Homework 1

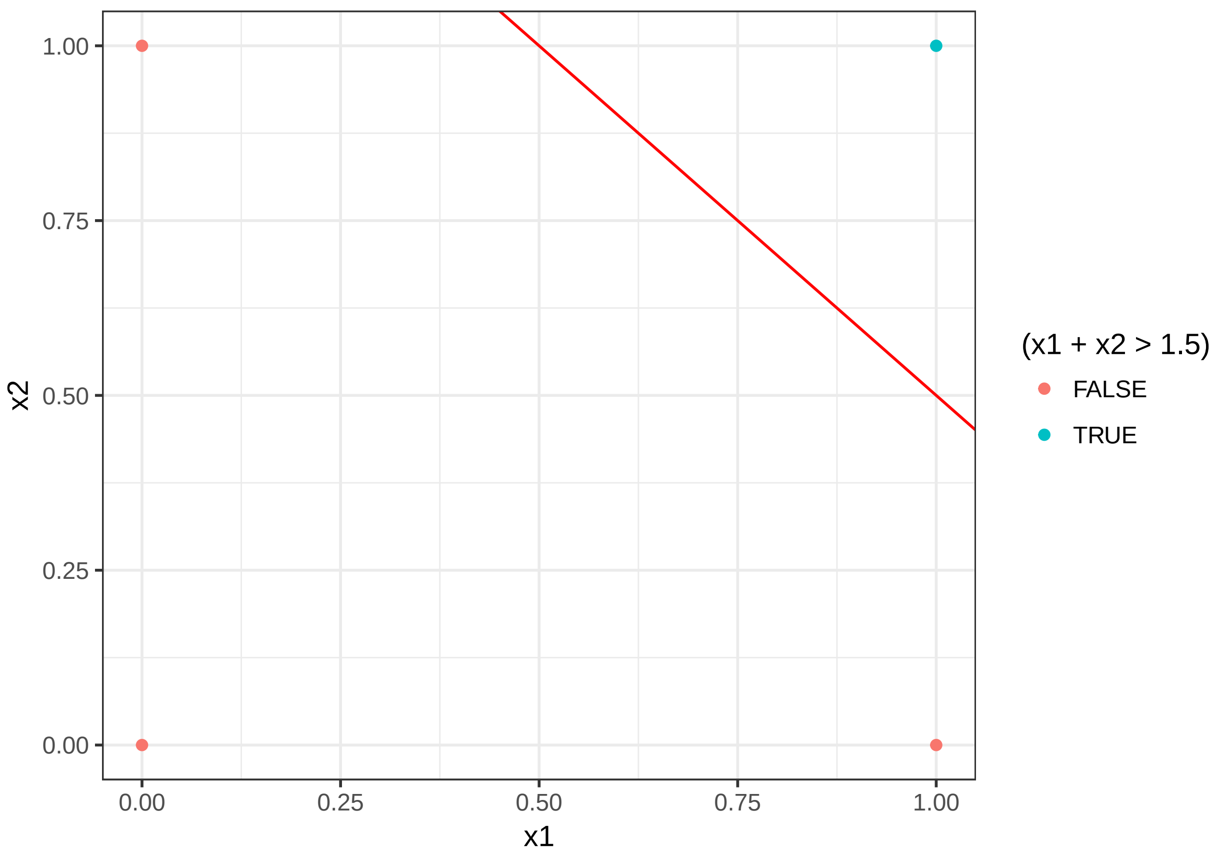
Kai Wang

2018/1/24

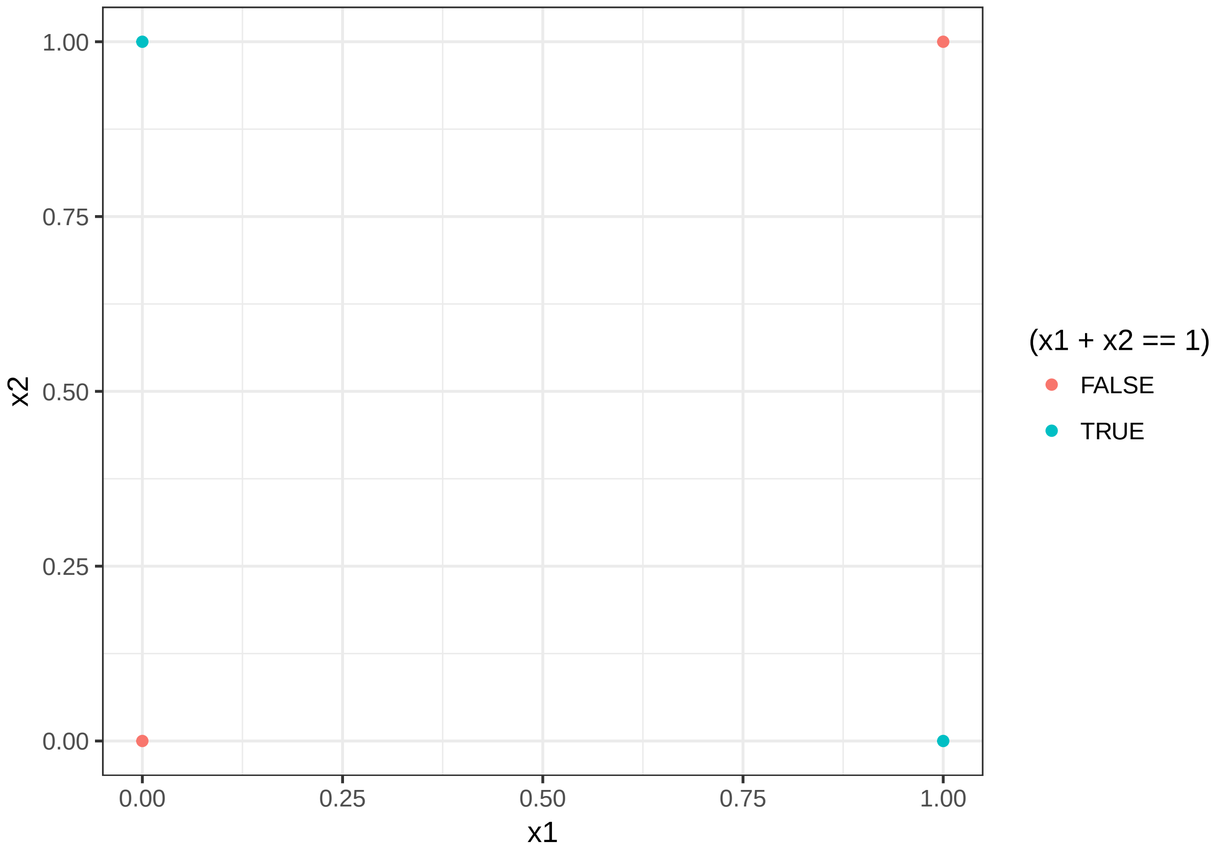
1 Perceptron Algorithm and