ML Fundamentals: Logistic Regression

February 2, 2021

1 Hypothesis

$$h_\theta(x) = g\left(\theta^T x\right) = \frac{1}{1+e^{-\theta^T x}}$$
$$g(z) = \frac{1}{1+e^{-z}} \text{ Where, } g(z) \text{ is a logistic or } \textbf{sigmoid} \text{ function}$$

2 Derivative of sigmoid function

Note, derivative of the sigmoid function is as below

$$\begin{aligned}
g'(z) &= \frac{d}{dz}\frac{1}{1+e^{-z}} \\
&= \frac{1}{\left(1+e^{-z}\right)^2}\left(e^{-z}\right) \\
&= \frac{1}{\left(1+e^{-z}\right)} \cdot \left(1 - \frac{1}{\left(1+e^{-z}\right)}\right) \\
&= g(z)(1-g(z))
\end{aligned}$$

3 Formulation

Lets assume, the probability of y=1 and y=0 as below

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

This can be written compactly as

$$p(y \mid x;\theta) = \left(h_\theta(x)\right)^y \left(1-h_\theta(x)\right)^{1-y}$$

4 Maximizing likelihood

We can write, Likelihood of the parameters as

$$
\begin{aligned}
L(\theta) &= p(\vec{y} \mid X; \theta) \\
&= \prod_{i=1}^{m} p\left(y^{(i)} \mid x^{(i)}; \theta\right) \\
&= \prod_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right)\right)^{y^{(i)}} \left(1 - h_\theta\left(x^{(i)}\right)\right)^{1-y^{(i)}}
\end{aligned}
$$

It is easier to maximize the log-likelihood of θ

$$
\begin{aligned}
\ell(\theta) &= \log L(\theta) \\
&= \sum_{i=1}^{m} y^{(i)} \log h\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - h\left(x^{(i)}\right)\right)
\end{aligned}
$$

5 Updating the parameters

We can update parameters as below

$$\theta := \theta + \alpha \nabla_\theta \ell(\theta)$$

$$
\begin{aligned}
\frac{\partial}{\partial \theta_j} \ell(\theta) &= \left(y \frac{1}{g\left(\theta^T x\right)} - (1-y) \frac{1}{1 - g\left(\theta^T x\right)}\right) \frac{\partial}{\partial \theta_j} g\left(\theta^T x\right) \\
&= \left(y \frac{1}{g\left(\theta^T x\right)} - (1-y) \frac{1}{1 - g\left(\theta^T x\right)}\right) g\left(\theta^T x\right)\left(1 - g\left(\theta^T x\right)\right) \frac{\partial}{\partial \theta_j} \theta^T x.
\end{aligned}
$$

[Note, lets use the rule for derivative of sigmoid function]

$$
\begin{aligned}
&= \left(y\left(1 - g\left(\theta^T x\right)\right) - (1-y) g\left(\theta^T x\right)\right) x_j \\
&= \left(y - h_\theta(x)\right) x_j
\end{aligned}
$$

Hence, the final equation to learn parameters is as below

$$\theta_j := \theta_j + \alpha\left(y^{(i)} - h_\theta\left(x^{(i)}\right)\right) x_j^{(i)}$$