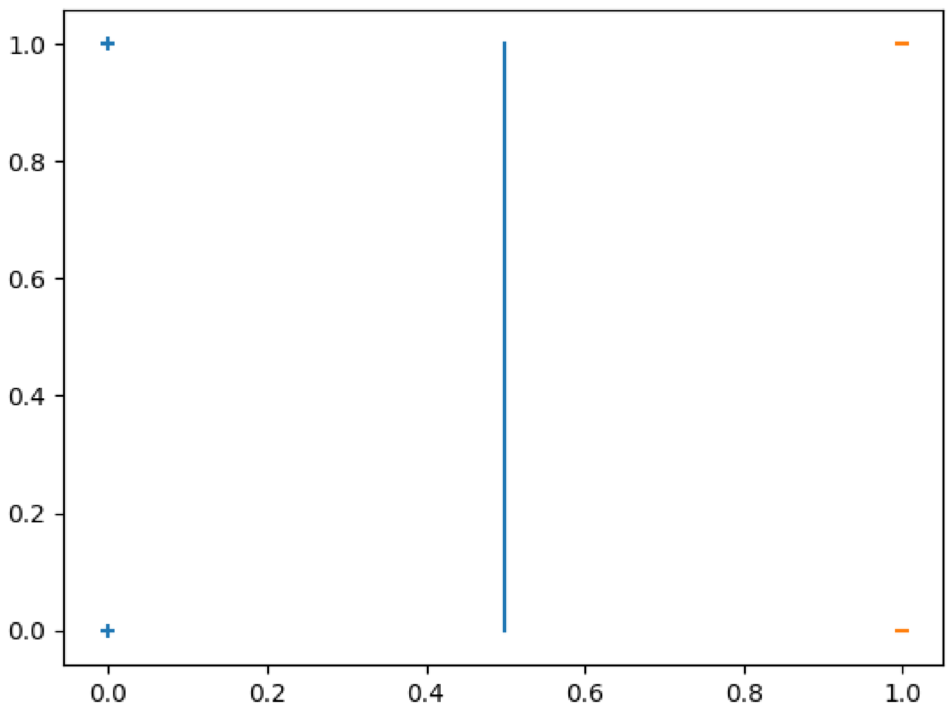
Homework 1

Weiyu Yan

1 Perceptron Algorithm and Convergence Analysis

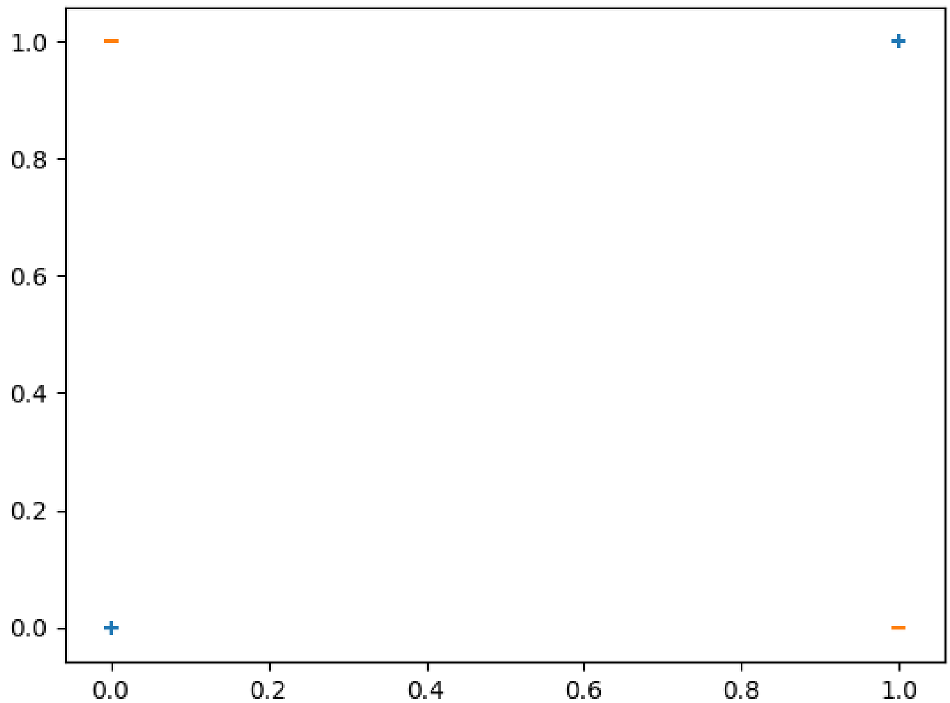
1. (a) $y = f(x_1, x_2) = x_1$



(b) A truth table for this function could be:

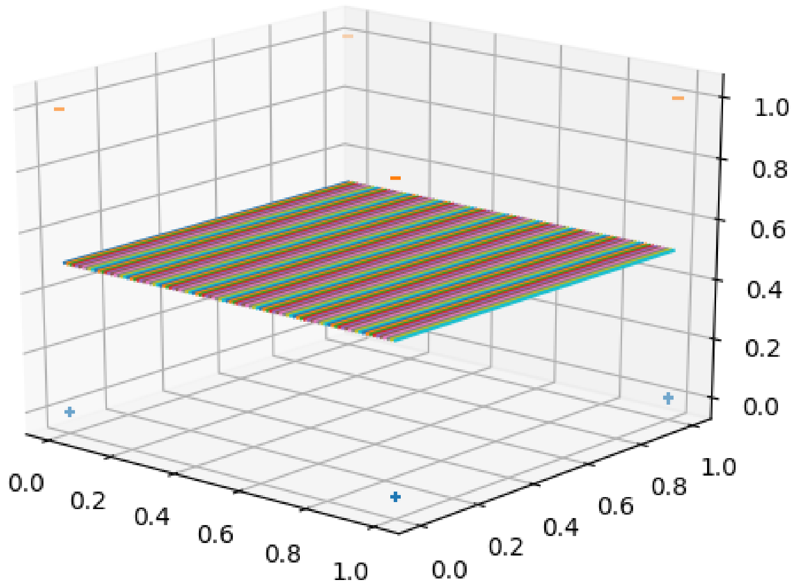
x_1	x_2	y
0	0	1
0	1	0
1	0	0
1	1	1

We can plot those points in a figure.



The perceptron algorithm gives us a weight vector which corresponds to its orthogonal hyperplane to divide points into 2 parts. In this figure since we can't find a hyperplane to divide the 2 kinds, this function can't be separated by single perceptron.

(c) $y = f(x_1, x_2, x_3) = x_3$



2. β is a weight vector perpendicular to the hyperplane the signed Euclidean Distance of the point x to the hyperplane is given by:

$$d = \frac{(x - x_0) \cdot \beta y}{\|\beta\|_2}$$

where

$$x_0 \in \{x | f(x) = 0\}$$

$$\because f(x_0) = \beta_0 + \beta^T x_0$$

$$\therefore \beta^T x_0 = -\beta_0$$

Therefore

$$\begin{aligned} d &= \frac{y(x \cdot \beta - x_0 \cdot \beta)}{\|\beta\|_2} \\ &= \frac{y(\beta^T x - \beta^T x_0)}{\|\beta\|_2} \\ &= \frac{y(\beta^T x + \beta_0)}{\|\beta\|_2} \\ &= \frac{yf(x)}{\|\beta\|_2} \end{aligned} \tag{1}$$

3. By perceptron convergence theorem we get that the perceptron algorithm makes at most $\frac{1}{\delta^2}$ mistakes where $\frac{1}{\delta^2}$ is the lower bound for $y_i(w^{sep^T} x_i)$ and w^{sep} is the "good separator":

$$T \leq \frac{1}{\delta^2}$$

where

$$y_i(w^{sep} x_i) \geq \delta$$

Since

$$y_i w^{sep} x_i = 1$$

we have

$$T \leq 1$$

Since $w^0 = 0$, $\|w^{sep}\| = 1$, we have:

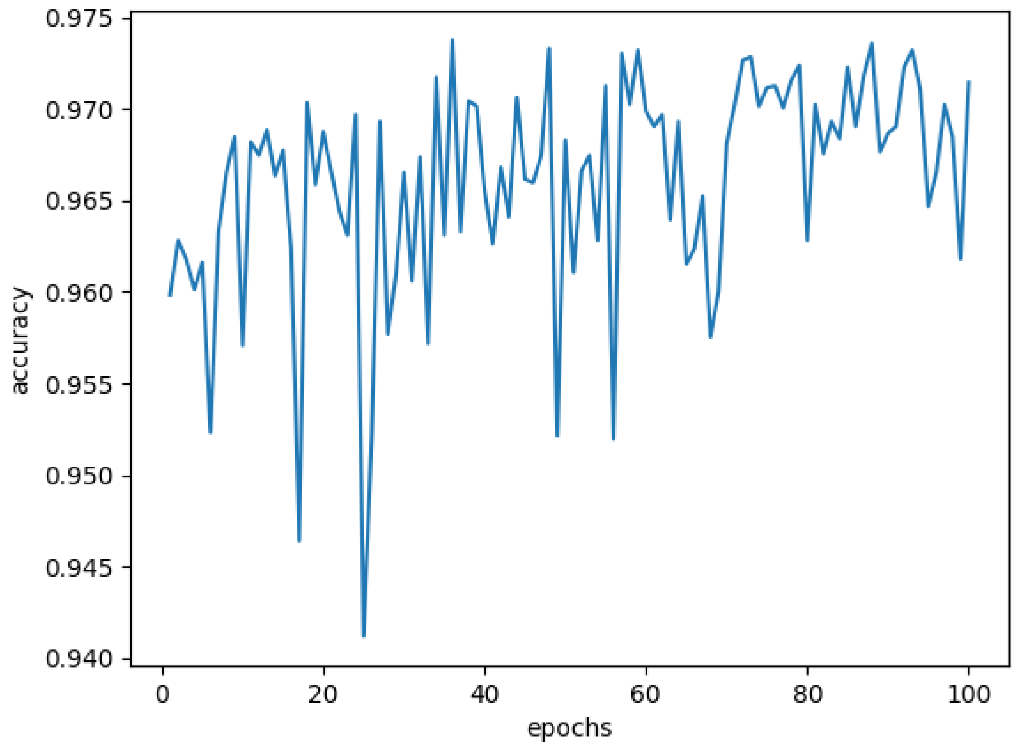
$$\|w^{(0)} - w^{sep}\|_2^2 = \|w^{sep}\|_2^2 = 1$$

