CS446: Machine Learning	**Fall 2015**
Problem Set 3	
Boyu Li	*Handed In: October 8, 2015*

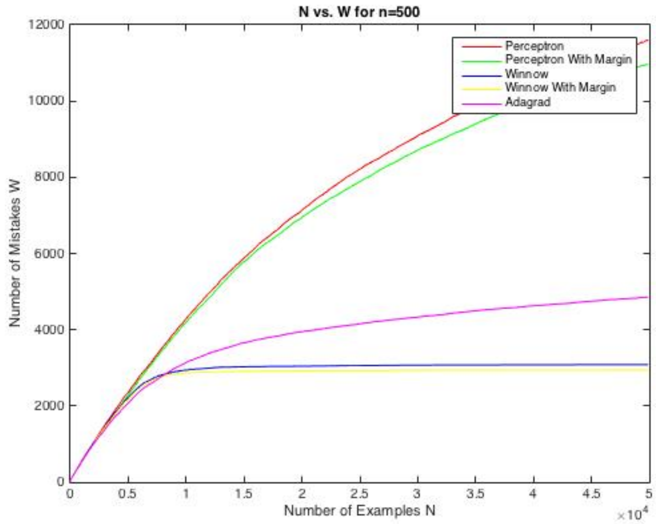
1. Number of Examples Versus Number of Mistakes

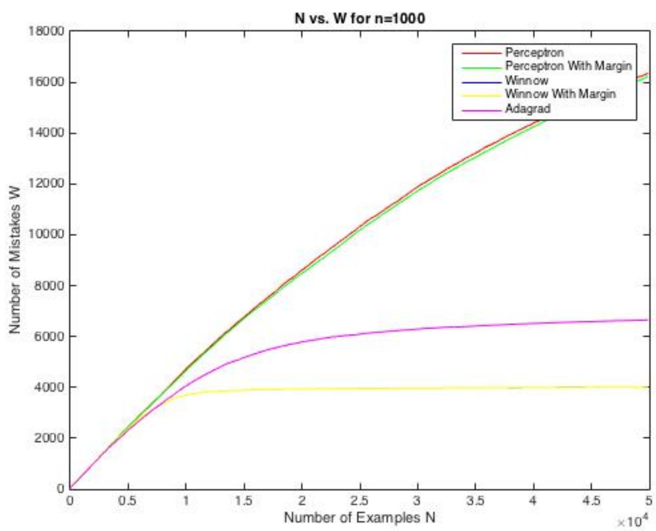
Algorithm	Parameters	Data Set	
		$n = 500$	$n = 1000$
Perceptron w/margin	η	0.05	0.03
Winnow	α	1.1	1.1
Winnow w/margin	α	1.1	1.1
	γ	2.0	0.04
AdaGrad	η	0.25	0.25

For this problem, I have 5 files that for the classify functions themselves, they are perceptron.m, perceptron_margin.m, winnow.m, winnow_margin.m, adagrad.m and other 5 files that count the mistakes, they are perceptron_mistake.m, perceptron_margin_mistake.m, winnow_mistake.m, winnow_margin_mistake.m and adagrad_mistake.m.
And I write a function in calaccuracy.m to calculate the accuracy to help me determine the best parameter. To test this part, I just track the max accuracy and corresponding parameter to different data size.And I write down the result in the file called *result.txt*
The data shows below
Where 1$\rightarrow$ Perceptron w/margin, 2$\rightarrow$winnow, 3$\rightarrow$ Winnow w/margin, 4$\rightarrow$AdaGrad

Algorithm	n	max accuracy	best parameter for rate	best parameter for margin
1	500	98.220 000	0.05	
1	1000	96.260 000	0.03	
2	500	99.680 000	1.1	
2	1000	99.720 000	1.1	
3	500	99.900 000	1.1	2.0
3	1000	99.720 000	1.1	0.04
4	500	98.240 000	0.25	
4	1000	99.300 000	0.25	

For this part, I separate the data set into small size one which each interval contain 100 samples, and plot the graph, this way makes my graph looks more smoothly.



