CS446: Machine Learning **Fall 2014**

Problem Set 3

Chen Zhang *Handed In: 10/11/2014*

1 Problem 1

In this problem, I implemented my code as required. I looped over all the parameter combinations and tuned the learning algorithm, and found best parameters and use them in the training process. The main result is the Table. 1 and Fig. 1 and Fig. 2. Depending on different l m n combination, the best parameter after tuning is also different.
From Fig.1 we can see that when $n = 500$, algorithms with margin behaves better than cases without margin(perceptron and winnow). But the advantage is not very obvious. Ada Grad, Winnow, Winnow with margin have a better behavior on the number of mistakes it make as number of example increases. These three algorithms tend to converge to a certain value, while perceptron and perceptron with margin does not have a very clear convergence to some value.
When $n = 1000$, we can see that the margin does not have a obvious effect on the behavior of the algorithms. The number of mistakes all the algorithms make are larger than the case of $n = 500$. But the convergence property of winnow, winnow with margin and ada grad still holds, while perceptron and perceptron still does not have a clear convergence.

Algorithm	Parameters	Data Set	
		$n = 500$	$n = 1000$
Perceptron w/margin	η	0.03	0.005
Winnow	α	1.1	1.1
Winnow w/margin	α	1.1	1.1
	γ	2	0.006
AdaGrad	η	0.25	0.25

Table 1: Problem 1 Table for the best Parameters

2 Problem 2

In this problem, I implemented my code as required. I looped over all the parameter combinations and tuned the learning algorithm, and found best parameters and use them in the training process, as is shown in Table. 2. I then used $R = 1000$ to calculate the number of mistakes it made before reaching the criteria.
From Fig. 3 we can see that winnow and winnow with margin has the best behavior. Perceptron has the largest number of mistakes before reaching the criteria. Perceptron with

