CS 434 Implementation Assignment 2 Report

Wenbo Hou Zhi Jiang Chongxian Chen

April 21, 2017

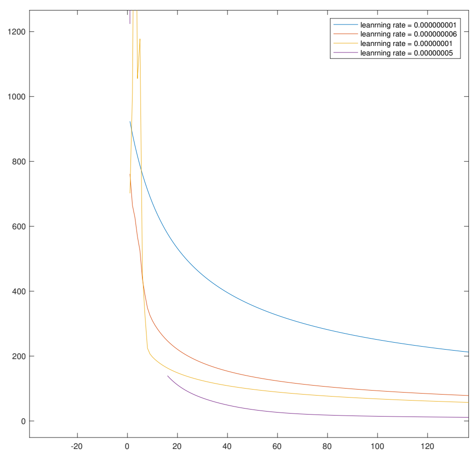
1 Introduction

In this assignment, we are required to train a logistic regression classifier with batch decent algorithm and L_2 regularization. First, we will determine a proper learning rate for the classifier. Second, we will use the learning rate to check the classifier's accuracy for both training data and testing data. Third, we will figure out the new gradient with L_2 regularization. In the end, we will use the new gradient and the found learning rate to rebuild the classifier and explore the relationship between the classifier's accuracy and different λ values.

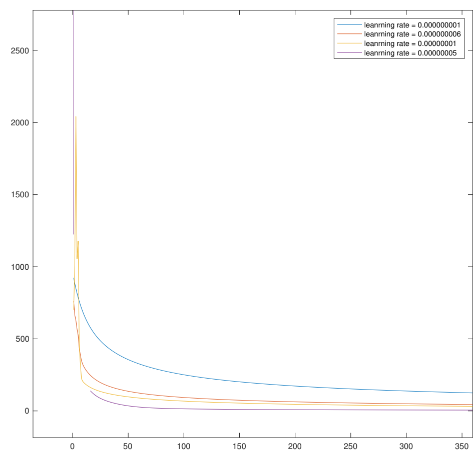
In this assignment, we did each problem separately. Each problems, except the third one, is a MATLAB script.

2 Problem 1

Following the pseudo-code from lecture slides, we first built our training program with the gradient of the loss function. The next step was to find a proper learning rate for the program. In this case, we iterated the learning process for 1000 times and ploted the loss for each iteration. If the learning rate is good for our classifier, the plot will be decent and eventually converge at low total loss. The first learning rate we guessed was 10^{-1}, but the loss got overflow. Then, we kept decreasing the exponent by 1 till the loss function got all real numbers during the iteration. Finally, we ended with 10^{-7}. To smooth the plot, we gently decreased the learning rate and found that 5×10^{-8} was the best one.



(a) Zoomed in plot to show the decent rate



(b) Zoomed in plot to show the convergence

Figure 1: The plot show how how the loss changes when the iteration goes

Above figures shows that the loss function will always converge no matter what the learning rate is. What the learning rate does to the loss function is the decent rate before convergence. From 1.a, we can see that large learning rate (6×10^{-7} and 10^{-7}) could lead to overflow (a gap on the plot) or jumps. Too small learning rate (10^{-8}) cannot achieve the convergence after the learning process. Thus, **we chose** 5×10^{-8} **as the learning rate for our program that has 2000 learning iterations**.

3 Problem 2

The second problem requires us to explore the classifier's accuracy for each iteration. As our professor stated, we consider $P(y^i = 1|X) >= 0.5$ as positive and $P(y^i = 1|X) < 0.5$ as negative. Then, we implemented our weight vector for both train data and test data. The following plot shows how the model's accuracy changes as the iteration goes.

