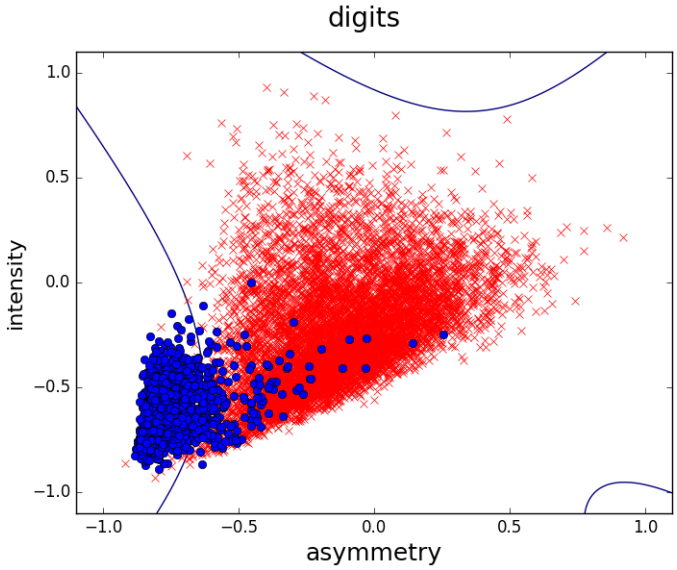
Machine Learning from Data HW9

Shane O'Brien

November 2017

Pre-Questions

Using all of the data, I normalized the digit data. I have plotted them here:

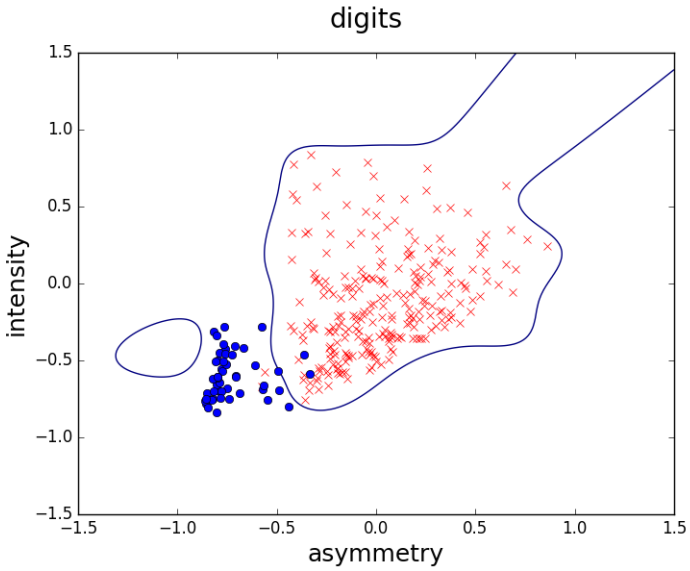


Question 1

We have 300 training points, and each training point has 45 dimensions. The dimensions of $\mathbf{Z}$ is (300x45)

Question 2

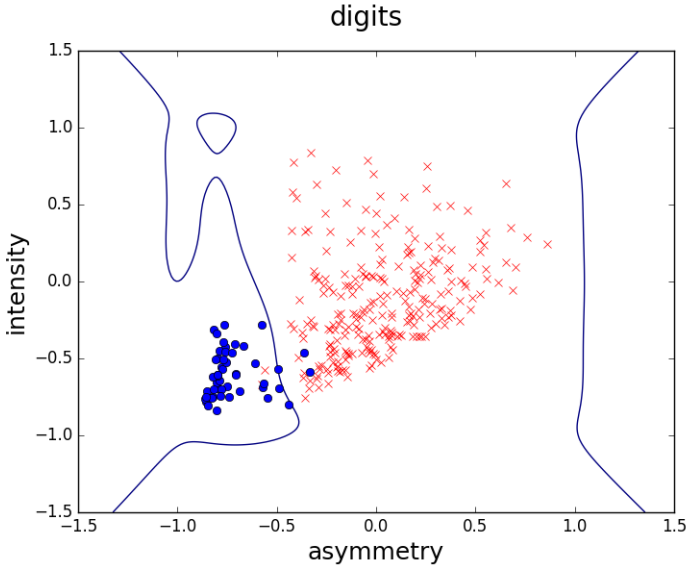
Below is a graph of the decision boundary for $\lambda = 0$.



