Linear Regression

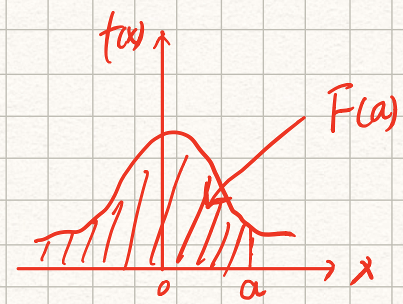
Cumulative Distribution Function (CDF):

$$F(a) = P(x \le a) = \int_{-\infty}^{a} f(x)\,dx$$

Possibility Density Function (pdf)

if $x \sim N(\mu, \delta^2)$

then $f(x) = \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(x-\mu)^2}{2\delta^2}}$



Hypothsis :

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots \theta_n x_n$$

Vetorization :

$$h_\theta(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T \cdot x$$

Error :

$$y^{(i)} = \theta^T x^{(i)} + \varepsilon^{(i)}$$

① ②

$\because$ $\varepsilon^{(i)} \sim N(0, \theta^2)$

(premise)

$\therefore$ $P(\varepsilon^{(i)}) = \frac{1}{\sqrt{2\pi}\delta} e^{\left(-\frac{(\varepsilon^{(i)})^2}{2\delta^2}\right)}$

$$P(y^{(i)} | x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi}\delta} e^{\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\delta^2}\right)}$$

③

Likelihood Function :

$$L(\theta) = \prod_{i=1}^{m} P(y^{(i)} | x^{(i)}; \theta) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\delta} e^{\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\delta^2}\right)}$$

④

Log-Likelihood Function :

(for convenience)

$$\log L(\theta) = \log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\delta} e^{\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\delta^2}\right)}$$

$$= m \log \frac{1}{\sqrt{2\pi}\delta} - \frac{1}{\delta^2} \times \frac{1}{2} \sum_{i=1}^{m} (y^{(i)} - \theta^T x^{(i)})^2$$

⑤

Loss Function :

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{m} (y^{(i)} - \theta^T x^{(i)})^2$$ — least square method

Gradient Descent : feature > 10,000

BGD/SGD/MBGD

Vetorization:

(modified) $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (y - \theta^T X)^2$

Gradient Descent :

$$\theta_j = \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j}$$

Partial Derivative

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{\partial \frac{1}{2m} \sum_{i=1}^{m} (y^{(i)} - h_\theta(x_\theta^{(i)}))^2}{\partial \theta_j}$$

$$= \frac{1}{m} \sum_{i=1}^{m} (y^{(i)} - h_\theta(x_\theta^{(i)})) \frac{\partial h_\theta(x^{(i)})}{\partial \theta_j}$$

$$= \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_\theta^{(i)}) - y^{(i)}) x_j^{(i)}$$

$$\theta_j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Normal Equation : feature < 10,000

Partial Derivative = 0 $\longrightarrow$ min

Vetorization:

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

$$= \frac{1}{2} (X\theta - y)^T (X\theta - y)$$

Partial Derivative:

let $\nabla J(\theta) = X^T X \theta - X^T y = 0$

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