Ex 3.29

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

Begin by calculating "a" in the case that X is a single column vector e.g.

$$X = [X_1] = \begin{bmatrix} X_{11} \\ \vdots \\ X_{1N} \end{bmatrix}$$

Now $\hat{\beta}^{ridge} = (\sum_i X_{1i}^2 + \lambda)^{-1} \sum_i X_{1i} y_i = \dfrac{\sum_i X_{1i} y_i}{\sum_i X_{1i}^2 + \lambda} = a$

Now consider the case where we have m copies of the column vector X_1 (including the case where $m=2$) e.g.

$$X = [\underbrace{X_1, X_1, \ldots, X_1}_{m \text{ times}}] = \begin{bmatrix} X_{11} & \cdots & X_{11} \\ \vdots & & \vdots \\ X_{1N} & & X_{1N} \end{bmatrix} \quad dim = N \times M$$

Now,

$$\hat{\beta}^{ridge} = (\sum X_{1i}^2 \cdot \mathbb{1}_{P \times P} + \lambda \mathbb{I})^{-1} \cdot \begin{bmatrix} \sum_i X_{1i} y_i \\ \vdots \\ \sum_i X_{1i} y_i \end{bmatrix}$$

Where $\mathbb{1}_{P \times P}$ is a $\underset{=}{\overset{m \times m}{}} P \times P$ matrix of 1's $\Rightarrow ee^T$ where e is a column vector of length p made up of 1's

To calculate the inverse we will use Sherman-Morrison:

$$(D + bee^T)^{-1} = D^{-1} + b \frac{D^{-1} ee^T D^{-1}}{1 + b e^T D^{-1} e}$$

Where D is a diagonal matrix and b is a scalar.

Therefore,

$$(\lambda \mathbb{I} + \sum_i x_{1i}^2 ee^T)^{-1} = 1/\lambda \mathbb{I} - \frac{\sum X_{1i}^2}{\lambda^2(1 + \sum X_i^2 \cdot P/\lambda)} \cdot e^T e$$

$$\Rightarrow 1/\lambda \left[\mathbb{I} - \frac{\sum X_{1i}^2}{\lambda + P\sum X_{1i}^2} e^T e \right] \begin{bmatrix} \sum X_{1i} y_i \\ \vdots \\ \sum X_{1i} y_i \end{bmatrix} = \hat{\beta}^{ridge}$$

It is clear now that all $m=P$ entries of $\hat{\beta}^{ridge}$ are identical. Any given entry is given by:

$$1/\lambda \left[1 - \frac{\sum X_{1i}^2}{\lambda + P\sum X_{1i}^2} - (P-1)\left(\frac{\sum X_{1i}^2}{\lambda + P\sum X_{1i}^2}\right) \right] \cdot \sum X_{1i} y_i$$

$$= 1/\lambda \left[1 - P \cdot \frac{\sum X_{1i}^2}{\lambda + P\sum X_{1i}^2} \right] \cdot \sum X_{1i} y_i$$

$$= \frac{1}{\lambda} \left[\frac{\lambda + P\sum X_{1i}^2 - P\sum X_{1i}^2}{\lambda + P\sum X_{1i}^2} \right] \cdot \sum X_{1i} y_i$$

$$= \frac{\sum X_{1i} y_i}{\lambda + P\sum X_{1i}^2}$$ (all elements of $\hat{\beta}^{ridge}$ are identically equal to this)