This is terribly overfitted. We have no regularization and a 44 dimension output function.

Question 3

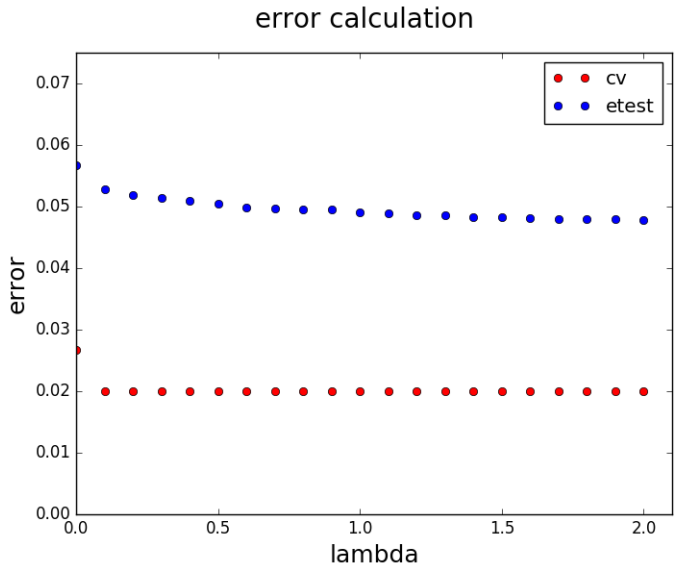
Below is a graph of the decision boundary for $\lambda = 2$.



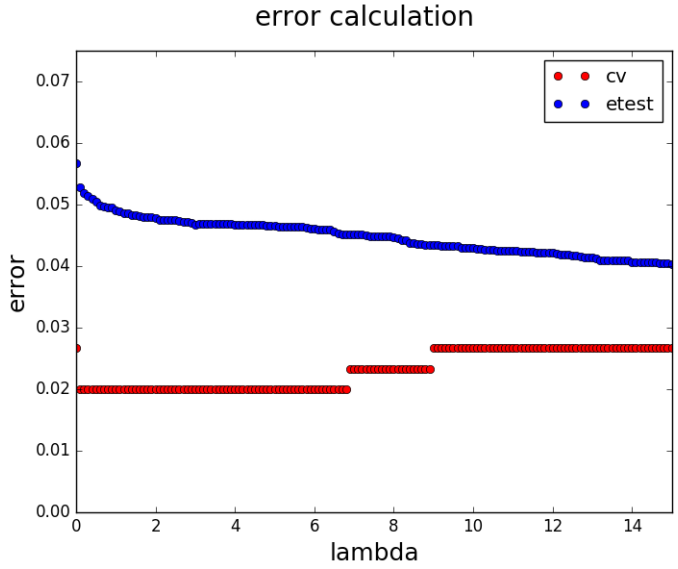
This is still overfitted, but much less so. I think it is still overfitting because of the large amount of dimensions in our output function.

Question 4

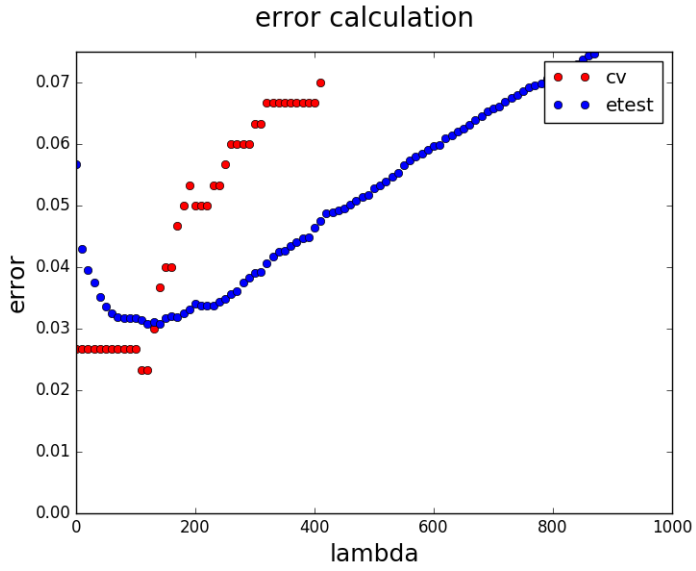
Below is a graph of λ changing from 0 to 2.



