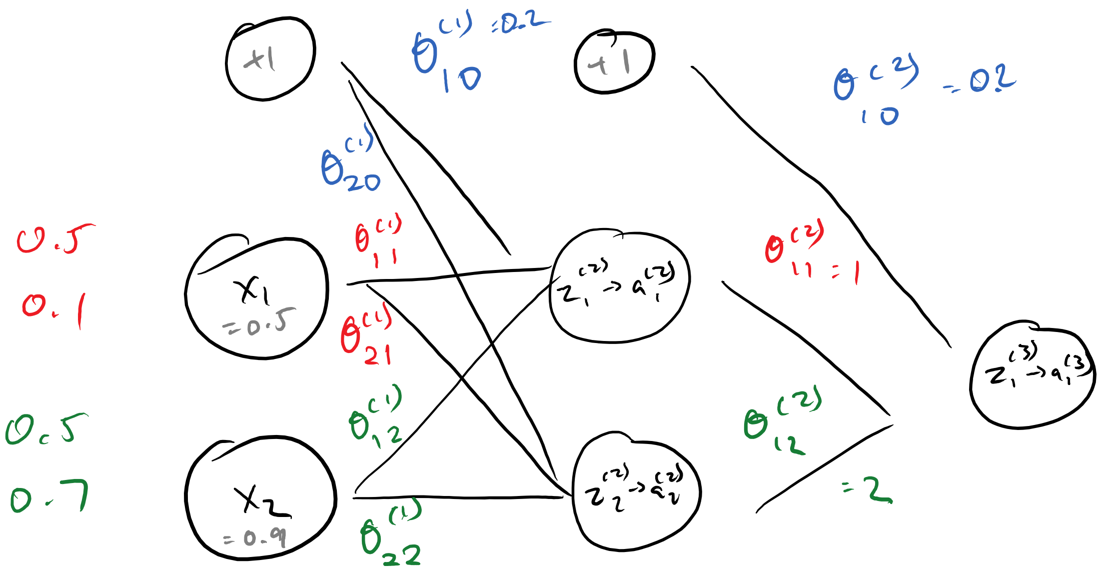
Machine Learning Assignment 3 - Julian Tsang

Friday, November 4, 2016 11:00 PM

3. **(GRADED 3 points)** We have a Neural Network with 1 hidden layer. Input and hidden layer both have 2 nodes. There is 1 output node. The values of theta for bias nodes are 0.2. The vector $\vartheta^{(1)}$ for layer 1 is: [0.5,0.1,0.5,0.7] and $\vartheta^{(2)}$ for layer 2 is [1,2].

   1. Calculate by hand the activations of all nodes for x1 = 0.5 and x2 = 0.9.
   2. Suppose the correct output is 1. Calculate the errors for all nodes and the updates of the weights (for 1 iteration).

1)



$$z_1^{(2)} = (\theta_{10}^{(1)} * 1) + (\theta_{11}^{(1)} * x_1) + (\theta_{12}^{(1)} * x_2) = (0.2 * 1) + (0.5 * 0.5) + (0.5 * 0.9) = 0.9$$
$$a_1^{(2)} = g\left(z_1^{(2)}\right) = \frac{1}{1 + e^{-z_1^{(2)}}} = \frac{1}{1 + e^{-0.9}} \approx 0.7109$$

$$z_2^{(2)} = (\theta_{20}^{(1)} * 1) + (\theta_{21}^{(1)} * x_1) + (\theta_{22}^{(1)} * x_2) = (0.2 * 1) + (0.1 * 0.5) + (0.7 * 0.9) = 0.88$$
$$a_2^{(2)} = g\left(z_2^{(2)}\right) = \frac{1}{1 + e^{-z_2^{(2)}}} = \frac{1}{1 + e^{-0.88}} \approx 0.70682$$

$$z_1^{(3)} = (\theta_{10}^{(2)} * 1) + (\theta_{11}^{(2)} * a_1^{(2)}) + (\theta_{12}^{(2)} * a_2^{(2)}) = (0.2 * 1) + (1 * 0.7109) + (2 * 0.70682) \approx 2.32454$$
$$a_1^{(3)} = g\left(z_1^{(3)}\right) = \frac{1}{1 + e^{-z_1^{(3)}}} = \frac{1}{1 + e^{-2.32454}} \approx 0.910889$$

2) Correct output = 1. Calculate all errors for nodes and update weights (theta values)

Calculate error between correct output and predicted value:

$$\delta_1^{(3)} = a_1^{(3)} - y^i = 0.910889\ - 1 = -0.08911$$

Using this error, calculate errors for nodes in hidden layer:

$$\delta_1^{(2)} = \theta_{11}^{(2)}\delta_1^{(3)} * a_1^{(2)}\left(1 - a_1^{(2)}\right) = (1 * -0.08911) * 0.7109(1 - 0.7109) = -0.018314$$
$$\delta_2^{(2)} = \theta_{12}^{(2)}\delta_1^{(3)} * a_2^{(2)}\left(1 - a_2^{(2)}\right) = (2 * -0.08911) * 0.70682(1 - 0.70682) = -0.03693$$

