Model Answer Exercise 1

The four points are:

$x^{(1)}$ =(2,4), $y^{(1)} = 1$
$x^{(2)}$ =(1,3), $y^{(2)} = 1$
$x^{(3)}$ =(4,2), $y^{(3)} = 0$
$x^{(2)}$ =(2,2), $y^{(4)} = 0$

The initial weights are
$\theta = (\theta_0, \theta_1, \theta_2)$ =(0.9, 1.3, 0.1)

For each vector $x = (x_1, x_2)$ we calculate

$$h_\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_1 + \theta_2 x_2)}}$$

So we have:

$$h_\theta\left(x^{(1)}\right) = 0.980$$

$$h_\theta\left(x^{(2)}\right) = 0.924$$

$$h_\theta\left(x^{(3)}\right) = 0.998$$

$$h_\theta\left(x^{(4)}\right) = 0.976$$

Now calculate the cost function for each of the four points (each point is indexed by i):

$$Cost\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) = -y^{(i)}\log\left(h_\theta\left(x^{(i)}\right)\right) - \left(1 - y^{(i)}\right)\log\left(1 - h_\theta\left(x^{(i)}\right)\right)$$

We have:

$$Cost\left(h_\theta\left(x^{(1)}\right), y^{(1)}\right) = 0.020$$

$$Cost\left(h_\theta\left(x^{(2)}\right), y^{(2)}\right) = 0.079$$

$$Cost\left(h_\theta\left(x^{(3)}\right), y^{(3)}\right) = 6.302$$

$$Cost\left(h_\theta\left(x^{(4)}\right), y^{(4)}\right) = 3.724$$

So the total initial cost function is the mean of the four above cost functions, that is:

$$J(\theta) = J(\theta_0, \theta_1, \theta_2) = 2.531$$

In order to update the three parameters $(\theta_0, \theta_1, \theta_2)$, we now need to calculate the gradient of the cost function at each point i in relation to each of the three parameters. For a given point $x^{(i)}$, we have:

$$\frac{\partial J(\theta)}{\partial \theta_0} = \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_0^{(i)} \qquad \frac{\partial J(\theta)}{\partial \theta_1} = \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_1^{(i)} \quad \frac{\partial J(\theta)}{\partial \theta_2} = \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_2^{(i)}$$

So for point $x^{(1)}$: $\frac{\partial J(\theta)}{\partial \theta_0} = -0.020 \quad \frac{\partial J(\theta)}{\partial \theta 1} = -0.040 \quad \frac{\partial J(\theta)}{\partial \theta_2}$ = -0.079

So for point $x^{(2)}$: $\frac{\partial J(\theta)}{\partial \theta_0} = -0.076 \quad \frac{\partial J(\theta)}{\partial \theta 1} = -0.076 \quad \frac{\partial J(\theta)}{\partial \theta_2}$ = -0.228

So for point $x^{(3)}$: $\frac{\partial J(\theta)}{\partial \theta_0} = 0.998 \quad \frac{\partial J(\theta)}{\partial \theta 1} = 3.993 \quad \frac{\partial J(\theta)}{\partial \theta_2} = 1.996$

So for point $x^{(4)}$: $\frac{\partial J(\theta)}{\partial \theta_0} = 0.976 \quad \frac{\partial J(\theta)}{\partial \theta 1} = 1.952 \quad \frac{\partial J(\theta)}{\partial \theta_2} = 1.952$

We add the gradients associated to each of the four data points, and we divide by the number of data points:

$$\frac{\partial J(\theta)}{\partial \theta_0} = 0.470 \quad \frac{\partial J(\theta)}{\partial \theta_1} = 1.457 \quad \frac{\partial J(\theta)}{\partial \theta_2} = 0.910$$

Now that we have the gradients we can modify the parameters $(\theta_0, \theta_1, \theta_2)$ using these gradients. If we denote $(\theta_0', \theta_1', \theta_2')$:

$$\theta_0' = \theta_0 - \alpha \frac{\partial J(\theta_0,\theta_1,\theta_2)}{\partial \theta_0} \qquad \theta_1' = \theta_1 - \alpha \frac{\partial J(\theta_0,\theta_1,\theta_2)}{\partial \theta_1} \quad \theta_2' = \theta_2 - \alpha \frac{\partial J(\theta_0,\theta_1,\theta_2)}{\partial \theta_2}$$

If we use a value of 0.1 for the learning rate α, we have:

$$\theta_0' = 0.9 - 0.1 \times 0.470 = 0.852$$
$$\theta_1' = 1.3 - 0.1 \times 1.457 = 1.154$$
$$\theta_2' = 0.1 - 0.1 \times 0.910 = 0.009$$

Now let us start with the second iteration, starting this time with the values:

$(\theta_0, \theta_1, \theta_2) = (0.852, 1.154, 0.009)$

We now have:

$$h_\theta(x^{(1)}) = 0.961$$

$$h_\theta(x^{(2)}) = 0.884$$

$$h_\theta(x^{(3)}) = 0.996$$

$$h_\theta(x^{(4)}) = 0.960$$

And

$$Cost(h_\theta(x^{(1)}), y^{(1)}) = 0.040$$

$$Cost(h_\theta(x^{(2)}), y^{(2)}) = 0.123$$

$$Cost(h_\theta(x^{(3)}), y^{(3)}) = 5.492$$

$$Cost(h_\theta(x^{(4)}), y^{(4)}) = 3.220$$

So the total cost function, averaged over the four points, is:

$$J(\theta_0, \theta_1, \theta_2) = 2.219$$

So for point $x^{(1)}$: $\frac{\partial J(\theta)}{\partial \theta_0} = -0.039$ $\quad \frac{\partial J(\theta)}{\partial \theta_1} = -0.079$ $\quad \frac{\partial J(\theta)}{\partial \theta_2}$ =-0.157

So for point $x^{(2)}$: $\frac{\partial J(\theta)}{\partial \theta_0} = -0.116$ $\quad \frac{\partial J(\theta)}{\partial \theta_1} = -0.116$ $\quad \frac{\partial J(\theta)}{\partial \theta_2} = -0.347$

So for point $x^{(3)}$: $\frac{\partial J(\theta)}{\partial \theta_0} = 0.996$ $\quad \frac{\partial J(\theta)}{\partial \theta_1} = 3.984$ $\quad \frac{\partial J(\theta)}{\partial \theta_2}$ =1.992

So for point $x^{(4)}$: $\frac{\partial J(\theta)}{\partial \theta_0} = 0.960$ $\quad \frac{\partial J(\theta)}{\partial \theta_1} = 1.920$ $\quad \frac{\partial J(\theta)}{\partial \theta_2} = 1.920$

By averaging over the four points we get:

$$\frac{\partial J(\theta)}{\partial \theta_0} = 0.450 \quad \frac{\partial J(\theta)}{\partial \theta_1} = 1.427 \quad \frac{\partial J(\theta)}{\partial \theta_2} = 0.852$$

And:

$$\theta_0'' = 0.852 - 0.1 \times 0.450 = 0.807$$
$$\theta_1'' = 1.154 - 0.1 \times 1.427 = 1.011$$
$$\theta_2'' = 0.009 - 0.1 \times 0.852 = -0.076$$