3 Expected Value of Empirical Risk for OLS

3.4 Question 1

The Bayes estimator under squared loss is $\mathbb{E}[y \mid X = x]$:

$$\begin{aligned} f^*(x) &= \mathbb{E}[y \mid X = x] && (1)\\ &= \mathbb{E}[X\theta + \xi \mid X = x] && (2)\\ &= x\theta + \mathbb{E}[\xi \mid X = x] && (3)\\ &= x\theta + 0 && (4)\\ &= x\theta && (5) \end{aligned}$$

Because $\mathbb{E}[\xi \mid X = x] = 0$.

The Bayes risk is defined as the expected squared loss with respect to the joint distribution of (X, Y):

$$\begin{aligned} R(f^*) &= \mathbb{E}[(Y - f^*(X))^2] && (6)\\ &= \mathbb{E}[(Y - X\theta)^2] && (7)\\ &= \mathbb{E}[(X\theta + \varepsilon - X\theta)^2] && (8)\\ &= \mathbb{E}[\varepsilon^2] && (9)\\ &= \sigma^2 && (10) \end{aligned}$$

We know from Proposition 1 that:

$$\mathbb{E}[R_n(\hat{\theta})] = \frac{n-d}{n}\sigma^2 \tag{11}$$

And we know that $\frac{n-d}{n} < 1$:

$$=> \mathbb{E}[R_n(\hat{\theta})] < \sigma^2.$$

Analysis:

- When $d < n$, the Empirical Risk is closer to the Bayes Risk.
- When d is large, the Empirical Risk is lower than the Bayes risk.

3.4 Question 2

We need to prove that:

$$\mathbb{E}[R_n(\hat{\theta})] = \mathbb{E}_\xi\left[\frac{1}{n}\|(I_n - X(X^TX)^{-1}X^T)\epsilon\|^2\right]. \tag{12}$$

We have:

$$\begin{aligned} \mathbb{E}[R_n(\hat{\theta})] &= \mathbb{E}\left[\frac{1}{n}\|y - X\hat{\theta}\|^2\right] && (13)\\ &= \mathbb{E}\left[\frac{1}{n}\|y - X(X^TX^{-1})X^Ty\|^2\right] && (14)\\ &= \mathbb{E}\left[\frac{1}{n}\|(I_n - X(X^TX^{-1})X^T)y\|^2\right] && (15)\\ &= \mathbb{E}\left[\frac{1}{n}\|(I_n - X(X^TX^{-1})X^T)(X\theta + \xi)\|^2\right] && (16)\\ &= \mathbb{E}\left[\frac{1}{n}\|(I_n - X(X^TX^{-1})X^T)X\theta + (I_n - X(X^TX^{-1})X^T)\xi\|^2\right] && (17)\\ & && (18) \end{aligned}$$

But

$$(I_n - X(X^T X^{-1})X^T)X\theta = 0 \tag{19}$$

So,

$$\begin{align}
\mathbb{E}[R_n(\hat{\theta})] &= \mathbb{E}\left[\frac{1}{n}\|0 + (I_n - X(X^T X^{-1})X^T)\xi\|^2\right] \tag{20}\\
&= \mathbb{E}\left[\frac{1}{n}\|(I_n - X(X^T X^{-1})X^T)\xi\|^2\right] \tag{21}
\end{align}$$

3.4 Question 3

Let $A \in \mathbb{R}^{n\times n}$ and $i, j \in [1, n]$.

$$\begin{align}
\mathrm{tr}(M) &= \sum_{i=1}^{n} M_{ii} \tag{22}\\
(A^\top A)_{jj} &= \sum_{k=1}^{n} (A^\top)_{ji} A_{ij} \tag{23}\\
& \tag{24}
\end{align}$$

But, $A^T_{ji} = A_{ij}$

$$\begin{align}
=> (A^T A)_{jj} &= \sum_{i=1}^{n} A_{ij} A_{ij} \tag{25}\\
&= \sum_{i=1}^{n} A_{ij}^2 \tag{26}
\end{align}$$

$$\begin{align}
\mathrm{tr}(A^T A) &= \sum_{j=1}^{n} (A^T A)_{jj} \tag{27}\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}^2 \tag{28}\\
&= \sum_{(i,j)\in[1,n]^2} A_{ij}^2 \tag{29}
\end{align}$$

