coursera

Menu

Optional Logistic Regression: Gradient

This is an optional reading where I explain gradient descent in more detail. Remember, previously I gave you the gradient update step, but did not explicitly explain what is going on behind the scenes.

The general form of gradient descent is defined as:

$$
\begin{aligned}
&Repeat\ \{ \\
&\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \\
&\}
\end{aligned}
$$

For all j. We can work out the derivative part using calculus to get:

$$
\begin{aligned}
&Repeat\ \{ \\
&\theta_j := \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} (h(x^{(i)}, \theta) - y^{(i)}) x_j^{(i)} \\
&\}
\end{aligned}
$$

A vectorized implementation is:

$\theta := \theta - \frac{\alpha}{m} X^T (H(X, \theta) - Y)$

Partial derivative of J(θ)

First calculate derivative of sigmoid function (it will be useful while finding partial derivative of J(θ)):

$$
\begin{aligned}
h(x)' &= \left(\frac{1}{1+e^{-x}}\right)' = \frac{-(1+e^{-x})'}{(1+e^{-x})^2} = \frac{-1' - (e^{-x})'}{(1+e^{-x})^2} = \frac{0 - (-x)'(e^{-x})}{(1+e^{-x})^2} = \frac{-(-1)(e^{-x}}{(1+e^{-x})^2} \\
&= \left(\frac{1}{1+e^{-x}}\right)\left(\frac{e^{-x}}{1+e^{-x}}\right) = h(x)\left(\frac{+1-1+e^{-x}}{1+e^{-x}}\right) = h(x)\left(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}
\end{aligned}
$$