Homework 9
CSCI4100

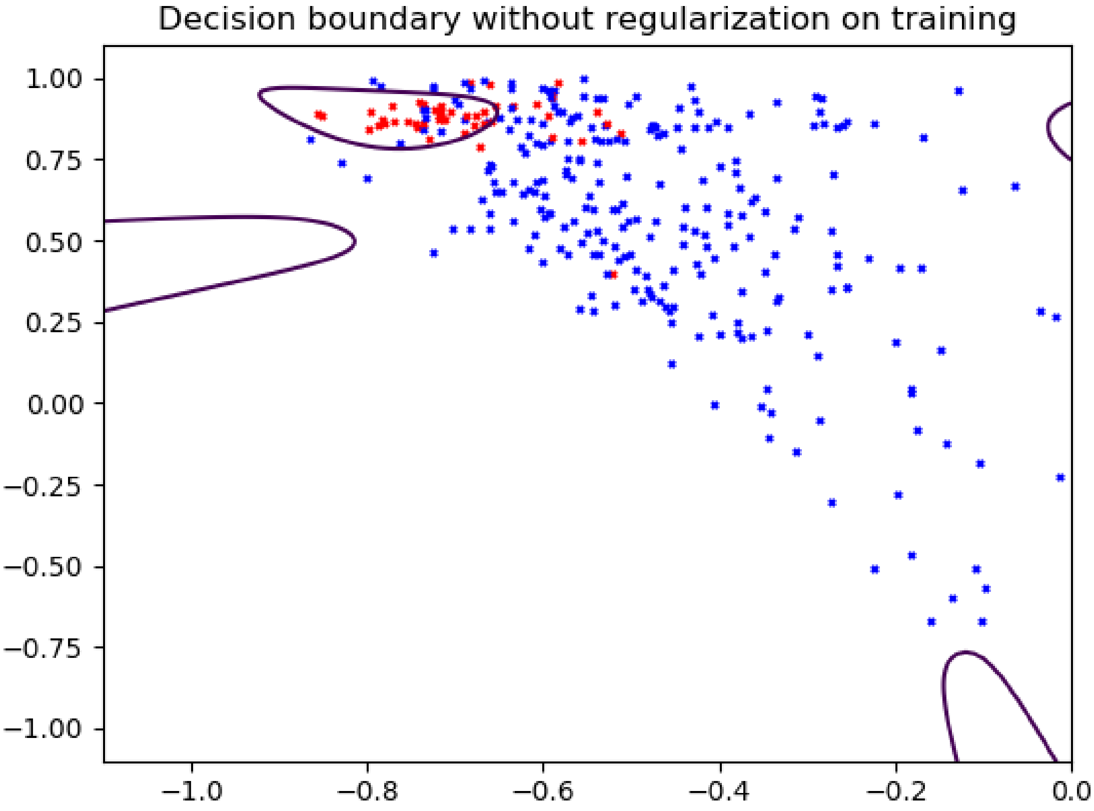
Han Hai
Rin:661534083
haih2@rpi.edu

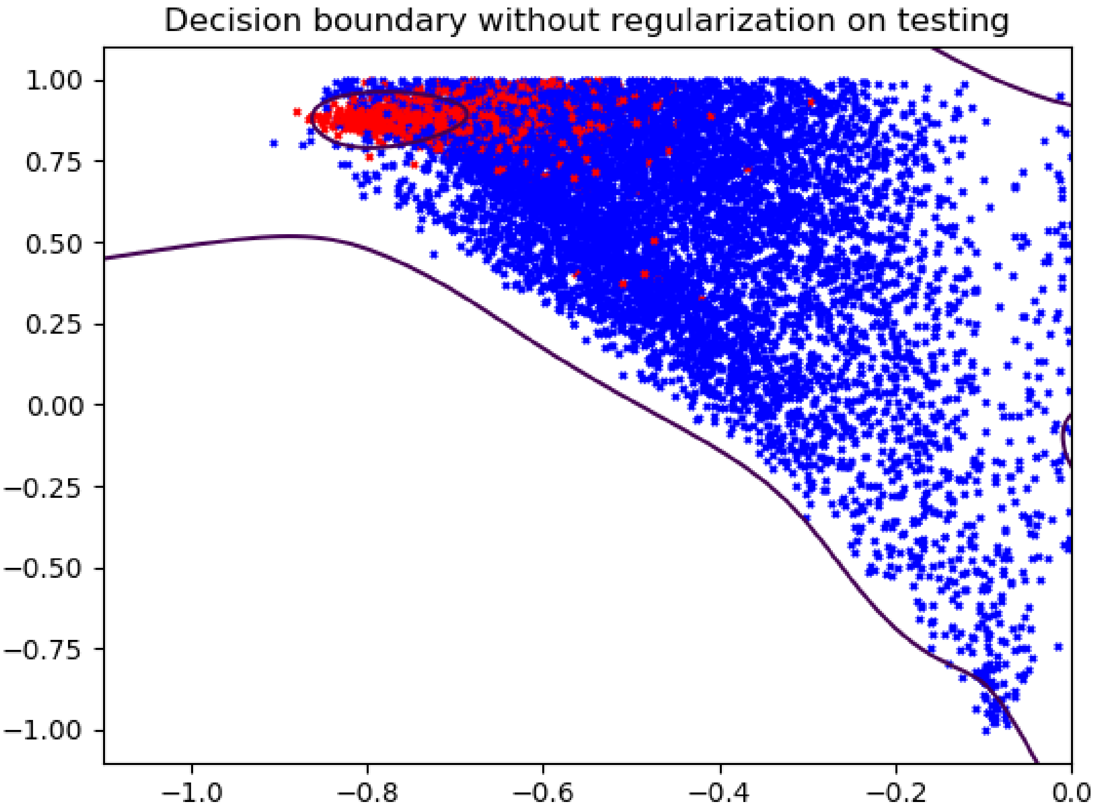
November 4th 2018

1. **8th order Feature Transform**
   Use the 8th order Legendre polynomial feature transform to compute Z. What are the dimensions of Z?

   Solution:
   After 8th order transform on my two features x1 x2, dimensions on input become 45, and since weed pick 300 random data Z is 45 × 300

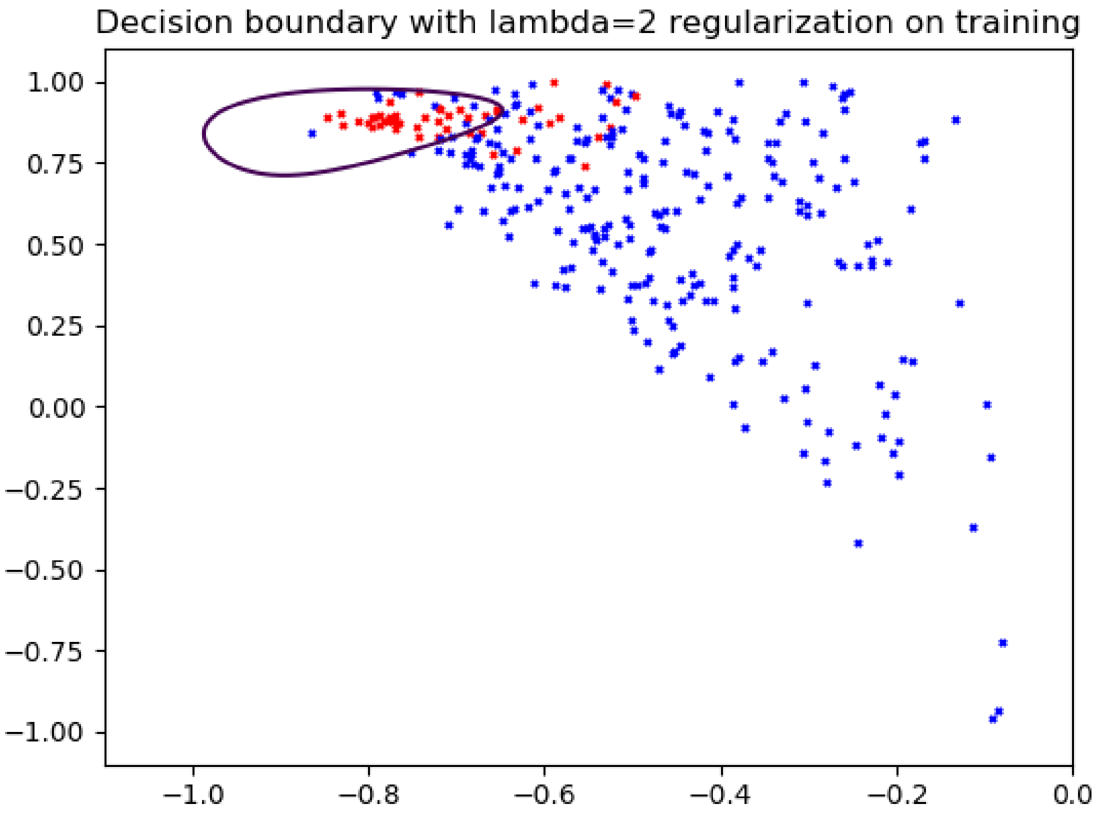
2. **Overfitting** Give a plot of the decision boundary for the resulting weights without any regularization ($\lambda = 0$). Do you think there is overfitting or underfitting?

