PROJECT REPORT: C S 487/519 Applied Machine Learning I

Project 1: Single-layer Linear Neural Networks

Written by

Omid Jafari

Department of Computer Science

New Mexico State University

September 2018

Accuracy of the prediction for different datasets:

We used a randomly chosen 75% of the data as the training set and the rest as the testing set.

Parameters			
	Eta	Iters	Seed
Iris	0.001	20	1
Winequality-red	0.0001	20	1

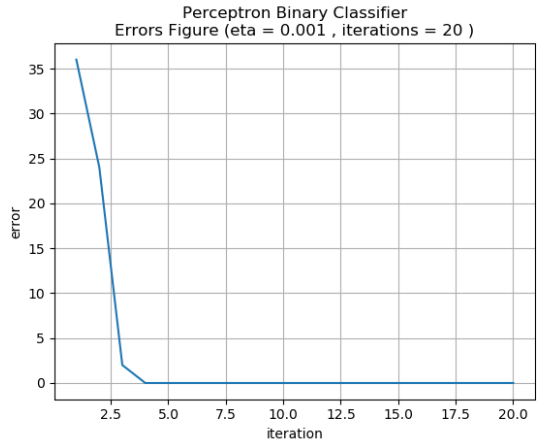
Accuracy		
	Iris	Winequality-red
Perceptron	100%	61.5%
Adaline	63.16%	55.75%
SGD	100%	57.5%
Multi-class SGD	100%	98.75%

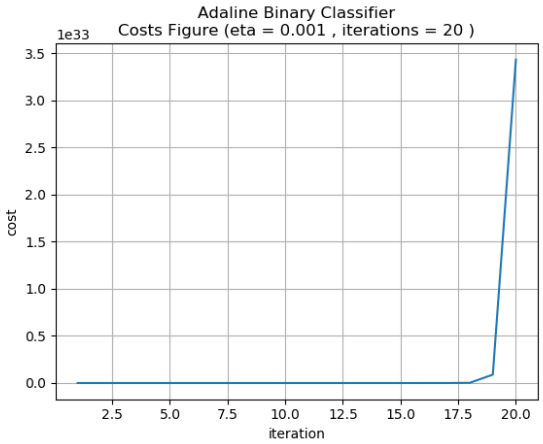
Errors/Costs plot for each iteration:

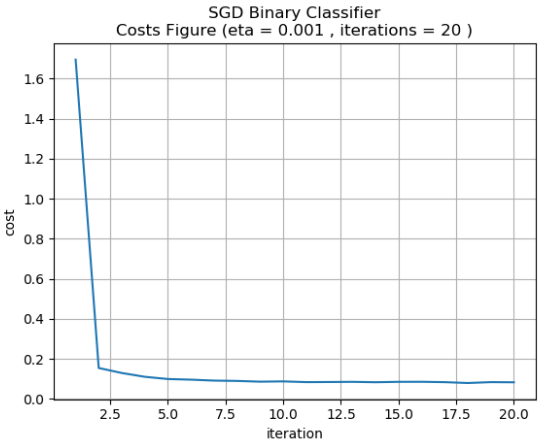
We used a randomly chosen 75% of the data as the training set and the rest as the testing set.

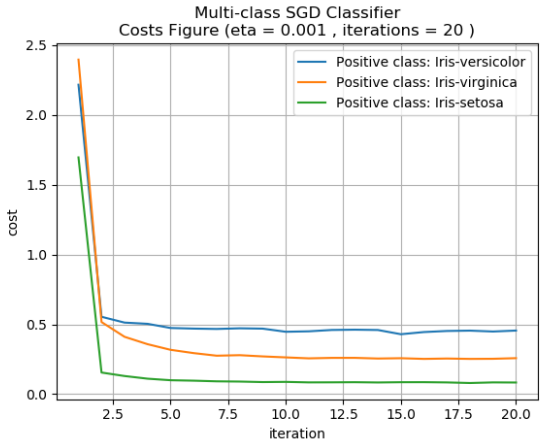
Parameters			
	Eta	Iters	Seed
Iris	0.001	20	1
Winequality-red	0.0001	20	1

Iris data set:

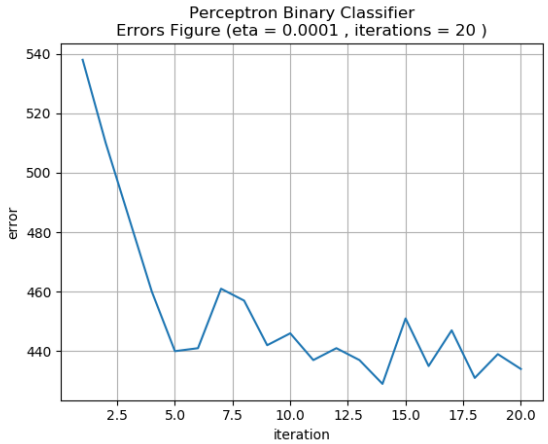


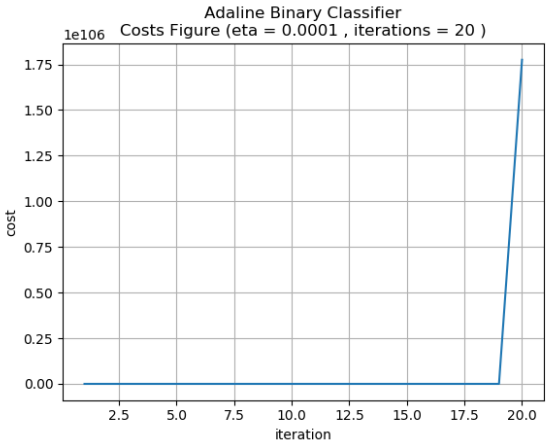


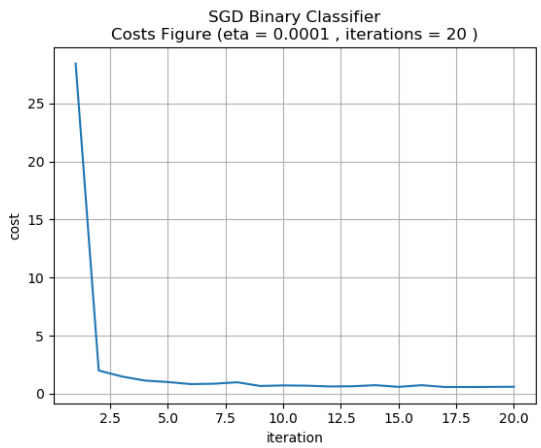


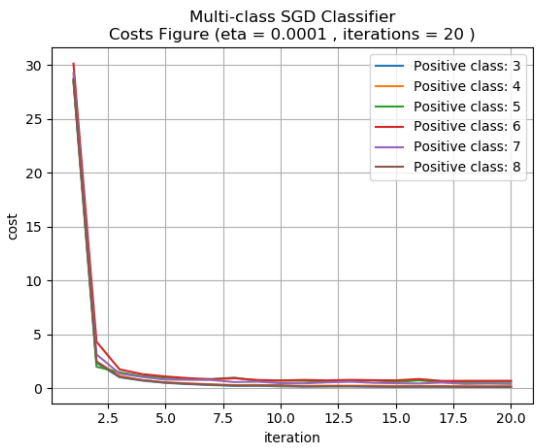


Wine data set:

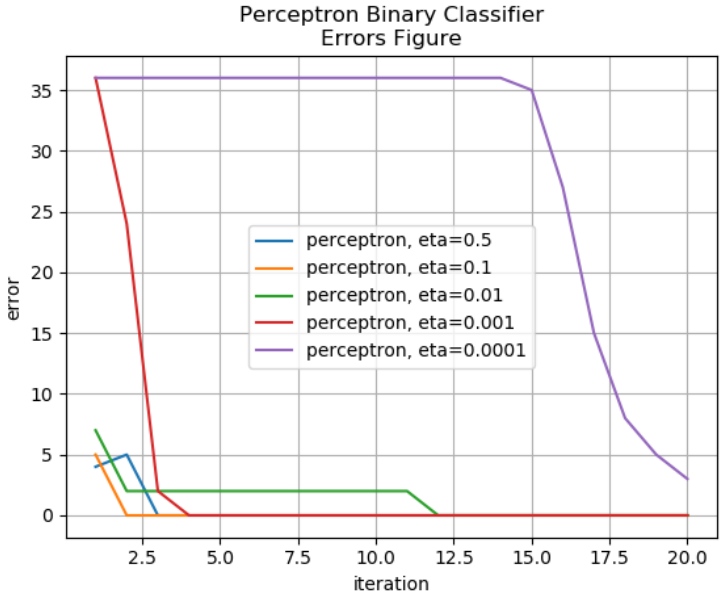






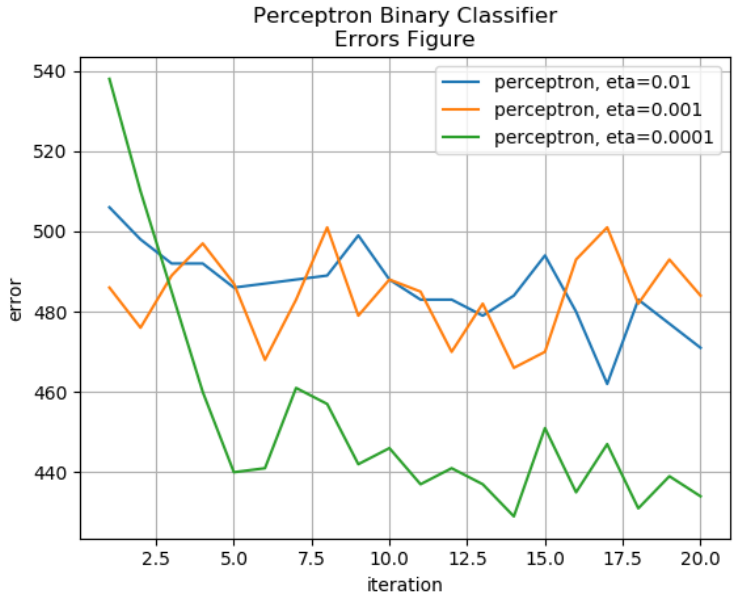


Perceptron analysis - Iris:



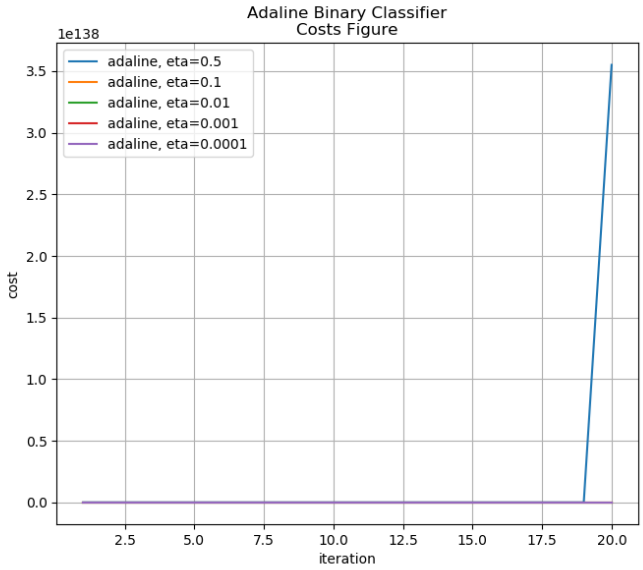
As shown in the picture, choosing the appropriate value for Eta is important for the Perceptron classifier to work properly. At first, it seems that choosing a smaller value would result in a faster converging; however, the purple line proves that if Eta becomes too small, the classifier will converge slower. Therefore, for this data set (iris), 0.1 is the best value for Eta.

Perceptron analysis - Wine:



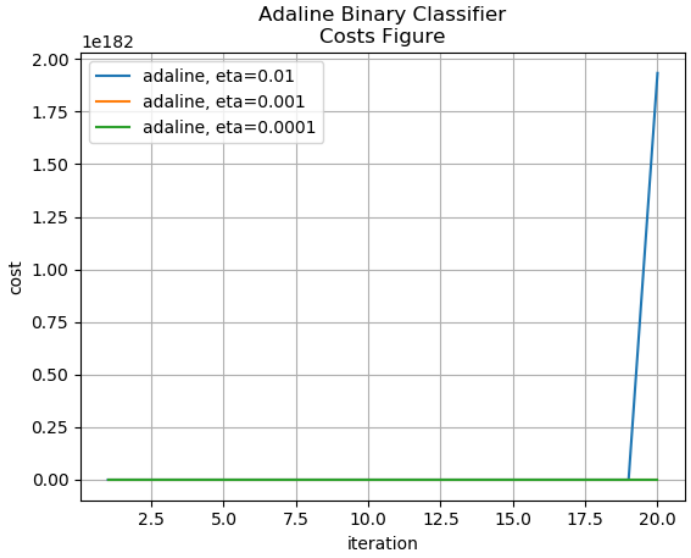
Since the Wine data set is much larger than the Iris data set, we are getting larger errors, and hence, we need more number of iterations to reach zero. However, from the above picture, it is clear that eta=0.0001 is reaching zero sooner than the other values. Moreover, the classifier never converges for Eta values more than 0.01.

