Machine Learning HW1 PA Report

B04505036
李慕家

For Problem 6, you will play with the PLA algorithm.

First, we use an artificial data set to study PLA. The data set is in

`http://www.csie.ntu.edu.tw/~htlin/course/mlfound19fall/hw1/hw1_6_train.dat`

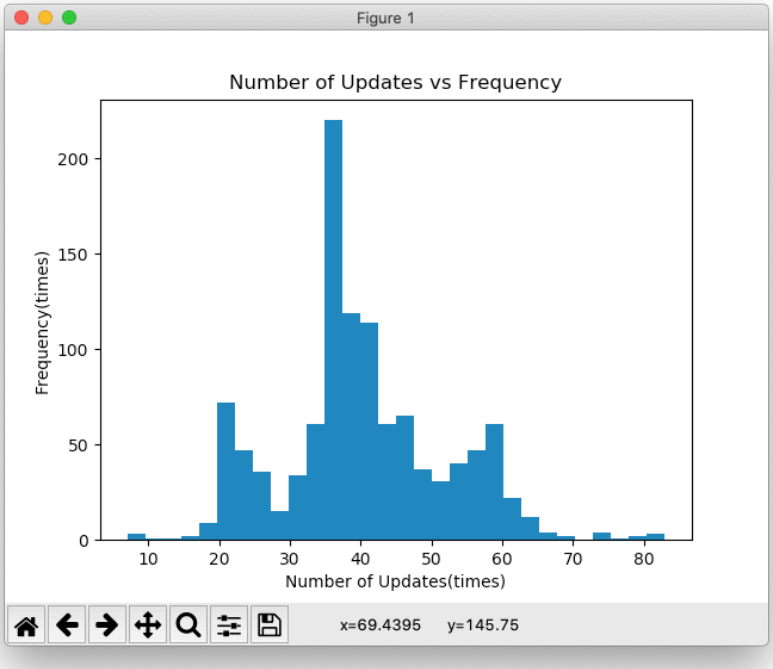
Note that the file is exactly the same as the one for Cousera Homework 1, Problem 15.

`https://www.csie.ntu.edu.tw/~htlin/mooc/datasets/mlfound_math/hw1_15_train.dat`

Each line of the data set contains one $(\mathbf{x}_n, y_n)$ with $\mathbf{x}_n \in \mathbb{R}^4$. The first 4 numbers of the line contains the components of $\mathbf{x}_n$ orderly, the last number is y_n. Please initialize your algorithm with $\mathbf{w} = \mathbf{0}$ and take sign(0) as -1. As a friendly reminder, remember to add $x_0 = 1$ as always!

6. (*, 20 points) Implement a version of PLA by visiting examples in fixed, pre-determined random cycles throughout the algorithm. Run the algorithm on the data set. Please repeat your experiment for 1126 times, each with a different random seed. What is the average number of updates before the algorithm halts? Plot a histogram (`https://en.wikipedia.org/wiki/Histogram`) to show the number of updates versus the frequency of the number.

6.




```
Average number of updates:40.22
```

Next, we play with the pocket algorithm. Modify your PLA in the previous problem by adding the 'pocket' steps to the algorithm. We will use