Convergence Analysis

1

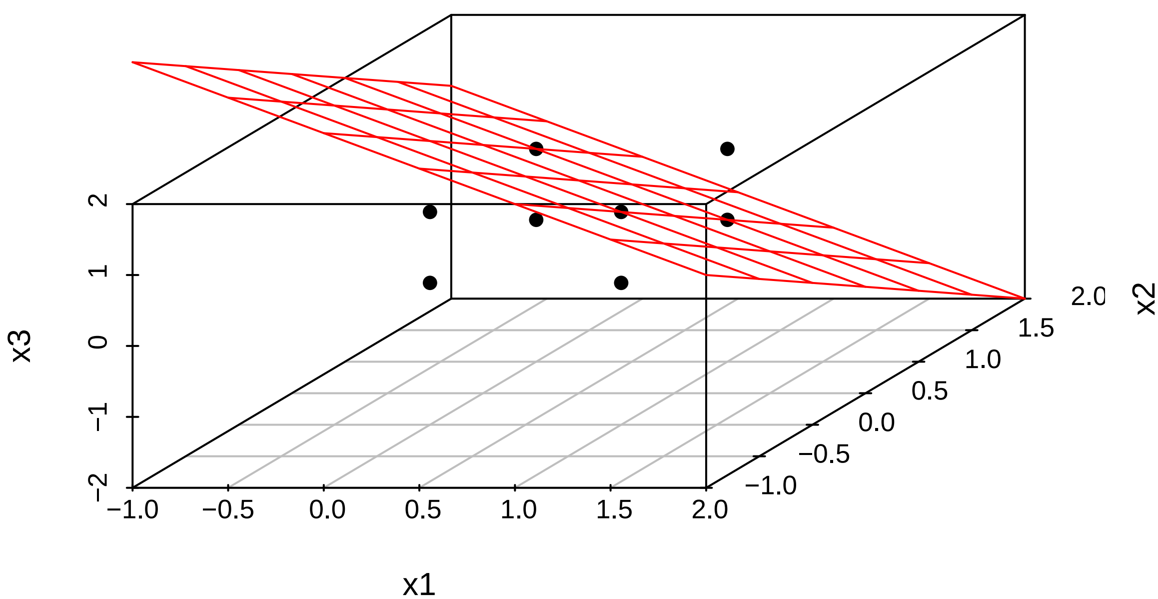
(a) The boolean function is $y = f(x_1, x_2) = \mathbb{I}_{x_1+x_2-1>0}$.