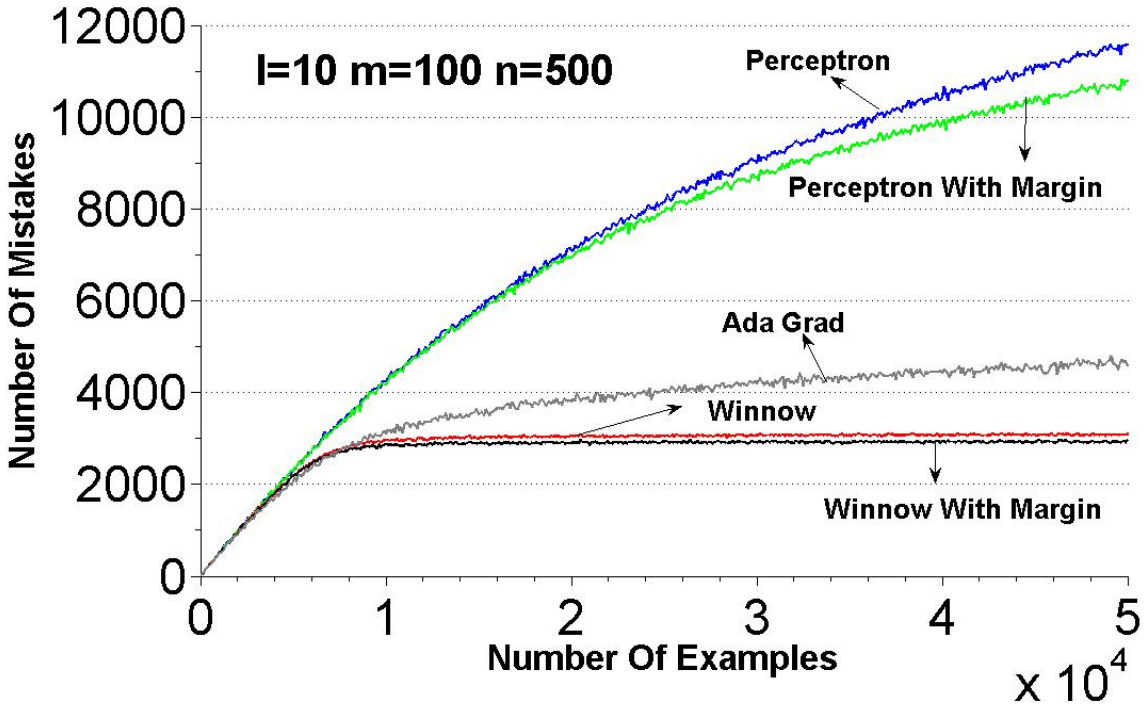

Figure 1: First Figure of Problem 1

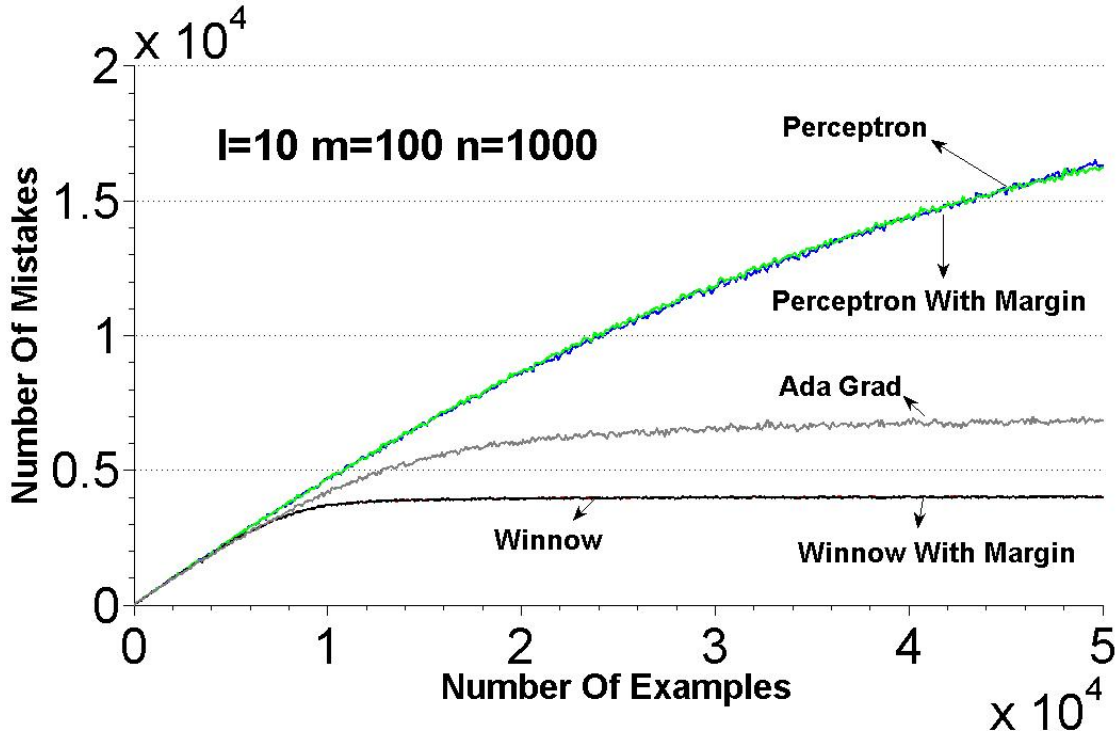

Figure 2: Second Figure of Problem 2

margin has better behavior than ada grad when n is small but worse behavior when n is large. Similar to the Fig. 1 and Fig. 2, winnow and winnow with margin also have a convergence trend, while the other three algorithms does not have a clear convergence evident.

Algorithm	Parameters	Data Set				
		$n = 40$	$n = 80$	$n = 120$	$n = 160$	$n = 200$
Perceptron w/margin	η	1.5	0.25	0.03	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	2	0.3	2	2
AdaGrad	η	1.5	1.5	1.5	1.5	1.5

Table 2: Problem 2 Table for the best Parameters

3 Problem 3

In this problem, I implemented my code as required. I generated data with noise and looped over all the parameter combinations and tuned the learning algorithm, and found best parameters and use them in the training process, as is shown in Table. 3. In the table, we could see that perceptron with margin and ada grad tend to behave the best, while in the problem 2, winnow and winnow with margin tend to behave best. This shows than when the data has noise, linear increment with margin would be the best algorithm to solve the problem

4 Bonus Problem

In this problem, I implemented my code as required. I generated unbalanced data, use perceptron for the three cases to calculate the accuracy. Then I tuned perceptron with margin, and used the best case to calculate the accuracy on the unbalanced data. The result is shown in the Table. 4.
We see that perceptron and perceptron with margin has extremely good behavior on the unbalanced data set. This shows that perceptron has extremely robust behavior on linearly separable data set.

