Written Assignment 3

Wendy Nieuwkamer

6 November 2016

1 Question 3

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 $\theta^{(1)}$ for layer 1 is: [0.5,0.1,0.5,0.7] and $\theta^{(2)}$ for layer 2 is [1,2].

1.1 Calculate by hand the activations of all nodes for $x_1 = 0.5$ and $x_2 = 0.9$.

We go layer by layer. For each node in a layer we first calculate the z value. We do this by taking the sum of all the inputs multiplied by their weights.

$$\begin{aligned} z_1^{(2)} &= \theta_{10}^{(1)} x_0^{(1)} + \theta_{11}^{(1)} x_1^{(1)} + \theta_{12}^{(1)} x_2^{(1)} \\ &= 0.2 \cdot 1 + 0.5 \cdot 0.5 + 0.5 \cdot 0.9 \\ &= 0.2 + 0.25 + 0.45 \\ &= 0.9 \end{aligned}$$

$$\begin{aligned} z_2^{(2)} &= \theta_{20}^{(1)} x_0^{(1)} + \theta_{11}^{(1)} x_1^{(1)} + \theta_{12}^{(1)} x_2^{(1)} \\ &= 0.2 \cdot 1 + 0.1 \cdot 0.5 + 0.7 \cdot 0.9 \\ &= 0.2 + 0.05 + 0.63 \\ &= 0.88 \end{aligned}$$

The we calculate the activation value for each node using the sigmoid function:

$$g(z) = \frac{1}{1 + e^{-z}}$$

$$
\begin{aligned}
a_1^{(2)} &= \frac{1}{1+e^{-z_1^{(2)}}} \\
&= \frac{1}{1+e^{-0.9}} \\
&= 0.7109
\end{aligned}
$$

$$
\begin{aligned}
a_2^{(2)} &= \frac{1}{1+e^{-z_2^{(2)}}} \\
&= \frac{1}{1+e^{-0.88}} \\
&= 0.7068
\end{aligned}
$$

The second and last layer will then be:

$$
\begin{aligned}
z_1^{(3)} &= \theta_{10}^{(2)} x_0^{(2)} + \theta_{11}^{(2)} x_1^{(2)} + \theta_{12}^{(2)} x_2^{(2)} \\
&= 0.2 \cdot 1 + 1 \cdot 0.7109 + 2 \cdot 0.7068 \\
&= 0.2 + 0.7109 + 1.4136 \\
&= 2.3245
\end{aligned}
$$

$$
\begin{aligned}
a_1^{(3)} &= \frac{1}{1+e^{-z_1^{(3)}}} \\
&= \frac{1}{1+e^{-2.3245}} \\
&= 0.9109
\end{aligned}
$$

Our final activation value is the hypothesis. Thus, in this case $a_1^{(3)} = h_\theta(x) = 0.9109$.

1.2 Suppose the correct output is 1. Calculate the errors for all nodes and the updates of the weights (for 1 iteration).

$\delta_j^{(l)}$ is the "error"
Calculation:

$$\delta_j^{(l)} = \theta^{(l)} \delta^{(l+1)}$$

Thus,

$$
\begin{aligned}
\delta_1^{(3)} &= y - a_1^{(3)} \\
&= 1 - 0.9109 \\
&= 0.0891
\end{aligned}
$$

$$
\begin{aligned}
\delta_1^{(2)} &= \theta_{11}^{(2)} \delta_1^{(3)} \cdot g'(z_1^{(2)}) \\
&= \theta_{11}^{(2)} \delta_1^{(3)} \cdot a_1^{(2)}(1 - a_1^{(2)}) \\
&= 1 \cdot 0.0891 \cdot 0.7109(1 - 0.7109) \\
&= 0.0183
\end{aligned}
$$

$$
\begin{aligned}
\delta_2^{(2)} &= \theta_{12}^{(2)} \delta_1^{(3)} \cdot a_2^{(2)}(1 - a_2^{(2)}) \\
&= 2 \cdot 0.0891 \cdot 0.7068(1 - 0.7068) \\
&= 0.0369
\end{aligned}
$$

Those are the deltas, now the updates. Choose the learning rate $\alpha = 1$ for convenience.

$$
\theta_{ij}^{(l)} := \theta_{ij}^{(l)} - \alpha a_j^{(l)} \delta_i^{(l+1)}
$$

Don't update the bias thetas. Layer 1:

$$
\begin{aligned}
\theta_{11}^{(2)} &:= \theta_{11}^{(2)} - \alpha a_1^{(2)} \delta_1^{(3)} \\
&= 1 - 1 \cdot 0.7109 \cdot 0.0891 \\
&= 0.9367
\end{aligned}
$$

$$
\begin{aligned}
\theta_{12}^{(2)} &:= \theta_{12}^{(2)} - \alpha a_2^{(2)} \delta_1^{(3)} \\
&= 2 - 1 \cdot 0.7068 \cdot 0.0891 \\
&= 1.9370
\end{aligned}
$$

Layer 2:

$$\begin{aligned}
\theta_{11}^{(1)} &:= \theta_{11}^{(1)} - \alpha x_1^{(1)} \delta_1^{(2)} \\
&= 0.5 - 1 \cdot 0.5 \cdot 0.0183 \\
&= 0.4909
\end{aligned}$$

$$\begin{aligned}
\theta_{12}^{(1)} &:= \theta_{12}^{(1)} - \alpha x_2^{(1)} \delta_1^{(2)} \\
&= 0.5 - 1 \cdot 0.9 \cdot 0.0183 \\
&= 0.4835
\end{aligned}$$

$$\begin{aligned}
\theta_{21}^{(1)} &:= \theta_{21}^{(1)} - \alpha x_1^{(1)} \delta_2^{(2)} \\
&= 0.1 - 1 \cdot 0.5 \cdot 0.0369 \\
&= 0.0816
\end{aligned}$$

$$\begin{aligned}
\theta_{22}^{(1)} &:= \theta_{22}^{(1)} - \alpha x_2^{(1)} \delta_2^{(2)} \\
&= 0.7 - 1 \cdot 0.9 \cdot 0.0369 \\
&= 0.6668
\end{aligned}$$