Home Credit Default Risk Challenge

Misbah Uddin

June 24, 2018

Machine Learning Models

Logisitic Regression:

Hypothesis: We want $0 \leq h_\theta x \leq 1$

So, we use *sigmoid function* (see Fig. 1): $h_\theta x = \frac{1}{1+e^{-\theta^T x}}$

Decision boundary:

predict $y = 1$, if $h_\theta x \geq 0.5$

predict $y = 0$, otherwise

* based on hyper parameters

Cost function:

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

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

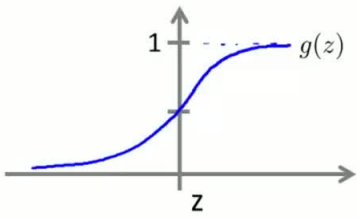
– *Gradient descent*: want $min_\theta \ J(\theta)$

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

Logistic regression

$h_\theta(x) = g(\theta^T x) = P(y=1 \mid x; \theta)$

$g(z) = \frac{1}{1+e^{-z}}$



Suppose predict "$y = 1$" if $h_\theta(x) \geq 0.5$

predict "$y = 0$" if $h_\theta(x) < 0.5$

Andrew Ng

Figure 1: Sigmoid function