Homework 4 Report

Peide Li

April 4, 2016

1 Logistic Regression: Experiment

In this experiment, I used Matlab to Apply logistic regression model to make classification. In my logistic_experiment function, the input is n, which denotes the size of the training set. That means I use first n rows of the training data to fit the model, and use the model to make prediction about the testing set. The function would return the weighted vector, the AUC value (used as accuracy of the model on the testing data set), the true positive rate and the false positive rate of the model when applying to the testing set. And when I apply Newton-Raphson iterative procedure in function logistic_ train, I use the update formula

$$w^{new} = (\Phi^T R \Phi)^{-1}\{R\Phi w^{old} - (Y - T)\}$$

The central part of the function is shown below:

```
while(abs(sum(w_new - w)) > epsilon & counts <= maxiter)
    w = w_new;
    z = phi * w;
    angle = zeros(1, length(z));
    for i = 1 : length(z)
        angle(i) = 1/ (1 + exp(-z(i)));
    end
    R = diag(angle);
    w_new = (phi' * R * phi)^(-1) * phi' * (R * phi * w - (z - t));
    counts = counts + 1;

end

weights = w_new;
```

Then, I change the value of n from 200 to 2000, and compute the value of AUC as the accuracy of the model. The relation between the Accuracy of the model and the training set size is shown in the Figure 1
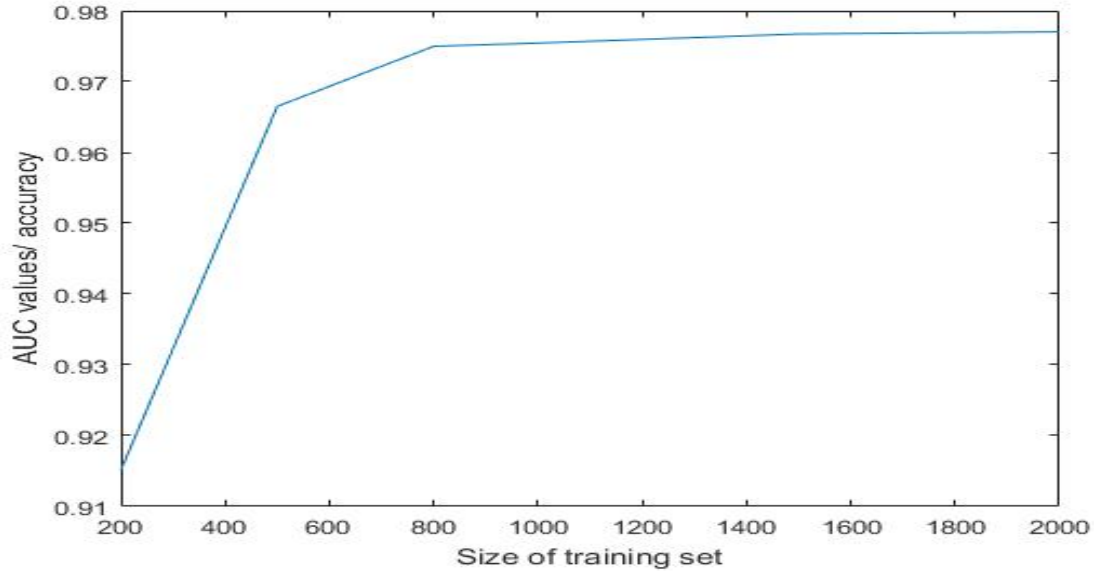
Figure 1: AUC against training set size

The figure shows that with larger training set, the prediction are more accurate.

2   Sparse Logistic Regression: Experiment

In this experiment, I use logisticR function to build sparse logistic regression model with l_1 regularization term. I wrote a function that can return the accuracy (AUC) of the model when applying to the prediction of the testing set. And I change different regularization parameters from 0 to 1 to check the tendency of accuracy of the models and the features of the models with different parameters. The Figure 2 shows the outcome:
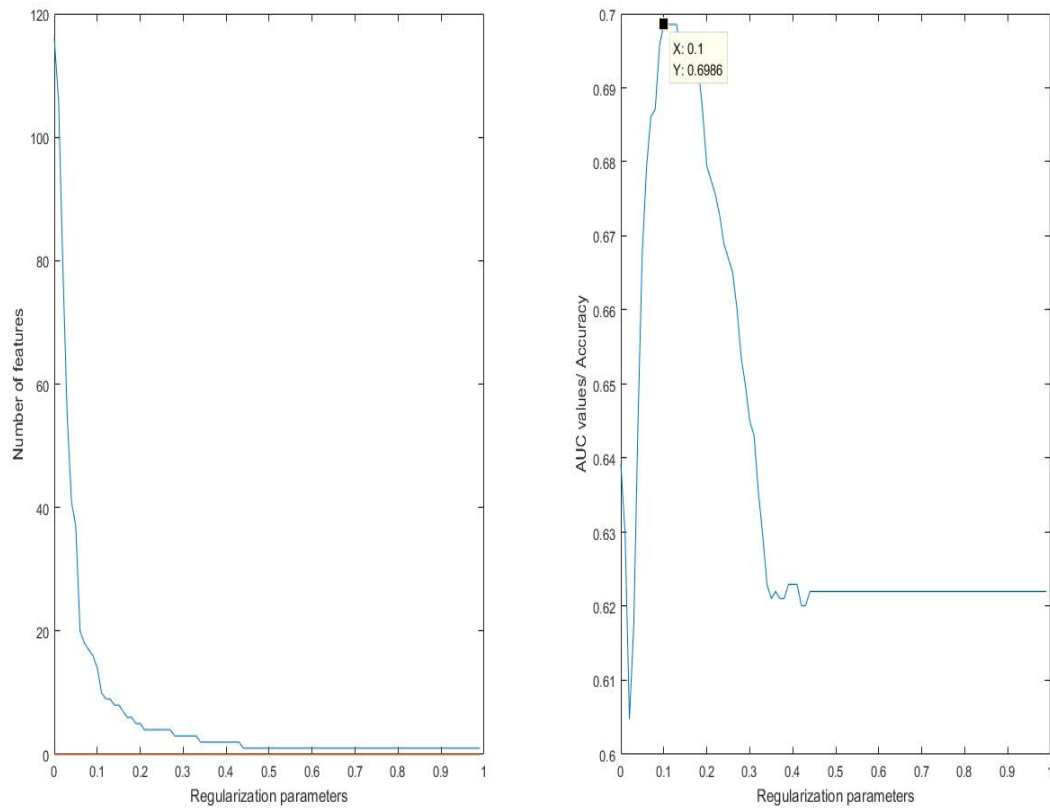
Figure 2: Accuracy & Features against parameters

The figure shows that the number of features is decreasing when the parameter is becoming larger. And when the parameter equal to about 0.1, the accuracy of the model is the largest.

The PDF report and the original Matlab code can be found at my github site: https://github.com/Kira233767/HomeWork-3.git