it is distributed $\sim N(\beta^*, \sigma^2(X^TX)^{-1})$
Now, since the data X is constant, $E(X\hat{\beta}) = E(X)E(\hat{\beta}) = X\beta^*$

2.

$$\begin{aligned}
E||e||_2^2 &= ETr(ee^T) \\
&= ETr((I_n - X(X^TX)^{-1}X^T)\epsilon\epsilon^T(I_n - X(X^TX)^{-1}X^T)) \\
&= \sigma^2Tr(I_n - X(X^TX)^{-1}X^T) \\
&\text{below using the cyclic property of trace} \\
&= \sigma^2Tr(I_n) - Tr(X^TX(X^TX)^{-1}) \\
&= \sigma^2Tr(I_n) - Tr(I_p) \\
&= \sigma^2(n - p)
\end{aligned}$$

3. optional

4. optional

5

1.

2.

3.

4.

5.

6.