Assignment 7

Name: Mohaiminul Al Nahian
RIN: 662026703
Course no.: CSCI 6100

Problem 1

a) For this problem logistic regression with gradient descent has been chosen with learning rate 5 and total 1000 iterations. The chosen features were intensity and symmetry of images.

The final weight vector is $w_{LR} = [6.71485034, -3.41471639, -13.95939419]$. The resultant plots are below

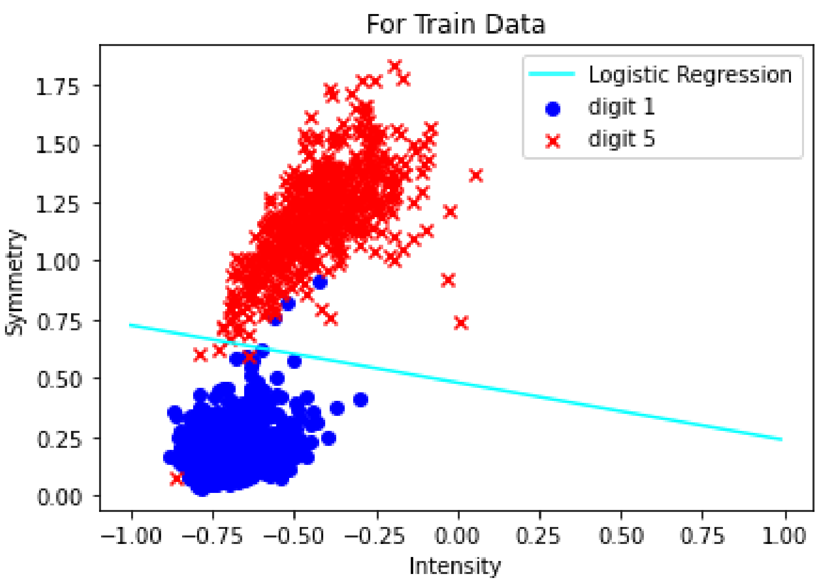

Figure 1: Logistic Regression on Training Data

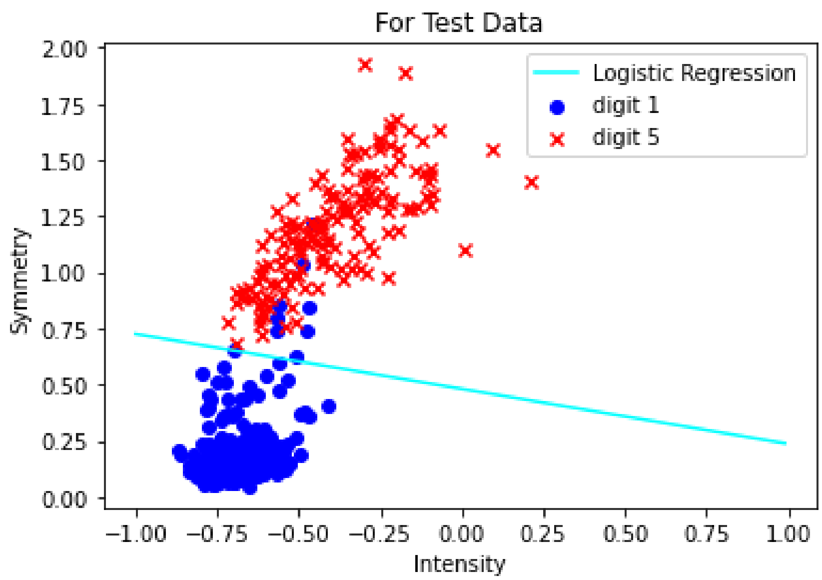

Figure 2: Logistic Regression boundary Test Data

b) From the code,

$$E_{in} = 0.0224$$
$$E_{test} = 0.08201$$

c) For the training set:

$$\begin{aligned}
\delta &= 0.05, \qquad N = 1561 \\
d_{vc} &= 2 + 1 = 3 \\
E_{out} &\leq E_{in} + \sqrt{\frac{8}{N} \ln\left(\frac{4((2N)^{d_{vc}} + 1)}{\delta}\right)} \\
&= E_{in} + \sqrt{\frac{8}{1561} \ln\left(\frac{4(3121^3 + 1)}{0.05}\right)} \\
&= 0.0224 + 0.3823 \\
E_{out} &\leq 0.4047
\end{aligned}$$

For the testing set:

$$\begin{aligned}
N &= 424, \qquad \delta = 0.05 \\
E_{out} &\leq E_{test} + \sqrt{\frac{1}{2N} \ln\frac{2M}{\delta}} \\
&= E_{test} + \sqrt{\frac{1}{848} \ln\frac{2}{0.05}} \\
&= 0.08201 + 0.065955 \\
E_{out} &\leq 0.1479
\end{aligned}$$

The bound based on E_{test} is a better bound.

