HW 4

Kevin Lin

2/6/2026

1

(i) Given design matrix $X \in \mathbb{R}^{n \times p}$ and response vector $y \in \mathbb{R}^n$, and that the columns of X are linearly dependent, we can prove that the OLS coefficient vector $\hat{\beta}$ is not unique as follows:
Since the columns of X are linearly dependent, there exists a non-zero vector $c \in \mathbb{R}^p$ such that $Xc = 0$. Let $\hat{\beta}$ be an OLS solution, i.e.,

$$\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|_2^2.$$

Now consider another vector $\hat{\beta}' = \hat{\beta} + c$. We have:

$$X\hat{\beta}' = X(\hat{\beta} + c) = X\hat{\beta} + Xc = X\hat{\beta} + 0 = X\hat{\beta}.$$

Therefore, the residuals for both $\hat{\beta}$ and $\hat{\beta}'$ are the same:

$$\|y - X\hat{\beta}'\|_2^2 = \|y - X\hat{\beta}\|_2^2.$$

This shows that there are infinitely many solutions to the OLS problem, proving that $\hat{\beta}$ is not unique.
The form of the solution set can be expressed as:

$$\{\hat{\beta} + c : c \in \mathrm{Null}(X)\},$$

where $\mathrm{Null}(X)$ is the null space of X.

(ii) If the columns of X are linearly independent, and the linear model $y = x^T\beta + \epsilon$ holds with $\epsilon \sim \mathcal{N}(0, \sigma^2)$, then the maximum variance over all unit-norm linear combinations of the fitted coefficients $\max_{\|a\|_2=1} \mathrm{Var}[a^T\hat{\beta}]$ can be derived as follows given $X = UDV^T$ (the SVD of X):
The OLS estimator is given by:

$$\hat{\beta} = (X^TX)^{-1}X^Ty.$$

The variance of $\hat{\beta}$ is:

$$\mathrm{Var}[\hat{\beta}] = \sigma^2(X^TX)^{-1}.$$

Using the SVD of X, we have:

$$X^T X = V D^2 V^T,$$

so:

$$(X^T X)^{-1} = V D^{-2} V^T.$$

Therefore:

$$\mathrm{Var}[\hat{\beta}] = \sigma^2 V D^{-2} V^T.$$

For any unit-norm vector a, we have:

$$\mathrm{Var}[a^T \hat{\beta}] = a^T \mathrm{Var}[\hat{\beta}] a = \sigma^2 a^T V D^{-2} V^T a.$$

Letting $b = V^T a$, we note that $\|b\|_2 = \|a\|_2 = 1$. Thus:

$$\mathrm{Var}[a^T \hat{\beta}] = \sigma^2 b^T D^{-2} b = \sigma^2 \sum_{i=1}^{p} \frac{b_i^2}{d_i^2},$$

where d_i are the singular values of X. To maximize this variance, we should allocate all the weight to the smallest singular value $d_{\min}$:

$$\max_{\|a\|_2=1} \mathrm{Var}[a^T \hat{\beta}] = \frac{\sigma^2}{d_{\min}^2}.$$

(iii) If the first $p-1$ columns of X are linearly independent, but the last column is a linear combination of the first $p-1$ columns, we can show that the maximum variance from part (ii) grows without bound as $\|z\|_2 \to \infty$ as follows:
If $z = 0$, then x_p lies exactly in the span of the first $p-1$ columns, so X has rank $p-1$. Therefore, the smallest singular value $d_{\min} = 0$, leading to an infinite variance:

$$\max_{\|a\|_2=1} \mathrm{Var}[a^T \hat{\beta}] = \frac{\sigma^2}{d_{\min}^2} = \infty.$$

When $z \neq 0$, still $\|z\|_2$ grows extremely small, and as $z \to 0$, likewise, the smallest singular value scales w/ the pertubation, which shows that the maximum variance grows without bound as $\|z\|_2 \to 0$.

2

See attached Jupyter notebook for code and plots.