Therefore,

$$T \leq \|w^{(0)} - w^{sep}\|_2^2$$

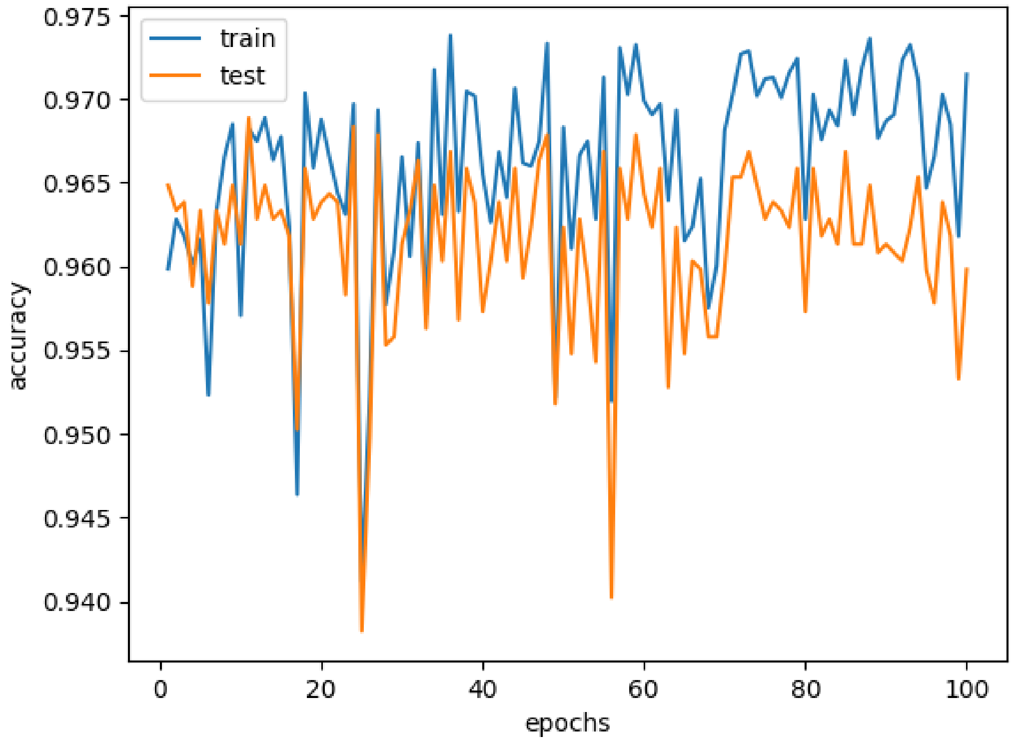
2 Programming Assignment

1. Please refer code in "perceptron.py"



(a)

Since first epoch the accuracy is already pretty good.
Then it shakes drastically at first(within a small bound, in general the accuracy is always above 0.94), then becomes more steady, with the moving average increasing slowly.
By increasing/decreasing number of epochs, the curve becomes denser/sparser – just revealing more/less information in the figure, the points' location won't change.

