STAT 2131
Orly Olbum

Homework 5

1. Deriving H in OLS

ⓐ $B \in \mathbb{R}^p$ and $f(B) = (Y - XB)^T(Y - XB)$

use properties (slide 12) to show

$$\hat{\beta} = \underset{B \in \mathbb{R}^p}{\operatorname{argmin}} f(B) = (X^TX)^{-1}X^TY$$

First, simplify $f(B)$

$$f(B) = (Y-XB)^T(Y-XB)$$
$$= (Y^T - X^TB^T)(Y-XB)$$
$$= Y^TY - B^TX^TY - Y^TXB + B^TX^TXB$$
$$= Y^TY - 2B^TX^TY + B^TX^TXB$$

Differentiate wrt B

$$\partial F(B)/\partial B = -2X^TY + 2X^TXB$$

Set equal to zero, solve for B

$$-2X^TY + 2X^TXB = 0$$
$$\hookrightarrow 2X^TXB = 2X^TY$$
$$\hookrightarrow B = \frac{2X^TY}{2X^TX} = (X^TX)^{-1}X^TY$$

ⓑ $\hat{Y} = X\hat{B}$, $\hat{e} = Y - \hat{Y}$ are predicted values & est. residuals

Show that $\hat{Y} = HY$ and $\hat{e} = QY$

If $H = (X^TX)^{-1}X^T$, $HY = X(X^TX)^{-1}X^TY$

$$\hat{Y} = X\hat{B} = \underbrace{X(X^TX)^{-1}X^T}_{H}Y$$

so $\hat{Y} = HY$

$$\hat{e} = Y - \hat{Y}$$
$$= Y - HY$$
$$= (I - H)Y$$
$$= QY$$

2. ⓐ Use definition of H to show
$$\text{im}(H) = \text{im}(X)$$
$H = X(X^TX)^{-1}X^T$
$\text{im}(X) = \{Av : v \in \mathbb{R}^d\} \subseteq \mathbb{R}^n$
First, we know that $\text{col}(H) = \text{col}(X)$
And $\text{rank}(X) = \text{rank}(X^TX) = p$
And $\text{rank}(H) = \text{rank}(X(X^TX)^{-1}X^T)$
$= \text{rank}((X^TX)^{-1}) = \text{rank}(X^TX) = p$ (by definition of rank)
Since $\text{col}(X) = \text{col}(H)$, or the span of X, we know that H is the orthogonal complement to X meaning
$$\text{im}(H) = \text{im}(X).$$

ⓑ Show that H & Q are symmetric, idempotent, and orthogonal

i) $H = H^T$ and $Q = Q^T$
$$H = H^T = (X(X^TX)^{-1}X^T)^T = X[(X^TX)^{-1}]^TX^T$$
$$= X[(X^TX)^T]^{-1}X^T = X(X^TX)^{-1}X^T = H$$
$$Q = Q^T = (I-H)^T = I^T - H^T = I - H$$

ii) $H^2 = H$ and $Q^2 = Q$
$$H^2 = HH = (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T)$$
$$= X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H$$
$$Q^2 = QQ = (I-H)(I-H) = I - 2H + HH = I - H$$

iii) $HQ = 0$ (left multiplication!)
$$HQ = H(I-H) = H - HH = H - H = 0$$
as shown above

2, continued

(c) Show that H projects vectors in $\mathbb{R}^n$ onto $im(X)$, and is an orthogonal projection matrix

i. If $v \in \mathbb{R}^n$, $Hv \in im(X)$. If $u \in im(X)$, $Hu = u$

ii. Let $v \in \mathbb{R}^n$. Hv is closest in $im(X)$ to v

$$Hv = \underset{u \in im(X)}{\operatorname{argmin}} \|v - u\|_2^2$$

(i) If $v \in \mathbb{R}^n$ and $Hv \in im(X)$, say $\vec{v_i}$ are the basis for X

If $u \in im(X)$, $Hu = u$, $\vec{u_i}$ are orthonormal basis for X.

