Homework 1

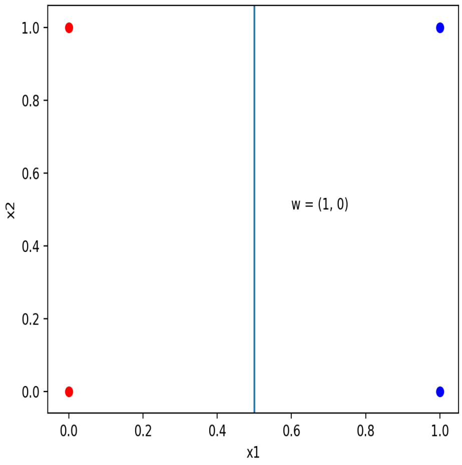
Xixiang Chen

1 Analysis

1.1

(a) Let $y = f(x_1, x_2) = x_1$, we have:

x_1	x_2	y
0	0	0
0	1	0
1	0	1
1	1	1



(b) Let $y = f(x_1, x_2) = x_1 \oplus x_2$, we have:

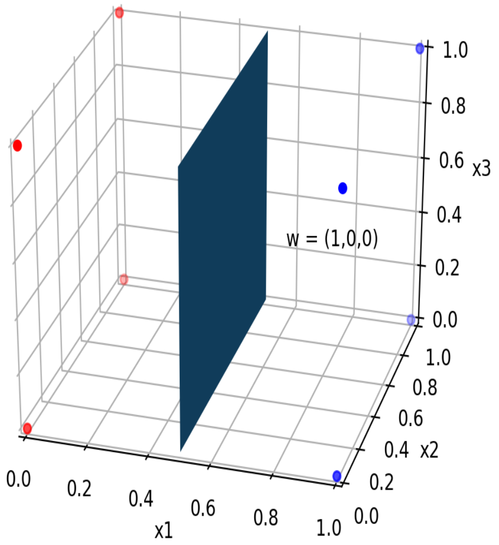
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Assume such perceptron $\overrightarrow{w} = (w_1, w_2)$ exists.
According to the boolean function, we have:

$$w_1 * 0 + w_2 * 0 + b < 0 \implies b < 0 \implies -b > 0 \tag{1}$$
$$w_1 * 0 + w_2 * 1 + b > 0 \implies w_2 + b > 0 \tag{2}$$
$$w_1 * 1 + w_2 * 0 + b > 0 \implies w_1 + b > 0 \tag{3}$$
$$w_1 * 1 + w_2 * 1 + b < 0 \implies w_1 + w_2 + b < 0 \tag{4}$$

from (2) + (3) we have $w_1 + w_2 + 2b > 0$, combined with (1) we have $w_1 + w_2 + 2b + (-b) > 0$, which is a contradiction with (4). Therefore such linear separator does not exist.

(c) Let $y = f(x_1, x_2, x_3) = x_1$, we have:



x_1	x_2	x_3	y
0	0	0	0
0	0	1	0
0	1	0	0
1	0	0	1
0	1	1	0
1	0	1	1
1	1	0	1
1	1	1	1

1.2

The Euclidean distance between point $\boldsymbol{x}$ and the hyperplane is the length of the vector, $\boldsymbol{v}$, from point $\boldsymbol{x}$ to the closest point on hyperplane x_0. Since $\boldsymbol{v}$ is parellel to β, we can write:

$$\begin{aligned} v &= x - x_0 \\ &= d \cdot y \cdot \frac{\beta}{\|\beta\|_2} \\ \implies x_0 &= x - d \cdot y \cdot \frac{\beta}{\|\beta\|_2} \end{aligned}$$

where d is the length of the vector $\boldsymbol{v}$. Since x_0 is on the hyperplane, we have:

$$\begin{aligned} & \beta^T x_0 + \beta_0 = 0 \\ \implies & \beta^T (x - d \cdot y \cdot \frac{\beta}{\|\beta\|_2}) + \beta_0 = 0 \\ & \beta^T x - \beta^T \cdot d \cdot y \cdot \frac{\beta}{\|\beta\|_2} + \beta_0 = 0 \\ & \beta^T x + \beta_0 = \|\beta\|_2 \cdot d \cdot y \\ & f(x) = \|\beta\|_2 \cdot d \cdot y \\ Since\ y \in \{-1, 1\}, we\ have: & \\ & d = \frac{1}{\|\beta\|_2} y f(x) \end{aligned}$$