(b) The boolean function is $y = f(x_1, x_2) = \mathbb{I}_{x_1+x_2=1}$. Such a boolean function cannot be represented by a perceptron.



(c) The boolean function is $y = f(x_1, x_2) = \mathbb{I}_{x_1+x_2+x_3>2}$. Note that (1,1,1) is the only positive point.



2

Let z be a random point on the decision boundary. We have $f(z) = \beta_0 + \beta^T z = 0$. So we have $\beta^T z = -\beta_0$. We know that the distance of a point to a vector equals to the product of those two vectors.

$$
\begin{aligned}
distance &= \frac{|\overrightarrow{\beta}(\overrightarrow{x} - \overrightarrow{z})|}{||\beta||_2} \\
&= \frac{|\overrightarrow{\beta}\,\overrightarrow{x} + \beta_0|}{||\beta||_2} \\
&= \frac{|f(x)|}{||\beta||_2} \\
&= \frac{1}{||\beta||_2} y f(x)
\end{aligned}
$$

Because y and $f(x)$ have the same sign, $yf(x) = |f(x)|$.

3

$$
w^{(T)} \cdot w^{(sep)} - w^{(T-1)} \cdot w^{(sep)} = y_i x_i w^{(sep)} \geq 1
$$

$$
(w^{(T)} - w^{(0)}) \cdot w^{(sep)} = \sum_{t=1}^{T}((w^{(T)} - w^{(T-1)}) \cdot w^{(sep)}) \geq T
$$

Hence,

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

Since we have $T > 1$ and $T \leq ||w^{(T)} - w^{(0)}||_2$, we have $T \leq ||w^{(T)} - w^{(0)}||_2^2$.

When the perceptron algorithm converges to a separating plane, we have $T \leq ||w^{(sep)} - w^{(0)}||_2^2$, which is equivalent to $T \leq ||w^{(0)} - w^{(sep)}||_2^2$

