

Machine Learning from Data Assignment 9

Greg Stewart

November 12, 2018

1. 8th order feature transform.

The 8th order transform for \mathbf{z} is given by

$$\mathbf{z} = (1 \quad L_1(x_1) \quad L_1(x_2) \quad L_1(x_1)L_1(x_2) \quad L_2(x_1) \quad \cdots \quad L_1(x_1)L_7(x_2) \quad L_8(x_2))^T$$

This is a 45 dimensional vector. Thus the transformed data matrix Z will be of dimensions $(300, 45)$.

2. Overfitting.

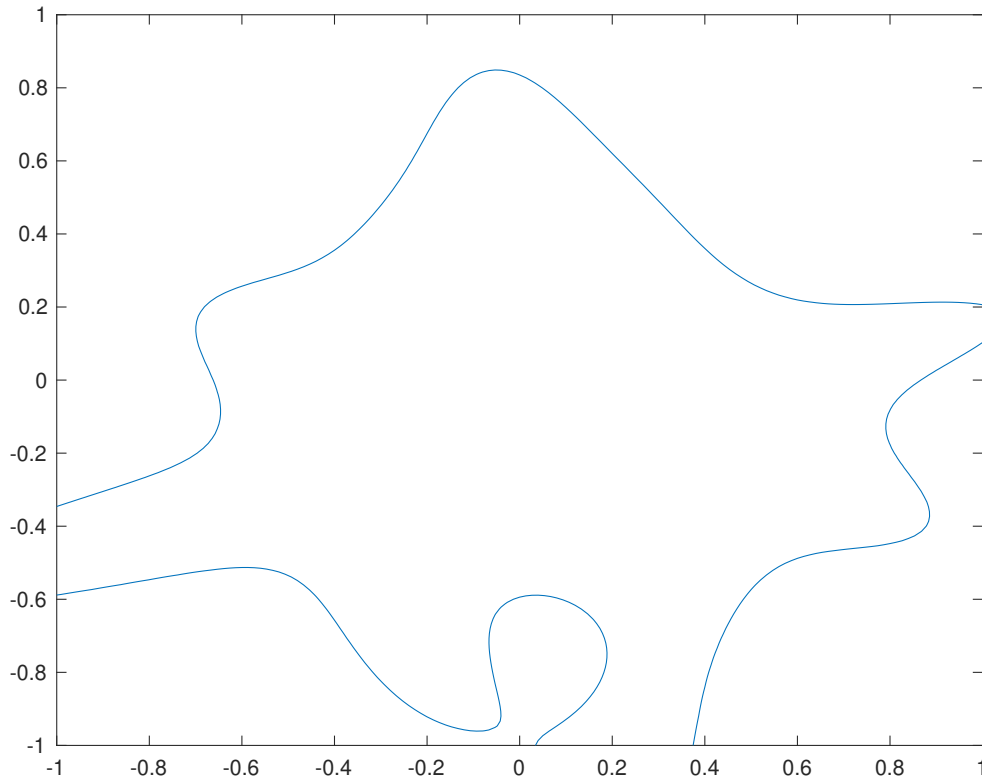


Figure 1: A plot of the decision boundary for $\lambda = 0$.

In the above graph, the horizontal axis is symmetry and the vertical axis intensity. There is no regularization here, and unsurprisingly the boundary appears to be overfit.

3. Regularization.

Now, we introduce a lot of regularization.

