coursera

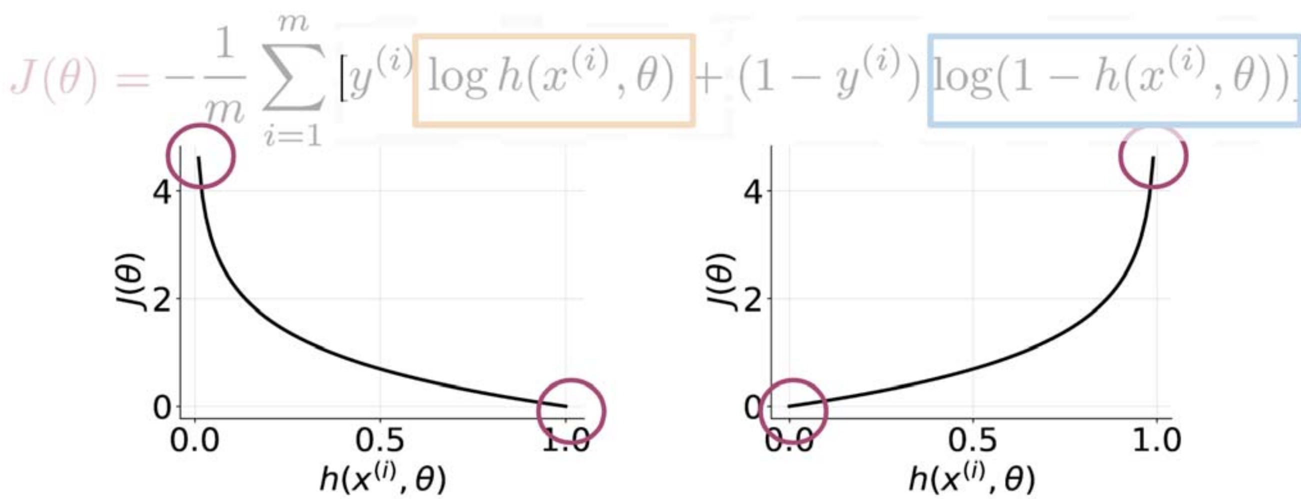
Menu

Optional Logistic Regression: Cost Function

This is an advanced optional reading where we delve into the details.. If you do not get the math, do not worry about it - you will be just fine by moving onto the next component. In this part, I will tell you about the intuition behind why the cost function is designed the way it is. I will then show you how to take the derivative of the logistic regression cost function to get the gradients.

The logistic regression cost function is defined as

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



As you can see in the picture above, if $y = 1$ and you predict something close to 0, you get a cost close to ∞. The same applies for then $y = 0$ and you predict something close to 1. On the other hand if you get a prediction equal to the label, you get a cost of 0. In either, case you are trying to minimize $J(\theta)$.

Math Derivation

To show you why the cost function is designed that way, let us take a step back and write up a function that compresses the two cases into one case.

$$P(y|x^{(i)},\theta) = h(x^{(i)},\theta)^{y^{(i)}}(1-h(x^{(i)},\theta))^{(1-y^{(i)})}$$