d) The 3rd order polynomial transform has been applied with features $(1, x_1, x_2, x_1^2, x_2^2, x_1x_2, x_1^3, x_2^3, x_1x_2^2, x_1^2x_2)$. The final weight vector is: $w = [6.08050417, -2.39105945, -2.89466422, 0.23416453, -4.28180022, 2.63854913, 0.7061569, -4.16443569, -2.44381095, 3.19048435]$. the resultant separator for training and testing data is given below:

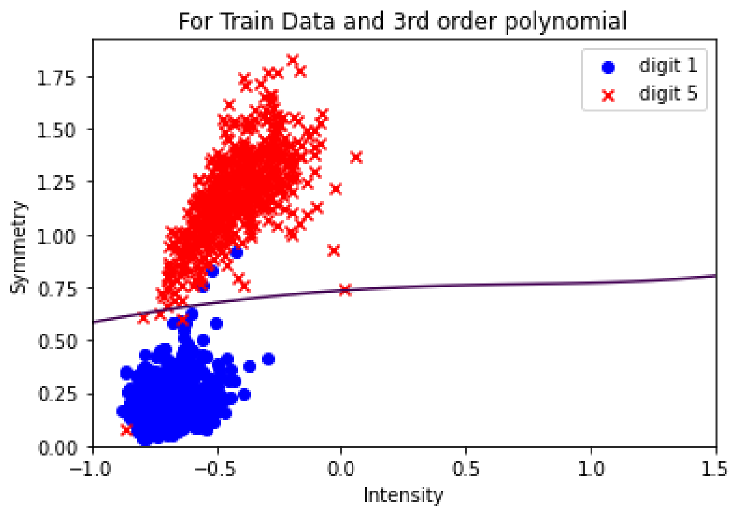

Figure 3: 3rd order polynomial on Training Data

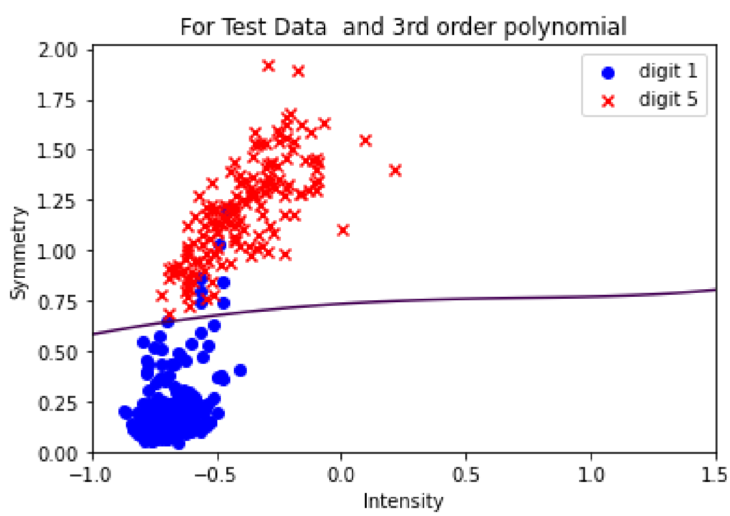

Figure 4: 3rd order polynomial on Testing Data

In this case, the error found from the code is

$$E_{in} = 0.01886$$

$$E_{test} = 0.0958$$

and for the training, testing E_{out} with a new $d_{vc} = 10$

$$\begin{aligned}
N &= 1561, \qquad \delta = 0.05 \\
d_{vc} &= 10 \\
E_{out} &\leq E_{in} + \sqrt{\frac{8}{N} \ln\left(\frac{4((2N)^{d_{vc}} + 1)}{\delta}\right)} \\
&\leq 0.01886 + \sqrt{\frac{8}{1561} \ln\left(\frac{4(3121^{10} + 1)}{0.05}\right)} \\
E_{out} &\leq 0.40116
\end{aligned}$$

For E_{out} using test set, we have

$$\begin{aligned}
N &= 424, \qquad \delta = 0.05 \\
E_{out} &\leq E_{test} + \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} \\
&\leq E_{test} + \sqrt{\frac{1}{848} \ln \frac{2}{0.05}} \\
&= 0.0958 + 0.065955 \\
E_{out} &\leq 0.1617
\end{aligned}$$

The bound based on E_{test} is a better bound.

e) The 3rd order polynomial transform gave worse performance on testing set and gives a higher bound of E_{out}. Also increasing VC dimension from 3 to 10 makes the model complicated and may fit noise. So I would choose the linear model without polynomial transform.

