

Exercise # 2

Q1) To begin we are given the equation for ridge regression. We will expand out the L2 norms of this expression to help simplify

$$\begin{aligned}
 &= \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2n} \|Xw + b1 - y\|_2^2 + \lambda \|w\|_2^2 \\
 &= \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2n} ((Xw + b1 - y)^T (Xw + b1 - y)) + \lambda (w^T w) \\
 &= \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2n} ((X^T w^T + b^T 1^T - y^T)(Xw + b1 - y)) + \lambda (w^T w) \\
 &= \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2n} ((X^T w^T Xw + X^T w^T b - X^T w^T y + b^T Xw + b^T b - b^T y - y^T Xw - y^T b + y^T y) + \lambda (w^T w)) \\
 &= \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2n} (X^T w^T Xw + X^T w^T b - X^T w^T y + b^T Xw + b^T b - b^T y - y^T Xw - y^T b + y^T y + 2n\lambda w^T w)
 \end{aligned}$$

From here we will split the elements into a matrix with 2 elements:

$$\begin{aligned}
 &= \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2n} \left(\begin{bmatrix} X^T w^T Xw + X^T w^T b - X^T w^T y + b^T Xw + b^T b - b^T y - y^T Xw - y^T b + y^T y \\ 2n\lambda w^T w \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) \\
 &= \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2n} \left(\begin{bmatrix} X^T w^T + b^T 1^T - y^T \\ \sqrt{2n\lambda} w^T \end{bmatrix} \begin{bmatrix} Xw + b1 - y \\ \sqrt{2n\lambda} w \end{bmatrix} \right) \\
 &= \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2n} \left(\begin{bmatrix} Xw + b1 - y \\ \sqrt{2n\lambda} w \end{bmatrix}^T \begin{bmatrix} Xw + b1 - y \\ \sqrt{2n\lambda} w \end{bmatrix} \right) \\
 &= \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2n} \left\| \begin{bmatrix} Xw + b1 - y \\ \sqrt{2n\lambda} w \end{bmatrix} \right\|_2^2 \\
 &= \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2n} \left\| \begin{bmatrix} Xw + b1 \\ \sqrt{2n\lambda} w \end{bmatrix} - \begin{bmatrix} y \\ 0_d \end{bmatrix} \right\|_2^2 \\
 &= \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{2n} \left\| \begin{bmatrix} X & 1_n \\ \sqrt{2n\lambda} & 0_d \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} - \begin{bmatrix} y \\ 0_d \end{bmatrix} \right\|_2^2
 \end{aligned}$$

Thus showing how we can transform the original equation into what was required.

Q2) We are now asked to calculate the derivative of the previous equation. Note that we can drop the minimum as it will have no impact on the derivative. We thus begin with the following equation:

$$\frac{1}{2n} (X^T w^T Xw + X^T w^T b - X^T w^T y + b^T Xw + b^T b - b^T y - y^T Xw - y^T b + y^T y + 2n\lambda w^T w)$$

We will then expand it out further to get:

$$\frac{X^T w^T Xw}{2n} + \frac{X^T w^T b}{2n} - \frac{X^T w^T y}{2n} + \frac{b^T Xw}{2n} + \frac{b^T b}{2n} - \frac{b^T y}{2n} - \frac{y^T Xw}{2n} - \frac{y^T b}{2n} + \frac{y^T y}{2n} + \lambda w^T w$$

To answer the first part of this question we will use this equation to calculate the derivative w.r.t w :

$$\begin{aligned}
\frac{\delta}{\delta w} &= \left(\frac{X^T w^T X w}{2n} + \frac{X^T w^T b}{2n} - \frac{X^T w^T y}{2n} + \frac{b^T X w}{2n} + \frac{b^T b}{2n} - \frac{b^T y}{2n} - \frac{y^T X w}{2n} - \frac{y^T b}{2n} + \frac{y^T y}{2n} + \lambda w^T w \right)' \\
&= \left(\frac{X^T w^T X w}{2n} \right)' + \left(\frac{X^T w^T b}{2n} \right)' - \left(\frac{X^T w^T y}{2n} \right)' + \left(\frac{b^T 1^T X w}{2n} \right)' - \left(\frac{y^T X w}{2n} \right)' + (\lambda w^T w)' \\
&= \left(\frac{2X^T X w}{2n} \right) + \left(\frac{X^T b 1}{2n} \right) - \left(\frac{X^T y}{2n} \right) + \left(\frac{b^T 1^T X}{2n} \right) - \left(\frac{y^T X}{2n} \right) + (2\lambda w) \\
&= \frac{X^T X w}{n} + \left(\frac{X^T b 1}{2n} + \frac{b^T 1^T X}{2n} \right) - \left(\frac{X^T y}{2n} + \frac{y^T X}{2n} \right) + 2\lambda w \\
&= \frac{X^T X w}{n} + \frac{X^T b 1}{n} - \frac{X^T y}{n} + 2\lambda w \\
&= \frac{1}{n} X^T (X w + b 1 - y) + 2\lambda w
\end{aligned}$$

Thus providing the derivative of w as required. We will now calculate the derivative w.r.t b :

$$\begin{aligned}
\frac{\delta}{\delta b} &= \left(\frac{X^T w^T X w}{2n} + \frac{X^T w^T b}{2n} - \frac{X^T w^T y}{2n} + \frac{b^T X w}{2n} + \frac{b^T b}{2n} - \frac{b^T y}{2n} - \frac{y^T X w}{2n} - \frac{y^T b}{2n} + \frac{y^T y}{2n} + \lambda w^T w \right)' \\
&= \left(\frac{X^T w^T b 1}{2n} \right)' + \left(\frac{b^T 1^T X w}{2n} \right)' + \left(\frac{b^T 1^T b 1}{2n} \right)' - \left(\frac{b^T 1^T y}{2n} \right)' - \left(\frac{y^T b 1}{2n} \right)' \\
&= \left(\frac{X^T w^T 1}{2n} \right) + \left(\frac{1^T X w}{2n} \right) + \left(\frac{1^T b 1}{n} \right) - \left(\frac{1^T y}{2n} \right) - \left(\frac{y^T 1}{2n} \right) \\
&= \left(\frac{X^T w^T 1}{2n} + \frac{1^T X w}{2n} \right) - \left(\frac{1^T y}{2n} + \frac{y^T 1}{2n} \right) + \left(\frac{b 1^T 1}{n} \right) \\
&= \left(\frac{1^T X w}{n} \right) - \left(\frac{1^T y}{n} \right) + \left(\frac{b 1^T 1}{n} \right) \\
&= \frac{1}{n} 1^T (X w + b 1 - y)
\end{aligned}$$

Thus proving the derivative of b as needed.