Cost Function

Let's first define a few variables that we will need to use

- L = total number of layers in the network
- s_l = number of units (not counting bias unit) in lay
- K = number of output units/classes

Recall that in neural networks, we may have many outpu hypothesis that results in the k^{th} output. Our cost funct generalization of the one we used for logistic regression logistic regression was:

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

For neural networks, it is going to be slightly more comp

$$J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)}\log((h_\Theta(x^{(i)}))_k) + (1-y_k^{(i)})\log$$

We have added a few nested summations to account fo of the equation, before the square brackets, we have ar through the number of output nodes.

In the regularization part, after the square brackets, we The number of columns in our current theta matrix is e layer (including the bias unit). The number of rows in ou number of nodes in the next layer (excluding the bias u square every term.

Note:

- the double sum simply adds up the logistic regress output layer
- the triple sum simply adds up the squares of all th