COMS 4030A: Assignment 2

Tamlin Love
1438243

Note: Where applicable, numbers have been rounded to two decimal points. Thus, for example, 0.075 is rounded to 0.08.

1 Question 1: K-Nearest Neighbour Classification

1.1

Record #	**Distance**	**Label**
9	1.98	0
Prediction	0	

1.2

Record #	**Distance**	**Label**
23	0.92	1
28	1.75	1
21	1.82	1
Prediction	1	

1.3

Record #	**Distance**	**Label**
32	0.81	1
39	0.95	1
21	1.00	1
28	1.22	1
23	2.20	1
Prediction	1	

1.4

K-NN	Training Error Rate	Test Error Rate
1-NN	0.00	0.05
3-NN	0.00	0.08
5-NN	0.00	0.05
7-NN	0.00	0.05
9-NN	0.05	0.08
11-NN	0.05	0.08
13-NN	0.05	0.03

2 Question 2: Logistic Regression

2.1

The range of $h_\Theta(x)$ is the interval $[0, 1]$.
The classifier cannot learn from an arbitrary data set if $\beta = 0$, as this causes $h_\Theta(x) = 0.5$ for all values of x.

2.2

We begin with the following definition for the cost function

$$J(\Theta) = \frac{1}{4N} \sum_{n=1}^{N} (h_\Theta(x^{(n)}) - y^{(n)})^4 \tag{1}$$

and the update rule

$$\theta_k \leftarrow \theta_k - \alpha \frac{\partial J(\Theta)}{\partial \theta_k} \tag{2}$$

and the activation function

$$h_\Theta(x^{(n)}) = \frac{1}{1 + e^{-\beta \Theta^T x^{(n)}}} \tag{3}$$

Now, by the chain rule, we have that

$$\frac{\partial J(\Theta)}{\partial \theta_k} = \frac{\partial J(\Theta)}{\partial h_\Theta(x^{(n)})} \frac{\partial h_\Theta(x^{(n)})}{\partial \theta_k} \tag{4}$$

The partial derivative of equation 1 with respect to $h_\Theta(x^{(n)})$ is

$$\frac{\partial J(\Theta)}{\partial h_\Theta(x^{(n)})} = \frac{1}{N} \sum_{n=1}^{N} (h_\Theta(x^{(n)}) - y^{(n)})^3 \tag{5}$$

The partial derivative of equation 3 with respect to θ_k is

$$\begin{aligned}
\frac{\partial h_\Theta(x^{(n)})}{\partial \theta_k} &= \frac{\partial}{\partial \theta_k}\left(\frac{1}{1+e^{-\beta\Theta^T x^{(n)}}}\right) \\
&= \frac{\partial}{\partial \theta_k}\left(\frac{1}{1+\prod_{i=0}^{d} e^{-\beta\theta_i x_i^{(n)}}}\right) \\
&= \frac{e^{-\Theta^T x^{(n)}}}{1+e^{-\Theta^T x^{(n)}}}\beta x_k^{(n)} \\
&= \beta h_\Theta(x^{(n)})(1-h_\Theta(x^{(n)}))x_k^{(n)}
\end{aligned} \tag{6}$$

Substituting equations 5 and 6 back into equation 4 yields

$$\frac{\partial J(\Theta)}{\partial \theta_k} = \frac{\beta}{N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)})^3 h_\Theta(x^{(n)})(1-h_\Theta(x^{(n)}))x_k^{(n)} \tag{7}$$

Finally, we substitute equation 7 into equation 2 to obtain the closed form update rule

$$\theta_k \leftarrow \theta_k - \frac{\alpha\beta}{N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)})^3 h_\Theta(x^{(n)})(1-h_\Theta(x^{(n)}))x_k^{(n)}$$

2.3

Testing Record #	**Classifier Output** $h_\Theta(x^{(n)})$	**Final Output**
5	0.01	0
10	0.06	0
15	0.13	0
20	0.09	0
25	0.63	1
30	0.87	1
35	0.58	1
40	0.57	1

