

STAT 527: Assignment #1

Philip Pham

April 16, 2020

Problem 1

(a) *Proof.* This follows from the definition of variance after some algebra.

$$\begin{aligned} \mathbb{E} \left[\left(\hat{f}(x) - f(x) \right)^2 \right] &= \mathbb{E} \left[\left[\left(\hat{f}(x) - \mathbb{E} [\hat{f}(x)] \right) + \left(\mathbb{E} [\hat{f}(x)] - f(x) \right) \right]^2 \right] \\ &= \mathbb{E} \left[\left(\hat{f}(x) - \mathbb{E} [\hat{f}(x)] \right)^2 \right] + 2 \mathbb{E} \left[\left(\hat{f}(x) - \mathbb{E} [\hat{f}(x)] \right) \left(\mathbb{E} [\hat{f}(x)] - f(x) \right) \right] + \mathbb{E} \left[\left(\mathbb{E} [\hat{f}(x)] - f(x) \right)^2 \right] \\ &= \text{var} \left(\hat{f}(x) \right) + \left(\mathbb{E} [\hat{f}(x)] - f(x) \right)^2 + 2 \left(\mathbb{E} [\hat{f}(x)] - f(x) \right) \mathbb{E} \left[\left(\hat{f}(x) - \mathbb{E} [\hat{f}(x)] \right) \right] \\ &= \left(\mathbb{E} [\hat{f}(x)] - f(x) \right)^2 + \text{var} \left(\hat{f}(x) \right). \end{aligned}$$

□

(b) *Proof.* The prediction error is

$$\begin{aligned} \mathbb{E} \left[\left(y_{\text{new}} - \hat{f}(x_{\text{new}}) \right)^2 \right] &= \mathbb{E} \left[\left(f(x_{\text{new}}) + \epsilon_{\text{new}} - \hat{f}(x_{\text{new}}) \right)^2 \right] \\ &= \mathbb{E} \left[\left(\epsilon_{\text{new}} + \left(f(x_{\text{new}}) - \hat{f}(x_{\text{new}}) \right) \right)^2 \right] \\ &= \mathbb{E} \left[\epsilon_{\text{new}}^2 \right] + 2 \mathbb{E} \left[\epsilon_{\text{new}} \left(f(x_{\text{new}}) - \hat{f}(x_{\text{new}}) \right) \right] + \left(\mathbb{E} [\hat{f}(x_{\text{new}})] - f(x_{\text{new}}) \right)^2 + \text{var} \left(\hat{f}(x_{\text{new}}) \right) \\ &= \sigma^2 + \left(\mathbb{E} [\hat{f}(x_{\text{new}})] - f(x_{\text{new}}) \right)^2 + \text{var} \left(\hat{f}(x_{\text{new}}) \right). \end{aligned}$$

There is an additional σ^2 term to account for the variance of our new observation.

□

(c) *Proof.*

$$\begin{aligned}
a &\geq \mathbb{E}[L] = \int_0^\infty t \, dL(t) \\
&= \int_0^{a/\epsilon} t \, dL(t) + \int_{a/\epsilon}^\infty t \, dL(t) \\
&\geq \int_0^{a/\epsilon} t \, dL(t) + \frac{a}{\epsilon} \int_{a/\epsilon}^\infty dL(t) \\
&\geq \frac{a}{\epsilon} \int_{a/\epsilon}^\infty dL(t) = \frac{a}{\epsilon} \mathbb{P}\left(L > \frac{a}{\epsilon}\right).
\end{aligned}$$

Thus, we have that

$$\frac{a}{\epsilon} \mathbb{P}\left(L > \frac{a}{\epsilon}\right) \leq a \Leftrightarrow \mathbb{P}\left(L > \frac{a}{\epsilon}\right) \leq \epsilon \Leftrightarrow \mathbb{P}\left(\frac{L}{a} > \frac{1}{\epsilon}\right) \leq \epsilon.$$

□

(d) *Proof.* Note that $y = X\beta + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, so

$$\begin{aligned}
\hat{\beta} &= (X^\top X)^{-1} X^\top y_i \\
&= (X^\top X)^{-1} X^\top (X\beta + \epsilon) \\
&= \beta + (X^\top X)^{-1} X^\top \epsilon \\
\hat{\beta} - \beta &= (X^\top X)^{-1} X^\top \epsilon.
\end{aligned}$$

Thus, we have that

$$\begin{aligned}
\mathbb{E}[\hat{\beta} - \beta] &= \mathbb{E}\left[(X^\top X)^{-1} X^\top \epsilon\right] = \mathbb{E}\left[(X^\top X)^{-1} X^\top\right] \mathbb{E}[\epsilon] = 0 \\
\text{var}(\hat{\beta} - \beta) &= \mathbb{E}\left[(X^\top X)^{-1} X^\top \epsilon \epsilon^\top X (X^\top X)^{-1}\right] = \sigma^2 \mathbb{E}\left[(X^\top X)^{-1}\right].
\end{aligned}$$

If the design was fixed ($\{x_i\} \subset \mathbb{R}^p$ were deterministic), then we simply have $\text{var}(\hat{\beta} - \beta) = \sigma^2 (X^\top X)^{-1}$, so

$$\hat{\beta} - \beta \sim \mathcal{N}\left(0, \sigma^2 (X^\top X)^{-1}\right)$$

With a random design, we appeal to Slutsky's theorem. $X^\top X = \sum_{i=1}^n x_i^\top x_i$, so by strong law of large numbers $(X^\top X)/n \xrightarrow{\text{a.s.}} \Sigma$, where Σ is a constant.

$X^\top \epsilon = \sum_{i=1}^n x_i \epsilon_i$, where the $x_i \epsilon_i$ are i.i.d. and $\text{var}(x_i \epsilon_i) = \sigma^2 \Sigma$.

By CLT, we have that

$$\sqrt{n} X^\top \epsilon \xrightarrow{d} \mathcal{N}(0, \sigma^2 \Sigma)$$

So, we can now apply Slutsky's theorem to show that

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &= \sqrt{n} (X^\top X)^{-1} X^\top \epsilon \xrightarrow{d} \sqrt{n} \Sigma^{-1} \mathcal{N}(0, \sigma^2 \Sigma) = \mathcal{N}(0, \sigma^2 \Sigma^{-1} \Sigma \Sigma^{-1}) \\
&= \mathcal{N}(0, \sigma^2 \Sigma^{-1})
\end{aligned}$$

as desired. □

Problem 2

(a)

(b)