CSCI 6100 Machine Learning From Data
Fall 2018

HOMEWORK 9
Daniel Southwick
661542908
southd@rpi.edu

1. 8th order Polynomial Feature Transform

For a kth polynomial feature transform, the number of features is $1+2+3+...+(k+1)$. So for an 8th order Polynomial Feature Transform, the number of feature is $1+2+3+4+5+6+7+8=45$. So $Z \in \mathbb{R}^{300\times 45}$.

2. Overfitting

With $\lambda = 0$, the algorithm overfitted the data, the boundary matches the training data too well and definitely has a large generalization error.

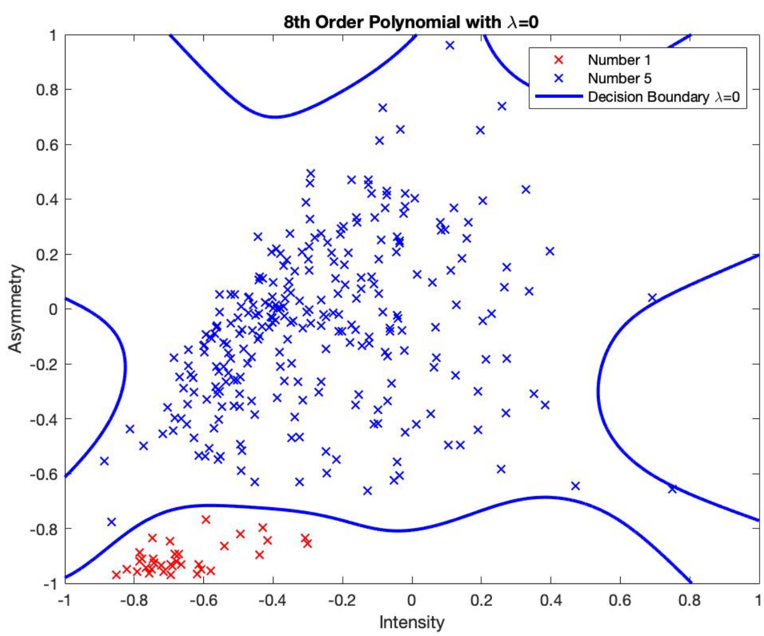

Figure 1: Decision Boundary with $\lambda = 0$

3. Regularization

Now, With $\lambda = 2$ regularization, the algorithm seems like it under fitted the data.

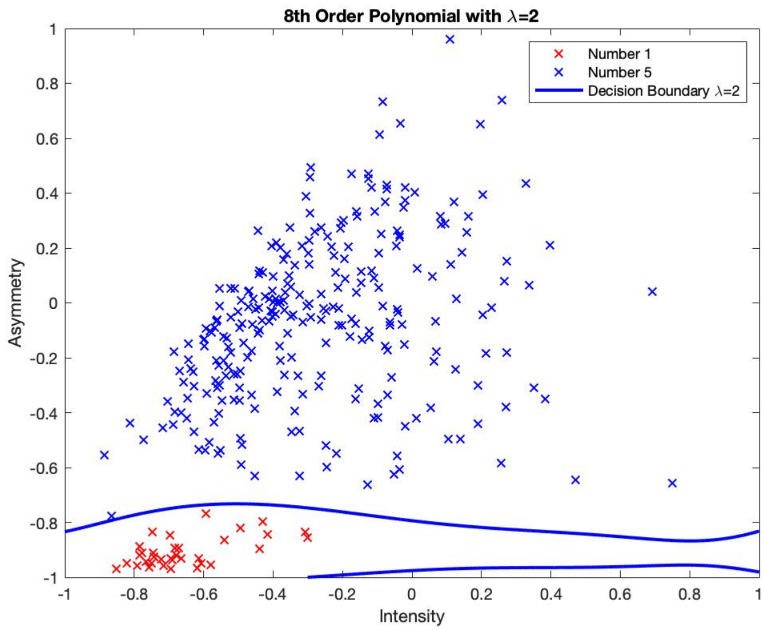

Figure 2: Decision Boundary with $\lambda = 2$

4. Cross Validation

The cross validation can be estimated as:

$$H(\lambda) = Z(Z^T Z + \lambda I)^{-1} Z^T$$
$$\hat{y} = H(\lambda) y$$
$$E_{cv} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\hat{y_n} - y_n}{1 - H_{nn}(\lambda)}\right)^2$$

We change the value of λ and plotted the cross validation error:

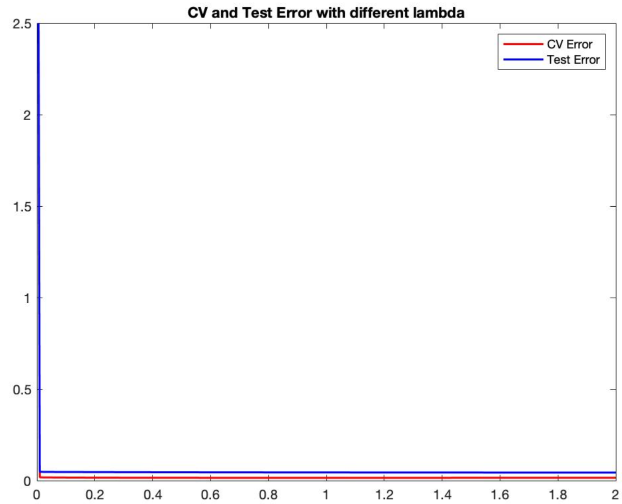

Figure 3: CV and Test Error for various λ

E_{cv} is always smaller than E_{test}, but it has the same trend as E_{test}, this means it reflects the changes of E_{test}

5. Pick λ^*

Now, we let $\lambda = \lambda^*$, which is the smallest E_{cv} corresponding λ, it is found to be 0.9. and using the same method as 2 and 3, we retain the result as follows:

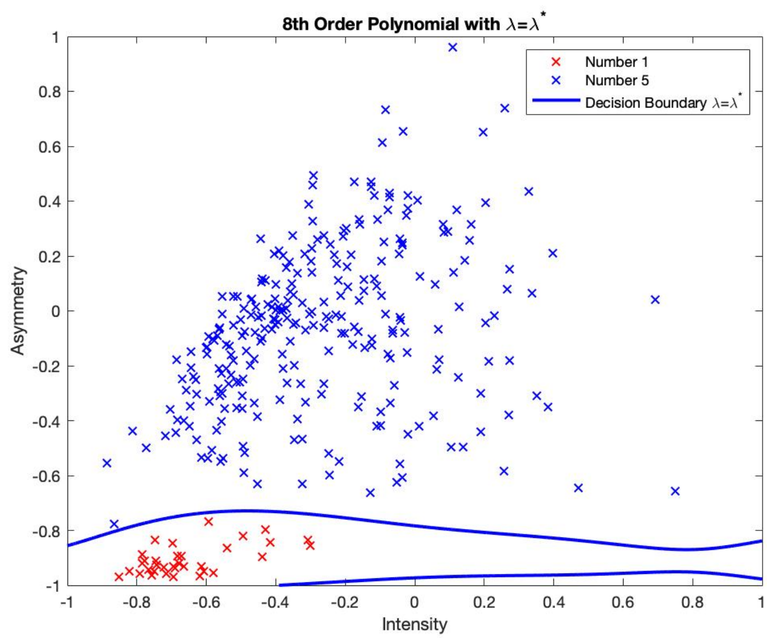

Figure 4: Decision Boundary with $\lambda = \lambda^*$

6. E_{out} Estimation

$\lambda^* = 0.9$ and there are 8998 data points within the test data and there are total of 109 points that got misclassified, and we set the tolerance to $\delta = 0.05$ so:

$$E_{test} = \frac{1}{N}\sum_{i=1}^{N} 1_{sign(\hat{y}_i \neq y_i)} = 0.048$$

$$\begin{aligned}
E_{out}(g) &\leq E_{test}(g) + \sqrt{\frac{1}{2N}\ln\frac{2M}{\delta}} \\
&\leq \frac{109}{8998} + \sqrt{\frac{1}{2\times 8998}\ln\frac{2\times 1}{0.05}} \\
&\leq 0.0118 + 0.0143 \\
E_{out}(g) &\leq 0.0261
\end{aligned}$$

Therefore we have 95% confidence to say E_{out} is less than 0.0261

7. E_{cv} Biasity

$E_{cv}(\lambda^*)$ is not an unbiased estimate of $E_{test}(w_{reg}(\lambda^*))$. E_{test} is the error from applying the hypothesis g on the test dataset. While E_{cv} is obtained with g^-, learned from $N-1$ data from the training dataset, since the λ^* is selected based on the E_{cv}, thus $E_{cv}(\lambda^*)$ is not an unbiased estimate of $E_{test}(w_{reg}(\lambda^*))$.

8. Data Snooping

$E_{test}(w_{reg}(\lambda^*))$ is not an unbiased estimate of $E_{out}(w_{reg}(\lambda^*))$. Since it chosen from the data set which was first split into training and testing dataset. λ^* is selected from the training set and $E_{test}(w_{reg}(\lambda^*))$ is computed after λ^* is computed thus data snooping occurs. To fix it, we should use separate data set while doing validation, let it be fixed instead of random generated in this problem.