CSCI-4100 Machine Learning
Yichuan Wang 661414395
Assignment 1

Exercises:
1.3a. Misclassification means y(t) and $w^T(t)*x(t)$ have different signs. The product of two numbers with different signs is negative.

1.3b. $y(t)w^T(t+l)x(t) = y(t)^T(t)x(t) + (x(t)y(t))^2$, and $(x(t)y(t))^2$ is a positive number. In this case $y(t)w^T(t+l)x(t)$ is larger than $y(t)^T(t)x(t)$.

1.3c. $w^T(t)x(t)$ is the output of our learning model, while y(t) is the answer. Larger $w^T(t)x(t)y(t)$ means it moves closer to a postive number, which means our learning model is closer to producing the same sign as the answer function does.

1.5a learning
1.5b design
1.5c learning
1.5d design
1.5e learning

1.6a) supervised learning. Training data: The user's bio profile combined with his book purchase history.
1.6b) reinforcement learning. Training data: Game player step inputs with corresponding game result. (The result can either be win or lose; it doesn't have to be correct)
1.6c) unsupervised learning. Training data: Some movies.
1.6d) supervised learning. Training data: Audio music clips or videos played by master musicians.
1.6e) supervised learning. Traing data: Historic customers' bio-profile and whether he/she paid the debt.

EXERCISE 1.7: All hypotheses have the same performance. (1 agree on 3 points; 3 agree on 2 points; 3 agree on 1 points; 1 agree on 0 points.)

1.7a
Always positive: 1 agree on 3 points; 3 agree on 2 points; 3 agree on 1 points; 1 agree on 0 points.

Always negative: 1 agree on 3 points; 3 agree on 2 points; 3 agree on 1 points; 1 agree on 0 points.

The two hypothesis are the same regarding their performances. (1 agree on 3 points; 3 agree on 2 points; 3 agree on 1 points; 1 agree on 0 points.)

1.7b
The two hypothesis are the same regarding their performances. (1 agree on 3 points; 3 agree on 2 points; 3 agree on 1 points; 1 agree on 0 points.)

1.7c
H has 1 agree on 3 points; 3 agree on 2 points; 3 agree on 1 points; 1 agree on 0 points.

1.7d
All hypotheses have the same performance. (1 agree on 3 points; 3 agree on 2 points; 3 agree on 1 points; 1 agree on 0 points.)

Problems:

P1.1

Second ball is black indicates the bag contains 2 black balls. Denote the bag with two black balls as B2 (two black balls) and the other bag as BW. (Black and white balls)

P(B2|black) = P(B2 and black)/P(black)
P(B2 and black) = 0.5
P(black) = (2+1)/(4) = 0.75
P(B2|black) = 0.5/0.75 = 0.667

Answer: 0.667

P1.2

(a) The equation is $w_0+w_1x_1+w_2x_2 = 0$, which can be transformed into

$$x_2 = (w_0+w_1x_1)/w_2$$
$$x_2 = (w_1/w_2)x_1 + w_0/w_2$$

And here we have:

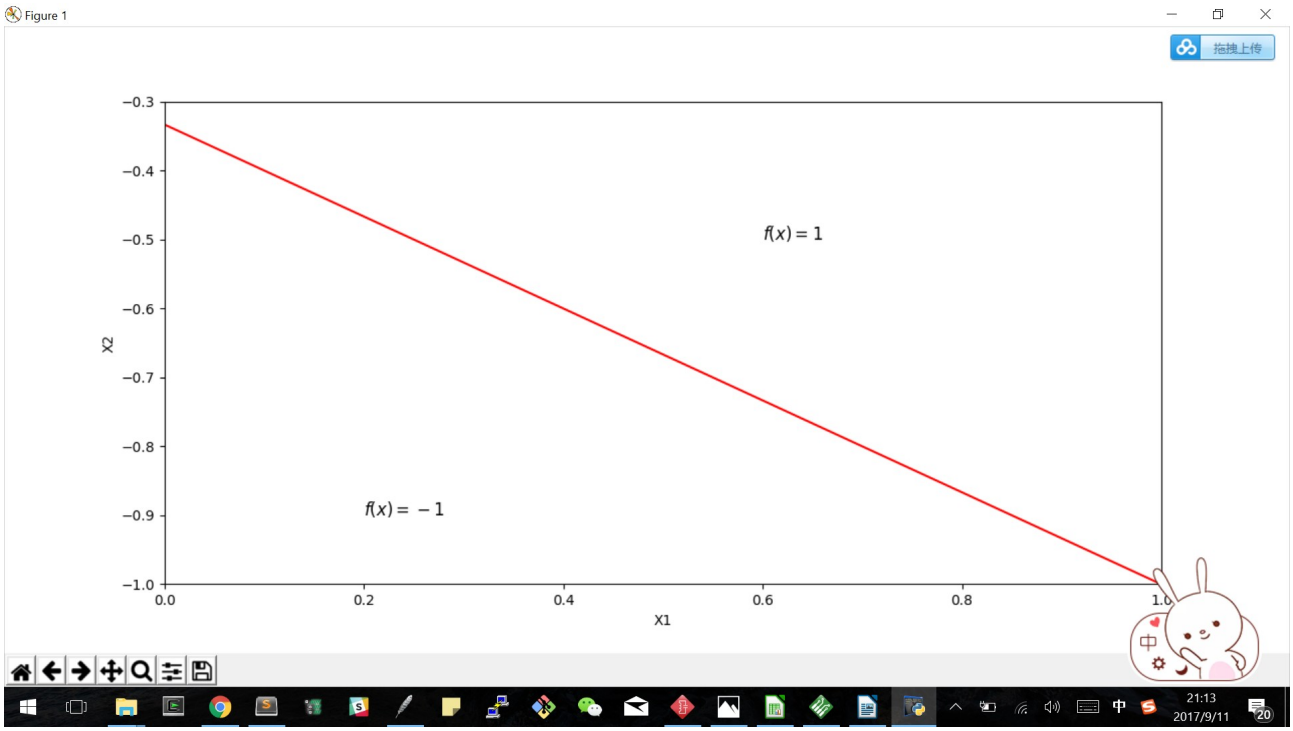
$$a = w_1/w_2 \text{ , } b = w_0/w_2$$

(b)

Two functions represents line on the graph but have opposite classifications on divided areas

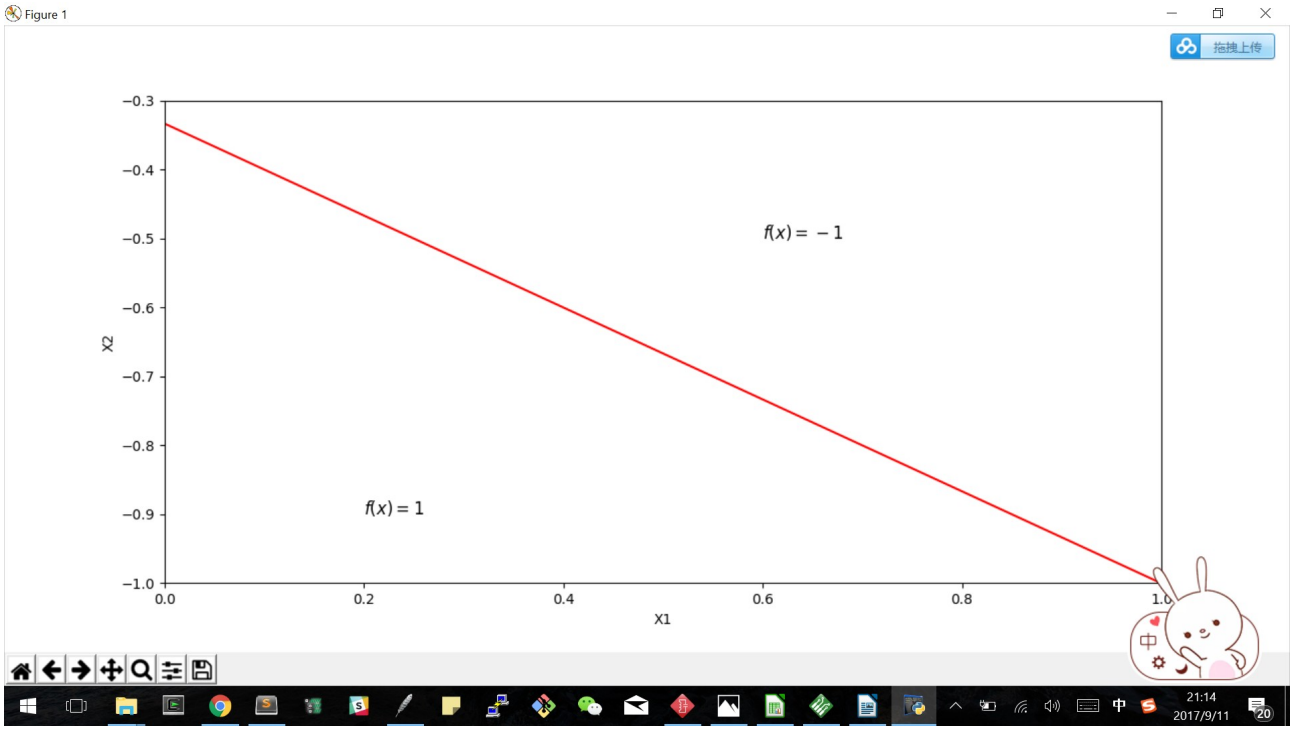
case[1,2,3]:

$$w^Tx = 1 + 2x_1 + 3x_2$$



case -[1,2,3]:

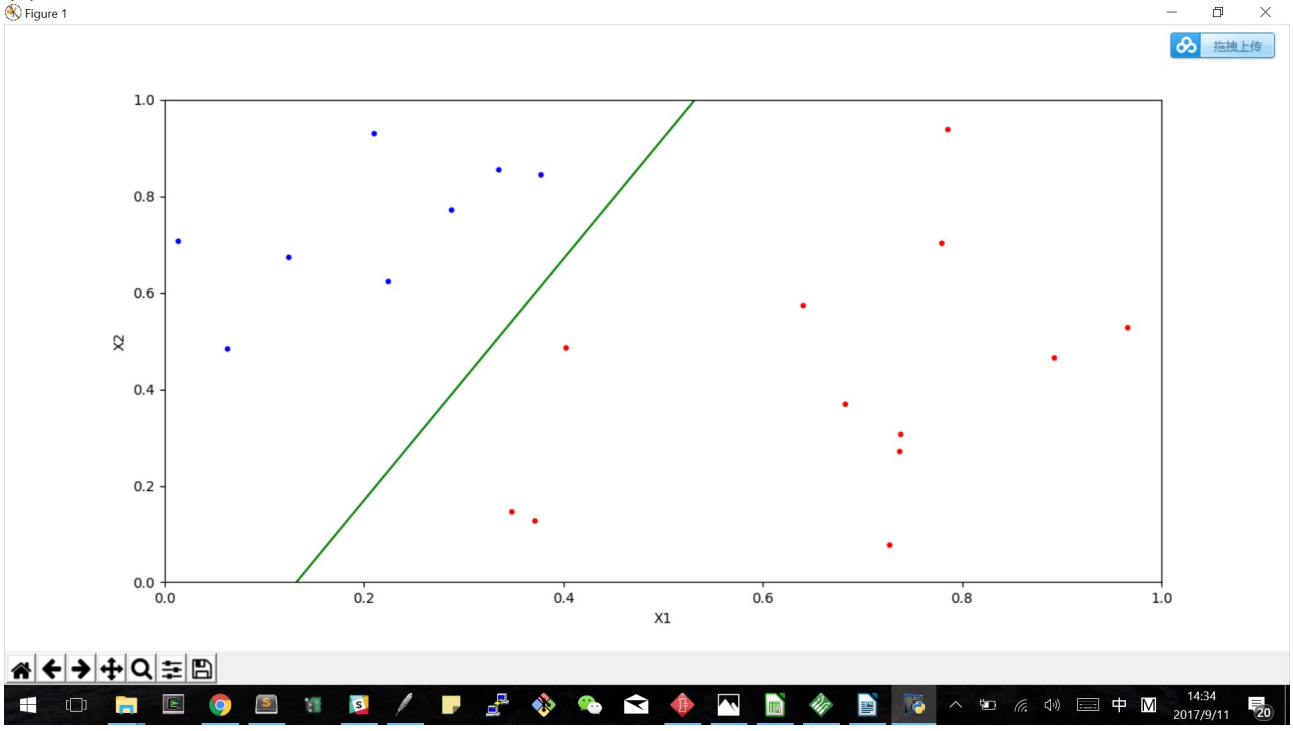
$$w^T x = -1 - 2x_1 - 3x_2$$



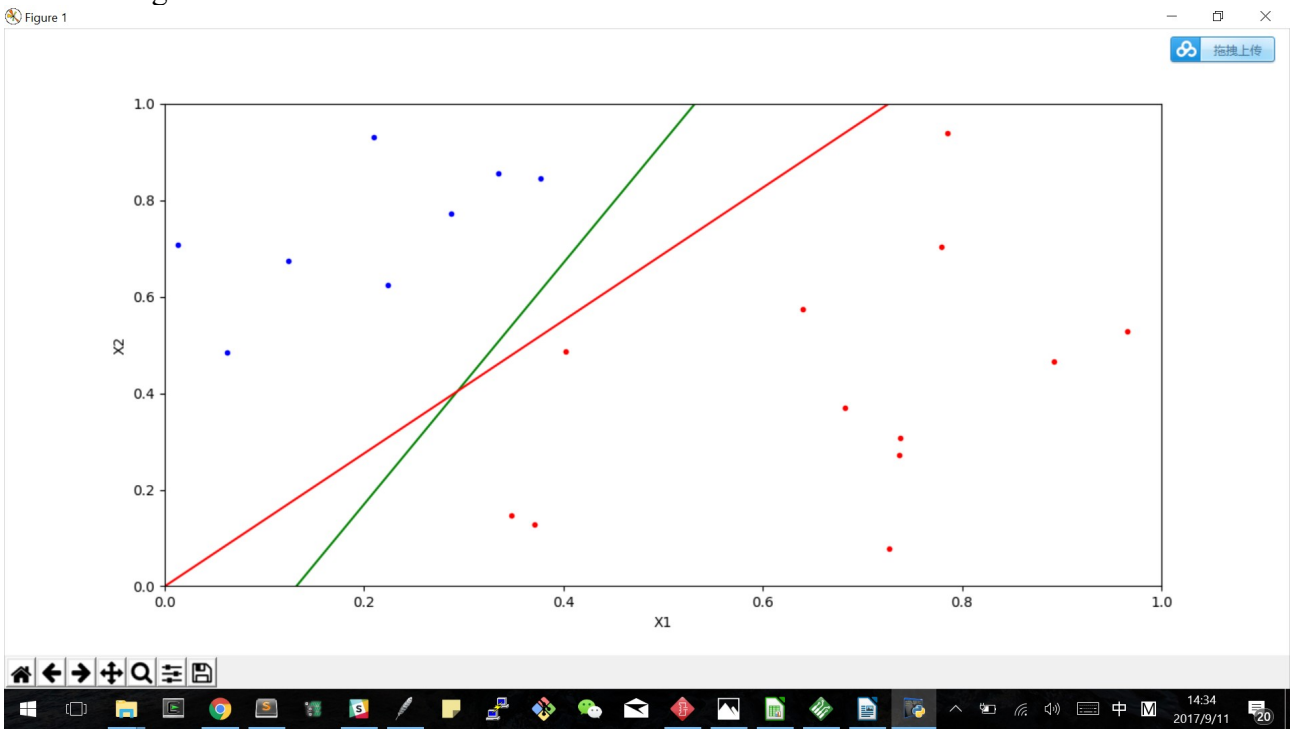
P1.4

NOTE: The green line denotes the target function, while the red line denotes the PLA output function. Red dots represent 1, while blue dots repesent -1.

