Logistic Resression

Logistic regression is used to deal with binary classification problem in mechine learning. for dataset $(x^{(i)}, y^{(i)})$, all output $y^{(i)} \in \{0, 1\}$, the model's prediction is

$$h_\theta(x) = g(\theta^T x)$$

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

that makes prediction in the interval (0,1), that means, if $h_\theta(x) = 0.7$, gives us a probability of 70% that our output is 1, so the predicted output is "1", that is to say,

$$y = \begin{cases} 0 & h_\theta(x) < 0.5 \\ 1 & h_\theta(x) \geq 0.5 \end{cases}$$

that is,

$$h_\theta(x) = P(y = 1|x; \theta) = 1 - P(y = 0|x; \theta)$$

Cost function

Our cost function of logistic regression is

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

and

$$Cost(h_\theta(x^{(i)}), y^{(i)}) = \begin{cases} -log(h_\theta(x)) & y = 1 \\ -log(1 - h_\theta(x)) & y = 0 \end{cases}$$

when $y = 1$ and $h_\theta(x)$ close to 1, $J(\theta)$ will be smaller, so do $y = 0$. In summary,

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

Gradient Descent

$$
\begin{gathered}
Repeat\{ \\
\theta_j = \theta_j - \alpha \cdot \frac{\partial}{\partial \theta_j} J(\theta) \\
= \theta_j - \frac{\alpha}{m} \cdot \sum_{i=1}^{m} ((h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}) \\
\}
\end{gathered}
$$

Reference

mechine learning