CS446: Machine Learning	**Fall 2016**

Problem Set 1

Shengliang Dai (sdai7) *Handed In: October 8, 2016*

The code and results are saved in hw3.ipynb, hw3.2.ipynb, hw3.3.ipynb, bonus.ipynb.

1. **[20 points] Number of examples versus number of mistakes**

Algorithm	Parameters	Dataset n=500	Dataset n=1000
Perceptron	NA	NA	NA
Perceptron w/margin	η	0.03	0.03
Winnow	α	1.1	1.1
Winnow w/margin	α, γ	1.1, 0.01	1.1, 2.0
AdaGrad	η	1.5	1.5

The plots of the cumulative mistakes are given in Fig.1 Fig.2. The number of mistakes of Perceptron, Perceptron with margin and adagrad go up linearly with n. It seems that the Winnows grow logarithmic as a function of n.

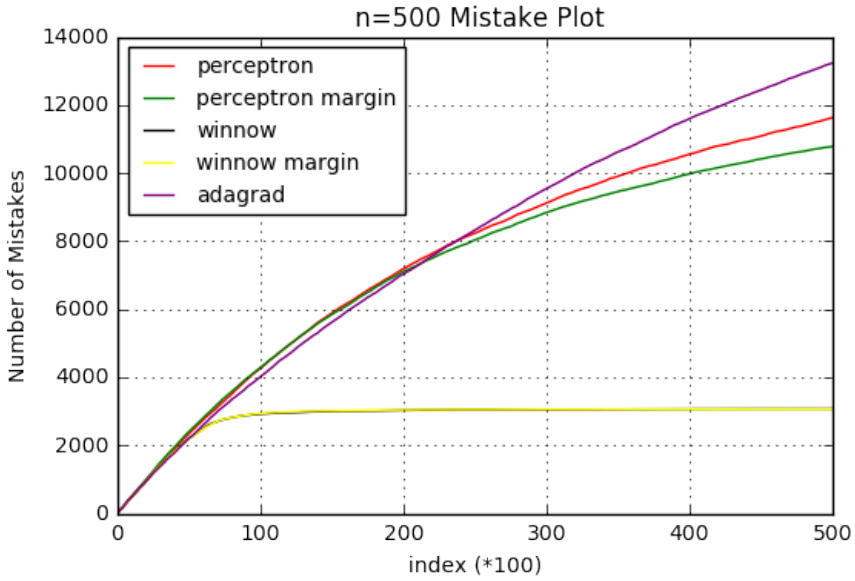

Figure 1: 3.1

Algorithm	Parameters	n=40	n=80	n=120	n=160	n=200
Perceptron	NA	NA	NA	NA	NA	NA
Perceptron w/margin	η	1.5	1.5	0.005	1.5	0.25
Winnow	α	1.1	1.1	1.1	1.1	1.1
Winnow w/margin	α, γ	1.1, 2	1.1, 2	1.1, 2	1.1, 2	1.1, 2
AdaGrad	η	1.5	1.5	1.5	1.5	1.5

Algorithm	m=100		m=500		m=1000	
	acc.	params.	acc.	params.	acc.	params.
Perceptron	98.05	NA	90.79	NA	75.38	NA
Perceptron w/margin	98.05	1.5	84.19	0.05	82.64	0.25
Winnow	94.8	1.1	87.75	1.1	72.74	1.1
Winnow w/margin	95.16	$\alpha = 1.1, \gamma = 0.3$	90.84	1.1, 0.01	71.46	1.1, 0.006
AdaGrad	93.23	1.5	77.16	1.5	82.51	1.5

2. **[35 points] Learning curves of online learning algorithms**

   See Fig. 3. As n grows, the mistakes always grows for all algorithm. Winnow and Winnow with margin always make the fewest errors before convergence. The Winnow with margin make fewer mistakes than Winnow, but the two are pretty close. Adagrad can perform better or worse than perceptron, but it always perform worse than winnow. Perceptron and perceptron with margin are very close.

3. **[45 points] Use online learning algorithms as batch learning algorithms**

   The best parameters are not so sensitive to changes in m.

   As m goes up, the overall accuracy is decreasing for all algorithms.

   The noise data affect the experiment results.

4. **Bonus Question** In Fig. 4, I run the experiment on n = 40, 80, 120. We can see that the mis-classification error of them are very close. The hinge loss will increase proportionally with n. As loop numbers increase, the overall loss will converge.

