Assignment 1

Ashraful Islam
RIN: 661854753

September 9, 2021

1 Question 1

1.1 a

As $x(t)$ is misclassified by $w(t)$, we have $w^T(t)x(t) < 0$ if $y(t) > 0 \implies y(t)w^T(t)x(t) < 0$, and $w^T(t)x(t) > 0$ if $y(t) < 0 \implies y(t)w^T(t)x(t) < 0$. For both cases, we get $y(t)w^T(t)x(t) < 0$

1.2 b

$$
\begin{aligned}
y(t)w^T(t+1)x(t) &= y(t)(w(t) + y(t)x(t))^T x(t) && (1)\\
&= y(t)(w^T(t) + y(t)x^T(t))x(t) && (2)\\
&= y(t)w^T(t)x(t) + y^2(t)x^T(t)x(t) && (3)\\
&= y(t)w^T(t)x(t) + y^2(t)||x(t)||^2 && (4)\\
&>= y(t)w^T(t)x(t) && (5)
\end{aligned}
$$

1.3 c

For a misclassified $x(t)$, if $y(t) > 0$, then $w^T(t)x(t) < 0$. However, we get from Equation 5, $w^T(t+1)x(t) = w^T(t)x(t) + y(t)||x(t)||^2$. Here, $y(t)||x(t)||$ is a positive number, hence $w^T(t+1)x(t)$ is increasing towards positive direction, which is the expected outcomes.

Similarly, if $y(t) < 0$, then $w^T(t)x(t) > 0$. However, we get from Equation 5, $w^T(t+1)x(t) = w^T(t)x(t) + y(t)||x(t)||^2$. Here, $y(t)||x(t)||$ is a negative number, hence $w^T(t+1)x(t)$ is decreasing towards negative direction, which is the expected outcomes.

2 Question 2

a. Design approach
b. Design approach

c. Learning approach
d. Design approach
e. Learning approach

3 Question 3

a) **"Recommending book"**. This fits **supervised learning** pradigm. The training data could be previous recommendation for all users, and how successful those recommendations were.
b) **"Playing tic-tac-toe"**. This fits with **reinforcement learning**. Reinforcement learning requires a success or reward function, which is suitable for games.
c) **"Categorizing movies"**. This can be learned with **"supervised learning"**. The training data consists of all movies in the catalogue and their categories. We could also fit it with **"unsupervised learning"**, in that case, we would do clustering on the collection of movies.
d) **"Learning to play music"**. This fits with **reinforcement learning**, as playing good music is a reward function for reinforcement learning.
e) **"Credit-limit"** fits with **Supervised learning**. The training data should be users' features (age, income, state, etc.) and credit limits.

4 Question 4

a) As 3 of the 5 outputs in training are '•', The learning algorithm will pick 'one that always return •'. Outside the training data, one($f1$) agrees with all three points, three functions agree with two points (f4, f6, f7), one agrees with one point, and one agrees with none.
b) As 3 of the 5 outputs in training are '∘', The learning algorithm will pick 'one that always return ∘'. One function agrees with all three (f1), three agree with two points (f2, f3, f5), three agree on one point (f4, f6, f7), one agrees with none (f8).
c) For XOR, the output of the testing set should be ∘, ∘, and •. In that case, one function agrees with three points (f2), three agree with two points (f1, f4, f6), three agree with one point (f3, f5, f8), one agrees with none (f7).
d) Here, f7 agrees with all training examples, and disagrees with XOR most. In that case, one agrees with all three points (f7), three agree in two points (f3, f5, f8), three agree in one point (f1, f4, f6), and one agrees with none (f2).

5 Question 5

We have to find $P(\text{Black}_2|\text{Black}_1) = \frac{P(\text{Black}_2 \cup \text{Black}_1)}{P(\text{Black}_1)}$.

$$P(\text{Black}_1) = P(\text{Black}_1|\text{Bag}_1)P(\text{Bag}_1) + P(\text{Black}_1|\text{Bag}_2)P(\text{Bag}_2) \tag{6}$$
$$= \frac{1}{2} + \frac{1}{4} \tag{7}$$
$$= \frac{3}{4} \tag{8}$$

Hence,

$$P(\text{Black}_2|\text{Black}_1) = \frac{P(\text{Black}_2 \cup \text{Black}_1)}{P(\text{Black}_1)} \tag{9}$$
$$= \frac{\frac{1}{2}}{\frac{3}{4}} \tag{10}$$
$$= \frac{2}{3} \tag{11}$$

6 Question 6

6.1 a

The separating line is $w^T x = 0 \implies w_0 x_0 + w_1 x_1 + w_2 x_2 = 0 \implies x_2 = (-w_1/w_2)x_1 + (-w_0/w_2)$

Hence, $a = -w_1/w_2$, and $b = -w_0/w_2$.

