Least Squares Method

$$J(\underline{\theta}) = \sum_n (y_n - \underline{\theta}^T \underline{x}_n)^2$$

$$= \sum_n \left[y_n^2 - 2(\underline{\theta}^T \underline{x}_n) y_n + (\underline{\theta}^T \underline{x}_n)(\underline{\theta}^T \underline{x}_n) \right]$$

$$= \sum_n \left[y_n^2 - 2(\underline{\theta}^T \underline{x}_n) y_n + \underline{\theta}^T \underbrace{\underline{x}_n \underline{x}_n^T}_{\text{Symmetric matrix}} \underline{\theta} \right]$$

$$\frac{dJ(\underline{\theta})}{d\underline{\theta}} = 0 \Rightarrow \sum_n y_n \underline{x}_n - \sum_n \underline{x}_n \underline{x}_n^T \underline{\theta} = 0$$

$$\frac{d(\underline{\theta}^T A \underline{\theta})}{d\underline{\theta}} = 2A\underline{\theta} \quad (A \text{ symmetric}).$$

$$\frac{d(A^T\underline{\theta})}{d\underline{\theta}} = A$$

$$\boxed{\sum_n \underline{x}_n \underline{x}_n^T \underline{\theta} = \sum_n y_n \underline{x}_n .}$$

<u>Using matrices:</u>

$$\mathcal{X} = \begin{bmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \vdots \\ \underline{x}_N^T \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1\ell} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & \cdots & x_{N\ell} & 1 \end{bmatrix} \qquad \underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Samples in tr. set

Components - features.

Compare $\underline{y}$ with $\mathcal{X}\underline{\theta}$

$$J(\underline{\theta}) = \|\underline{y} - X\underline{\theta}\|^2 = (\underline{y} - X\underline{\theta})^T(\underline{y} - X\underline{\theta})$$

$$= \underline{y}^T\underline{y} + \underline{\theta}^T X^T X \underline{\theta} - \underline{y}^T X \underline{\theta} - \underline{\theta}^T X^T \underline{y}$$

symmetric matrix

$$\frac{dJ}{d\underline{\theta}} = 0 \quad \Rightarrow \quad X^T X \underline{\theta} - X^T \underline{y} = 0.$$

$$\Rightarrow X^T X \underline{\theta} = X^T \underline{y} \quad \Rightarrow \quad \boxed{\hat{\underline{\theta}} = (X^T X)^{-1} \cdot X^T \underline{y}}$$

Cross term in MSE of estimator:

$$E\left[\left(\hat{\theta}-E[\hat{\theta}]\right)\left(E[\hat{\theta}]-\theta_0\right)\right]=$$

$$= E\left[\hat{\theta}E(\hat{\theta}) - \hat{\theta}\theta_0 - \left(E[\theta]\right)^2 + \theta_0 E[\theta]\right]$$

$$= E[\hat{\theta}]E[\hat{\theta}] - \theta_0 E[\hat{\theta}] - \left(E[\theta]\right)^2 + \theta_0 E[\theta]$$

$$= 0.$$