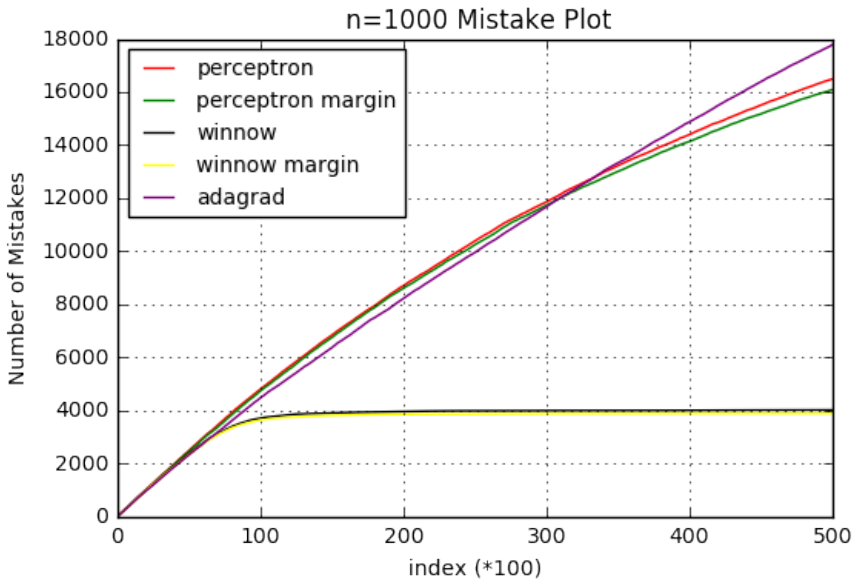

Figure 2: 3.1

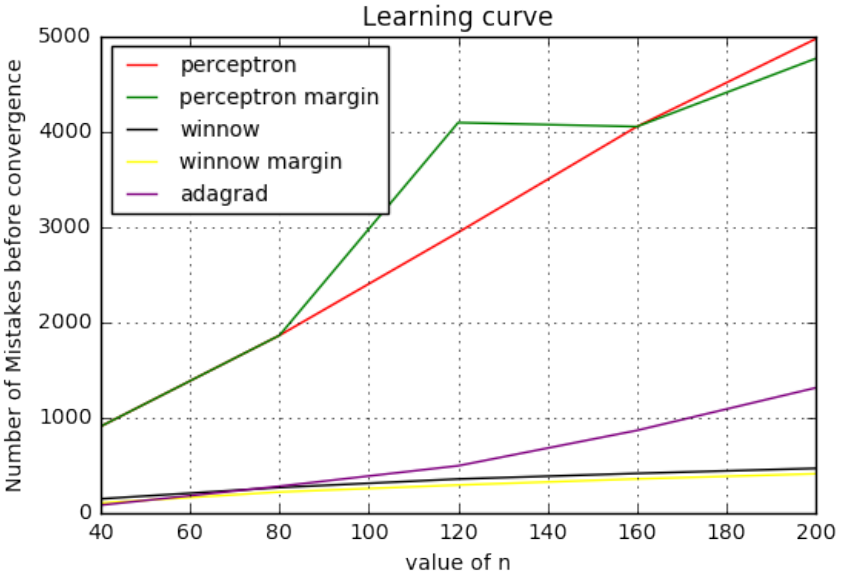

Figure 3: 3.2

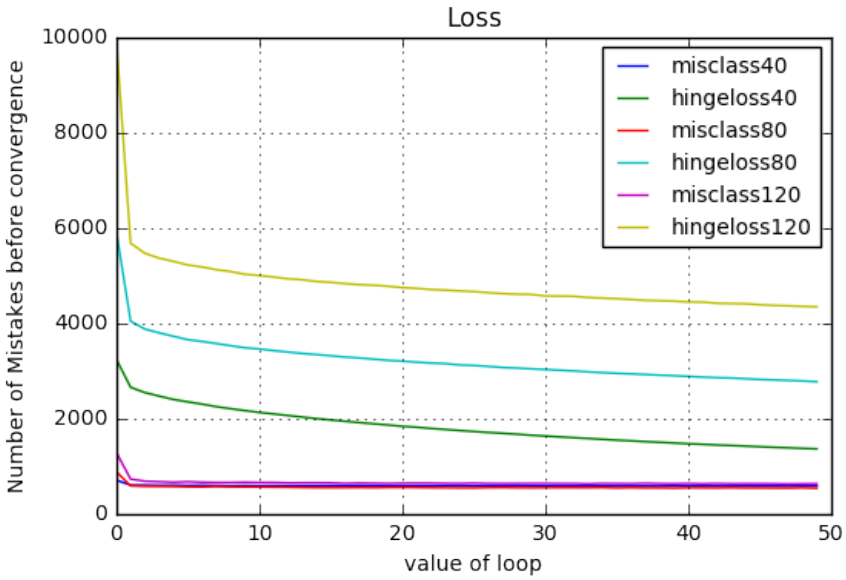

Figure 4: bonus