Ridge Regression (Regularized MSE Regression)

$$J(\underline{w}) = \frac{1}{N}\|\underline{\underline{X}}\underline{w} - \underline{y}\|_2^2 + \underbrace{\lambda\|\underline{w}\|_2^2}$$

$\|\underline{w}\|_2$ can't be too large
So regression won't be overfitting.

$$\therefore \nabla_{\underline{w}} J(\underline{w}) = \frac{1}{N}\left[2\underline{\underline{X}}^T\underline{\underline{X}}\underline{w} - 2\underline{\underline{X}}^T\underline{y}\right] + 2\lambda\underline{w} = 0$$

$$\Rightarrow \quad (\underline{\underline{X}}^T\underline{\underline{X}} + N\lambda\underline{\underline{I}})\hat{\underline{w}} = \underline{\underline{X}}^T\underline{y}$$

$$\boxed{\hat{\underline{w}} = (\underline{\underline{X}}^T\underline{\underline{X}} + N\lambda\underline{\underline{I}})^{-1}\underline{\underline{X}}^T\underline{y}}$$

← Algebraic Solution

Gradient Descent (Sequential)

$$J(\underline{w}) = \frac{1}{N}\sum_{n=1}^{N}(\underline{w}^T\underline{x}_n - y_n)^2 + \lambda\|\underline{w}\|_2^2$$

← augmented notation

$$\therefore J_n(\underline{w}) = \frac{1}{N}(\underline{w}^T\underline{x}_n - y_n)^2 + \lambda\|\underline{w}\|_2^2$$

$$\nabla_{\underline{w}} J_n(\underline{w}) = \frac{2}{N}\left[\underline{w}^T\underline{x}_n - y_n\right]\cdot\underline{x}_n + 2\lambda\underline{w}$$

weight update $\boxed{\underline{w}(i+1) = \underline{w}(i) - \eta(i)\left[\frac{2}{N}\left(\underline{w}(i)^T\underline{x}_n - y_n\right)\underline{x}_n + 2\lambda\underline{w}(i)\right]}$