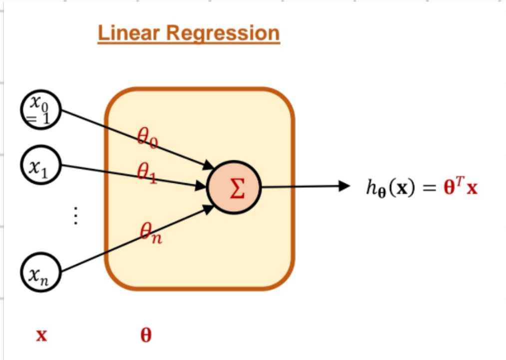
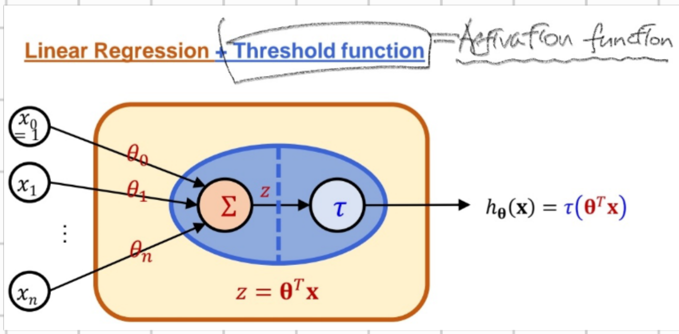
[ML_05] Logistic_Regression

Linear regression 과 비교





Model Representation

$$z = \underline{\theta}^T \underline{x} = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

$$= \sum_{i=0}^{n} \theta_i x_i$$

$$h_\theta(\underline{\theta}^T \underline{x}) = h_\theta(z) = \sigma(z) = \frac{1}{1+e^{-z}} \rightarrow \text{Sigmoid function}$$

Cross-Entropy Loss

Logistic Regression 에서는 Loss Function 으로 Cross-Entropy 를 사용한다.

classic form:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left[y^i \log(h_\theta(x^i)) + (1-y^i) \log(1-h_\theta(x^i)) \right]$$

vector form:

$$J(\underline{\theta}) = -\frac{1}{m} \left(\underline{y}^T \log(\underline{h}) + (1-\underline{y})^T \log(1-\underline{h}) \right)$$

Gradient of CELoss

LR Model $h_\theta(x) = \sigma(\underline{\theta}^T \underline{x}) = \dfrac{1}{1+e^{-\underline{\theta}^T \underline{x}}}$

→ Gradient

classic form:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^{m} \left(\sigma(\underline{\theta}^T \underline{x}^i) - y^i \right) x_j^i$$

Vector form:

$$\nabla J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left(\sigma(\underline{\theta}^T \underline{x}^i) - y^i \right) \underline{x}^i = \frac{1}{m} \underline{X}^T (\sigma(\underline{X}\underline{\theta}) - \underline{y})$$

Linear regression

$$: \nabla J(\underline{\theta}) = \frac{1}{m} \underline{X}^T (\underline{X}\underline{\theta} - \underline{y})$$

proof

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \text{cost}(h_\theta(x^i), y^i)$$

$$\text{cost}(h_\theta(x), y) = -y \log(h_\theta(x)) - (1-y) \log(1-h_\theta(x))$$

$$\frac{\partial \text{cost}}{\partial \theta_j} = -y \frac{1}{h_\theta(x)} \cdot \frac{\partial h_\theta(x)}{\partial \theta_j} + (1-y) \times \frac{1}{1-h_\theta(x)} \times \frac{\partial h_\theta(x)}{\partial \theta_j}$$

$$= \left(-y \frac{1}{h_\theta(x)} + (1-y) \times \frac{1}{1-h_\theta(x)} \right) \times \frac{\partial h_\theta(x)}{\partial \theta_j}$$

$$\frac{\partial h_\theta(x)}{\partial \theta_j} = \frac{\partial \sigma(z)}{\partial \theta_j} = \frac{\partial \sigma(z)}{\partial z} \times \frac{\partial z}{\partial \theta_j}$$

$$= \left(\frac{1}{1+e^{-z}} \right)' \times \frac{\partial (\theta_1 x_1 + \theta_2 x_2 + \cdots \theta_n x_n)}{\partial \theta_j}$$

$$= \frac{e^{-z}}{(1+e^{-z})^2} \times x_j$$

$$= \frac{1}{(1+e^{-z})} \times \left(1 - \frac{1}{(1+e^{-z})} \right) \times x_j$$

$$= \sigma(z) \times (1-\sigma(z)) \times x_j$$

$$= h_\theta(x) \times (1-h_\theta(x)) \times x_j$$

$$\rightarrow \frac{\partial \text{cost}}{\partial \theta_j} = \left(-y \times \frac{1}{h_\theta(x)} + (1-y) \frac{1}{1-h_\theta(x)} \right) \times h_\theta(x)(1-h_\theta(x))\, x_j$$

$$= -y(1-h_\theta(x)) + (1-y) h_\theta(x)\, x_j$$

$$= (h_\theta(x) - y)\, x_j$$

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