2.

$$Y_{(N \times P)}$$

$$X_{(N \times d)}$$

$$W_{(d \times P)}$$

$$\Rightarrow Err_{(N \times P)} = Y_{(N \times P)} - X_{(N \times d)} W_{(d \times P)}$$

$$\Rightarrow loss =$$

$$J(W)_{(P \times P)} = \frac{1}{N} (Err)^T (Err)$$

$$= \frac{1}{N} (Y - Xw)^T (Y - Xw)$$

$$= \frac{1}{N} (Y^T Y + w^T X^T X w - 2w^T (X^T Y))$$

to minimize loss w.r.t. w,

$$\frac{\partial (J(w))}{\partial w} = 0$$

$$\Rightarrow \frac{1}{N} \left(2 X^T X w - 2(X^T Y) \right) = 0$$

$$\Rightarrow X^T X \omega - X^T Y = 0$$

$$\Rightarrow X^T X \omega = X^T Y$$

$$\Rightarrow \cancel{\omega = X^T Y}$$

$$\Rightarrow (X^T X)^{-1} (X^T X) \omega = (X^T X)^{-1} (X^T X) X^T Y$$

$$\Rightarrow I \omega = (X^T X)^{-1} X^T Y$$

$$\Rightarrow \omega = (X^T X)^{-1} X^T Y$$

again where

$$Y_{N \times P}$$

$$X_{N \times d}$$

$$\omega_{d \times P}$$