# Exercise 2

We start by calculating the derivative:

$$\begin{aligned}\frac{\partial}{\partial\beta}\sum_{i=1}^{N}(y_i^* - X_i\beta)^2 &= \sum_{i=1}^{N} -2X_i^{\mathrm{T}}(y_i^* - X_i\beta) \\ &= -2\left(\sum_{i=1}^{N} X_i^{\mathrm{T}} y_i^* - (\sum_{i=1}^{N} X_i^{\mathrm{T}} X_i)\beta\right). \qquad (1)\end{aligned}$$

Because the two classes are balanced, the first sum can be rewritten in the following way:

$$\begin{aligned}\sum_{i=1}^{N} X_i^{\mathrm{T}} y_i^* &= \sum_{y_i^*=1} X_i^{\mathrm{T}} - \sum_{y_i^*=-1} X_i^{\mathrm{T}} \\ &= \frac{N}{2}(\mu_1 - \mu_{-1})^{\mathrm{T}} \qquad (2)\end{aligned}$$

This will be the right-hand side of the desired equation. To derive the left-hand side, we will work backwards:

$$\begin{aligned}N\Sigma &= \sum_{y_i^*=1}(X_i - \mu_1)^{\mathrm{T}}(X_i - \mu_1) + \sum_{y_i^*=-1}(X_i - \mu_{-1})^{\mathrm{T}}(X_i - \mu_{-1}) \\ &= \sum_{y_i^*=1}(X_i^{\mathrm{T}} X_i - X_i^{\mathrm{T}}\mu_1 - \mu_1^{\mathrm{T}} X_i + \mu_1^{\mathrm{T}}\mu_1) + \sum_{y_i^*=-1}(X_i^{\mathrm{T}} X_i - X_i^{\mathrm{T}}\mu_{-1} - \mu_{-1}^{\mathrm{T}} X_i + \mu_{-1}^{\mathrm{T}}\mu_{-1}) \\ &= \sum_{i=1}^{N} X_i^{\mathrm{T}} X_i - \left(\sum_{y_i^*=1} X_i^{\mathrm{T}}\right)\mu_1 - \mu_1^{\mathrm{T}}\left(\sum_{y_i^*=1} X_i\right) + \frac{N}{2}\mu_1^{\mathrm{T}}\mu_1 \\ &\quad - \left(\sum_{y_i^*=-1} X_i^{\mathrm{T}}\right)\mu_{-1} - \mu_{-1}^{\mathrm{T}}\left(\sum_{y_i^*=-1} X_i\right) + \frac{N}{2}\mu_{-1}^{\mathrm{T}}\mu_{-1} \\ &= \sum_{i=1}^{N} X_i^{\mathrm{T}} X_i - \frac{N}{2}\left(\mu_1^{\mathrm{T}}\mu_1 + \mu_{-1}^{\mathrm{T}}\mu_{-1}\right).\end{aligned}$$

We add the second term of the left-hand side, namely $\frac{1}{4}(\mu_1-\mu_{-1})^{\mathrm{T}}(\mu_1-\mu_{-1})$:

$$
\begin{aligned}
&\Sigma + \frac{1}{4}(\mu_1 - \mu_{-1})^{\mathrm{T}}(\mu_1 - \mu_{-1}) \\
&= \frac{1}{N}\sum_{i=1}^{N} X_i^{\mathrm{T}} X_i - \frac{1}{2}\left(\mu_1^{\mathrm{T}}\mu_1 + \mu_{-1}^{\mathrm{T}}\mu_{-1}\right) + \frac{1}{4}(\mu_1^{\mathrm{T}}\mu_1 - \mu_1^{\mathrm{T}}\mu_{-1} - \mu_{-1}^{\mathrm{T}}\mu_1 + \mu_{-1}^{\mathrm{T}}\mu_{-1}) \\
&= \frac{1}{N}\sum_{i=1}^{N} X_i^{\mathrm{T}} X_i - \frac{1}{4}(\mu_1^{\mathrm{T}}\mu_1 + \mu_1^{\mathrm{T}}\mu_{-1} + \mu_{-1}^{\mathrm{T}}\mu_1 + \mu_{-1}^{\mathrm{T}}\mu_{-1}) \\
&= \frac{1}{N}\sum_{i=1}^{N} X_i^{\mathrm{T}} X_i - \frac{1}{4}(\mu_1 + \mu_{-1})^{\mathrm{T}}(\mu_1 + \mu_{-1}) \\
&= \frac{1}{N}\sum_{i=1}^{N} X_i^{\mathrm{T}} X_i. \qquad (3)
\end{aligned}
$$

For the last step, note that because the data are assumed to be centered, we find that $\mu_1 + \mu_{-1} = \frac{2}{N}\sum_{i=1}^{N} X_i = 0$, thus the last term vanishes. Putting everything together, we get the desired equation:

$$
\begin{aligned}
\frac{\partial}{\partial\beta}\sum_{i=1}^{N}(y_i^* - X_i\beta)^2 &\stackrel{!}{=} 0 \\
\stackrel{(1)}{\Rightarrow} \sum_{i=1}^{N} X_i^{\mathrm{T}} y_i^* &= (\sum_{i=1}^{N} X_i^{\mathrm{T}} X_i)\beta \\
\stackrel{(2),(3)}{\Longrightarrow} \frac{N}{2}(\mu_1 - \mu_{-1})^{\mathrm{T}} &= N\left(\Sigma + \frac{1}{4}(\mu_1 - \mu_{-1})^{\mathrm{T}}(\mu_1 - \mu_{-1})\right)\beta \\
\Rightarrow \Sigma\beta + \frac{1}{4}(\mu_1 - \mu_{-1})^{\mathrm{T}}(\mu_1 - \mu_{-1})\beta &= \frac{1}{2}(\mu_1 - \mu_{-1})^{\mathrm{T}}.
\end{aligned}
$$