Programming Assignment 1 (Finished)

Linear Regression

Data Matrix

$$\mathcal{X} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix}$$

Observation Vector

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Weight Vector

$$\underline{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}$$

Prediction Vector

$$\underline{\hat{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_N \end{bmatrix}$$

$$\hat{y}_i = \underline{x_i^T \, w}$$

$$\underline{\hat{y}} = \mathcal{X} \, \underline{w}$$

$$y_i = x_{i0} w_0 + x_{i1} w_1 + \dots + x_{id} w_d$$

Loss Function

$$E_{in}(w) = \| \underline{Y} - \underline{\hat{Y}} \|^2 = \| \underline{Y} - \mathcal{X} \underline{w} \|^2$$

$$\underline{w^*} = \underset{\underline{w} \in \mathbb{R}^{d+1}}{\operatorname{argmin}} E_{in}(w) \qquad \underline{\hat{y}} = \mathcal{X} \, \underline{w^*}$$

$$= (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T \underline{y}$$

Least Square Solution

Geometric Derivation of Least Squares

$$\underline{\hat{y}} = \mathcal{X}\,\underline{w}$$

$\mathcal{X}$: Data Matrix $\in \mathbb{R}^{N\times(d+1)}$

$$\mathcal{X} = \begin{bmatrix} Q_0 & Q_1 & Q_2 & \cdots & Q_d \end{bmatrix}$$

Length N Column Vector

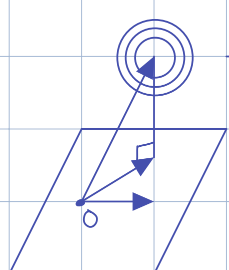
$$\underline{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}$$

$$\underline{\hat{y}} = \begin{bmatrix} Q_0 & Q_1 & Q_2 & \cdots & Q_d \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}$$

$$= w_0 \underline{Q_0} + w_1 \underline{Q_1} + \cdots + w_d \underline{Q_d}$$

$$\in \text{span}\{\underline{Q_0}, \underline{Q_1}, \ldots, \underline{Q_d}\} \equiv \text{column span}\{\mathcal{X}\}$$

$$\underline{Y} \in \mathbb{R}^{N+1} \qquad = P \subset \mathbb{R}^N$$



$P \subseteq \mathbb{R}^N$ minimize $\|\underline{Y} - \underline{\hat{Y}}\|^2$

Solution, project $\underline{Y}$ onto subspace $\mathcal{P}$

$\longrightarrow$ $\underline{Y} - \underline{\hat{Y}}$ is perpendicular to every vector in $\mathcal{P}$ and in particular to $\underline{Q_0}, \underline{Q_1}, \ldots, \underline{Q_d}$

$Q_i^T(\underline{Y} - \underline{\hat{Y}}) = 0$ for all i

$$\begin{bmatrix} Q_0 \\ Q_1 \\ \vdots \\ Q_d \end{bmatrix} (\underline{Y} - \underline{\hat{Y}}) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\mathcal{X} = [Q_0 \mid Q_1 \mid Q_2 \mid \cdots \mid Q_d]$$

$$\mathcal{X}^T(\underline{Y} - \mathcal{X}\underline{w}) = \underline{0}$$

$$\mathcal{X}^T\underline{Y} = \mathcal{X}^T\mathcal{X}\underline{w}$$

$$\underline{w} = (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T Y \longrightarrow \text{Least Square Solution}$$

Regularized Least Square

$$\underset{\underline{w} \in \mathbb{R}^{d+1}}{\operatorname{argmin}} \Big\{ \underbrace{\|\underline{Y} - \mathcal{X}\underline{w}\|^2}_{E_{in}(\underline{w})} + \lambda \|\underline{w}\|^2 \Big\} = \underline{w_\lambda^*}$$

$\lambda \geq 0$, $\lambda \equiv$ Regularization Coefficient

Prevents This

$$\hat{y} = w_0 + w_1 x$$

$w_0 \to 10{,}000$, $w_1 \to -10{,}000$

$$f(\underline{w}) = \| \underline{Y} - \mathcal{X}\underline{w} \|^2 + \lambda \|\underline{w}\|^2$$

$f(\cdot)$ is also a convex function

$$\nabla f(\underline{w}) = 0$$

$$\nabla_{\underline{w}} \left(\| \underline{Y} - \mathcal{X}\underline{w} \|^2 + \lambda \|\underline{w}\|^2 \right) = 0 \qquad I$$

$$= 2\mathcal{X}^T\mathcal{X}\underline{w} - 2\mathcal{X}^T\underline{Y} + \lambda \nabla_{\underline{w}} \left(\underline{w^T A w} \right) = 0$$

$$= 2\mathcal{X}^T\mathcal{X}\underline{w} - 2\mathcal{X}^T\underline{Y} + 2\lambda \underline{w}$$

$$(\mathcal{X}^T\mathcal{X} + \lambda)\underline{w} = \mathcal{X}^T\underline{Y}$$

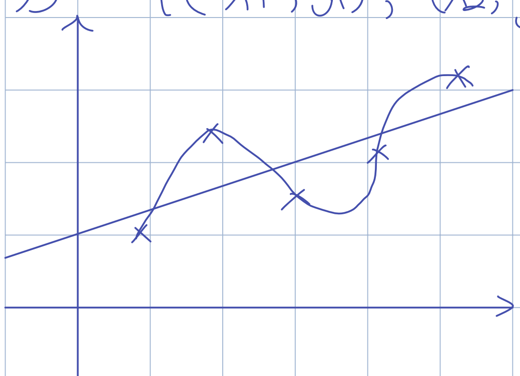
$$\underline{w_\lambda^*} = (\mathcal{X}^T\mathcal{X} + \lambda)^{-1}\mathcal{X}^T\underline{Y}$$

Regularized Least Square

$$\nabla(\underline{w^T A w}) = 2A\underline{w}$$

Example, Polynomial Fitting

$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ $\quad d = 1$



$$\hat{y} = w_0 + w_1 x + w_2 x^2 + \dots + w_d x^d$$

$$x_1 \longrightarrow \underline{z_1} = (1, x_1^1, x_1^2, \dots, x_1^d)$$

$$x_2 \longrightarrow \underline{z_2} = (1, x_2^1, x_2^2, \dots, x_2^d)$$

$\vdots$

$$\mathcal{D}' = \{(\underline{z}_1, y_1), \ldots, (\underline{z}_N, y_N)\}$$

$$\mathcal{X} = \begin{bmatrix} \underline{z}_1^T \\ \vdots \\ \underline{z}_N^T \end{bmatrix} \qquad \underline{\hat{y}} = \underline{w^T z}$$

Compute $\underline{w_\lambda^*} = (\mathcal{X}^T \mathcal{X} + \lambda)^{-1} \mathcal{X}^T \underline{Y}$

$$x \longrightarrow \underline{z} = (1, x, x^2, \ldots, x^d)$$

$$\underline{\hat{y}} = \underline{w^T} \, \underline{z} = w_0 + w_1 x + \ldots + w_d x^d$$