We know H has columns $\vec{u_i}$ and is symmetric and idempotent.

For some R, $n \times n$ invertible linearly independent square matrix,

$\vec{v_i} = \sum R_{ij} \vec{u_i}$ and $X = HR$

we find the orthogonal projection matrix, P, onto $im(X)$ to be:

$$\begin{aligned}
P &= X(X^T X)^{-1} X^T \\
&= (HR)((HR)^T(HR))^{-1} (HR)^T \\
&= HR(R^T H^T HR)^{-1} R^T H^T \\
&= HR(R^T R)^{-1} R^T H^T \\
&= HRR^{-1}(R^T)^{-1} R^T H^T \\
&= HH^T \\
&= H
\end{aligned}$$

If the columns of H $(\vec{u_i})$ are the orthonormal basis for X, H is the orthogonal projection matrix onto X

2 ⓒ, continued

(ii) Let $v \in \mathbb{R}^n$. Hv is closest in im(x) to v

$$Hv = \underset{u \in im(x)}{\operatorname{argmin}} \|v-u\|_2^2$$

(write $v-u = H(v-u) + Q(v-u)$, expand)

$$\|v-u\|_2^2 = \|v-Hv+Hv-u\|_2^2$$
$$= \|(v-Hv)+(Hv-u)\|_2^2$$
$$= [(v-Hv)+(Hv-u)]^T[(v-Hv)+(Hv-u)]$$
$$= \underset{①}{(v-Hv)^T(v-Hv)} + \underset{②}{(v-Hv)^T(Hv-u)} + \underset{③}{(Hv-u)^T(v-Hv)} + \underset{④}{(Hv-u)^T(Hv-u)}$$

① $(v-Hv)^T(v-Hv) = \|v-Hv\|_2^2$

② $(v-Hv)^T(Hv-u)$
$$= v^THv - v^Tu - v^TH^THv + v^TH^Tu$$
$$= v^THv - v^Tu - v^THv + v^Tu$$
$$= (v^THv - v^THv) + (v^Tu - v^Tu) = 0$$

and $H^TH = HH = H$
$H^Tu = Hu = u$ (from i)

③ $(Hv-u)^T(v-Hv)$
$$= v^TH^Tv - v^THHv - u^Tv - u^THv$$
$$= v^THv - v^THv - u^Tv + u^Tv$$
$= 0$ because $HH = H$, $u^TH = (Hu)^T = u^T$

④ $(Hv-u)^T(Hv-u) = \|Hv-u\|_2^2$

so,
$$\|v-u\|_2^2 = ① + \cancel{②}^{\,0} + \cancel{③}^{\,0} + ④$$
$$= \|v-Hv\|_2^2 + \|Hv-u\|_2^2 \geq \|v-Hv\|_2^2$$

making Hv the closest in im(x) to v

→

3. H eigenvalues: $\lambda_1, \dots, \lambda_n$

ⓐ Show that $\lambda_i = 0$ or 1 for all λ_n

In terms of n & p, how many are 1, how many 0?

We know that H is idempotent ($HH = H$) and $n \times n$. we say v is the eigenvector of H with eigenvalues λ for $v \neq 0$, so

$$Hv = \lambda v$$

we also have $n \times n$ matrix P with columns v_i ($i = 1, \dots, n$) and Λ diagonal matrix containing λ_i ($i = 1, \dots, n$) as diagonal values

meaning $tr(\Lambda) = \sum_i \lambda_i$ (sum eigenvalues)

$tr(H) = \sum_i \lambda_i$ because H is symmetric.

Now we have

$$\lambda v = \underbrace{Hv}_{\text{from above}} = HHv = \lambda Hv = \lambda^2 v$$

and since $v \neq 0$, $\lambda - \lambda^2 = \lambda(1-\lambda) = 0$,

so λ can only be 0 or 1

then, $H = P \Lambda P^T$, so $HH = P\Lambda P^T P \Lambda P^T = P \Lambda^2 P^T$

and $\Lambda^2 = \Lambda\Lambda = \Lambda$.

