Name: Hieu Khuong

TUID: 915399644

Programming Assignment 3

An English description of your algorithms:

In this programming assignment I implemented 2 machine learning algorithms: Logistic Regression and Soft-Margin SVM.

Logistic Regression: is a linear model trained with the logistic loss without the regularization term. The logistic loss will be computed by this formula:

$$\text{logistic\_loss} = log(1 + e^{-innerProduct})$$

where innerProduct is the inner product of the label vectors of the training data set and the test data set.

Soft-margin SVM: is (equivalent to) a linear model trained with the hinge loss with the l2 regularization term. The hinge loss will be computed by this formula:

$$\text{hinge\_loss} = max(0, 1 - innerProduct)$$

where innerProduct is the inner product of the label vectors of the training data set and the test data set. Also, the l2 regularization will be computed by this:

$$\text{l2\_reg} = ||\text{w}|| = w^T w$$

In **both** algorithms, we will need to calculate the loss of the prediction using weight w:

$$L(w) = \lambda R(w) + loss(w)$$

With this loss value, we will update the weight w using gradient descent:

$$\Delta w_j = \frac{\delta L(w)}{\delta w_j} = \frac{L(w_0, w_1, .., w_j + h, ...) - L(w_0, ...)}{h}$$

(where h is a very small number to use for numerical differentiation).

Then we will update the weight:

$$w = w - \eta \, \Delta w$$

(where η is the learning rate)

All the training and testing data will be normalized by using: (dataX - mean)/std

Normalizing data will help the training model to be more accurate by turning large data point to have less impact on the prediction.

Discussion of the comparison of the different classifiers on this problem.

Difference between 2 classifier:

Logistic Regression: use logistic loss as training criteria, without regularization.

Accuracy for Logistic Regression with different learning rates:

Learn rate 0.1: Accuracy 0.6074074074074074

Learn rate 0.01: Accuracy 0.4829629629629629

Learn rate 0.001: Accuracy 0.5996296296296297

Learn rate 0.0001: Accuracy 0.5907407407407408

Learn rate 0.00001: Accuracy 0.6255555555555555

So the highest accuracy that Logistic Regression reached is 62.55%.

Soft-margin SVM: use hinge loss as training criteria along with L2 regularization.

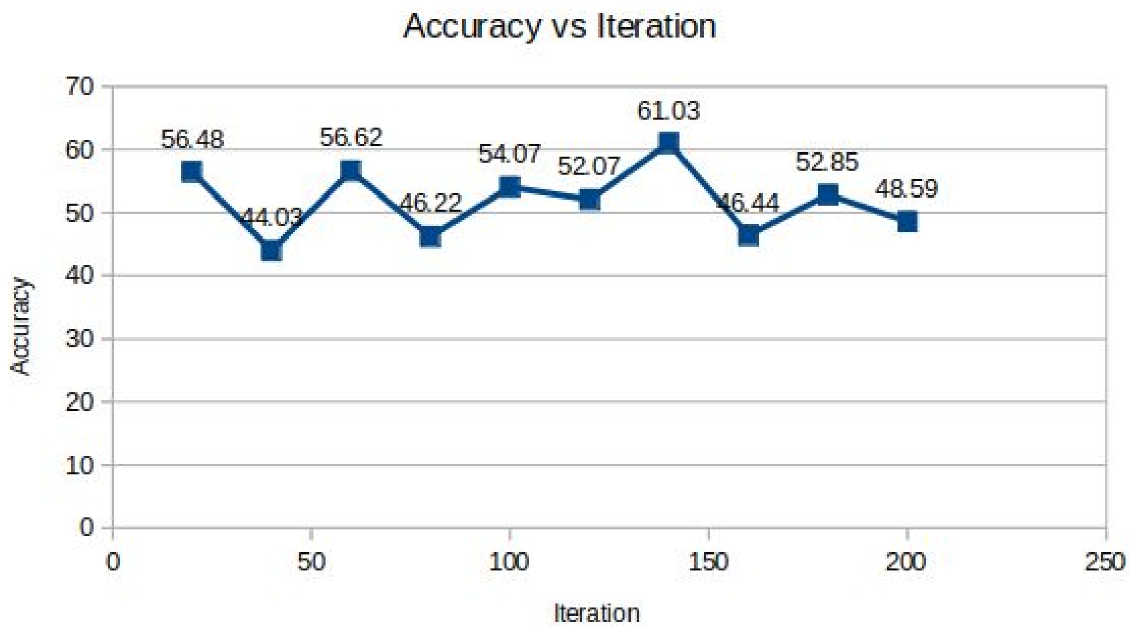
Accuracy for Soft-margin SVM with different learning rates and lambda values:

Learn rate: 0.00001	Lambda: 0.1: Accuracy 0.6037037037037036
	Lambda: 0.01: Accuracy 0.5981481481481482
	Lambda: 0.001: Accuracy 0.6018518518518519
Learn rate: 0.0001	Lambda: 0.1: Accuracy 0.5874074074074074
	Lambda: 0.01: Accuracy 0.5933333333333334
	Lambda: 0.001: Accuracy 0.5892592592592594
Learn rate: 0.001	Lambda: 0.1: Accuracy 0.5781481481481483
	Lambda: 0.01: Accuracy 0.6114814814814815
	Lambda: 0.001: Accuracy 0.5707407407407408

So the highest accuracy that Softmargin SVM reached is 61.14%.

In conclusion, Logistic Regression is slightly better than Soft-margin SVM.

<u>**For one experiment, include a plot of the loss as a function of the number of iterations:**</u>



For this graph, I chose numbers of iteration to be 20,40,60,80,100,120,140,160,180,200} and I used logistic regression with learning rate being 0.01. We can see that the highest accuracy reached is 61.03% with number of iterations being 140.