3 Question 3: Artificial Neural Networks

3.1

$$\begin{aligned} h_\Theta(x) &= g_3(g_2(g_1(\sum_{i=0}^{4}\theta_i x_i)\theta_6 + \theta_5 x_0)\theta_7 + \theta_8 x_0) \\ &= \frac{\theta_6\theta_7}{1+e^{-\sum_{i=0}^{4}\theta_i x_i}} + (\theta_5\theta_7 + \theta_8)x_0 \end{aligned} \tag{8}$$

3.2

We define the cost function to be

$$J(\Theta) = \frac{1}{2N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)})^2 \tag{9}$$

we also note that the update rule remains the same as in equation 2, and that the value of $h_\Theta(x^{(n)})$ is the same as in equation 8.
Using the chain rule in equation 4, and substituting the appropriate values for $J(\Theta)$ and $h_\Theta(x^{(n)})$, we have the partial derivative with respect to $h_\Theta(x^{(n)})$

$$\frac{\partial J(\Theta)}{\partial h_\Theta(x^{(n)})} = \frac{1}{N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)}) \tag{10}$$

Now we take the partial derivative with respect to θ_k for $k = 5, 6, 7, 8$

$$\begin{aligned} \frac{\partial h_\Theta(x^{(n)})}{\partial\theta_5} &= \theta_7 x_0^{(n)} \\ \frac{\partial h_\Theta(x^{(n)})}{\partial\theta_6} &= \theta_7 g_1(\sum_{i=0}^{4}\theta_i x_i^{(n)}) = \frac{\theta_7}{1+e^{-\sum_{i=0}^{4}\theta_i x_i}} \\ \frac{\partial h_\Theta(x^{(n)})}{\partial\theta_7} &= \theta_5 x_0^{(n)} + \theta_6 g_1(\sum_{i=0}^{4}\theta_i x_i^{(n)}) \\ &= \theta_5 x_0^{(n)} + \frac{\theta_6}{1+e^{-\sum_{i=0}^{4}\theta_i x_i^{(n)}}} \\ \frac{\partial h_\Theta(x^{(n)})}{\partial\theta_8} &= x_0^{(n)} \end{aligned} \tag{11}$$

Substituting equations 10 and 11 into equation 4, we arrive at

$$\frac{\partial J(\Theta)}{\partial \theta_5} = \frac{1}{N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)})\theta_7 x_0^{(n)}$$

$$\frac{\partial J(\Theta)}{\partial \theta_6} = \frac{1}{N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)})\theta_7 g_1(\sum_{i=0}^{4}\theta_i x_i^{(n)})$$

$$\frac{\partial J(\Theta)}{\partial \theta_7} = \frac{1}{N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)})(\theta_5 x_0^{(n)} + \theta_6 g_1(\sum_{i=0}^{4}\theta_i x_i^{(n)}))$$

$$\frac{\partial J(\Theta)}{\partial \theta_8} = \frac{1}{N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)})x_0$$

which, when substituted into equation 2, yields the closed form update rules for the selected θ_i

$$\theta_5 \leftarrow \theta_5 - \alpha\frac{1}{N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)})\theta_7 x_0^{(n)}$$

$$\theta_6 \leftarrow \theta_6 - \alpha\frac{1}{N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)})\theta_7 g_1(\sum_{i=0}^{4}\theta_i x_i^{(n)})$$

$$\theta_7 \leftarrow \theta_7 - \alpha\frac{1}{N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)})(\theta_5 x_0^{(n)} + \theta_6 g_1(\sum_{i=0}^{4}\theta_i x_i^{(n)}))$$

$$\theta_8 \leftarrow \theta_8 - \alpha\frac{1}{N}\sum_{n=1}^{N}(h_\Theta(x^{(n)}) - y^{(n)})x_0$$