Assignment 9

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

1. 8th order Feature Transform.

For a kth Legendre feature transform, the number of features is $(k+1)(k+2)/2$. So for an 8th order Polynomial Feature Transform, the number of feature is $(8+1)(8+2)/2 = 45$. So Z has dimension of (300,45).

2. Overfitting

decision boundary for the resulting weights produced by the linear regression algorithm without any regularization ($\lambda = 0$) is given below:

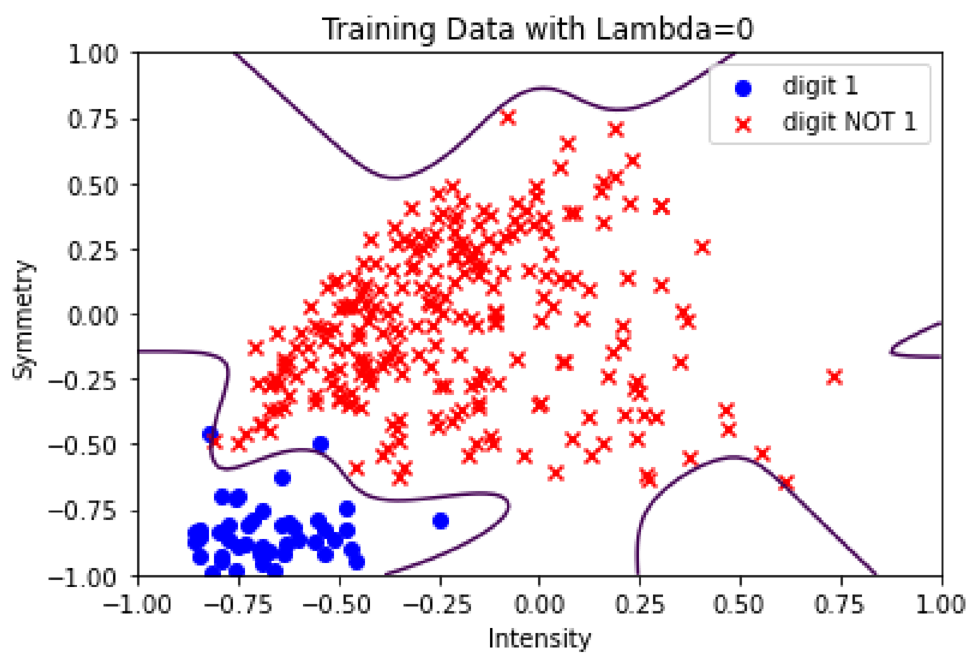

Figure 1: Decision Boundary on training data with $\lambda = 0$

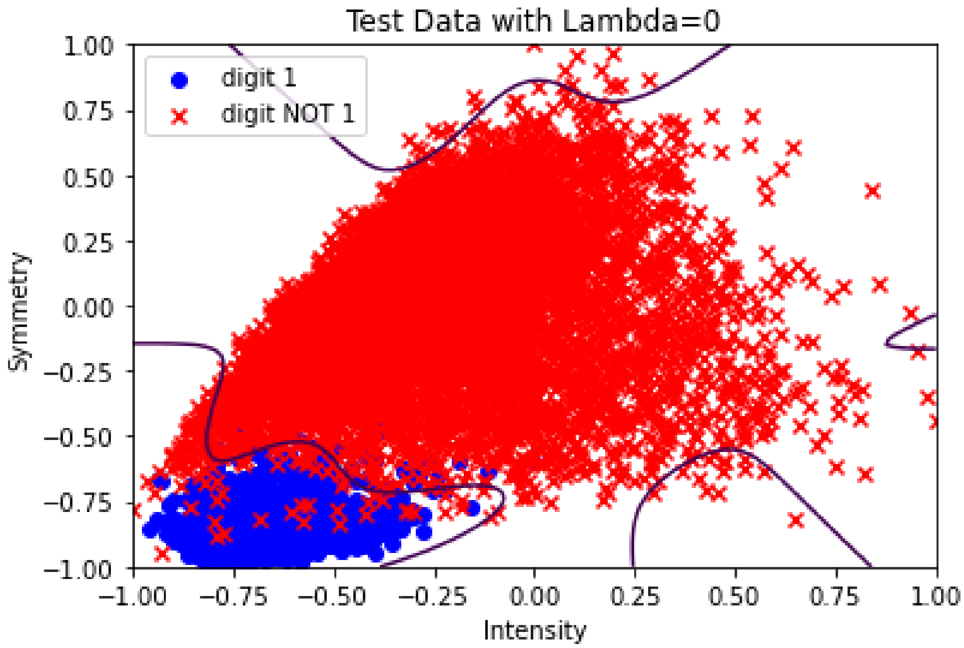

Figure 2: Decision Boundary on test data for $\lambda = 0$

the algorithm overfitted the training data too well and the boundary matches the training data. so, more generalization error occurs in test data

