Homework 4

Solutions

1. (Condition number)

   (a) It is equal to the maximum singular value of A^{-1}, which equals $1/s_n$.

   (b)

$$
\begin{aligned}
\|w - x_{\text{true}}\| &= \|A^{-1}(y+z) - A^{-1}y\| && (1)\\
&= \|A^{-1}z\| && (2)\\
&\leq \|A^{-1}\|\|z\| && (3)\\
&= \frac{\|z\|}{s_n}. && (4)
\end{aligned}
$$

   (c) If $z = 0$ then $x = w$ and the inequality holds (as an equality). Otherwise $h \neq 0$ and we have

$$
\begin{aligned}
\frac{\|w - x_{\text{true}}\|\|y\|}{\|x_{\text{true}}\|\|z\|} &= \frac{\|A^{-1}z\|\|Ax_{\text{true}}\|}{\|x_{\text{true}}\|\|z\|} && (5)\\
&\leq \|A^{-1}\|\|A\| && (6)\\
&= \frac{s_1}{s_n}. && (7)
\end{aligned}
$$

2. (Simple linear regression)

   (a) In this case $X \in n \times 1$ is just a row vector. Let $\tilde{z}$ denote the vector of iid noise. The OLS estimate equals

$$
\begin{aligned}
\tilde{\beta}_{OLS} &= (XX^T)^{-1}X(X^T\beta + \tilde{z}) && (8)\\
&= \beta + \frac{X\tilde{z}}{\sum_{i=1}^n x_i^2}. && (9)
\end{aligned}
$$

   By Theorem 8.6 in the notes on PCA this is a Gaussian with mean β and variance

$$
\begin{aligned}
\frac{X\Sigma_{\tilde{z}}X^T}{(\sum_{i=1}^n x_i^2)^2} &= \frac{X\sigma^2 I X^T}{(\sum_{i=1}^n x_i^2)^2} && (10)\\
&= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{(\sum_{i=1}^n x_i^2)^2} && (11)\\
&= \frac{\sigma^2}{\sum_{i=1}^n x_i^2} && (12)\\
&= \frac{1}{n\,\text{SNR}}. && (13)
\end{aligned}
$$

   (b) If n or the SNR tend to infinity, then the variance is zero, which means that the estimate is perfect. This happens even if the noise is not Gaussian, because the mean is still β and the variance is still equal to $\frac{1}{n\,\text{SNR}}$.

(c) If the noise is dependent, then the estimate can behave very differently. For example, let all the noise realizations be the same, i.e. $z_i = z_0$ for all i. Then, if $\vec{1}$ is a vector of ones, we have

$$\tilde{\beta}_{OLS} = (XX^T)^{-1}X(X^T\beta + z_0\vec{1}) \tag{14}$$

$$= \beta + \frac{X\vec{1}z_0}{\sum_{i=1}^n x_i^2} \tag{15}$$

$$= \beta + \frac{z_0 \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2}. \tag{16}$$

In this case, the estimate is only unbiased if the sample mean of $x_1, \ldots, x_n$ is zero.

3. (Best unbiased estimator)

(a)

$$\mathrm{E}[C\tilde{y}] = \mathrm{E}[C(X^T\beta + \tilde{z})] = CX^T\beta + C\mathrm{E}[\tilde{z}] = CX^T\beta.$$

(b)

$$\begin{aligned}
\mathrm{E}[(C\tilde{y})(C\tilde{y})^T] - \mu\mu^T &= C\mathrm{E}[\tilde{y}\tilde{y}^T]C^T - \mu\mu^T \\
&= C\mathrm{E}[X^T\beta\beta^T X + zz^T]C^T - \mu\mu^T \\
&= CX^T\beta\beta^T XC^T + C(\sigma^2 I)C^T - \mu\mu^T \\
&= \sigma^2 CC^T.
\end{aligned}$$

(c) By part (a) we have

$$\mathrm{E}[C\tilde{y}] = CX^T\beta = ((XX^T)^{-1}X + D)X^T\beta = (I + DX^T)\beta = \beta + DX^T\beta.$$

Thus we require $DX^T = 0$.

(d) By part (b) we have

$$\begin{aligned}
\Sigma_C &= \sigma^2 CC^T \\
&= \sigma^2((XX^T)^{-1}X + D)((XX^T)^{-1}X + D)^T \\
&= \sigma^2 \left[(XX^T)^{-1}XX^T(XX^T)^{-1} + DD^T + (XX^T)^{-1}XD^T + DX^T(XX^T)^{-1}\right] \\
&= \sigma^2((XX^T)^{-1} + DD^T) \\
&= \Sigma_{\mathrm{OLS}} + \sigma^2 DD^T.
\end{aligned}$$

Since DD^T is positive semidefinite, the result follows.