`http://www.csie.ntu.edu.tw/~htlin/course/mlfound19fall/hw1/hw1_7_train.dat`

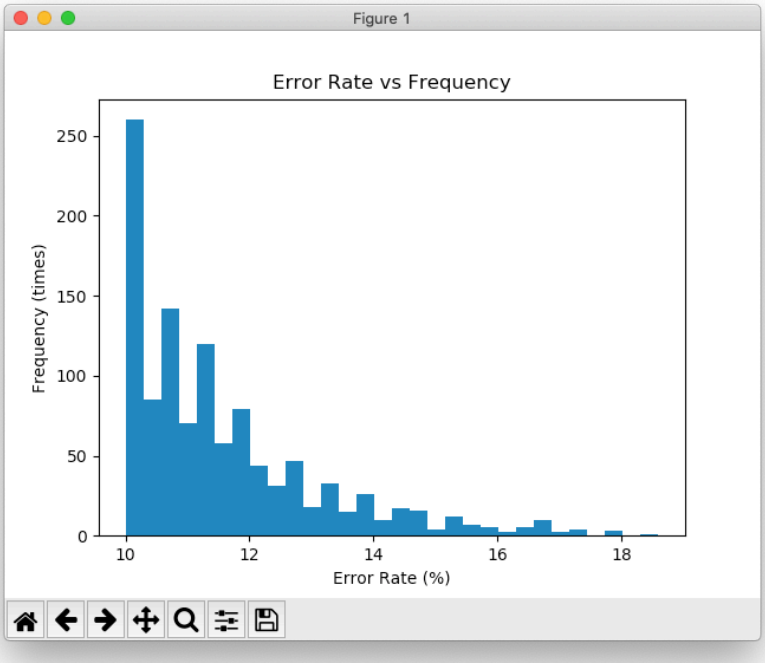
as the training data set $\mathcal{D}$, and

`http://www.csie.ntu.edu.tw/~htlin/course/mlfound19fall/hw1/hw1_7_test.dat`

as the test set for "verifying" the g returned by your algorithm (see lecture 4 about verifying). The sets are of the same format as the previous one.

7. (*, 20 points) Run the pocket algorithm with a total of 100 updates on $\mathcal{D}$, and verify the performance of $\mathbf{w}_{\text{POCKET}}$ using the test set. Please repeat your experiment for 1126 times, each with a different random seed. What is the average error rate on the test set? Plot a histogram to show the error rate versus frequency.

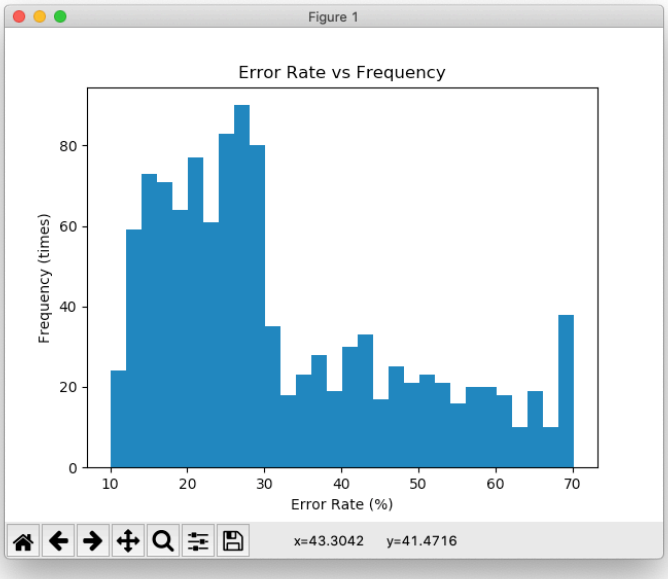
7.




```
Average error rate: 11.6%
```

> **8.** (*, 20 points) Modify your algorithm in the previous problem to return $\mathbf{w}_{100}$ (the PLA vector after 100 updates) instead of $\hat{\mathbf{w}}$ (the pocket vector) after 100 updates. Run the modified algorithm on $\mathcal{D}$, and verify the performance using the test set. Please repeat your experiment for 1126 times, each with a different random seed. What is the average error rate on the test set? Plot a histogram to show the error rate versus frequency. Compare your result to the previous problem and briefly discuss your findings.

8.




```
Average error rate: 32.1%
```

顯然使用[第 100 次 update 後的 weight vector]比[前 100 次 update 中最佳(error 最少)的 weight vector]平均效果要差了不少，而且穩定度變差(變異數變大)很多，average error rate 上升了 176.7%，這也是為什麼我們要記錄最佳的 weight vector 放在 pocket 裡面。