a/

$$MSE_{train}(w) = \frac{1}{N}\|w - w^*\|_R^2 + \frac{1}{N} y^T (I_N - XX^\dagger) y$$

we can rewrite $\frac{1}{N}\|w - w^*\|_R^2$ as:

$$\frac{1}{N}(w - w^*)^T R (w - w^*) = \frac{1}{N}(w^T - w^{*T}) R (w - w^*)$$

$$= \frac{1}{N}(w^T - w^{*T})(Rw - Rw^*) = \frac{1}{N}(w^T R w - w^T R w^* - w^{*T} R w + w^{*T} R w^*)$$

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

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

and rewrite $\frac{1}{N} y^T (I_N - XX^\dagger) y$ as:

$$\frac{1}{N}(y^T y - y^T X X^\dagger y) = \frac{1}{N}(y^T y - y^T X R^{-1} X^T y)$$

Then $\frac{1}{N}\|w - w^*\|_R^2 + \frac{1}{N} y^T (I_N - XX^\dagger) y$

$$\Rightarrow \frac{1}{N}\left[w^T R w - w^T X^T y - y^T X (R^{-1})^T R w + \cancel{y^T X (R^{-1})^T X^T y} + y^T y - \cancel{y^T X R^{-1} X^T y}\right]$$

$$= \frac{1}{N}\left(w^T R w - w^T x^T y - y^T x w + y^T y\right)$$

$$= \frac{1}{N}\left(w^T R w - 2 w^T x^T y + y^T y\right)$$

$$= \frac{1}{N}\left(w^T x^T x w - 2 w^T x^T y + y^T y\right)$$

$$= \frac{1}{N}\left(y - X\tilde{w}\right)\left(y - X\tilde{w}\right)$$

$$\boxed{= \frac{1}{N}\| y - X\tilde{w}\|_2^2 = MSE_{train}(w)}$$

b/

$$y = w^* x + b$$

$$\frac{1}{N}\sum_{n=1}^{N} y_n = w^* \frac{1}{N}\sum_{n=1}^{N} x_n + b$$

Assuming $b=0$, the optimal hyperplane would yield the optimal MSE_{train} which we denote as -

$$MSE_{train}(w^*) = \frac{1}{N}\sum_{n=1}^{N}\left(y_n - w^{*T} x_n\right)^2$$

$$\boxed{MSE_{train}(w^*) = \frac{1}{N}\|\underline{y} - \underline{X}\underline{w}^*\|_2^2}$$

c/ $RMSE_{train}(\tilde{w}|\mu) \triangleq \frac{1}{N}\|y - \tilde{X}\tilde{w}\|_2^2 + \frac{\mu}{N}\|w\|_2^2$

$$= \frac{1}{N}(\tilde{w}^T\tilde{X}^T\tilde{X}\tilde{w} - 2\tilde{w}^T\tilde{X}^Ty + y^Ty) + \frac{\mu}{N}w^Tw$$

$$\frac{\partial(RMSE_{train}(\tilde{w}|\mu))}{\partial\tilde{w}} = \frac{1}{N}(2\tilde{X}^T\tilde{X}\tilde{w} - 2\tilde{X}^Ty) + \frac{\mu}{N}2\tilde{w}^T$$

$$\frac{2}{N}\tilde{X}^T(\tilde{X}\tilde{w} - y) + \frac{2\mu}{N}\tilde{w}^T = \frac{2}{N}\left[\tilde{X}(\tilde{X}\tilde{w} - y) + \frac{\mu\tilde{w}^T}{N}\right]$$

for $\frac{\partial(RMSE_{train})}{\partial\tilde{w}}$ to equal zeor

$$\tilde{X}^T(\tilde{X}\hat{w} - y) + \frac{\mu\tilde{w}^T}{N} = 0$$

$$\tilde{X}^T\tilde{X}\tilde{w} - \tilde{X}^Ty + \frac{\mu\tilde{w}^T}{N} = 0$$

$$\tilde{X}^T\tilde{X}\tilde{w} + \frac{\mu\tilde{w}^T}{N} = \tilde{X}^Ty$$

$$\left(\tilde{X}^T\tilde{X} + \frac{\mu}{N}\right)\tilde{w} = \tilde{X}^Ty$$

$$\boxed{\tilde{w} = \tilde{X}^Ty\left(\tilde{X}^T\tilde{X} + \frac{\mu}{N}\right)^{-1} = w^*(\mu)}$$

Then

$$\boxed{BMSE_{train}(w^*(\mu)) = \frac{1}{N} \| y - X(\tilde{X}^T y (\tilde{X}^T \tilde{X} + \frac{\mu}{N})^{-1}) \|}$$