1.3

Let T be the number of steps perceptron algorithm takes to converge, and w^T the hyperplane at iteration T:

$$\begin{aligned}\left\|w^T - w^{sep}\right\|_2^2 &= \left\|(w^{T-1} + y_i x_i) - w^{sep}\right\|_2^2 \\ &= \left\|(w^{T-1} - w^{sep}) + y_i x_i\right\|_2^2 \\ &= \left\|w^{T-1} - w^{sep}\right\|_2^2 + y_i^2 \left\|x_i\right\|_2^2 + 2(w^{T-1} - w^{sep})^T y_i x_i \\ &= \left\|w^{T-1} - w^{sep}\right\|_2^2 + y_i^2 \left\|x_i\right\|_2^2 + 2w^{T-1^T} y_i x_i - 2w^{sep^T} y_i x_i\end{aligned}$$

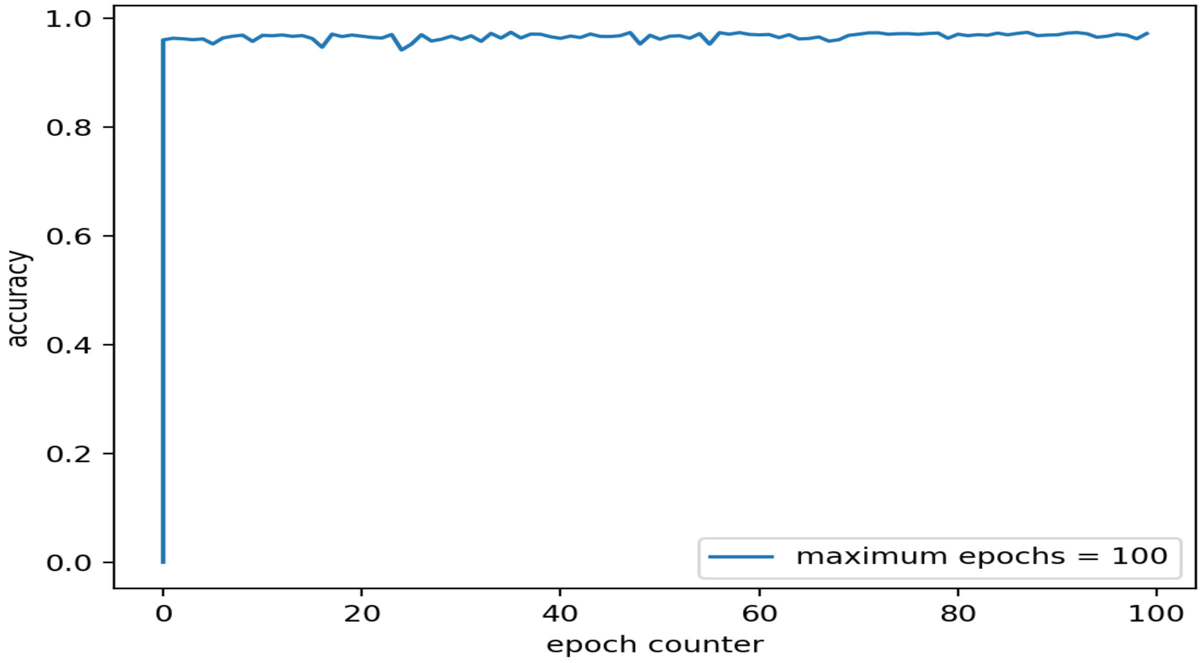
Since we have $\left\|x_i\right\|_2 \leq 1$, and

