Derive the optimal least squares parameter value $\hat{\underline{w}}$ for the total training loss

$$L = \sum_{n=1}^{N} (t_n - \underline{w}^T \underline{x}_n)^2$$

$$= (\underline{t} - \underline{X}\underline{w})^T (\underline{t} - \underline{X}\underline{w})$$

$$= \underline{t}^T\underline{t} - 2\underline{w}^T\underline{X}^T\underline{t} + \underline{w}^T\underline{X}^T\underline{X}\underline{w}$$

$$\frac{\partial L}{\partial \underline{w}} = -2\underline{X}^T\underline{t} + 2\underline{X}^T\underline{X}\underline{w}$$

Set partial derivative to 0.

$$\therefore \quad \underline{X}^T\underline{t} = \underline{X}^T\underline{X}\underline{w}$$

$$\therefore \quad \underline{w} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{t}$$

It's the same as the average loss. Only differs by a constant.