6.2 b

The plot is shown in Fig. 1.

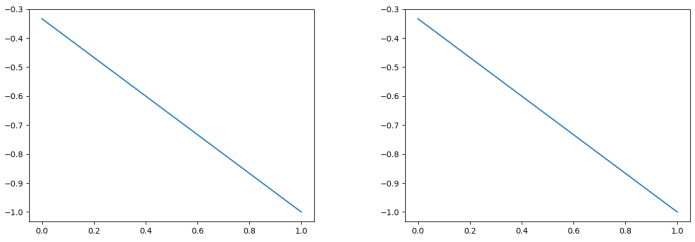

Figure 1: 6.2.b

7 7

7.1 a

The dataset is shown in Fig 2.

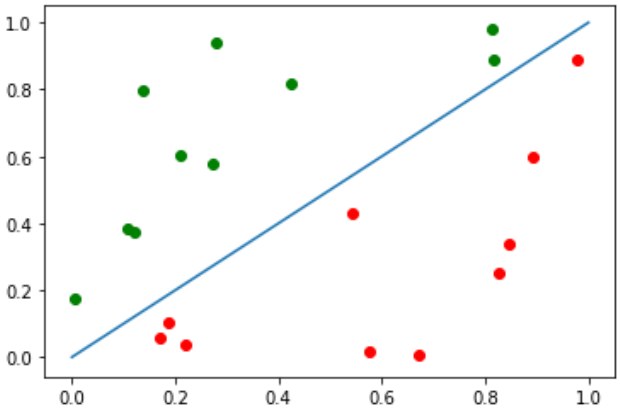

Figure 2:

7.2 b

Number of updates: 24

The final w: [0,2.11872628,-2.08444632]

The PLA algorithm provides the following plot in Fig. 3.

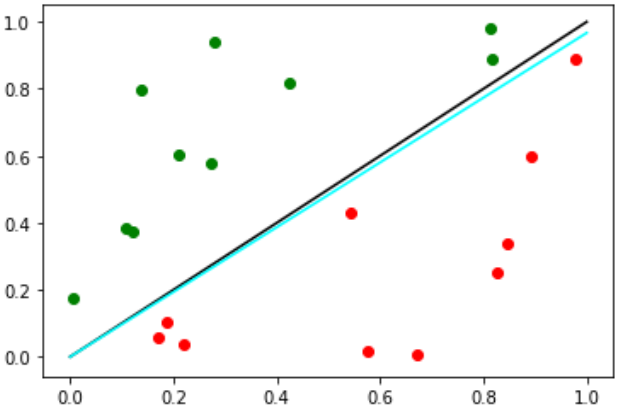

Figure 3: 1.4.b

7.3 c

The dataset is shown in Fig 4. The PLA algorithm provides the following plot in Fig. 5. Number of updates for the PLA is 103.

d)