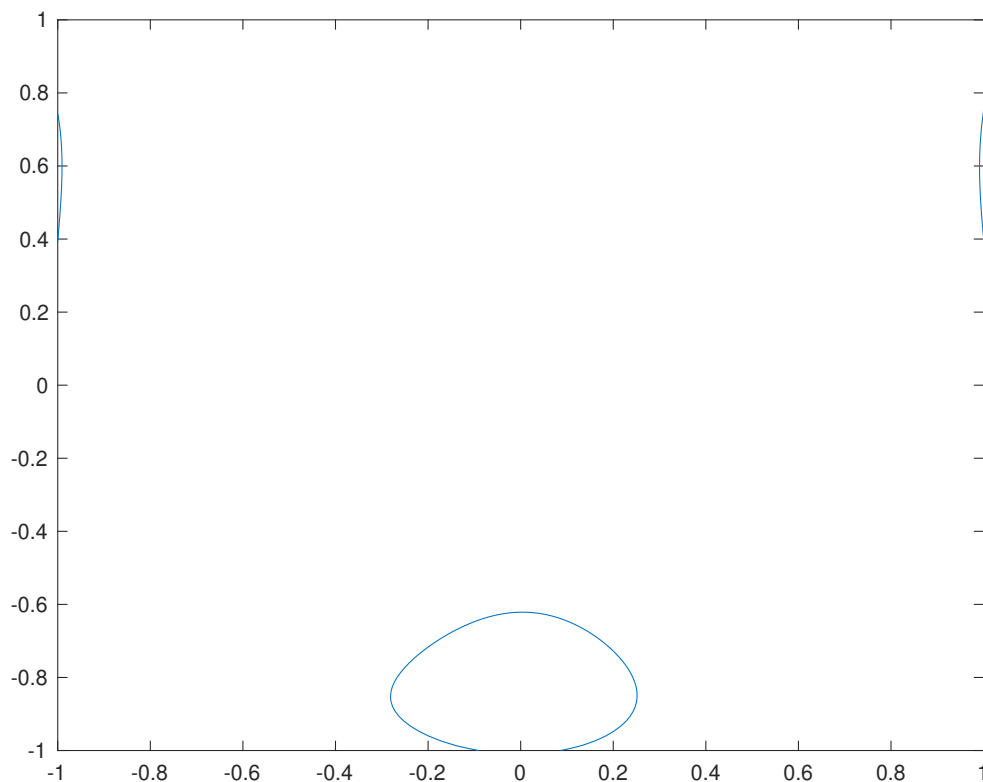


Figure 2: A plot of the decision boundary for $\lambda = 2$.

As expected, there appears to be some underfitting here, with so much regularization.

4. Cross Validation.

Graphing the cross validation and test errors for each lambda in the range $[0, 2]$, with steps of 0.01, we have the graph in Figure 3.

As one might guess, the cross validation error is lower than the test error, given that leave-one-out cross validation still uses essentially the training data. Both have minimums, but at slightly different lambda values, and both increase somewhat as lambda increases beyond its apparently optimal value.

5. Pick λ .

The value chosen for lambda based on the error measures is

$$\lambda^* = .04$$

Figure 4 plots the decision boundary given by $\mathbf{w}_{reg}(\lambda^*)$.

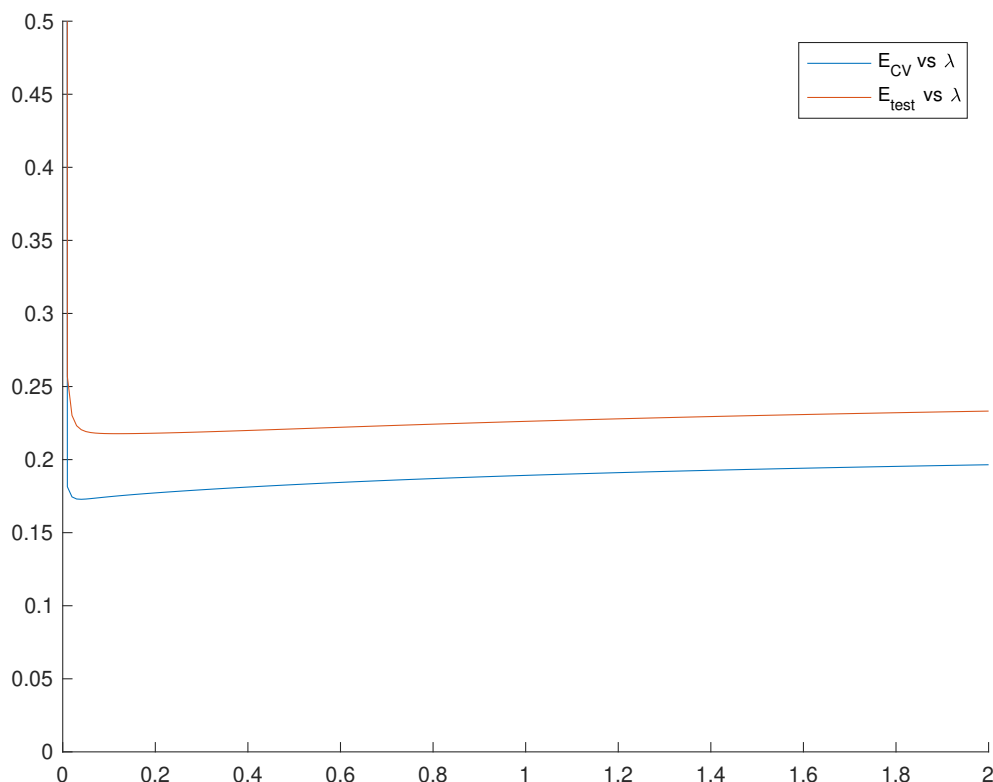


Figure 3: The two error measures calculated for $\lambda \in 0, 0.01, 0.02, \dots, 2$.

