Ex. 3.12

We will denote this new augmented Matrix as:

$$X_a = \begin{bmatrix} X \\ \sqrt{\lambda}\,\mathbb{I}_p \end{bmatrix}, \text{ where dimensions are: } \begin{aligned} X &= N \times P \\ \mathbb{I}_p &= P \times P \\ \Rightarrow X_a &= (N+P) \times P \end{aligned}$$

Now we will calculate the least squares estimates of this augmented data:

$$\hat{\beta}_{ls} = (X_a^T X_a)^{-1} X_a^T y_a$$

Where $y_a = \begin{bmatrix} y \\ 0_p \end{bmatrix}$, where y and 0_p are vectors of length N and P denoting the labels of X and the 0 Vector respectively.

Now,

$$X_a^T X_a = [X^T, \sqrt{\lambda}\,\mathbb{I}_p] \begin{bmatrix} X \\ \sqrt{\lambda}\,\mathbb{I}_p \end{bmatrix} = X^T X + \lambda \mathbb{I}_p$$

and

$$X_a^T y_a = [X^T, \sqrt{\lambda}\,\mathbb{I}_p] \begin{bmatrix} y \\ 0_p \end{bmatrix} = X^T y$$

So $\hat{\beta}_{ls} = (X^T X + \lambda \mathbb{I}_p)^{-1} X^T y$

and since X has been Centered this is simply the ridge regression estimates of X and y