Problem 2

$$f(x, y) = x^2 + 2y^2 + 2\sin(2\pi x)\sin(2\pi y)$$

The gradient has been calculated as follows

$$Grad[f(x, y)] = \begin{bmatrix} 2x + 4\pi\cos(2\pi x)\sin(2\pi y) \\ 4y + 4\pi\sin(2\pi x)\cos(2\pi y) \end{bmatrix}$$

The weight has been updated in each of 50 iterations as

$$\begin{aligned}
x_{i+1} &= x_i - \eta[2x_i + 4\pi\cos(2\pi x_i)\sin(2\pi y_i)] \\
y_{i+1} &= y_i - \eta[4y_i + 4\pi\sin(2\pi x_i)\cos(2\pi y_i)]
\end{aligned}$$

a) For $\eta = 0.01$ and $\eta = 0.1$ following curve has been obtained

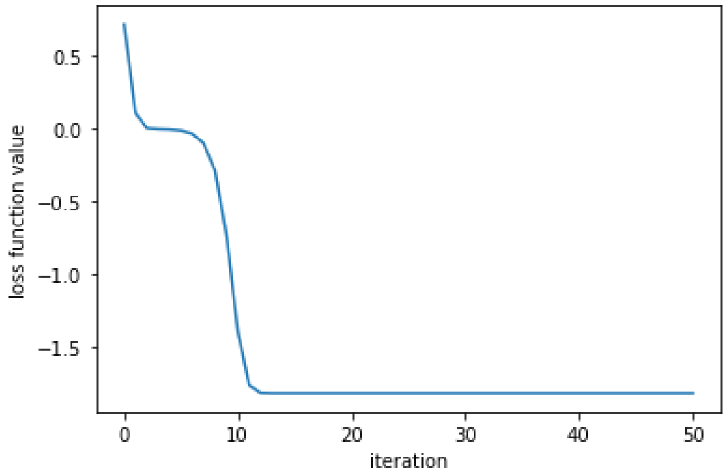

Figure 5: $\eta = 0.01$

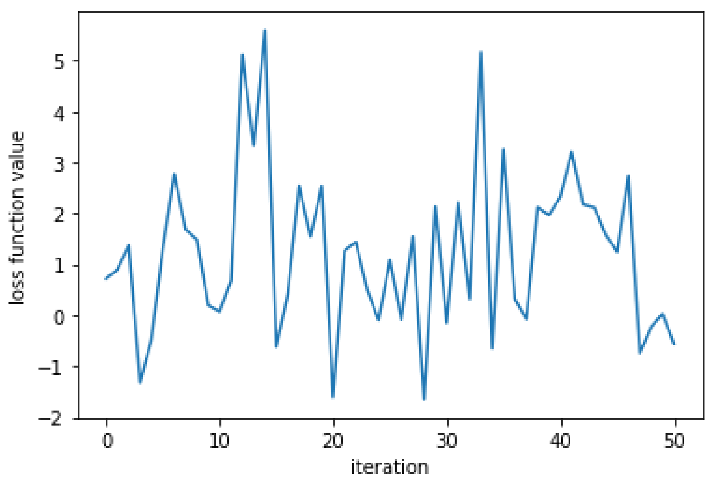

Figure 6: $\eta = 0.1$

With $\eta = 0.1$ the loss function value moves back and forth and cannot find the local minima.

(b) The desired table is given below:

Learning Rate	Initial Point	MinimumValue of Function	Location of Minimum (x,y)
0.01	0.1, 0.1	-1.820078542	0.2438 , -0.2379
0.01	1, 1	0.593269374	1.21807, 0.7128
0.01	-0.5, -0.5	-1.332481062	-0.7313 , -0.2378
0.01	-1, -1	0.593269374	1.21807, -0.7128
0.1	0.1, 0.1	-1.645221618	0.2445 , -0.3031
0.1	1, 1	-1.699744167	0.1973, -0.2668
0.1	-0.5, -0.5	-1.396466643	0.2863 , -0.3315
0.1	-1, -1	-1.699744167	-0.1973 , 0.2668

Figure 7: Table

We can see that different initial point leads to different local minima and it is hard to find global minima

Problem 3.16

a) $P[y = +1|\mathbf{x}] = g(\mathbf{x}) = P[Correct]$, so, $P[y = -1|\mathbf{x}] = 1 - g(\mathbf{x}) = P[Intruder]$. So:

$$\begin{aligned}\text{cost(accept)} &= P[Correct] * 0 + P[Intruder] * C_a = (1 - g(\mathbf{x}))c_a \\ \text{cost(reject)} &= P[Correct] * C_r + P[Intruder] * 0 = g(\mathbf{x})c_r\end{aligned}$$

b) If cost(reject) $\geq$ cost(accept), the person is accepted. So, the condition of acceptance is

$$\begin{aligned}g(\mathbf{x})c_r &\geq (1 - g(\mathbf{x}))c_a \\ g(\mathbf{x})(c_r + c_a) &\geq c_a \\ g(\mathbf{x}) &\geq \frac{c_a}{c_r + c_a} \\ Given, g(\mathbf{x}) &\geq \kappa \\ \text{Thus } \kappa &= \frac{c_a}{c_r + c_a}\end{aligned}$$

c) For super market, $C_a = 1$ and $C_r = 10$. So,

$$\kappa = \frac{1}{1 + 10} = 0.0909$$

For CIA, $C_a = 1000$ and $C_r = 1$. So,

$$\kappa = \frac{1000}{1000 + 1} = 0.9990$$

The supermarket has high cost of False reject, so it will accept a false person with relatively low threshold. But for CIA, intruders may cause serious safety breach. So, CIA has very high cost of false accept. Thus the threshold is very high in order to be very sure that only the right person is being accepted.