CSCI 4100 Fall 2018
Assignment 9 Answers

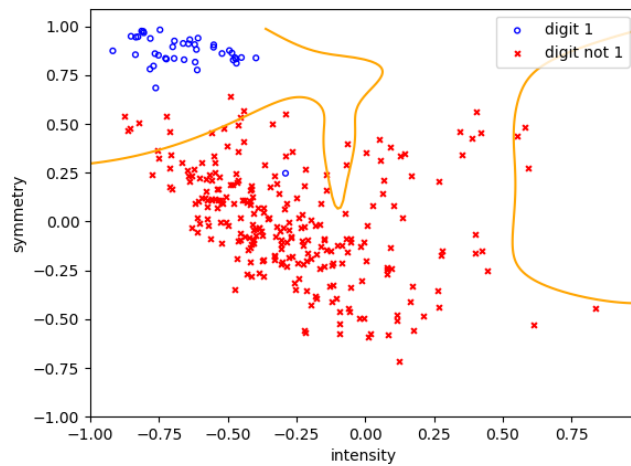Damin Xu
661679187

November 12, 2018

Problem 1
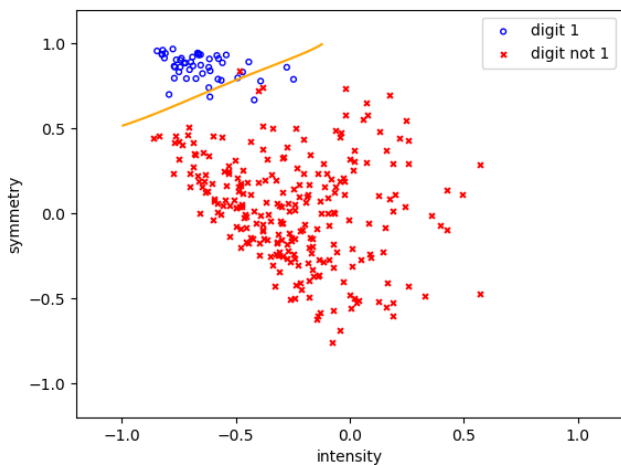
Dimension of Z after 8th order transformation is 300*45.
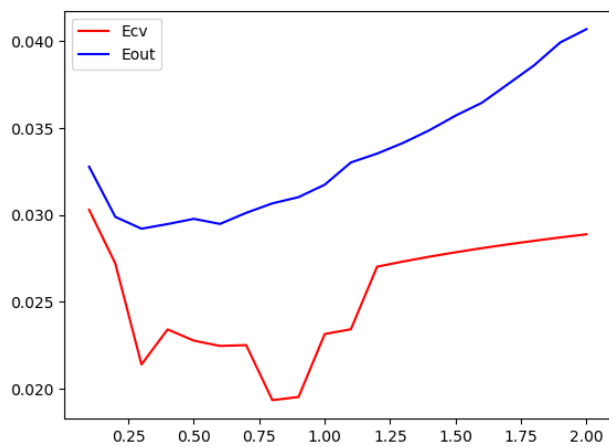
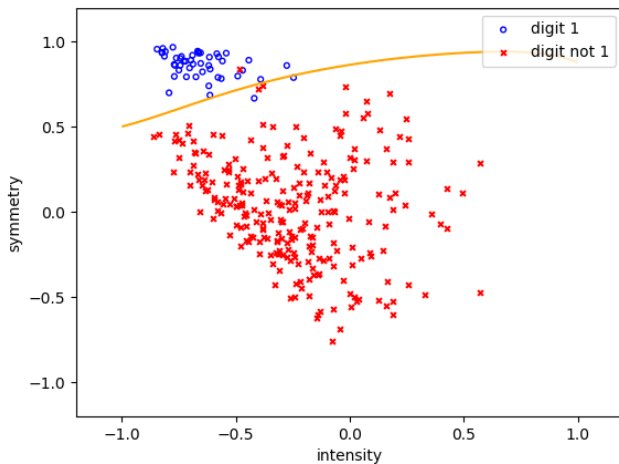**Problem 2**

There is overfitting.

Problem 3

There is underfitting.

Problem 4

When E_{CV} is at the minimum, E_{test} is almost at the minimum. So E_{CV} indicates the best λ which causes a relatively very small Etest.

Problem 5

The best λ is 0.8.

Problem 6

The estimate $E_{out} = 0.030669072830919387 + 0.00248 = 0.0331$.

Problem 7

E_{CV} is a biased estimate of E_{test}, because λ^* is chosen from a range of λ by test among the train set of 300 data, which means this E_{CV} is close to E_{in}. Since E_{in} is not unbiased, E_{CV} is also unbiased.

Problem 8

$E_{test}(w_{reg}(\lambda^*))$ is a biased estimate of $E_{out}(w_{reg}(\lambda^*))$, because here the train set of data is randomly chosen from all data, this random chosen set may contain some bad data which cannot be classified well. The process of random pick continues until a better result is given. This causes data snopping.

To fix this probelm, I would like to do the normalization after spliting train set from test set.