

Logistic Function

$$\hat{y} = \frac{1}{1 + e^{-z}}$$

→ \hat{y} : predicted probability
→ z : linear combination of input features

$$z = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

→ b_0 : intercept

→ b_1, b_2, \dots, b_n are coefficients.

Cost function

(measures the error)

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

→ $y^{(i)}$ is the actual label

→ $\hat{y}^{(i)}$ predicted probability.

→ m is the no. of training examples.

Gradient Descent (optimization)

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

→ α is learning rate.

$$\frac{\partial J(\theta)}{\partial b_j} = \frac{1}{m} \sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)}) x_j^{(i)}$$

Decision

$$\hat{y} = \frac{1}{1 + e^{-z}}$$

→ $\hat{y} \geq 0.5$ predicts class 1

→ $\hat{y} < 0.5$ predicts class 0.