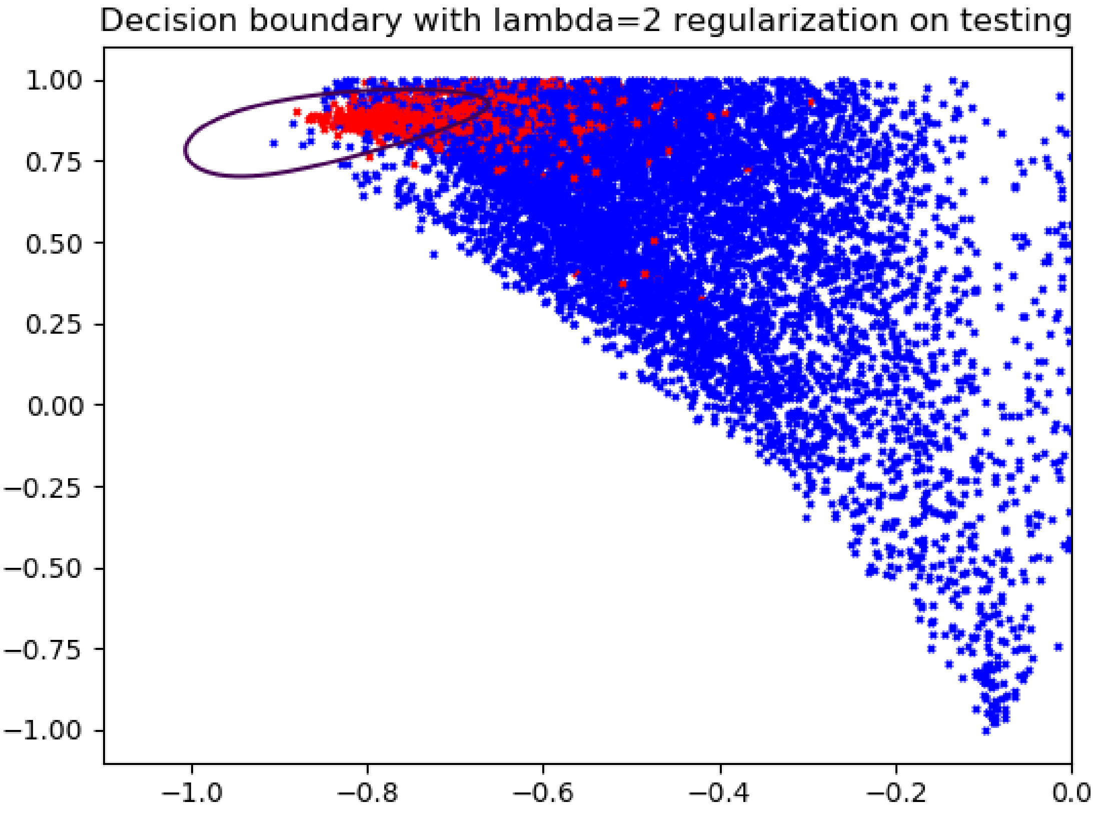




E_{test} here is 0.108
As we can see, overfitting occurs since unnecessary extra boundaries appear. Some data are miscalssified due to those extra boundaries.