(e) By part (a) we have

$$\mathrm{E}[\hat{\beta}] = (XX^T)^{-1}XX^T\beta + (XX^T)^{-1}XZ^Tw = \beta + (XX^T)^{-1}XZ^Tw.$$

Thus to have an unbiased estimator we require $XZ^T = 0$ (since $(XX^T)^{-1}$ is invertible). Stated differently, our estimator is unbiased if the missing features are orthogonal to the known features. If not, there are choices of w causing a biased estimate of β.

4. (Distributions of β)

   (a) Figure 1

   (b) Figure 2

   (c) Yes, they align here. They align when you assume that $y = \beta_{true}^T x + \eta$ where η is the noise term. The covariance matrix of η has the form $\Sigma = \sigma^2 I$. Refer to theorem 4.3 in the notes on linear regression for more details.

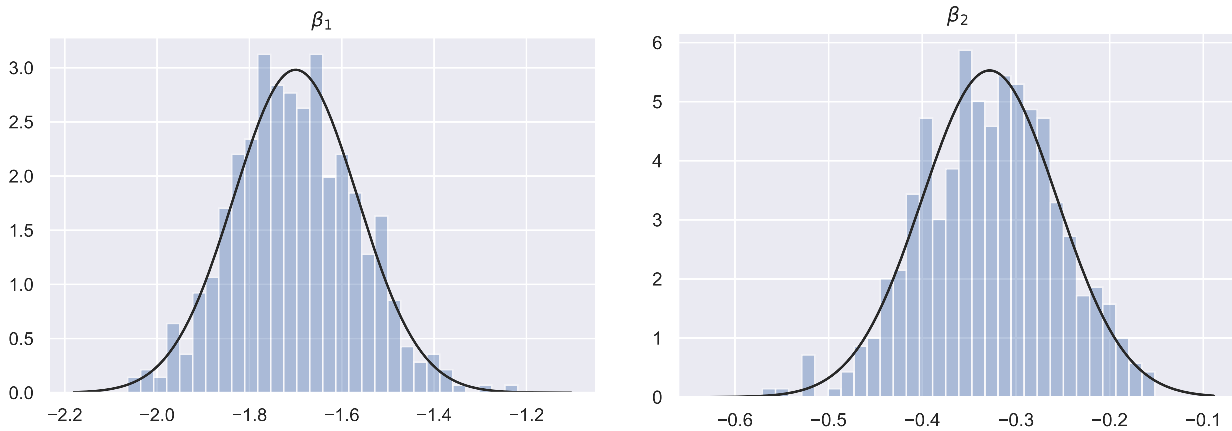

Figure 1: 4(a)

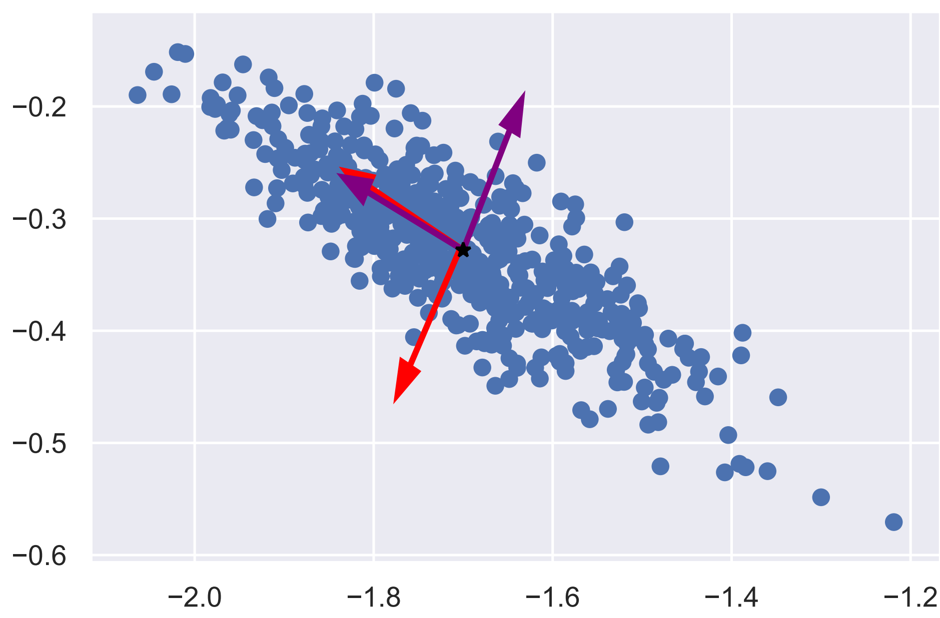

Figure 2: 4(b)