Michael Lee

CSE 446

HW2

1 a) False, is not typically sparse. This is because by taking the 2-norm of the vector we don’t select sparse solutions optimally. For example:

If we calculate the L2 norm of the vectors:

Versus if we calculate the L1 norm:

In the L2 case we consider where as in the L1 case we consider

, hence we don’t optimize for solutions of sparsity.

b) True, because as the weights increase with lambda = 0 our algorithm will optimize for the reduction in training error. And be increasing the values of the weights the training will subsequently decrease.

c) False, the increase in lambda will cause the coefficients of to decrease thereby increasing the error on the training set. Hence, this will cause the likelihood to decrease.

d) At lambda=0 there will be overfitting and high error as a result. Then as lambda increases the test error will decrease to due to an extent, but the error will then increase with high lambda due to under fitting.

e)

i)

ii) The relationship between is that is accounting for some other weight given to the same data point .

iii) Would it change if using L2 regularization <ie. adding a penalty>

2) Boosting

a)

b) show:

c)

i)

= 0

..

3. [attached photos]

4.2

a) – logistic regression.