Q2. (1)

$$\min_{w\in\mathbb{R}^d,\, b\in\mathbb{R}} \overbrace{\underbrace{\frac{1}{2n}\|Xw + b\mathbf{1} - y\|_2^2}_{\text{error}} + \lambda\|w\|_2^2}^{\text{loss}}, \tag{1}$$

where $X \in \mathbb{R}^{n\times d}$ and $y \in \mathbb{R}^n$ are the given dataset and $\lambda \geq 0$ is the regularization hyperparameter. If $\lambda = 0$, then this is the standard linear regression problem. Observe the distinction between the *error* (which does not include the regularization term) and the *loss* (which does).

Prove (1) is equivalent to

$$\min_{w\in\mathbb{R}^d,\, b\in\mathbb{R}} \frac{1}{2n}\left\|\begin{bmatrix} X & \mathbf{1}_n \\ \sqrt{2\lambda n}\, I_d & 0_d \end{bmatrix}\begin{bmatrix} w \\ b \end{bmatrix} - \begin{bmatrix} y \\ 0_d \end{bmatrix}\right\|_2^2$$

Simplifying the expression inside the norm,

$$\begin{aligned}
&\begin{bmatrix} X & \mathbf{1}_n \\ \sqrt{2\lambda n}\, I_d & 0_d \end{bmatrix}\begin{bmatrix} w \\ b \end{bmatrix} - \begin{bmatrix} y \\ 0_d \end{bmatrix} \\
&= \begin{bmatrix} Xw + b\mathbf{1}_n \\ \sqrt{2\lambda n}\, I_d\, w + 0_d \cdot b \end{bmatrix} - \begin{bmatrix} y \\ 0_d \end{bmatrix} \\
&= \begin{bmatrix} Xw + b\mathbf{1}_n - y \\ \sqrt{2\lambda n}\, w - 0_d \end{bmatrix} \\
&= \begin{bmatrix} Xw + b\mathbf{1}_n - y \\ \sqrt{2\lambda n}\, w \end{bmatrix}
\end{aligned}$$

So the full expression becomes

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

(2) We have that the loss function is:

$$L(w,b) = \frac{1}{2n}||Xw + b1 - y||_2^2 + \lambda||w||_2^2$$

We will first calculate the partial derivative with respect to w:

$$\frac{\partial}{\partial w}L(w,b) = \frac{\partial}{\partial w}\left[\frac{1}{2n}\|Xw + b\mathbf{1} - y\|_2^2 + \lambda\|w\|_2^2\right]$$

$$
\begin{aligned}
&= \frac{\partial}{\partial w}\left[\frac{1}{2n}(Xw + b\mathbf{1} - y)^T(Xw + b\mathbf{1} - y) + \lambda(w)^T(w)\right] \\
&= \frac{1}{2n}\frac{\partial}{\partial w}\left[(Xw + b\mathbf{1} - y)^T(Xw + b\mathbf{1} - y)\right] + \lambda\frac{\partial}{\partial w}\left[(w)^T(w)\right] \\
&= \frac{1}{2n}2(Xw + b\mathbf{1} - y)\frac{\partial}{\partial w}(Xw + b\mathbf{1} - y)^T + 2\lambda w \\
&= \frac{1}{n}X^T(Xw + b\mathbf{1} - y) + 2\lambda w
\end{aligned}
$$

Now, we will calculate the partial derivative with respect to b:

$$
\begin{aligned}
\frac{\partial}{\partial b}L(w, b) &= \frac{\partial}{\partial b}\left[\frac{1}{2n}\|Xw + b\mathbf{1} - y\|_2^2 + \lambda\|w\|_2^2\right] \\
&= \frac{\partial}{\partial b}\left[\frac{1}{2n}(Xw + b\mathbf{1} - y)^T(Xw + b\mathbf{1} - y) + \lambda(w)^T(w)\right] \\
&= \frac{1}{2n}\frac{\partial}{\partial b}\left[(Xw + b\mathbf{1} - y)^T(Xw + b\mathbf{1} - y)\right] + \lambda\frac{\partial}{\partial b}\left[(w)^T(w)\right] \\
&= \frac{1}{2n}2(Xw + b\mathbf{1} - y)\frac{\partial}{\partial b}(Xw + b\mathbf{1} - y)^T + \lambda 0 \\
&= \frac{1}{n}\mathbf{1}^T(Xw + b\mathbf{1} - y)
\end{aligned}
$$