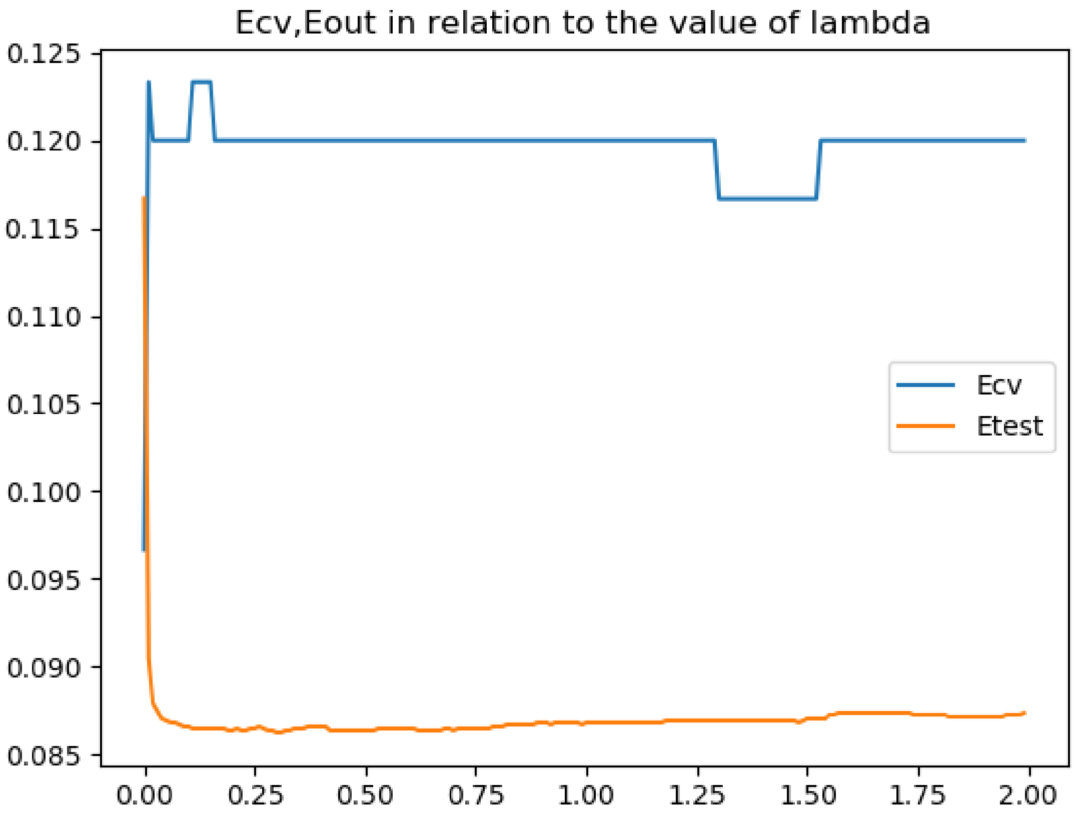
3. **Regularization**Give a plot of the decision boundary for the resulting weights with $\lambda = 2$. Do you think there is overfitting or underfitting?