3. Regularization
With $\lambda = 2$ regularization, the resulting decision boundary is shown below

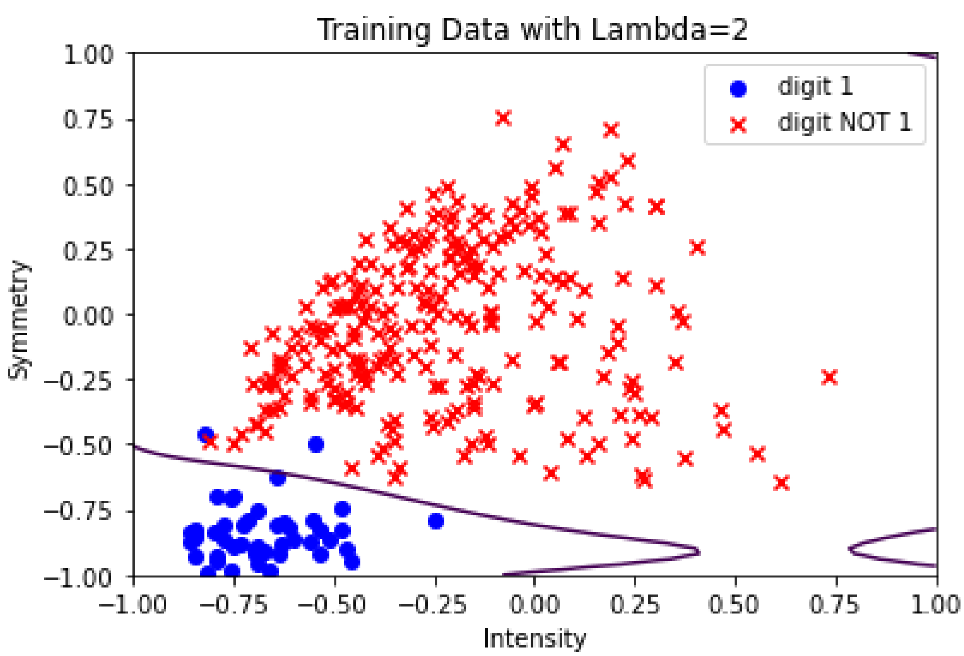

Figure 3: Decision Boundary on Training data with $\lambda = 2$

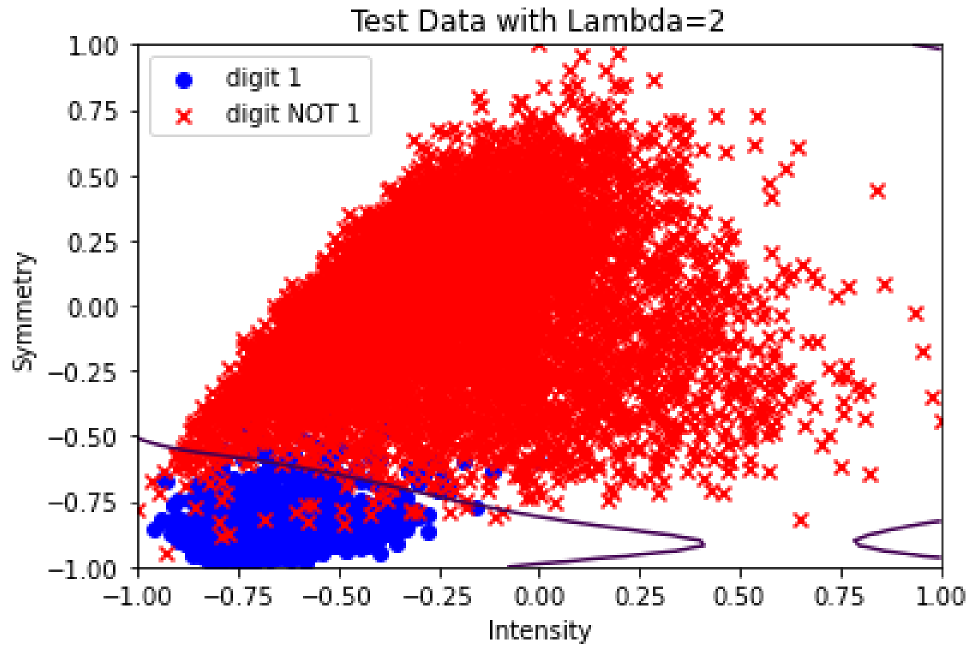

Figure 4: Decision Boundary on Test data for $\lambda = 2$

Here, a clear boundary is seen between the two classes and it does not overfit or underfit. So, generalization error is less.

4. Cross Validation