Gradient to update theta values

Since there is only 1 iteration, $\Delta_{ij}^{(L)} = 0$

$$\Delta^{(1)} \coloneqq \begin{bmatrix} a_1^{(1)}\delta_1^{(2)} & a_2^{(1)}\delta_1^{(2)} \\ a_1^{(1)}\delta_2^{(2)} & a_2^{(1)}\delta_2^{(2)} \end{bmatrix} = \begin{bmatrix} (0.5)(-0.018314) & (0.9)(-0.018314) \\ (0.5)(-0.03693) & (0.9)(-0.03693) \end{bmatrix} = \begin{bmatrix} -0.009157 & -0.0164826 \\ -0.018465 & -0.033237 \end{bmatrix}$$

$$\Delta_{ij}^{(l)} \coloneqq \Delta_{ij}^{(l)} + a_j\delta_i^{(l+1)}$$

Update Weights:

$$\theta_{11}^{(1)} \coloneqq \theta_{11}^{(1)} + a_1^{(1)}\delta_1^{(2)} = 0.5 - 0.009157 = 0.490843$$
$$\theta_{21}^{(1)} \coloneqq \theta_{21}^{(1)} + a_1^{(1)}\delta_2^{(2)} = 0.1 - 0.018465 = 0.081535$$
$$\theta_{12}^{(1)} \coloneqq \theta_{12}^{(1)} + a_2^{(1)}\delta_1^{(2)} = 0.5 - 0.0164826 = 0.4835174$$
$$\theta_{22}^{(1)} \coloneqq \theta_{22}^{(1)} + a_2^{(1)}\delta_2^{(2)} = 0.7 - 0.033237 = 0.666763$$
$$\theta_{11}^{(2)} \coloneqq \theta_{11}^{(2)} + a_1^{(2)}\delta_1^{(3)} = 1\ - (0.7109)(-0.08911) = 1.0633483$$
$$\theta_{12}^{(2)} \coloneqq \theta_{12}^{(2)} + a_2^{(2)}\delta_1^{(3)} = 2\ - (0.70682)(-0.08911) = 2.06298473$$

4. **(GRADED 2 point)** Mitchell chapter 4 (p. 124):

In this exercise you will use a perceptron (instead of sigmoid function), that is explained in the slides of Mitchell. You can find those slides here. Please refer to pages 78 and 79.

Exercise 4.1

What are the values of weights w_0, w_1 and w_2 for the perceptron whose decision surface is illustrated in figure b on page 79?
Assume the surface crosses the x_1 axis at -1 and the x_2 axis at 2.

Exercise 4.2

(a) Design a two-input perceptron that implements the Boolean function A AND (NOT B).

(b) Design a two-layer network of perceptrons that that implements A XOR B.

Exercise 4.1

A line that crosses the x_1 axis at -1 and the x_2 at 2 has the equation $x_2 = 2x_1 + 2$, which can be rearranged to: $0 = -x_2 + 2x_1 + 2$.
Since the notation is $w_0 + w_1x_1 + w_2x_2 = 0$
Then:
$w_0 = -1$
$w_1 = 2$
$w_2 = 2$

Exercise 4.2

a) Truth Table for *A and (NOT B)*

x_1	x_2	$h_\theta(x)$	Representation
0	0	-1	$x_0 < 0$
0	1	-1	$x_0 + x_2 < 0$
1	0	1	$x_0 + x_1 > 0$
1	1	-1	$x_0 + x_1 + x_2 < 0$

Weights:

$x_0 = -10$
$x_1 = 15$
$x_2 = -10$

b) XOR, exclusive OR

x_1	x_2	$a_1^{(2)}$ OR	$a_2^{(2)}$ $NOT\ x_1\ OR\ NOT\ x_2$	$h_\theta(x)$ AND
0	0	0	1	0
0	1	1	1	1
1	0	1	1	1
1	1	1	0	0

Representation for OR

$x_0 < 0$
$x_0 + x_2 > 0$
$x_0 + x_1 > 0$
$x_0 + x_1 + x_2 > 0$

Weights:

$x_0 = -5$
$x_1 = 10$
$x_2 = 15$

Representation for $NOT\ x_1\ OR\ NOT\ x_2$

$x_0 > 0$
$x_0 + x_2 > 0$
$x_0 + x_1 > 0$
$x_0 + x_1 + x_2 < 0$

Weights:

$x_0 = 10$
$x_1 = -5$
$x_2 = -6$

Representation for AND

$x_0 + x_2 < 0$

$x_0 + x_1 + x_2 > 0$
$x_0 + x_1 + x_2 > 0$
$x_0 + x_1 < 0$

Weights:

$x_0 = -10$
$x_1 = 5$
$x_2 = 6$