Decision boundary with lambda=2 regularization on training
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.0
−0.8
−0.6
−0.4
−0.2
0.0



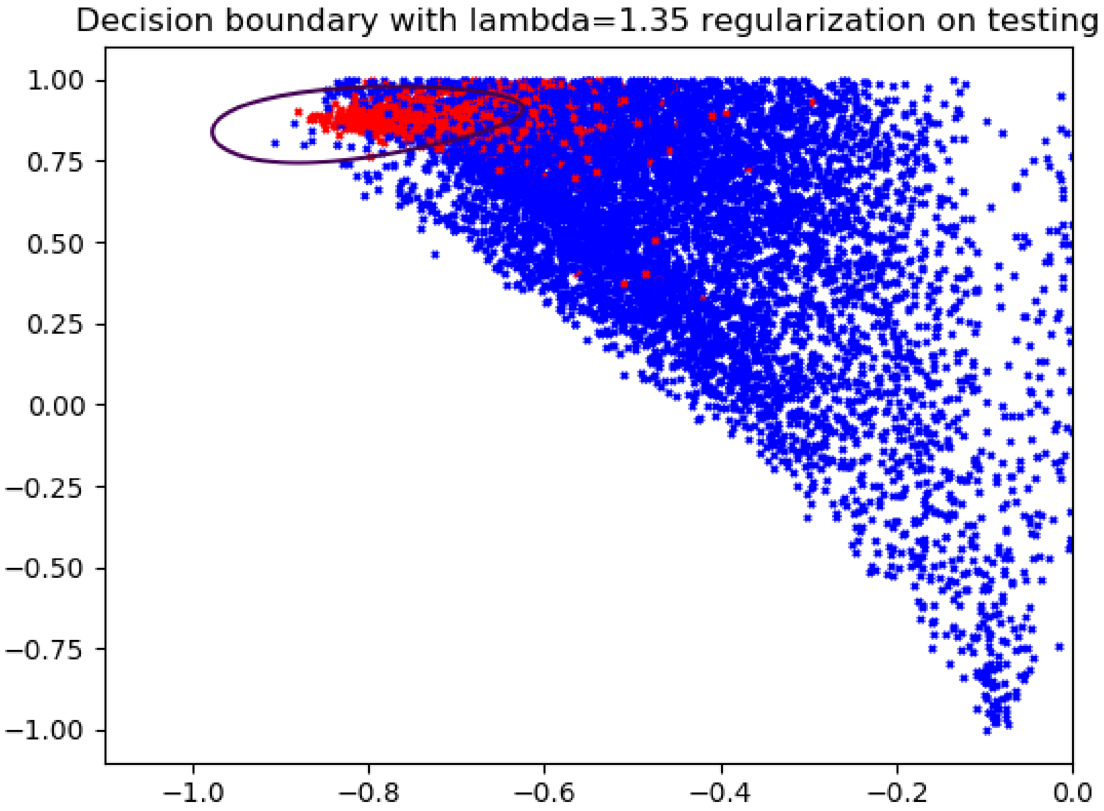
E_{test} here is 0.0896
As we can see extra boundaries due to overfitting disappear, it is neither overfitting nor underfitting.
E_{test} has also improved significantly.

4. **Cross Validation.**
   X axis is the value of lambda while y axis is the error



   As we can see, at lower value lambda, both Ecv and Etest has high error, the error converges to low at $\lambda = 1.35$.

5. **Pick λ**
   Use the cross validation error to pick the best value of λ, call it λ^* . Give a plot of the decision boundary for the weights. Use the optimal lambda $\lambda^* = 1.35$



$E_{test} = 0.083$

6. **Estimate Eout**

   $E_{out} < E_{test} + \sqrt{\frac{1}{2N} ln(\frac{4(2N)^{dvc}+1}{\epsilon})}$
   N=9298-300=8998 use epsilon=0.05,
   $E_{out} < 0.083 + 0.014 = 0.097$

7. **Is Ecv biased?**
   No, when each cross validation occurs, the validation point is independent of the training set, therefore Ecv is unbiased.

8. **Data Snooping.** Is E_{test} unbiased? How can we fixed this?
   Data snooping occurred, so E_{test} is biased. When we picked lambda, we used data that are used to calculate Etest to find a smallest Etest, therefore it becomes biased. To avoid this, do not use Etest to pick hypothesis, instead, only consider Ein.