So, $tr(H) = tr(P\Lambda P^T) = tr(\Lambda P P^T) = tr(\Lambda) = \sum_i \lambda_i$

and all diagonal entries in Λ are 0 or 1 with the number of eigenvalues being 1 equal to $tr(\Lambda)^{(p)}$, and the remaining being zero $(n-p)$

ⓑ Find eigenvalues of Q

$Q = I - H$ is symmetric, $HQ = H(I-H) = 0$

Rank$(I-H) = n-p$

Since I has only 1's diagonal, the eigenvalues are all 1, so eigenvalues for $Q = I - H$ are $\{1 - \lambda_i\}$ for $i = 1, \dots, n$

$n-p$ eigenvalues are 1, p are 0

3, continued

ⓒ Show that $tr(H)=p$ and $tr(Q)=n-p$

We know, by nature of I, that $tr(I)=n$

$tr(H) = tr(X(X^TX)^{-1}X^T)$

$= tr((X^TX)^{-1}XX^T)$

$= tr((X^TX)^{-1})$

and $(X^TX)^{-1}$ is an I $p \times p$ matrix, so

$= tr(I_{p\times p})$

$= p$

$tr(I-H) = tr(I) - tr(H)$

$= n - p$

↳ from above

4. $Y = X\beta + \epsilon$ for $\beta \in \mathbb{R}^p$, mean $\epsilon = 0$ and $var(\epsilon) = \sigma^2 I_n$

For simple linear regression, $p=2$.

ⓐ use definition of $\hat{\beta}$ from 1ⓐ to show $E(\hat{\beta}) = \beta$

$\hat{B} = (X^TX)^{-1}X^TY$

$E(\hat{B}) = E[(X^TX)^{-1}X^TY]$

$= E[(X^TX)^{-1}X^T(XB+\epsilon)]$ Replace $Y = XB+\epsilon$

$= E[(X^TX)^{-1}X^TXB + (X^TX)^{-1}X^T\epsilon]$

$= E[B + (X^TX)^{-1}X^T\epsilon]$

$\boxed{\text{because } \underbrace{(X^TX)^{-1}}_{A^{-1}}\underbrace{X^TX}_{A} = 1}$

$= E(B) + E[(X^TX)^{-1}X^T\epsilon]$

$= E(B) + (X^TX)^{-1}X^TE(\epsilon)$

$= E(B) + 0$ $(E(\epsilon)=0)$

$= E(B)$

$= B$

(next page)

→

4, continued

(b) Recall that for $W, Z \in \mathbb{R}^n$

$$cov(W,Z) = E[\{W - E(W)\}\{Z - E(Z)\}] \in \mathbb{R}^{n \times n}$$

Use 2b to show $cov(\hat{Y}, \hat{\epsilon}) = 0$

$cov(\hat{Y}, \hat{\epsilon}) = E[(\hat{Y} - E(\hat{Y}))(\hat{\epsilon} - E(\hat{\epsilon}))]$

We know $E(\hat{Y}) = E(HY) = HE(Y) = HXB + HE(\epsilon)$

$= X(X^TX)^{-1}X^TXB + 0 = XB$, $\hat{\epsilon} = Y - \hat{Y}$,

and $E(\hat{\epsilon}) = (I-H)(XB + E(\epsilon)) = XB - \hat{X}B = 0$

$Cov(\hat{Y}) = cov(HY) = H cov(Y) H^T = H(\sigma^2 I)H = \sigma^2 H$

And $var(\epsilon) = \sigma^2 I_n$ $\quad \hookrightarrow E(\hat{Y}) = E(HY) = H\cancel{X}B = XB = E(Y)$

$cov(\hat{Y}, \hat{\epsilon}) = E[(\hat{Y} - E(\hat{Y}))(\hat{\epsilon} - E(\hat{\epsilon}))]$

$= E[(\hat{Y} - E(Y))(\hat{\epsilon} - 0)]$

$= E[(XB - XB)(\hat{\epsilon})]$

$= 0$