Sean McGlincy

Machine Learning

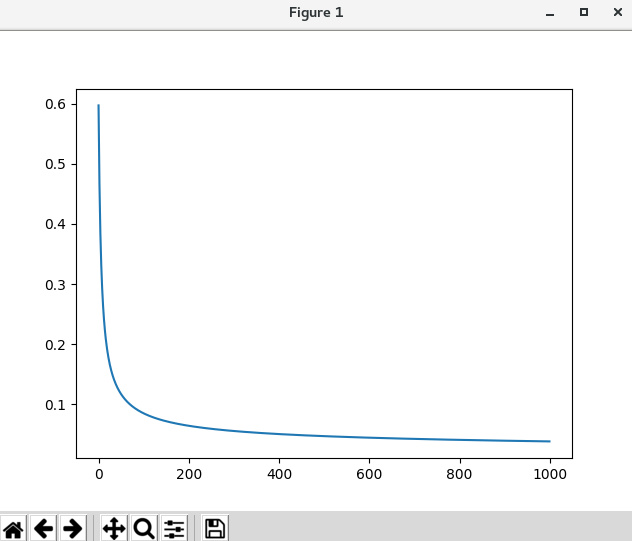
HW 3: Logistic Regression

For this assignment I imported the MNIST data set which consisted of pixel data with a label of 6 & 8. I used the sklearn to implement k-folding. The library provided an array of values that acted as pointers.

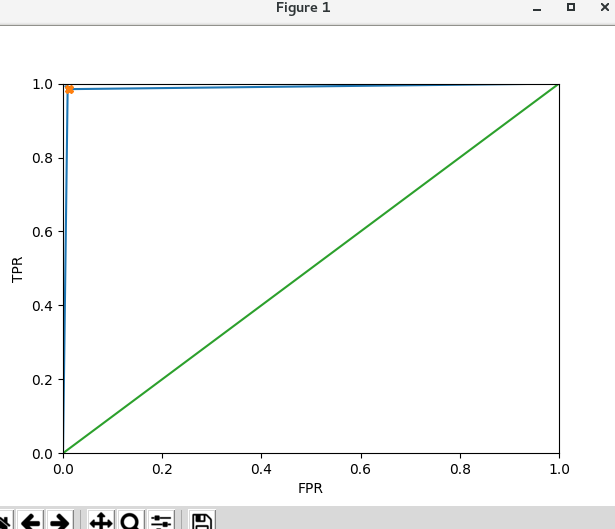
For each fold I created test data and training data arrays and separated the labels. Data and labels were normalized to treat the data for the gradient ascent algorithm. The gradient was calculated by dot(X.T , y – prediction). Predictions were calculated using dot product of weights and features, that were fed into the sigmoid functions. A cost curve was calculated using an equation which calculated the cost of labels 0 and 1 independently. The results were then added together, this assured that the equation produced a convex cost function.

The TPR and FPR were then calculated using the results and determining if the algorithms prediction was a true positive, false positive, true negative, false negative. The array was then mapped to produce a ROC curve. TPR and FPR were averaged and marked with an 'X' on the ROC curve.

Cost Curve



ROC Curve



TPR: [0.9898734177215189, 0.9875930521091811, 0.9848866498740554, 0.9879227053140096, 0.9803439803439803, 0.980440097799511, 0.9776674937965261, 0.9906976744186047, 0.9926108374384236, 0.9849624060150376]

FPR: [0.009411764705882352, 0.016786570743405275, 0.01182033096926714, 0.007389162561576354, 0.012106537530266344, 0.014598540145985401, 0.011990407673860911, 0.01282051282051282, 0.004830917874396135, 0.009501187648456057]

Avg TPR: 0.9856998314830848

Avg FPR: 0.011125593267360878