This result makes a lot of sense. Our current process is prone to overfitting, so we will see that regularizing it improves our E_{test}. Furthermore, our E_{cv} is always a bit lower than our E_{test}. This is consistent with the textbook's predictions. Below, I have increased the scope of λ all the way up to 15.



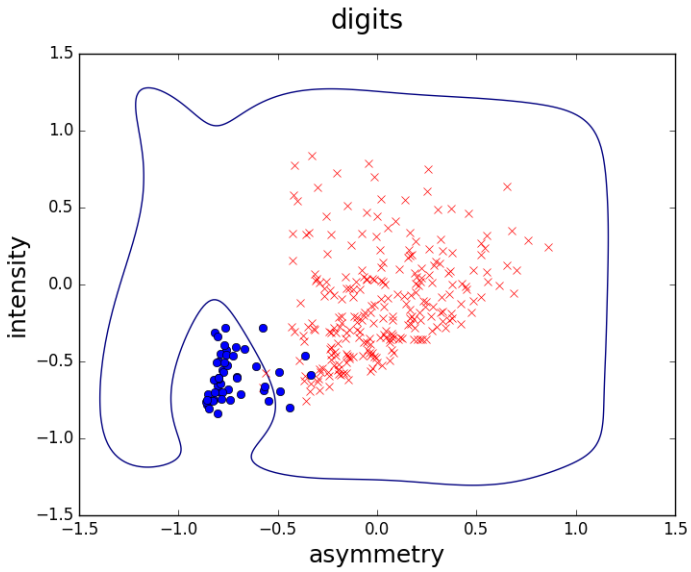
This still improves our E_{test}. I wanted to see at which point our regularization would lead to underfitting, so I kept increasing λ. Below, I have a graph of λ from 0 to 1000.



The best λ value seems to be around 150. This number for λ is very high due to the huge dimensionality of our function. Also, there is perhaps a scaling difference in my code, as λ is typically much smaller.

Question 5

As stated in the previous question, λ of 150 is the best value for my code. The value of $E_{test}(\mathbf{w}(\lambda))$ is 0.03178. I have graphed that here:



All of the results make sense looking at this graph. The decision boundary looks perfect without the [-1,1] area.

Question 6

Using E_{test} to calculate our E_{out}, we get:

$$\mathbf{E}_{out} \leq \mathbf{E}_{test} + \sqrt{\frac{1}{2N} * ln\frac{2H}{\delta}}$$

$$\mathbf{E}_{out} \leq \mathbf{E}_{test} + \sqrt{\frac{1}{2 * 300} * ln\frac{2 * 1}{0.05}}$$

$$\mathbf{E}_{out} \leq 0.03178 + 0.07841$$

$$\mathbf{E}_{out} \leq 0.11019$$

Question 7

No, this is not an unbiased estimate. There is an element of choice here. We have picked (based on $E_{cv}(\lambda^*)$) our λ that we used to get $\mathbf{w}_{reg}$. This element of choice makes $E_{cv}(\lambda^*)$ a biased estimate for $E_{out}(\mathbf{w}_{reg}(\lambda^*))$

Question 8

No, this is not an unbiased estimate. This is because of the pre-processing. We used both our test data and our training data for our normalization. For it to be unbiased, we can't use our testing data to normalize our training data. If we want to normalize, the normalization HAS to be done separately.