Adaline analysis - Iris:



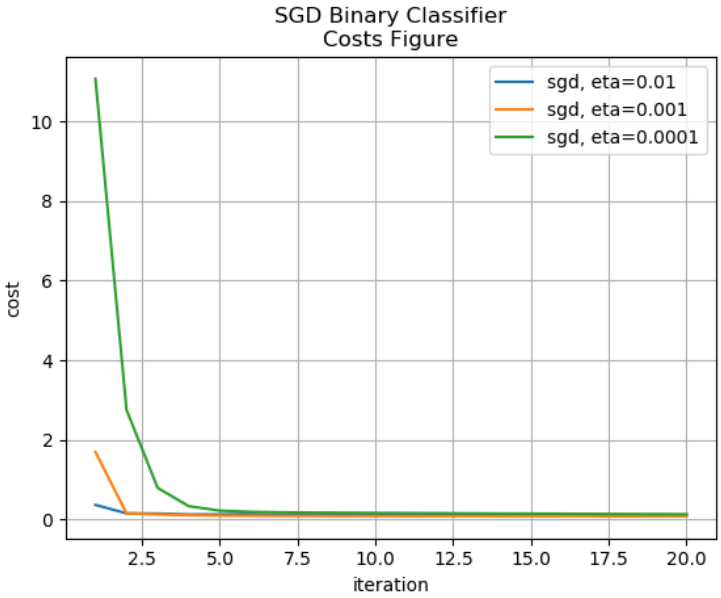
For the Adaline classifier, the costs are always zero for the small values of Eta. However, choosing a larger value will result the costs to go higher after a certain amount of iterations.

Adaline analysis - Wine:



In this experiment, we got the same results as the Iris dataset. Therefore, the same analysis holds true.

SGD analysis - Iris:

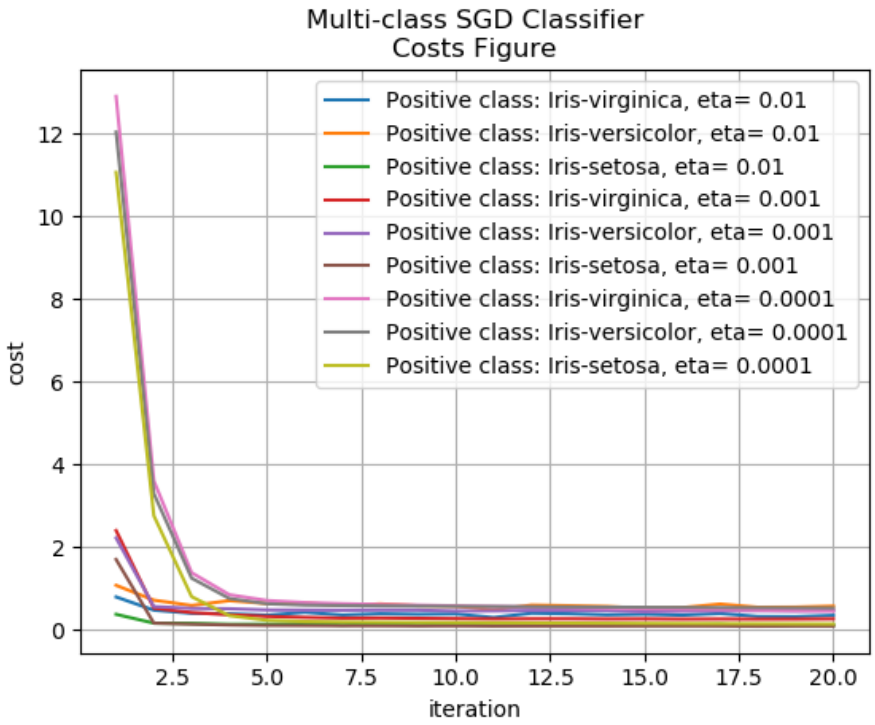


SGD classifier never converges for Eta values greater than 0.01. For the iris data set, choosing 0.01 is the best option for Eta as it minimizes the cost as soon as possible.

SGD analysis - Wine:

SGD classifier for the Wine data set just converged with eta=0.0001. Same analysis applies.

Multi-class analysis - Iris:



Since the underlying classifier for our multi-class analysis is SGD, the same logic applies in here too. Eta value of 0.01 is the best option as it minimizes the cost as soon as possible.

Multi-class analysis - Wine:

Only converged for eta=0.0001. Same analysis applies.

Analysis on other aspects:

Since the datasets given were balanced, we never tried analyzing the results based on the different training and testing sets. However, for the unbalanced datasets it is better to use a method called Stratifying to wisely choose training and testing sets. Moreover, accuracy is not always the best metric to choose the classifiers. As we may have false positives and false negatives, it is better to differentiate them by using other metrics such as precision, recall, or other methods.

The other important metric that we missed is the execution time. We need to investigate the execution time based on the size of the datasets and using different classifiers.

For all the datasets we used, feature scaling and normalization was not important. However, if the effect of attributes is different in the class label, we should use these methods to get more accurate results.