Algorithm	Parm & Accy	Data Set		
		$m = 100$	$m = 500$	$m = 1000$
Perceptron	Accy %	91.19	88.99	64.01
Perceptron w/margin	η	0.03	1.5	0.25
	Accy%	98.34	88.99	64.8
Winnow	α	1.1	1.1	1.1
	Accy%	91.62	84.76	69.37
Winnow w/margin	α	1.1	1.1	1.1
	γ	0.006	0.3	0.04
	Accy%	96.54	88.74	69.9
AdaGrad	η	0.25	0.25	1.5
	Accy%	99.96	96.78	76.05

Table 3: Problem 3 Table for the best Parameters

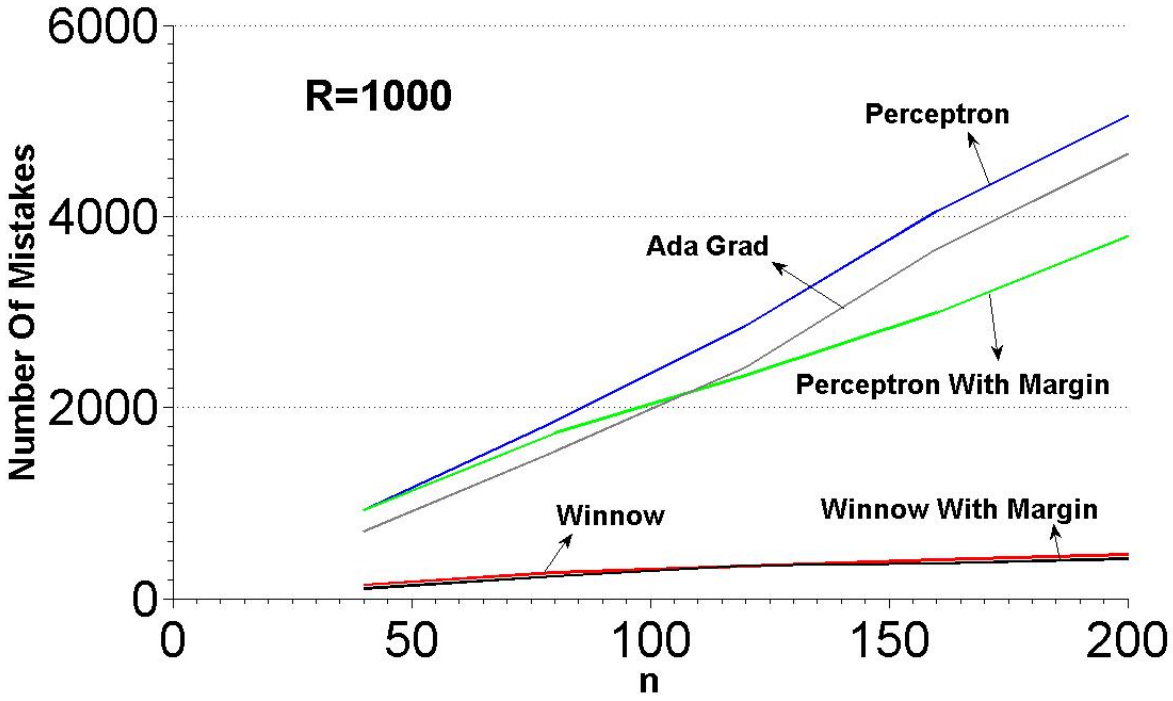

Figure 3: First Figure of Problem 2

Algorithm	Parm & Accy	Data Set		
		$m = 100$	$m = 500$	$m = 1000$
Perceptron	Accy %	99.85	99.79	99.98
Perceptron w/margin	η	0.03	0.03	0.03
	Accy%	99.53	99.98	100

Table 4: Bonus Problem Table for the best Parameters