Taking the value of λ from 0 to 10 with an increment of 0.01, the plot of cross validation error ($E_{cv}(\lambda)$) vs λ and $E_{test}(w_{reg}(\lambda))$ vs λ is given below:

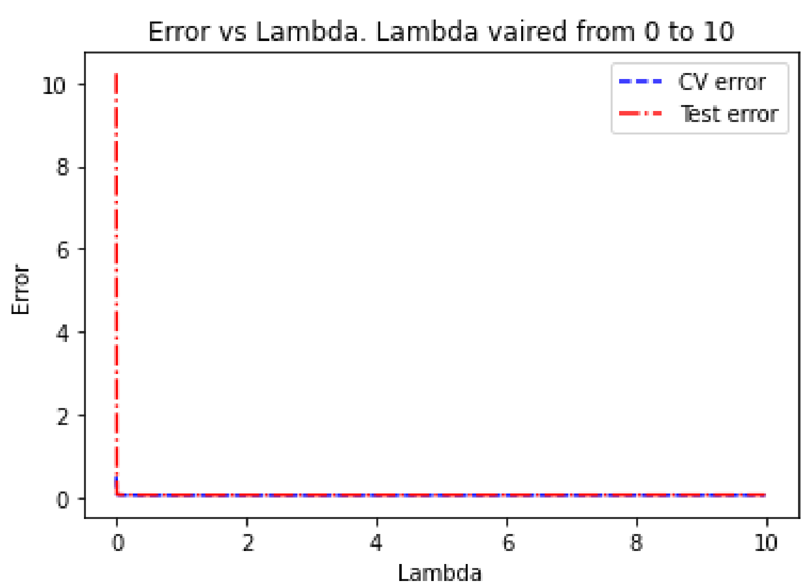

Figure 5: Cross Validation and Test Error for $\lambda \in [0, 10]$

For, $\lambda = 0$, the $E_{CV}(\lambda)$ and $E_{test}(w_{reg})$ is very high, so we cannot visualize the trend properly from the above plot. So, Another plot for $\lambda \in [0.01, 10]$ is given below for a better view.