$$\begin{aligned} & y_i w^{sep^T} x_i \geq 1 and \\ & y_i w^{k^T} x_i < 0 \; for \; iteration \; 0 < k < T, \\ Therefore \;\; \left\|w^T - w^{sep}\right\|_2^2 & \leq \left\|w^{T-1} - w^{sep}\right\|_2^2 + 1 + 0 - 2 \\ &= \left\|w^{T-1} - w^{sep}\right\|_2^2 - 1 \\ &\leq \left\|w^{T-2} - w^{sep}\right\|_2^2 - 1 - 1 \\ & \ldots \\ &\leq \left\|w^{0} - w^{sep}\right\|_2^2 - T \\ so \;\; \left\|w^T - w^{sep}\right\|_2^2 + T &\leq \left\|w^{0} - w^{sep}\right\|_2^2 \\ \implies T &\leq \left\|w^{0} - w^{sep}\right\|_2^2 \end{aligned}$$

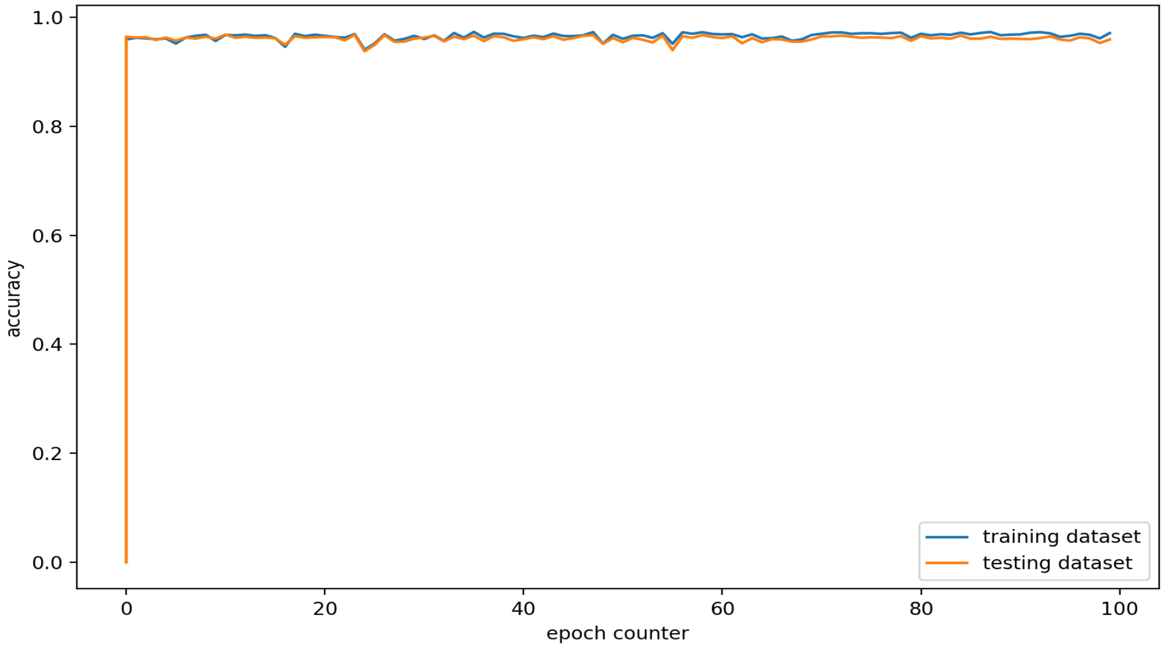
2 Programming

2.1

(a) The accuracy rises to around 0.96 pretty quickly, and it bounces around 0.97 for further runs. Increasing or decreasing the maximum number of epochs does not change the trend much.