3.4 Question 4

We should prove that:

$$\mathbb{E}_\xi\left[\frac{1}{n}\|A\xi\|^2\right] = \frac{\sigma^2}{n}\mathrm{tr}(A^T A) \tag{30}$$

$$\begin{align}
\mathbb{E}_\xi\left[\frac{1}{n}\|A\xi\|^2\right] &= \frac{1}{n}\mathbb{E}_\xi\left[\|A\xi\|^2\right] \tag{31}\\
&= \frac{1}{n}\mathbb{E}_\xi\left[(A\xi)^T(A\xi)\right] \tag{32}\\
&= \frac{1}{n}\mathbb{E}_\xi\left[\xi^T A^T A\xi\right] \tag{33}\\
&= \frac{1}{n}\mathbb{E}_\xi\left[\xi^T Q\xi\right] \tag{34}\\
& \tag{35}
\end{align}$$

With Q $= A^T A$

$$
\begin{aligned}
=> \mathbb{E}_\xi\left[\frac{1}{n}\|A\xi\|^2\right] &= \frac{1}{n}\sigma^2 \mathrm{tr}(Q) && (36)\\
&= \frac{\sigma^2}{n}\mathrm{tr}(A^T A) && (37)
\end{aligned}
$$

3.4 Question 5

We know that:

$$
\begin{aligned}
A &= I_n - X(X^T X)^{-1}X^T. && (38)\\
A^T &= (I_n - X(X^T X)^{-1}X^T)^T && (39)\\
&= I_n^T - X^T((X^T X)^{-1})^T(X^T)^T && (40)\\
&= I_n - X^T((X^T X)^T)^{-1}X && (41)\\
&= I_n - X^T(X^T X)^{-1}X && (42)\\
A^T A &= (I_n - X^T(X^T X)^{-1}X)(I_n - X(X^T X)^{-1}X^T) && (43)\\
&= In - In.X(X^T X)^{-1}X^T - X^T(X^T X)^{-1}X.In + X^T(X^T X)^{-1}XX(X^T X)^{-1}X^T && (44)\\
&= In - X(X^T X)^{-1}X^T - X^T(X^T X)^{-1}X + X(X^T X)^{-1}X^T && (45)\\
&= In - X^T(X^T X)^{-1}X && (46)\\
&= In - X(X^T X)^{-1}X^T && (47)\\
&= A && (48)
\end{aligned}
$$

3.4 Question 6

$$
\begin{aligned}
\mathbb{E}_\xi[R_n(\hat{\beta})] &= \mathbb{E}_\xi\left[\frac{1}{n}\|A\xi\|^2\right] && (49)\\
&= \frac{\sigma^2}{n}\mathrm{tr}(A^T A) && (50)\\
&= \frac{\sigma^2}{n}\mathrm{tr}(A) && (51)
\end{aligned}
$$

Let's compute $\mathrm{tr}(A)$.
We know that $A = I_n - X(X^T X)^{-1}X^T$ from Question 5. So,

$$
\begin{aligned}
\mathrm{tr}(A) &= \mathrm{tr}(I_n - X(X^T X)^{-1}X^T) && (52)\\
&= \mathrm{tr}(I_n) - \mathrm{tr}(X(X^T X)^{-1}X^T) && (53)\\
&= n - \mathrm{tr}(X(X^T X)^{-1}X^T) && (54)\\
&= n - d && (55)
\end{aligned}
$$

The $\mathrm{tr}(X(X^T X)^{-1}X^T)$ equals d, the rank of the projection matrix $X(X^T X)^{-1}X^T$.
Finally, we can write:

$$
\begin{aligned}
\mathbb{E}[R_n(\hat{\theta})] &= \frac{\sigma^2}{n}(n-d) && (56)\\
&= \left(1 - \frac{d}{n}\right)\sigma^2 && (57)
\end{aligned}
$$

3.5 Question 7

$$\mathbb{E}\left[\frac{\|Y - X\hat{\theta}\|^2}{n-d}\right] = \frac{1}{n-d}.E\left[\|Y - X\hat{\theta}\|^2\right] \tag{58}$$

$$\tag{59}$$

We know that,

$$\mathbb{E}[R_n(\hat{\theta})] = \frac{n-d}{n}\sigma^2 \tag{60}$$

$$and \tag{61}$$

$$\mathbb{E}[R_n(\hat{\theta})] = \mathbb{E}\left[\frac{1}{n}\|Y - X\hat{\theta}\|^2\right] \tag{62}$$

$$=> \mathbb{E}\left[\frac{1}{n}\|Y - X\hat{\theta}\|^2\right] = \frac{n-d}{n}\sigma^2 \tag{63}$$

$$=> \mathbb{E}\left[\|Y - X\hat{\theta}\|^2\right] = (n-d)\sigma^2 \tag{64}$$

$$=> \mathbb{E}\left[\frac{\|Y - X\hat{\theta}\|^2}{n-d}\right] = \frac{(n-d)\sigma^2}{n-d} \tag{65}$$

$$= \sigma^2 \tag{66}$$

This shows that $\frac{\|Y - X\hat{\theta}\|^2}{n-d}$ is an unbiased estimator of σ^2.