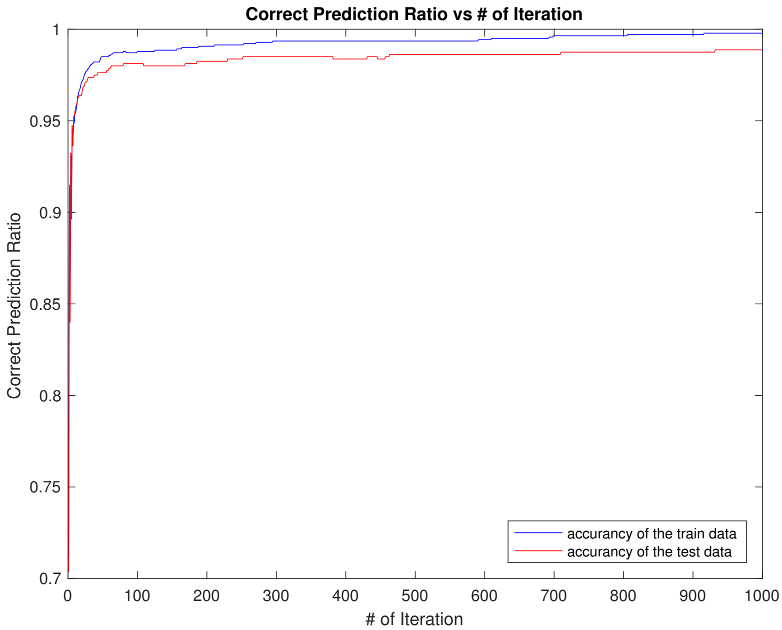


Figure 2: Classifier's accuracy changes for each iteration

We could concluded that the classifier's accuracy get converged. The classifier's accuracy for the training data approaches to 1.0. The classifier's accuracy from the testing data become stable around 0.97. The observation of loss in problem can prove our observation is reasonable.

As the iteration goes, the classifier gets more and more features from the training data and become more and more fit the training data. However, this fact has little effect on the classifier's accuracy for the testing data. We generate two hypothesis for this observation. First, the learning rate is very small, and unique features from training set cannot affect the classifier very much. Second, the tolerance ($P(y^i = 1|X) = 0.5$) of the classifier is large, which makes the classifier ignore some unique features.

4 Problem 3

The problem 3 asks us to implement L_2 regularization to the logistic regression classifier. The only thing we need to do is to find the derivative of the regularization term and add it to old gradient. Here is the new gradient:

$$\nabla l(g(W^T X^i), y^i) = -(y^i - g(W^T X^i))X^i + \lambda W \tag{1}$$

Here is the new pseudo-code:

Given: traninning examples $(X^i, y^i), i = 1, 2, 3..N$
Let $W \leftarrow (0, 0, ..., 0)$
loop
  if Convergence **then** Stop
  Let $d \leftarrow (0, 0, ..., 0)$
  for i = 1 to N **do**
    $error = y^i - \frac{1}{1+e^{-WX^i}}$
    $d = d + error \cdot X^i$
  Let $W \leftarrow W + \eta(d + \lambda W)$

5 Problem 4

Problem requires us to explore the relationships between the classifier's accuracy and λ values. The following plot shows that larger λ makes the convergence much better in the situation where the learning iteration is fixed. The larger λ value let the classifier keep more features from previous learning iteration, which allows the learning process take advantages from last learning and become stable. Especially, the stabilizing process become dominated when the weight vector starts to converge.

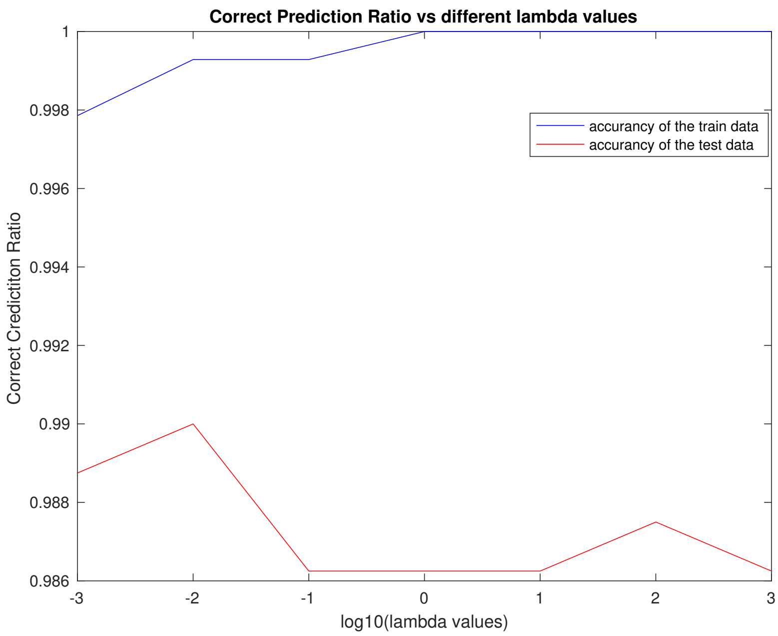


Figure 3: Classifier's accuracy changes vs log10(lambda)