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 are accounting for the same feature. Since are dealing with the same feature the algorithm will optimize such that these two weights are reduced equally. In comparison to the first data set would have a value larger value as it wouldn’t have its data point duplicated.

iii) My answer would not change because the effect would be the same.

2) Boosting

a)

b) show:

We start by rewriting the form of

-- identifying the pattern

--using the pattern

\*Note: If we multiply by and we arrive at the definition of .

Hence,

c)

i)

= 0

..

3. [attached photos]

4.2

a) – logistic regression.

– linear regression

b) i)

ii)

Logistic Linear

|  |  |  |  |
| --- | --- | --- | --- |
| Iteration | L2 Norm | Iteration | L2 Norm |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
| 5 |  |  |  |
| 6 |  |  |  |
| 7 |  |  |  |
| 8 |  |  |  |
| 9 |  |  |  |
| 10 |  |  |  |

iii)