Errata Week 5

Errata in video "Backpropagation Algorithm"

- Note: In this video, the NN diagrams omit the bias units in the input and hidden layers.

- At time 0:14-1.0, the indices for Θ should be $\Theta_{ij}^{(l)}$ for both in the cost function and in the partial derivative.

- At 1:30, the first step of forward propagation omits adding the bias unit. The bias units are shown for a(2) and a(3), but not for a(1).

Errata in video "Backpropagation Intuition"

- At time 4:39, the last term for the calculation for z_1^3 (three-color handwritten formula) should be a_2^2 instead of a_1^2.

- At time 6:08 and after, the equation for cost(i) is incorrect. The first term is missing parentheses for the log() function, and the second term should be $(1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))$

- At time 7:10, the statement is given $\delta_j^{(l)} = \frac{\partial}{\partial z_j^{(l)}} \text{cost}(t)$ This statement is not strictly correct, and is provided as an intuition for how the backpropagation process works. This video does not attempt to provide mathematical proofs.

- At time 8:50, Prof Ng writes on the slide that $\delta^{(4)} = y - a^{(4)}$. This is incorrect, it should be $\delta^{(4)} = a^{(4)} - y$

- At time 9:40, the descriptions of the δ_3 and δ_2 values are not correct. Again, this video provides intuitions, and is not intended to be used for either proofs or implementation of your NN. See the video "Backpropagation Algorithm" for the correct implementation.

- At time 11:09, Professor Ng correctly writes Θ but mistakenly says delta.

Errata in video "Implementation Note: Unrolling Parameters"

- Starting at 2:03, the image in the upper right corner of the slide is incorrect - it is missing a one of the hidden layers. The text of this lesson discusses a NN with two hidden layers.

Errata in video "Gradient Checking"

Errata in video "Random Initialization"

- At 1:00 Prof Ng provides the example of $\Theta_{ij} = 0$ but his mathematical reasoning assumes $\Theta_{ij} = n \neq 0$. Otherwise he would use that $a_i^{(2)} = 0.5$, since the logistic function outputs 0.5 at input 0.

Errata in video "Putting It Together"

- In minute 11 while Prof Ng is explaining gradient descent, the vertical axis on the graph of the cost function has a range (-3,+3) but the cost function is positive by definition

Errata in the lecture slides (Lecture9.pdf)

- On page 5: The final term in the expression for J(θ) has a subscript i missing i.e. $\theta_j^{(l)}$ becomes $\theta_{ij}^{(l)}$, and i,j index allows every element in array l to contribute to the matrix norm. This matches the final equation on page 3.

- On page 6: The first line of forward propagation omits adding the bias units.

- On page 8: The equation for D when j ≠ 0 should include $\dfrac{\lambda}{m} \Theta$.

- On page 8: Name collision! The loop/training example index "i" is overloaded with the node index for the next layer.

Errata in ex4.pdf

- On page 3: Below Figure 1 text says "...5000 training examples in ex3data1.mat". The text should say "...in ex4data1.mat".

- On page 9: In Step 5, the text says "...by dividing the accumulated gradients by $\frac{1}{m}$:". The text should say "... by multiplying...".

Errata in the programming exercise scripts

- In ex4.m at line 114 and 115, the vector of test values for sigmoidGradient() should start with '-1', not '1'.

- In ex4.m at line 168, the fprintf() statement is hard-coded to output "lambda = 10", even though the variable lambda is set to 3.

- checkNNGradients.m: Line 41 should read "'(**Right**-Your Numerical Gradient, **Left**-Analytical Gradient)\n\n']);"

- randInitializeWeights : line 19 "Note: The first row of W corresponds to the parameters for the bias units" it is column not row. also it's bias unit.