CMPS 142: Homework Assignment 4

Jeffrey Petersen - 1329242
Peter Czupil - 1317993
Raymond Colebaugh - 1377877

May 19, 2015

1. (a) The probability that a point randomly drawn from p is located somewhere in the interval (z_ϵ, θ) is equal to ϵ. Thus the probability that a point falls outside this interval is the complement of the previous probability. Therefore, $p((0, z_\epsilon]) = 1 - \epsilon$.

   (b) Assuming that the training set may contain duplicate x values, the probability that all points lie outside the interval $(z_\epsilon, \theta]$ is the product of the probability from part (a) for all x's in the training set. Therefore $p(X \notin (z_\epsilon, \theta]) = \prod_{i=1}^{N} 1 - \epsilon = (1-\epsilon)^N$

   (c) Since, $p((z_\epsilon, \theta])$ is equal to ϵ, the probability that $\hat{\theta}$ has an error rate of at least ϵ is the probability that $\hat{\theta} \leq z_\epsilon$. This is the area under our density function, on the interval $[0, z_\epsilon]$. We had computed this area previously in part (b) which turned out to be equal to $(1-\epsilon)^N$ where N is the cardinality of the training set. Therefore the probability that $\hat{\theta}$ has an error rate of at least ϵ is equal to $(1-\epsilon)^N$.

   (d) To find the smallest N s.t. $\epsilon \leq \textit{error rate of } \hat{\theta} \leq \delta$, we can take our result from (c) and use the following inequality:

   $$(1-\epsilon)^N \leq \delta$$

   Taking the log of both sides gives us:

   $$N log(1-\epsilon) \leq log(\delta)$$

   which implies:

   $$N \geq \frac{log(\delta)}{log(1-\epsilon)}$$

2. First we calculate the a_j and z_j values in the network:

$$a_3 = w_{13} * 1 + w_{23} * 2 = 0 * 1 + 0 * 2 = 0$$

$$z_3 = \sigma(a_3) = 1/(1 + e^{-0}) = \frac{1}{2}$$

$$a_4 = w_{14} * 1 + w_{24} * 2 = 0 * 1 + 0 * 2 = 0$$

$$z_4 = \sigma(a_4) = 1/(1 + e^{-0}) = \frac{1}{2}$$

Then we can calculate the $\frac{\partial E}{\partial a_5}$ value at the output node:

$$\frac{\partial E}{\partial a_5} = a_5 - t = 0 - 1 = -1 = \delta_5$$

Now we can begin to backpropagate, and calculate the derivative of the error with respect to the weights on the internal nodes. First, on the edges which terminate on 5:

$$\frac{\partial E}{\partial w_{54}} = \frac{\partial E}{\partial a_5}\frac{\partial a_5}{\partial a_{54}} = \delta_5 - z_4 = -\frac{1}{2}$$

$$\frac{\partial E}{\partial w_{53}} = \frac{\partial E}{\partial a_5}\frac{\partial a_5}{\partial a_{53}} = \delta_5 - z_3 = -\frac{1}{2}$$

Then, for the edges which terminate on 4:

$$\frac{\partial E}{\partial a_4} = \left(\sum_{k \in U_4} \frac{\partial E}{\partial a_k} w_{k4}\right) z_4(1 - z_4) = \frac{1}{4} w_{54} * \delta_5 = 0 = \delta_4$$

$$\frac{\partial E}{\partial w_{41}} = \frac{\partial E}{\partial a_4}\frac{\partial a_4}{\partial w_{41}} = \delta_4 * z_1 = 0$$

$$\frac{\partial E}{\partial w_{42}} = \frac{\partial E}{\partial a_4}\frac{\partial a_4}{\partial w_{42}} = \delta_4 * z_2 = 0$$

Next, for the edges which terminate on 3:

$$\frac{\partial E}{\partial a_3} = \left(\sum_{k \in U_3} \frac{\partial E}{\partial a_k} w_{k3}\right) z_3(1 - z_3) = \frac{1}{4} w_{53} * \delta_5 = 0 = \delta_3$$

$$\frac{\partial E}{\partial w_{31}} = \frac{\partial E}{\partial a_3}\frac{\partial a_3}{\partial w_{31}} = \delta_3 * z_1 = 0$$

$$\frac{\partial E}{\partial w_{32}} = \frac{\partial E}{\partial a_3}\frac{\partial a_3}{\partial w_{32}} = \delta_3 * z_2 = 0$$

Finally, we update each weight according to the gradient and the learning rate $\eta = 0.1$:

$$w_{13} := w_{13} - \eta \frac{\partial E}{\partial w_{13}} = 0 - 0.1 * 0 = 0$$

$$w_{14} := w_{14} - \eta \frac{\partial E}{\partial w_{14}} = 0 - 0.1 * 0 = 0$$

$$w_{23} := w_{23} - \eta \frac{\partial E}{\partial w_{23}} = 0 - 0.1 * 0 = 0$$

$$w_{24} := w_{24} - \eta \frac{\partial E}{\partial w_{24}} = 0 - 0.1 * 0 = 0$$

$$w_{35} := w_{35} - \eta \frac{\partial E}{\partial w_{35}} = 0 - 0.1 * -0.5 = 0.05$$

$$w_{45} := w_{45} - \eta \frac{\partial E}{\partial w_{45}} = 0 - 0.1 * -0.5 = 0.05$$