CSE 847 (Spring 2016): Machine Learning— Homework 4
Ding Wang

1 Logistic Regression: Experiment

I had implement my own logistic regression with MATLAB. I set up the maximum number of iterations and also convergence criterion. However, there is still one issue with matrix R. The diagonal element of R is $y_n(1 - y_n)$ which should be always positive. However, y_n might become 0 or 1 because of Matlab calculation precision. This will make H become a singular matrix. I resolve the issue by stop the iterations if this case happened.

My logistic regression was trained using the Newton-Raphson (IRLS) iterative procedure. I trained the classifier on the first n rows of the training data where $n \in [200, 500, 800, 1000, 1500, 2000]$. The accuracy on the test data as a function of n was shown in Figure 1.
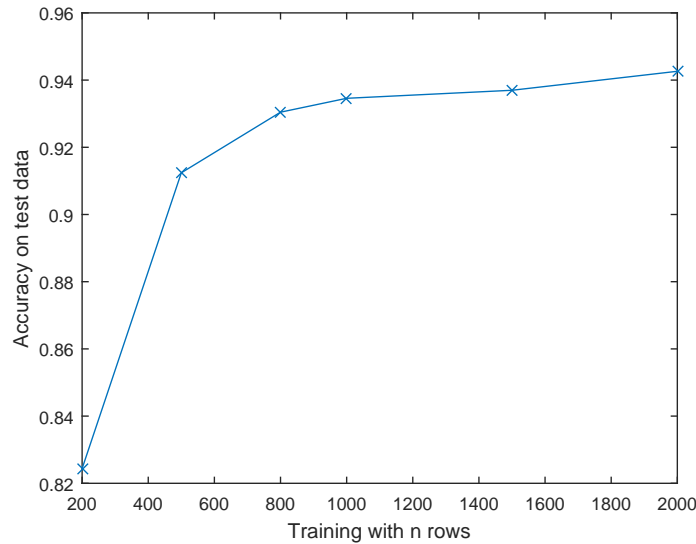

Figure 1: Accuracy on the test data when training logistic regression classifier on the first n rows of the training data.

It is clearly that with more training data, we can built more accurate model.

2 Sparse Logistic Regression: Experiment

I use SLEP to perform experiments using sparse logistic regression. I tried different values of regularization parameter of $[0, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$. Base on the note, I changed the criteria in line 73 of LogisticR to allow zero as the regularization coefficient. The AUC as a function of L1 parameter was shown in Figure 2. The number of features selected as a function of L1 parameter was shown in Figure 3.

Since AUC shows how good the classifier is, we can see the sparse logistic regression works best when we choose 0.1 as L1 parameter. The number of features selected keep decreasing when we increase the L1 parameter. Therefore, choosing a proper L1 parameter is very important when using sparse logistic regression.

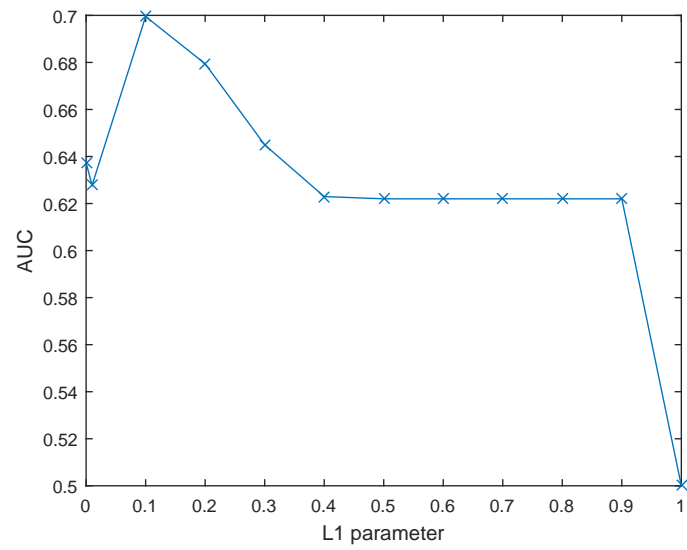Matlab code is uploaded to GitHub `https://github.com/cqwangding/CSE847`.
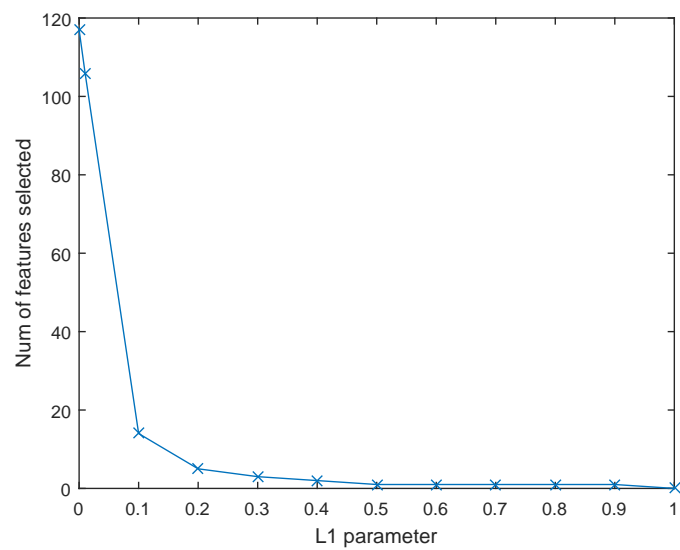
Figure 2: AUC with different L1 parameter.

Figure 3: Num of features selected with different L1 parameter.