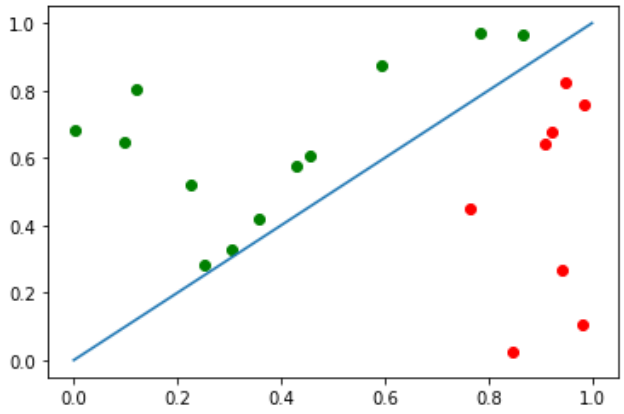

Figure 4: 1.4.c

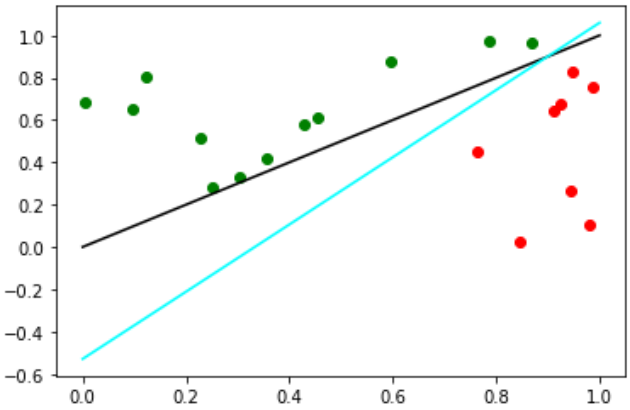

Figure 5: 1.4.c

The dataset is shown in Fig 6. The PLA algorithm provides the following plot in Fig. 7. Number of updates for the PLA is 30.

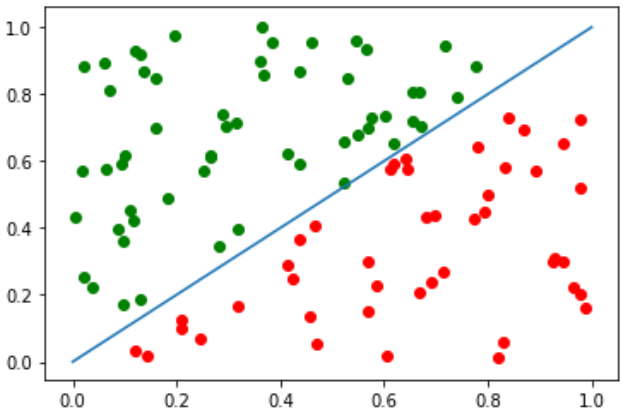

Figure 6: 1.4.d

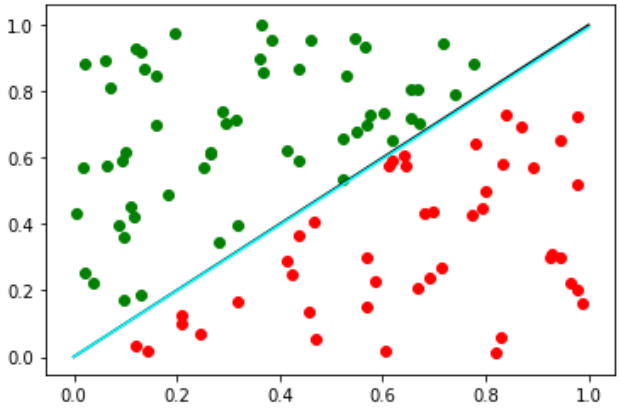

Figure 7: 1.4.d

e) The dataset is shown in Fig 8. The PLA algorithm provides the following plot in Fig. 9. Number of updates for the PLA is 604.

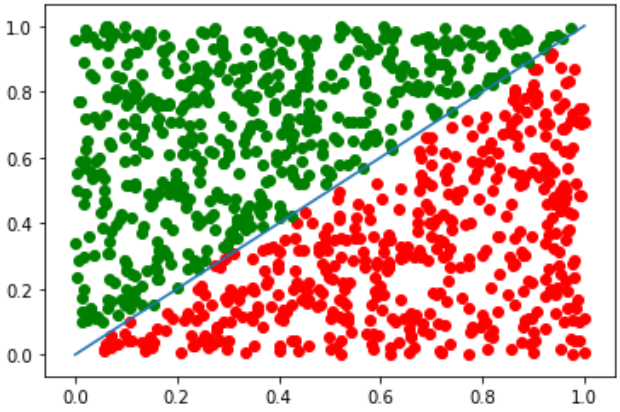

Figure 8: 1.4.e

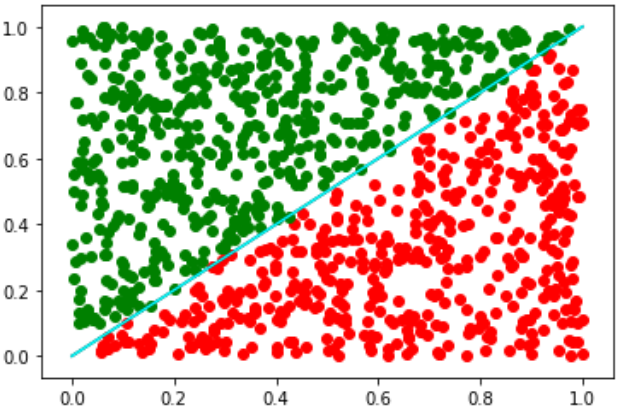

Figure 9: 1.4.e