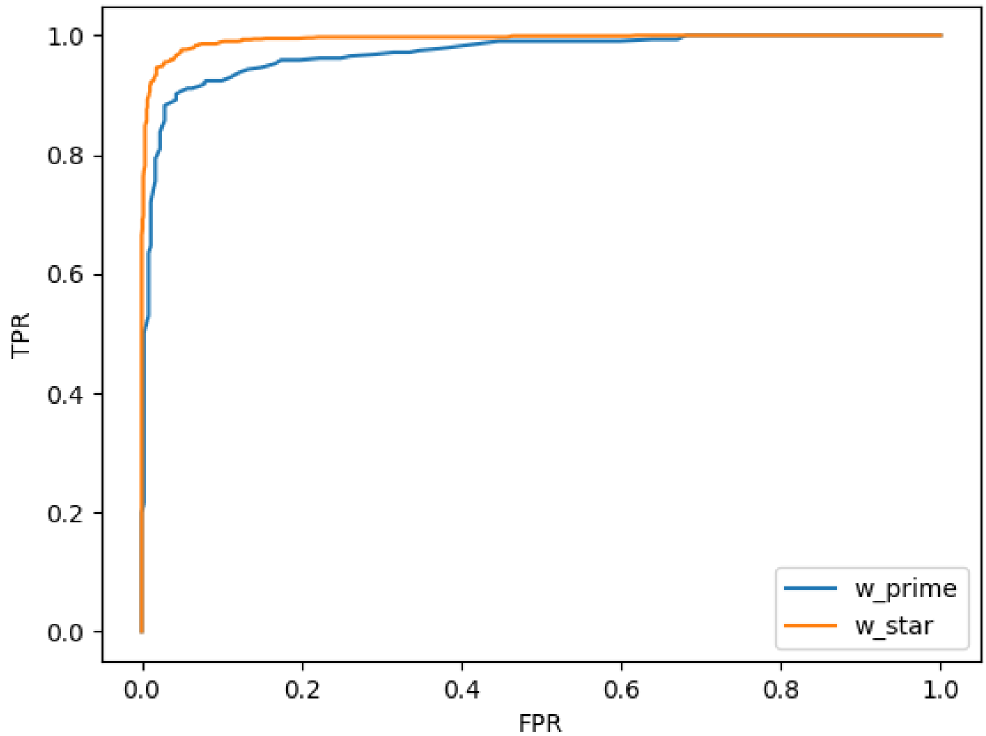


(b)

The curve for test set roughly follows the same pattern of traning set. With the epoch number going up, it becomes lower than the training set.

(c) accuracy: 0.99397287795077 85
confusion matrix:

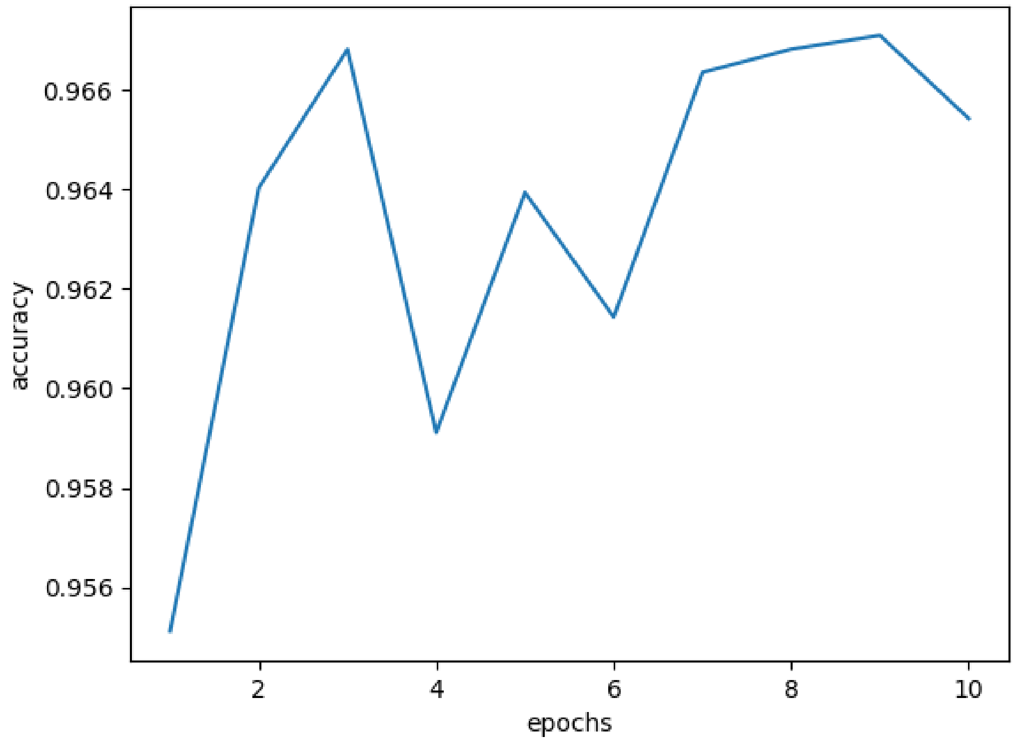
$$\begin{bmatrix} 968 & 39 \\ 41 & 943 \end{bmatrix}$$



(d)

(e) AUC(w') = 0.9877518533678995
AUC(w^*) = 0.9999666948582919 The greater the AUC value, the bigger area that is under the ROC curve, the 'upper' that the ROC curve goes.

2. (a) Evolution for training set after 10 epochs:



For test set:
accuracy: 0.9653440482169764
confusion matrix:

$$\begin{bmatrix} 980 & 40 \\ 29 & 942 \end{bmatrix}$$

(b)