CMPS 142 - Homework 5

Aidan Gadberry, Christopher Hsiao, Tom Burch
{agadberr, chhsiao, tburch1}@ucsc.edu

June 8, 2017

1 Back Propagation

1.1 Feed Foward

First, we will perform the feed forward process to calculate the output of our neural network: z_5.
Note that we will be using $\sigma(a_j)$ to represent the logistic sigmoid function on the activations of nodes $j = 1...5$.

Input Layer - 1, 2

Given an input of $(1, 2)$, $a_1 = 1$ and $a_2 = 2$, we can calculate the outputs:

$$z_1 = \sigma(a_1) = \sigma(1) \approx 0.73$$
$$z_2 = \sigma(a_2) = \sigma(2) \approx 0.88$$

Hidden Layer - 3, 4

First, we'll calculate the activations a_3, a_4 for the hidden nodes.

$$a_3 = z_1\ w_{3,1} + z_2\ w_{3,2} \approx 0.536$$
$$a_4 = z_1\ w_{4,1} + z_2\ w_{4,2} \approx 0.536$$

And we can now calculate the outputs of these nodes:

$$z_3 = \sigma(a_3) = \sigma(0.536) \approx 0.631$$
$$z_4 = \sigma(a_4) = \sigma(0.536) \approx 0.631$$

Output - 5

And now to calculate the output activation a_5. Recall that for the output node, $a_5 = z_5$.

$$a_5 = z_3\ w_{5,3} + z_4\ w_{5,4} = 0$$
$$\text{thus, } a_5 = z_5 = 0$$

1.2 Backpropagation

Now we'll begin calculating the $\frac{\delta E}{\delta a_j}$ values, which we'll represent as δ_j, starting with δ_5.

Output Node: δ_5

Since the error function is:

$$E = \frac{1}{2}(a_5 - t)^2$$

then...

$$\delta_5 = \frac{\delta E}{\delta a_5} = (a_5 - t) = (1 - 2) = -1$$

Now, recall that the general equation for other δ_j values is:

$$\frac{\delta E}{\delta a_j} = \delta_j = \left(\sum_{k \in U_j} \delta_k \, w_{k,j} \right) z_j(1 - z_j)$$

Where U_j represented outgoing arcs from node j.

Hidden Layer - 3, 4

Using δ_5, we'll use the equation above to calculate δ_4 and δ_3.

$$\delta_4 = (-1 \cdot 1)(0.631)(0.369) \approx -0.233$$

$$\delta_3 = (-1 \cdot -1)(0.631)(0.369) \approx 0.233$$

Input Layer - 1, 2

And now to find δ_1 and δ_2.

$$\delta_2 = [(\delta_3 \cdot \frac{1}{3}) + (\delta_4 \cdot \frac{1}{3})] \, (0.88)(1 - 0.88) = 0$$

$$\delta_1 = [(\delta_3 \cdot \frac{1}{3}) + (\delta_4 \cdot \frac{1}{3})] \, (0.73)(1 - 0.73) = 0$$

Updating the Weights - $w_{j,i}$'s

Recall that the full step/update equation in SGD for each weight is:

$$w_{j,i} \coloneqq w_{j,i} - \eta \frac{\delta E}{\delta w_{j,i}}$$

$$\text{where: } \frac{\delta E}{\delta w_{j,i}} = \delta_j \, z_i$$

Thus, the updates for the weights are as follows:

$$w_{5,3} = -1 - (0.1)(-1)(0.631) = -0.937$$
$$w_{5,4} = 1 - (0.1)(-1)(0.631) = 1.063$$
$$w_{4,1} = \frac{1}{3} - (0.1)(-0.233)(0.73) = 0.35$$
$$w_{4,2} = \frac{1}{3} - (0.1)(-0.233)(0.88) = 0.354$$
$$w_{3,1} = \frac{1}{3} - (0.1)(0.233)(0.73) = 0.316$$
$$w_{3,2} = \frac{1}{3} - (0.1)(0.233)(0.88) = 0.313$$

2 Clustering

Given: 1, 3, 6, 10, 15, 21, 28, 36, K=3, Given mean=1,3,6

Iteration	K1	K1-mean	K2	K2-mean	K3	K3-mean
#1	1	1	3	3	6, 10, 15, 21, 28, 38	19.3
#2	1,3	1.5	6	6	10, 15, 21, 28, 36	22
#3	1,3,6	3.3	10	10	15, 21, 28, 36	25
#4	1,3,6,10	5	15	15	21, 28, 36	28.3
#5	1,3,6,10,15	7	21	21	28, 38	32
#6	1,3,6,10,15,21	9.3	28	28	36	36

Given: 1, 3, 6, 10, 15, 21, 28, 36, K=3, Given mean=21,28,36

Iteration	K1	K1-mean	K2	K2-mean	K3	K3-mean
#1	1,3,6,10,15,21	9.3	28	28	36	36

Yes, we believe starting with the same given set 1, 3, 6, 10, 15, 21, 28, 36, the K-cluster algorithm will output the same clustering when comparing the initial means [1,3,6] and [21,28,36]. Although the final clustering is identical, the difference is the execution of the K-means cluster algorithm, where for the first set of initial means will result in the algorithm processing 6 iterations, while the the second set of means will allow the algorithm to complete in one iteration.