6. Estimate E_{out} .

The estimate for E_{out} is simply given by E_{test} , for which we have the value 0.0670. Thus,

$$E_{out} \approx .0670 = 6.70\%$$

7. Is E_{CV} unbiased?

E_{CV} is very close to an unbiased estimate of $E_{test}(\mathbf{w}_{reg}(\lambda^*))$. Because of the leave-one-out cross validation, it is a nearly unbiased estimate of E_{test} for each one left out, and as we're taking the average of a large set (for each data point left out), we get an unbiased estimate.

8. Data Snooping.

$E_{test}(\mathbf{w}_{reg}(\lambda^*))$ is not an unbiased estimate of $E_{out}(\mathbf{w}_{reg}(\lambda^*))$ due to data snooping in the normalization of features. Normalization of features can only be done when the features are decided and the values known, so the data is essentially altered before any learning is done. This is a form of data snooping and results in bias in the estimate. To fix this problem, you could either not normalize the features, or normalize them afterward, not looking at the data before hand, resulting in an unbiased estimate.

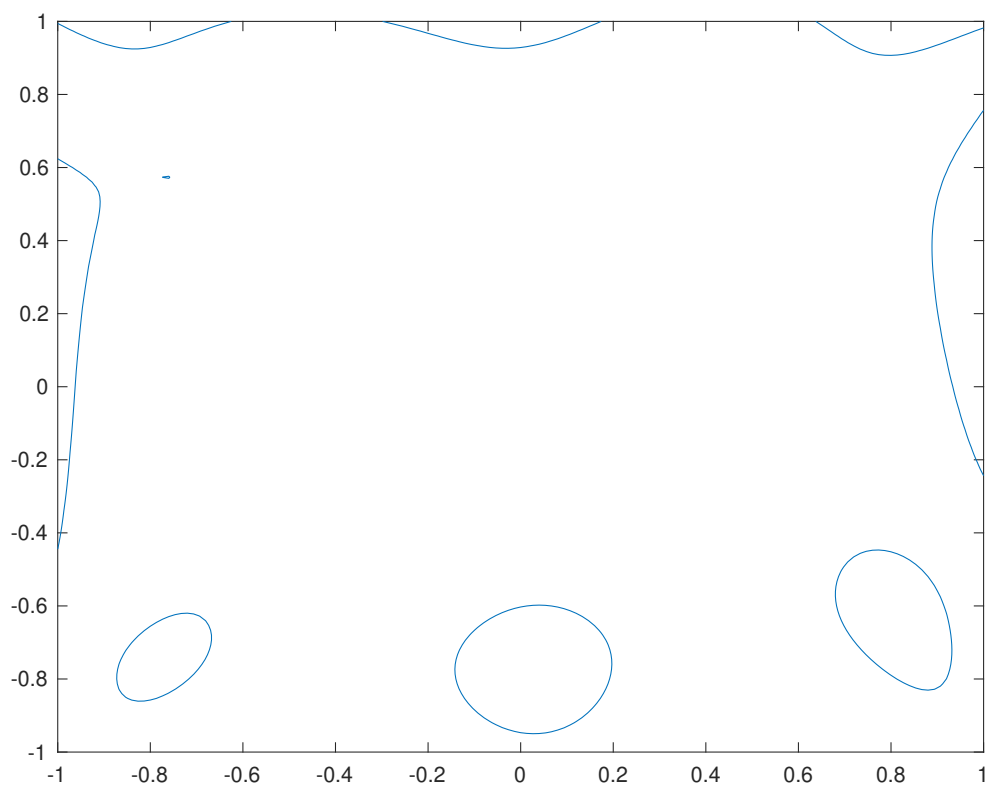


Figure 4: The final decision boundary, using the λ^* obtained from the error