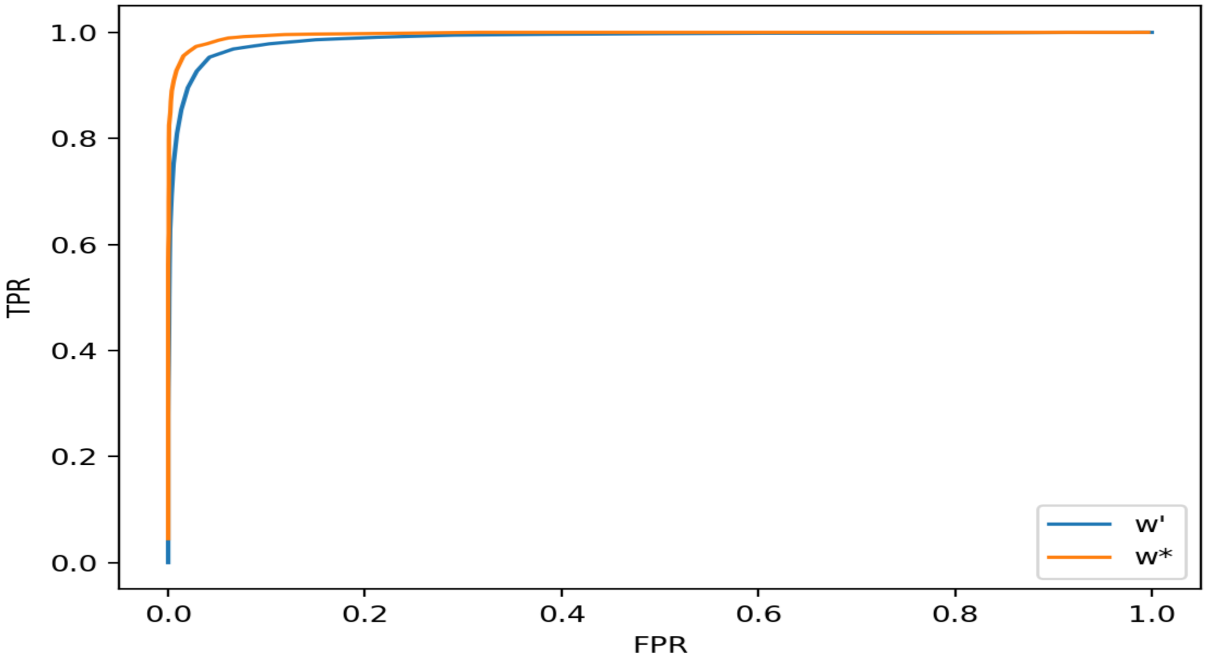


(b) The accuracy curve on testing dataset is very similar to that on training dataset.



(c)
TP = 943 FP = 41
FN = 39 TN = 968
accuracy = 0.9598191863385234

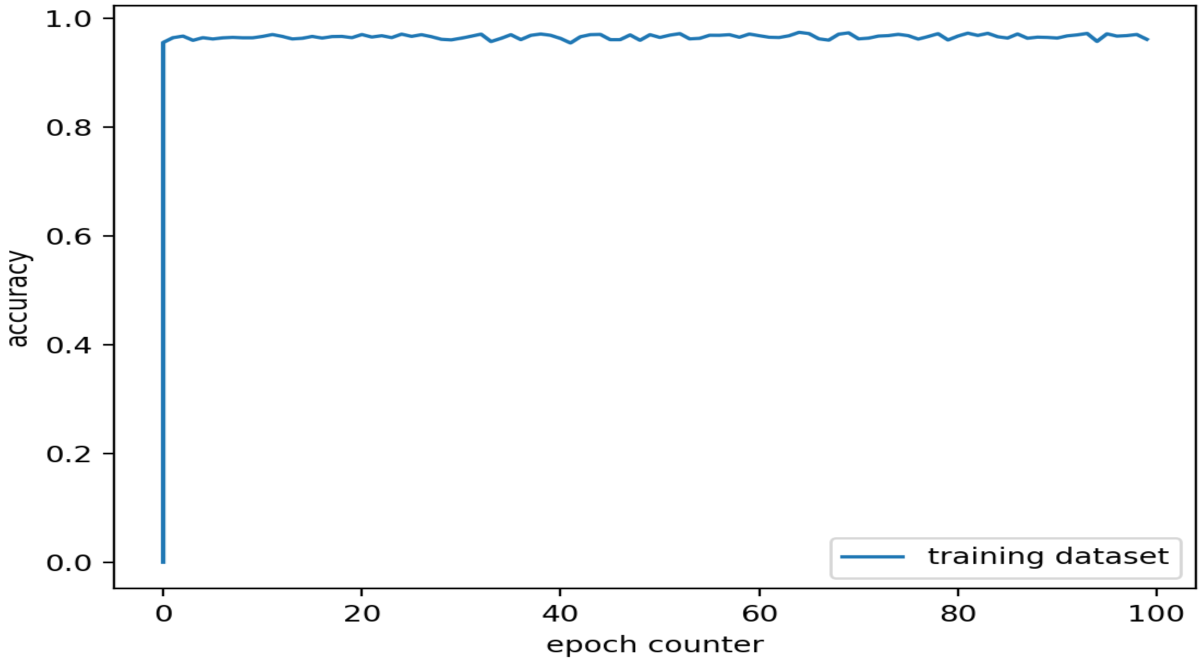
(d) The ROC curve of w* converges to 1 faster than that of w', therefore w* yields a better decision boundary.



