Linear Regression

$$\hat{y} = X\theta \tag{1}$$

$$\left[\hat{y}\right]_{N\times 1} = \left[1 \cdots \quad \cdots x_{N,M}\right]_{N\times(M+1)} \left[\theta_M\right]_{(M+1)\times 1}$$

$\mathrm{y}_i = \hat{y}_i + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim N(0, \sigma^2)$
y_i : denotes the ground truth for the ith sample
$\hat{y}_i$: denotes the prediction for the ith sample, where $\hat{y}_i = x_i^T\theta$
ε_i : denotes the error/residual for the ith sample
θ_0, θ_1 : The parameters of the linear regression
$\varepsilon_i = y_i - \hat{y}_i$
$\varepsilon_i = y_i - (\theta_0 + x_i \cdot \theta_1)$

$|1|, |2|, |3|, \ldots$ should be small.
Minimize $2\,\sqrt{1} + 2\sqrt{2} + \ldots + 2\sqrt{N}$ L2 Norm
Minimize $|1| + |1| + \ldots + |1|$ L1 Norm

$$Y = X\theta + \varepsilon \tag{2}$$

To learn θ : Objective: Minimize Minimize $\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_N^2$

Derivation of Normal Equation

$\varepsilon = y - X\theta$
T = (y - Xθ)^T $= y^T - \theta^T X^T$
T = (y$^T - \theta^T X^T)(y - X\theta)$
T = y$^T y - \theta^T X^T y - y^T X\theta + \theta^T X^T X\theta$
T = y$^T y - 2y^T X\theta + \theta^T X^T X\theta$

$$\frac{\partial \varepsilon^T \varepsilon}{\partial \theta} = 0 \tag{3}$$

$$\frac{\partial (y^T y)}{\partial \theta} = 0$$

$$\frac{\partial(-2y^T X\theta)}{\partial \theta} = (-2y^T X)^T = -2X^T y \tag{5}$$

$$\frac{\partial(\theta^T X^T X\theta)}{\partial \theta} = 2X^T X\theta \tag{6}$$

$$0 = -2X^T y + 2X^T X\theta \tag{7}$$

$$X^T y = X^T X\theta \tag{8}$$

$$\theta_{\text{hat}} = (X^T X)^{-1} X^T y \tag{9}$$