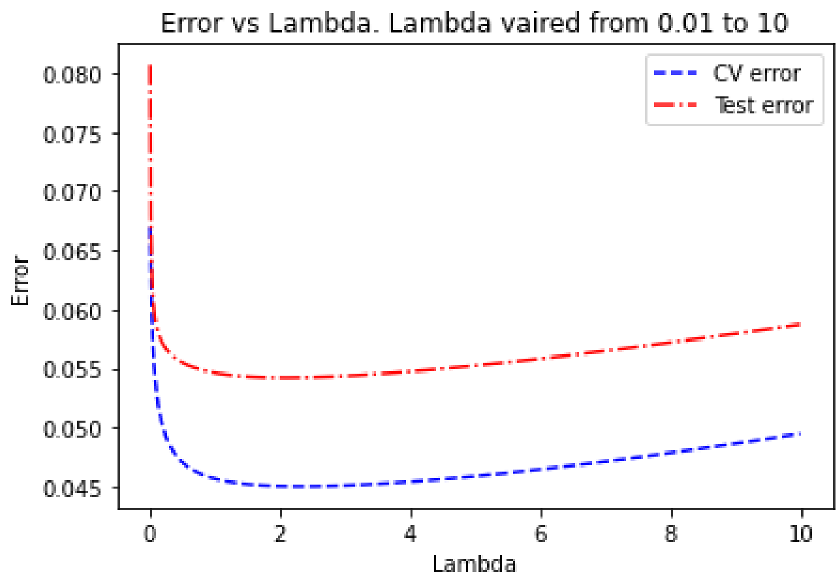

Figure 6: Cross Validation and Test Error for $\lambda \in [0.01, 10]$

It can be seen that E_{cv} is smaller than E_{test} in all cases. Also, it follows the same trend as E_{test}, i.e, as λ increases, both E_{cv} and E_{test} decreases first and then again increases. So, E_{cv} in a way represents same behaviour as E_{test}.

5. Pick λ^*

$\lambda = \lambda^*$, has been chosen for which the cross validation error is the smallest,i.e $E_{cv}(\lambda^*) = min(E_{cv}(\lambda))$. Corresponding λ^* is found to be 2.32. Using this value we get the following decision boundary:

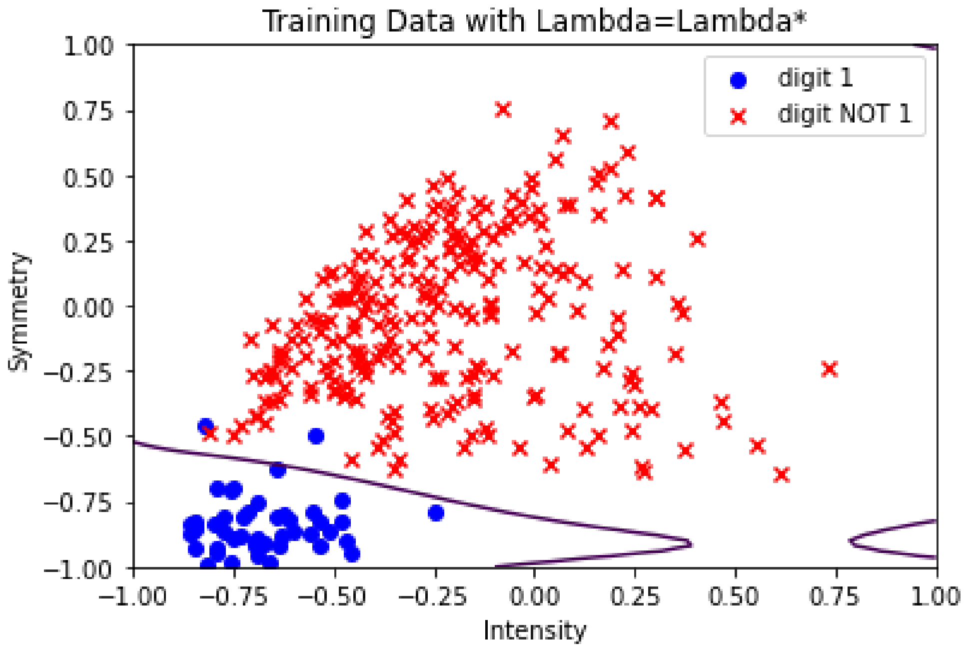

Figure 7: Decision Boundary on Training Data with $\lambda = \lambda^*$

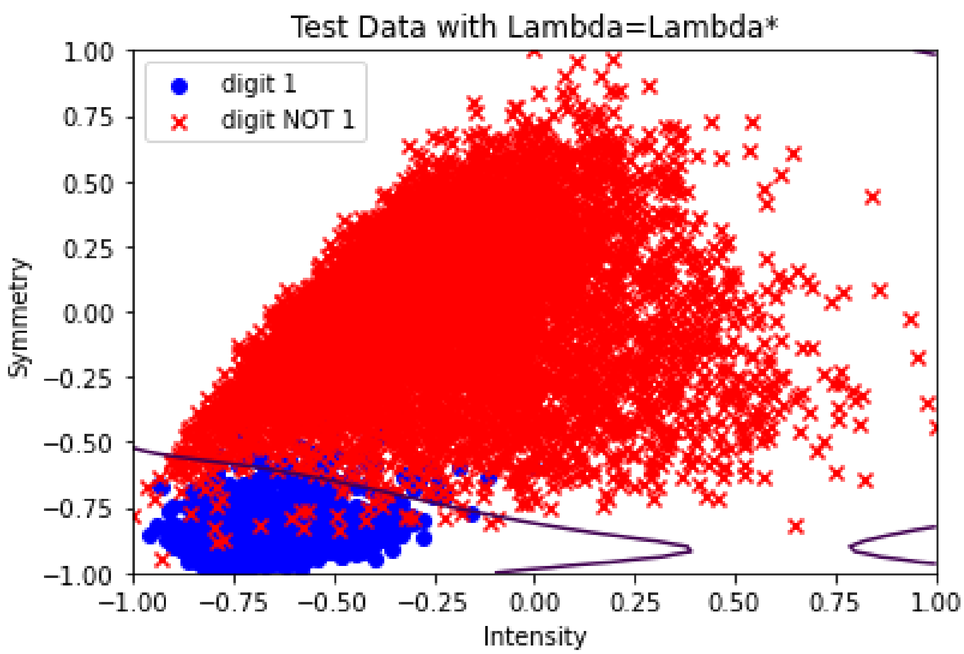

Figure 8: Decision Boundary on Test Data for $\lambda = \lambda^*$

6.Estimate the Classification Error $E_{out}(W_{reg}(\lambda^*)$

Here, $\lambda^* = 2.32$ and there are 8998 test data points. A total of 138 points that got misclassified from the test set, and we set the tolerance to $\delta = 0.01$ so:

$$
\begin{aligned}
E_{test-classification} &= \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i \neq y_i) = \frac{138}{8998} = 0.01534 \\
So, E_{out}(g) &\leq E_{test}(g) + \sqrt{\frac{1}{2N}\ln\frac{2M}{\delta}} \\
&\leq 0.01534 + \sqrt{\frac{1}{2\times 8998}\ln\frac{2\times 1}{0.01}} \\
&\leq 0.01534 + 0.01716 \\
E_{out}(g) &\leq 0.0325
\end{aligned}
$$

So, we have 99% confidence to say E_{out} is less than 0.0325

7. Is E_{cv} Biased?

$E_{cv}(\lambda^*)$ is an unbiased estimate of $E_{test}(w_{reg}(\lambda^*))$. Because, for each leave-one-out cross validation, no test data is involved in the training process, so all $E_{cv}(\lambda)$ is unbiased estimate of $E_{test}(w_{reg}(\lambda))$. So, $E_{cv}(\lambda^*)$ is an unbiased estimate of $E_{test}(w_{reg}(\lambda^*))$.

8. Data Snooping

$E_{test}(w_{reg}(\lambda^*)$ is not an unbiased estimate of $E_{out}(w_{reg}(\lambda^*)$. Because, we normalized (scaled and shifted) the entire datapoints first and then split them into training and testing set. So, information of test set datapoints has been contained in the training set as well. So, data snooping has occured. So, it is not unbiased.
In order to fix data snooping, we have to first split the entire data into train and test set. Then normalize the training data. During testing, the same scale and shift parameters will be used to normalize test data. In this way, we can fix data snooping.