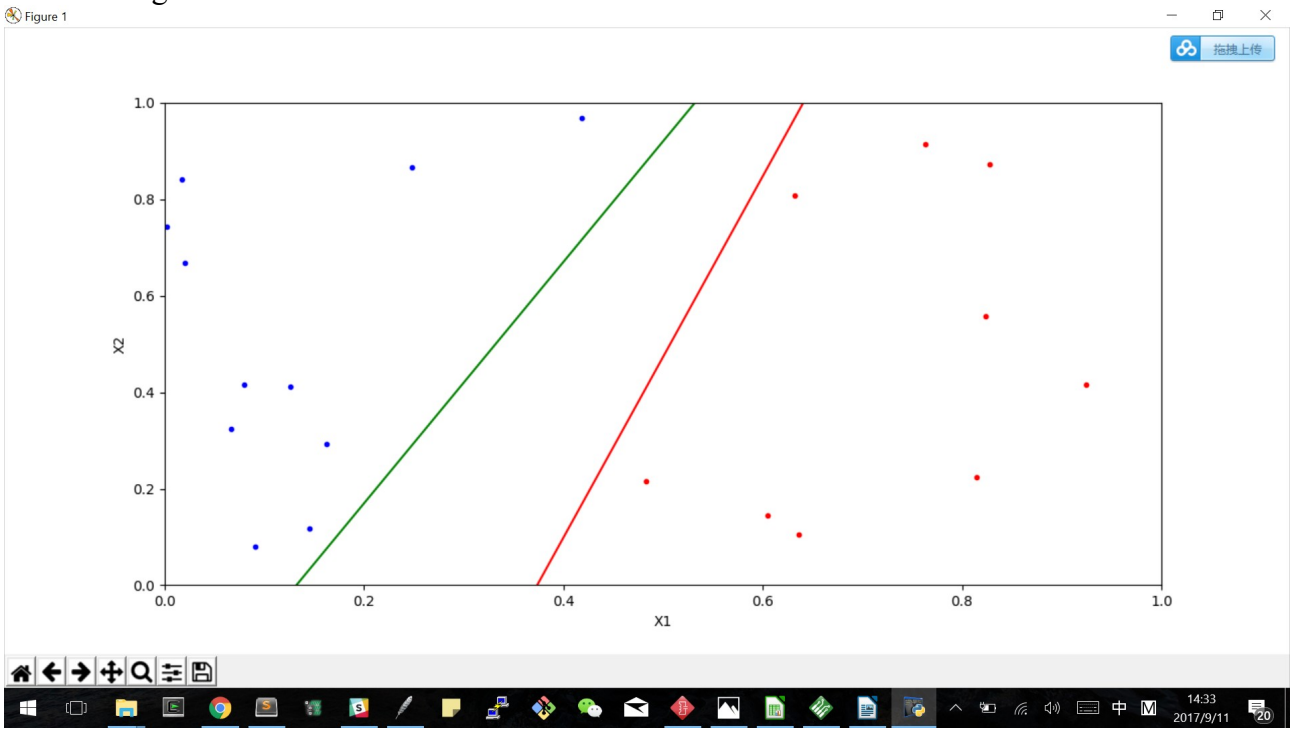
(a) Data size:20



(b) Data size: 20
PLA takes 4 updates to converge. Target function f and output g are in the similar orientation but not close together.

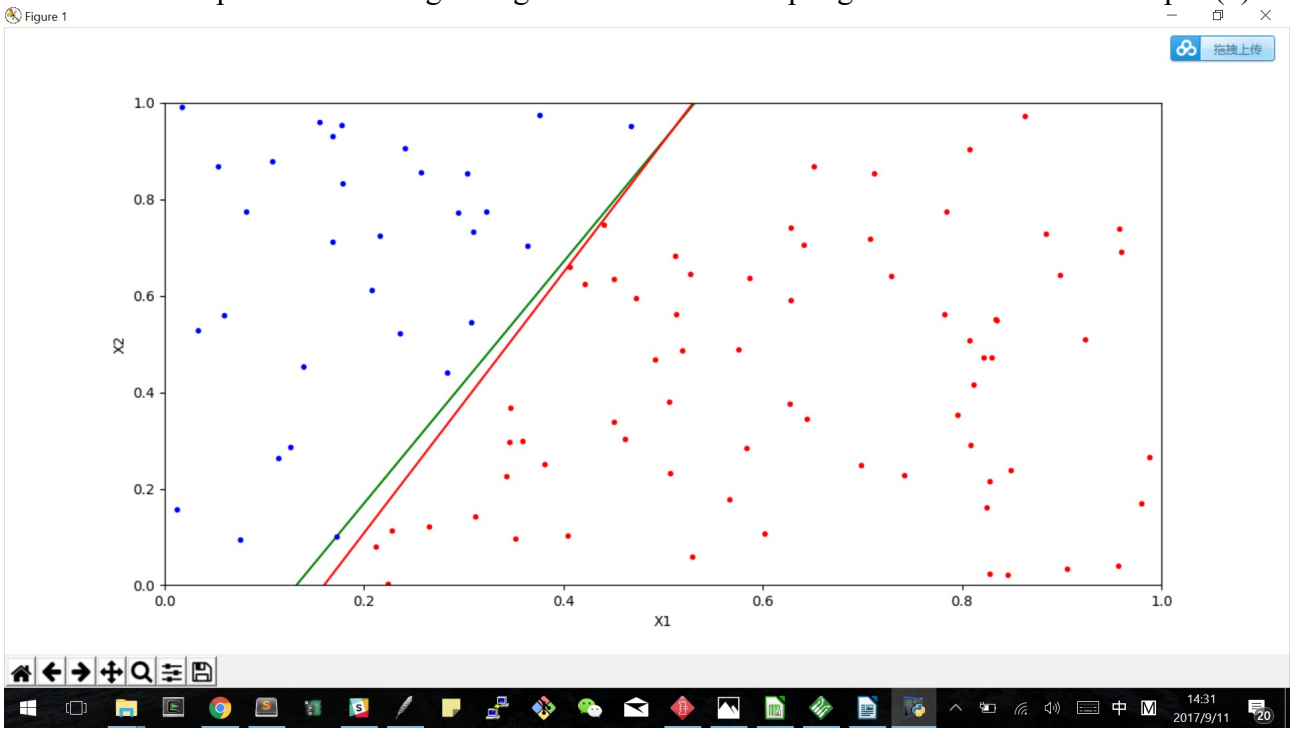


(c) Data size: 20
PLA takes 13 updates to converge. Target function f and output g are in the similar orientation but not close together.



(d) Data size: 100

PLA takes 168 updates to converge. Target function f and output g are closer than those in part(b)



(e) Data size: 1000

PLA takes 1532 updates to converge. Target function f and output g overlapped, and from the graph one cannot tell the difference.