Exercise 5

2 LDA-Derivation from the Least Squares Error (16 points)

$$\begin{aligned}\frac{\partial}{\partial\beta}\sum_{i=1}^{N}(y^*-X_i\beta)^2 &= \sum_{i=1}^{N}\Big(-2X_i^\intercal\,(y^*-X_i\beta)\Big)\\ &= \sum_{i=1}^{N}(-2X_i^\intercal y^* + 2X_i^\intercal X_i\beta)\\ &= -2(\sum_{i:y_i^*=1}X_i^\intercal - \sum_{i:y_i^*=-1}X_i^\intercal) + 2\beta X_i^2\end{aligned}$$

$\because N_1 = N_{-1} = N/2$

$\therefore \sum_{i:y_i^*=1}X_i^\intercal = N_1\mu_1 = \frac{N}{2}\mu_1, \sum_{i:y_i^*=-1}X_i^\intercal = N_{-1}\mu_{-1} = \frac{N}{2}\mu_{-1}$

$\because \mathbb{E}(X^2) = \frac{\sum_{i=1}^{N}X_i^2}{N}$

$\therefore \sum_{i=1}^{N}X_i^\intercal X_i = N\cdot\mathbb{E}(X^2)$

$$\begin{aligned}\therefore\ \frac{\partial}{\partial\beta}\sum_{i=1}^{N}(y^*-X_i\beta)^2 &= -2(\frac{N}{2}\mu_1 - \frac{N}{2}\mu_{-1})^\intercal + 2\beta N\,\mathbb{E}(X^2)\\ &= -N(\mu_1-\mu_{-1})^\intercal + 2\beta N\,\mathbb{E}(X^2)\end{aligned}$$

Next, we infer the relationship between $\mathbb{V}(X)$ and Σ.

Frist we need some intermediate result.

$\because \sum_{i=1}^{N}X_i = 0$

$\therefore \mathbb{V}(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i-0)^2 = \frac{1}{N}\sum_{i=1}^{N}X_i^2$

Notice that under our assumption, $\mu_{-1}+\mu_1 = 0$.

$$\begin{aligned}\Sigma &= \frac{1}{N}\Big(\sum_{i:y_i^*=-1}(X_i-\mu_{-1})^2 + \sum_{i:y_i^*=1}(X_i-\mu_1)^2\Big)\\ &= \frac{1}{N}\Big(\sum_{i=1}^{N}(X_i-0)^2 + 2\sum_{i:y_i^*=-1}X_i\mu_{-1} - \sum_{i:y_i^*=-1}\mu_{-1}^2 + 2\sum_{i:y_i^*=1}X_i\mu_1 - \sum_{i:y_i^*=1}\mu_1^2\Big)\\ &= \mathbb{V}(X) + \frac{1}{N}(2\mu_{-1}\sum_{i:y_i^*=-1}X_i + 2\mu_1\sum_{i:y_i^*=1}X_i - \sum_{i=1}^{N}\mu_1^2)\\ &= \mathbb{V}(X) + \frac{1}{N}(2\mu_{-1}\cdot\frac{N}{2}\mu_{-1} + 2\mu_1\cdot\frac{N}{2}\mu_1 - N\cdot\mu_1^2)\\ &= \mathbb{V}(X) + \mu_{-1}^2 + \mu_1^2 - \mu_1^2\\ &= \mathbb{V}(X) + \mu_{-1}^2\end{aligned}$$

$\because \mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

also $\because \mathbb{E}(X) = 0$

$$
\begin{aligned}
\mathbb{E}(X^2) &= \mathbb{V}(X) + \mathbb{E}(X)^2 \\
\therefore \qquad &= \mathbb{V}(X) \\
&= \Sigma - \mu_{-1}^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \beta} \sum_{i=1}^{N} (y^* - X_i \beta)^2 &= -N(\mu_1 - \mu_{-1})^\intercal + 2\beta N \, \mathbb{E}(X^2) \\
&= -N(\mu_1 - \mu_{-1})^\intercal + 2\beta N(\Sigma - \mu_{-1}^2) \\
&= -N(\mu_1 - \mu_{-1})^\intercal + 2\beta N\Sigma - 2\beta N \mu_{-1}^2
\end{aligned}
$$

Now we only need to obtain the $(\mu_1 - \mu_{-1})^\intercal(\mu_1 - \mu_{-1}) = (\mu_1 - \mu_{-1})^2$ part.

$$
\begin{aligned}
(\mu_1 - \mu_{-1})^2 &= \mu_1^2 + \mu_{-1}^2 - 2\mu_1 \mu_{-1} \\
&= \mu_1^2 + \mu_{-1}^2 - 2\mu_1(-\mu_1) \\
&= 4\mu_1^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \beta} \sum_{i=1}^{N} (y^* - X_i \beta)^2 &= -N(\mu_1 - \mu_{-1})^\intercal + 2\beta N\Sigma - 2\beta N \mu_{-1}^2 \\
&= -N(\mu_1 - \mu_{-1})^\intercal + 2\beta N\Sigma - \frac{1}{2}\beta N(4\mu_1^2) \\
&= -N(\mu_1 - \mu_{-1})^\intercal + 2\beta N\Sigma - \frac{1}{2}\beta N(\mu_1 - \mu_{-1})^2 \overset{!}{=} 0
\end{aligned}
$$

Dividing by $2N$

$$
\begin{aligned}
&\Rightarrow -\frac{1}{2}(\mu_1 - \mu_{-1})^\intercal + \beta\Sigma - \frac{1}{4}\beta(\mu_1 - \mu_{-1})^2 = 0 \\
&\Rightarrow \beta\Sigma - \frac{1}{4}\beta(\mu_1 - \mu_{-1})^2 = \frac{1}{2}(\mu_1 - \mu_{-1})^\intercal
\end{aligned}
$$