Michael Lee

CSE 446

HW1

1. Let T mean “Tested Positive”

Let H mean “Actually have the disease”

* 1. a)

=

=

b)

In our case,

* 1. The multinomial distribution of three variables follows the function:

\*Log of multinomial

Where represent respectively and represent days clear, days cloud, and days rainy respectively. To account for the constraint , we use the method of LaGrange multipliers, where we optimize for a solution under constraint. Hence, we subtract the function to account for our original constraint and solve by taking partial derivatives.

Since

So,

Plugging in:

The actual estimates then are

3. 1) To estimate we need and

2) To estimate we need

3)

By symmetry of the definitions between x and y we attain the expression for y:

4)

=

] -factoring terms out of sum

Therefore, we have shown

Deriving a similar expression for

5) Two examples where it would be beneficial to do in online learning is when our data is extremely large and training over the data set would be infeasible. The second situation would arise when our data is dynamic and there is need to update our predictor routinely.

4.1 Lasso Regression

1.

1. The error on the training set will decrease because the focus of the algorithm will be to minimize the residual sum of squares on the training set.
2. The error on the testing set will be large because the model will over fit to the training set and lack prediction accuracy.
3. will have large coefficients because there is no penalty for the magnitude of coefficients.
4. All the elements of will be likely be non-zero.

2.

a) The error on the training set will be large because the focus of the algorithm will be to penalize the coefficients of our prediction. Hence, our coefficients will go to zero. Therefore, our predictions will be very poor when used on the data set, thus increasing the error.

b) The error on the test set will again be large as the predictor likely has all zero coefficients and our residual sum of squares calculation will be the square of each value in the data set as a result.

c) The coefficients of will be small, if not zero, because the function will penalize large coefficients.

d) There will be many non-zero elements of , if not all the elements.

4.2 Ridge Regression

3. When ridge regression pushes further away and when lasso pushes further away.

5 Programming Question

Code and plots are all part of the zip, each figure is labelled appropriately.

6. Selecting lamda

7. Max coefficient: 0.07741645, with label ‘PctIlleg’

Least coefficient: -0.064852304, with label ‘PctKids2Par’