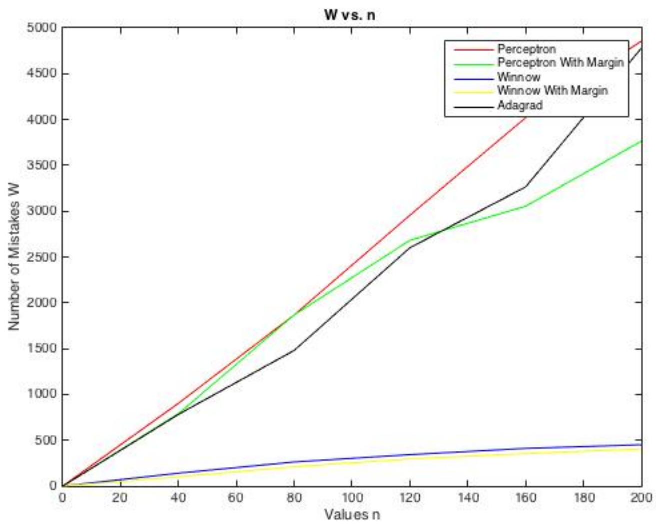


This part almost need 30-40 minutes to finish the run,it is really a long time.

2. Learning Curves of Online Learning Algorithms

   The basic algorithm to determine the best parameter is same with previous one by using max accuracy.
   The difference is that we set a converge R=1000 in our function. If the number reach the convergence,we just stop.
   I run the this part in test2.m and store the result in result1.txt file.

Algorithm	Parameters	Data Set				
		$n = 40$	$n = 80$	$n = 120$	$n = 160$	$n = 200$
Perceptron w/margin	η	0.25	1.5	0.25	0.03	0.03
Winnow	α	1.1	1.1	1.1	1.1	1.1
Winnow w/margin	α	1.1	1.1	1.1	1.1	1.1
	γ	2.0	2.0	2.0	2.0	2.0
AdaGrad	η	1.5	1.5	1.5	1.5	1.5



In this graph, it is obviously that the number of winnow and winnow with margin grow stable as the value of n grow and almost coverage at the end. The perceptron, perceptron with margin and adagrad algorithm seems make more mistakes. At the beginning,the perceptron and perceptron with margin seems grow with the same rate but at the some point for n, the rate for perceptron with margin decrease, and the adagrad grow almost same rate with perceptron.

The data for this test shows below
Where 1→ Perceptron w/margin, 2→winnow, 3→ Winnow w/margin, 4→AdaGrad

Algorithm	n	max accuracy	best parameter for rate	best parameter for margin
1	40	100.000 000	0.25	
1	80	100.000 000	1.5	
1	120	100.000 000	0.25	
1	160	100.000 000	0.03	
1	200	99.800 000	0.03	
2	40	100.000 000	1.1	
2	80	100.000 000	1.1	
2	120	100.000 000	1.1	
2	160	100.000 000	1.1	
2	200	99.980 000	1.1	
3	40	100.000 000	1.1	2.0
3	80	100.000 000	1.1	2.0
3	120	100.000 000	1.1	2.0
3	160	100.000 000	1.1	2.0
3	200	100.000 000	1.1	2.0
4	40	100.000 000	1.5	
4	80	100.000 000	1.5	
4	120	100.000 000	1.5	
4	160	99.360 000	1.5	
4	200	97.420 000	1.5	

3. Use Online Learning Algorithms As Batch Learning Algorithms

   For this part, I use the similar method to get the best parameter and I finish this test in test3 and store the result in result2.txt.

Algorithm	Parm & Accy	Data Set		
		$m = 100$	$m = 500$	$m = 1000$
Perceptron	Accy %	80.140000	58.080000	69.180000
Perceptron w/margin	η	0.005000	0.005000	0.030000
	Accy%	81.400000	61.900000	71.180000
Winnow	α	1.1	1.1	1.1
	Accy%	88.700000	76.980000	72.860000
Winnow w/margin	α	1.1	1.1	1.1
	γ	0.006000	0.040000	0.001000
	Accy%	88.780000	79.020000	73.260000
AdaGrad	η	0.25	1.5	1.5
	Accy%	87.520000	75.440000	72.840000

In general, we see that as m increases, the accuracy decreases.
We can calculate the average accuracy for different across the difference m, and we can rank it in decreasing which we will get: Adagrad, winnow with margin, winnow, perceptron with margin and perceptron. And the average for winnow and winnow with margin is pretty close,so we may think it is really stable and the perceptron is much worse that perceptron with margin.

4. Bonus

Since in this problem we will have unbalanced data, so what i did is the change the margin based on the ratio.If the negative data occupy more we just lower the margin for it corresponding ratio with positive example.
The classify function is in perceptron_margin_modified.m and the test file is in test_bouns.m. and the result is stored in bouns.txt
The result shows below where 1 $\rightarrow$ perceptron_modified, and 2 $\rightarrow$ perceptron_margin.

Algorithm	m	max accuracy	best parameter for rate
1	100	99.630 000	0.03
1	500	99.450 000	0.001
1	1000	99.390 000	1.5
2	100	99.860 000	0.03
2	500	90.000 000	0.001
2	1000	99.590 000	0.25