(e)
AUC for w': 0.987965
AUC for w*: 0.996461
The values agree with the ROC curves in part (c): higher accuracy in the beginning gives larger AUC.

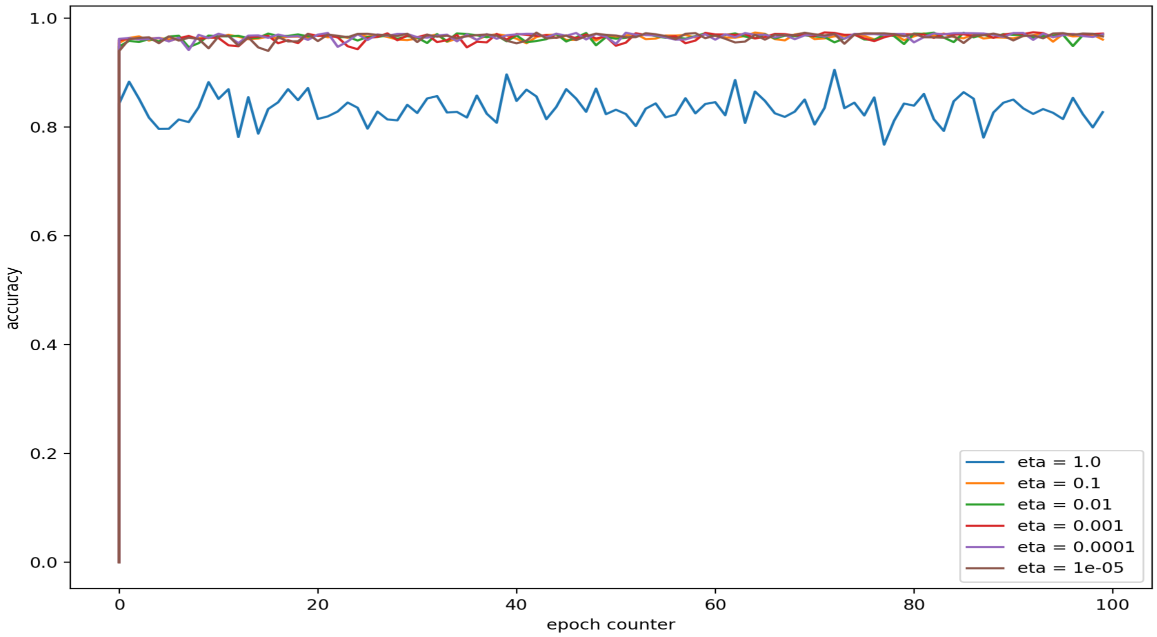
2.2

(a)



TP = 963 FP = 63
FN = 19 TN = 946
accuracy = 0.9588146659969864

(b) As η gets smaller, the accuracy curve gets more stable. But after a certain point (as in the figure, 0.0001), it is pretty insignificant difference when η decreases further more. Meanwhile we want the margin to be large. So 0.0001 would be a good choice for η.


1.0
0.8
0.6
0.4
0.2
0.0
accuracy
0
20
40
60
80
100
epoch counter
eta = 1.0
eta = 0.1
eta = 0.01
eta = 0.001
eta = 0.0001
eta = 1e-05