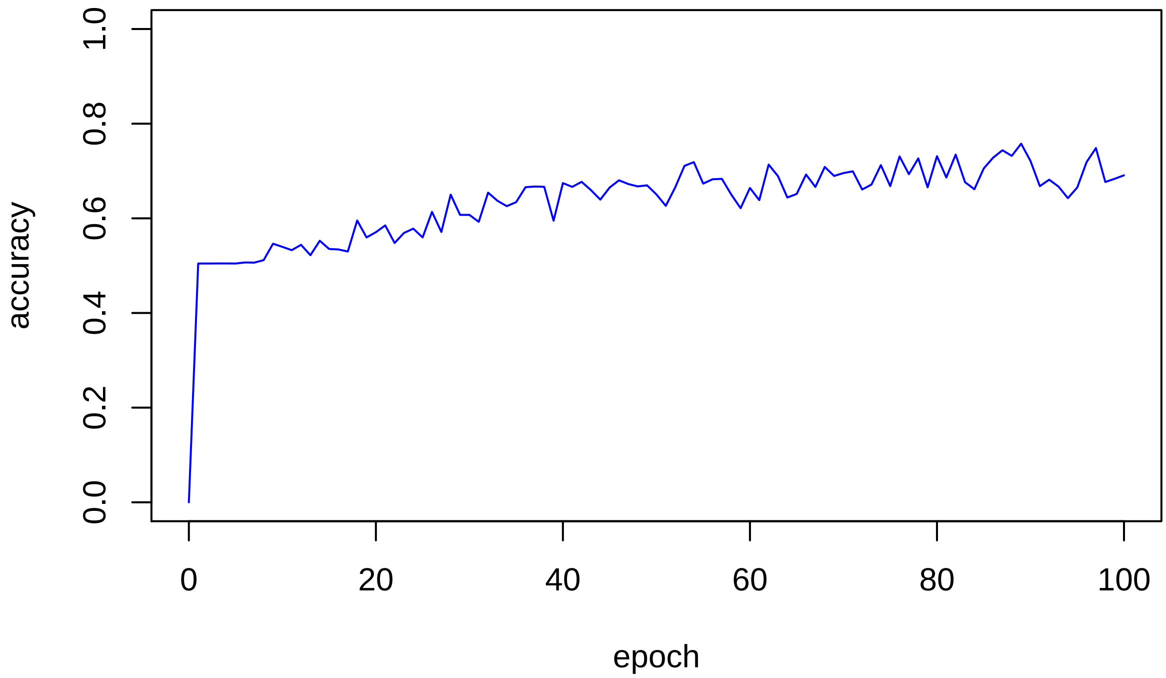
2 Programming Assignment

1

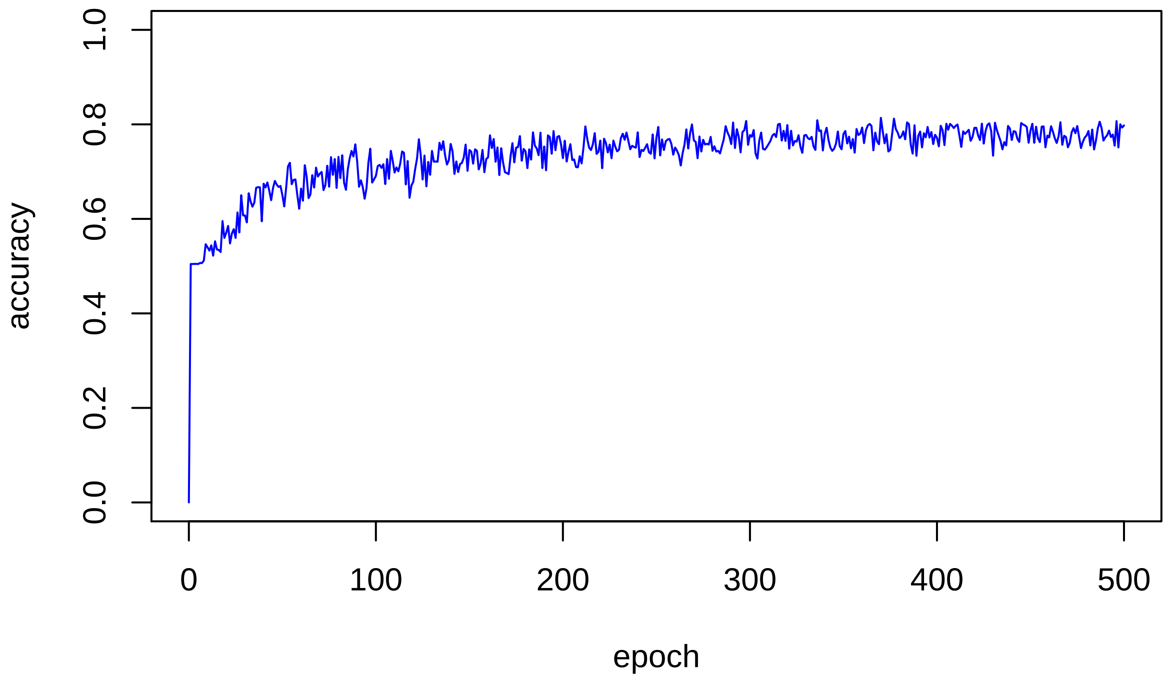
```
## INFO [2018-01-25 01:32:10] MNIST data set already available, nothing left to do.

## [1] TRUE
```

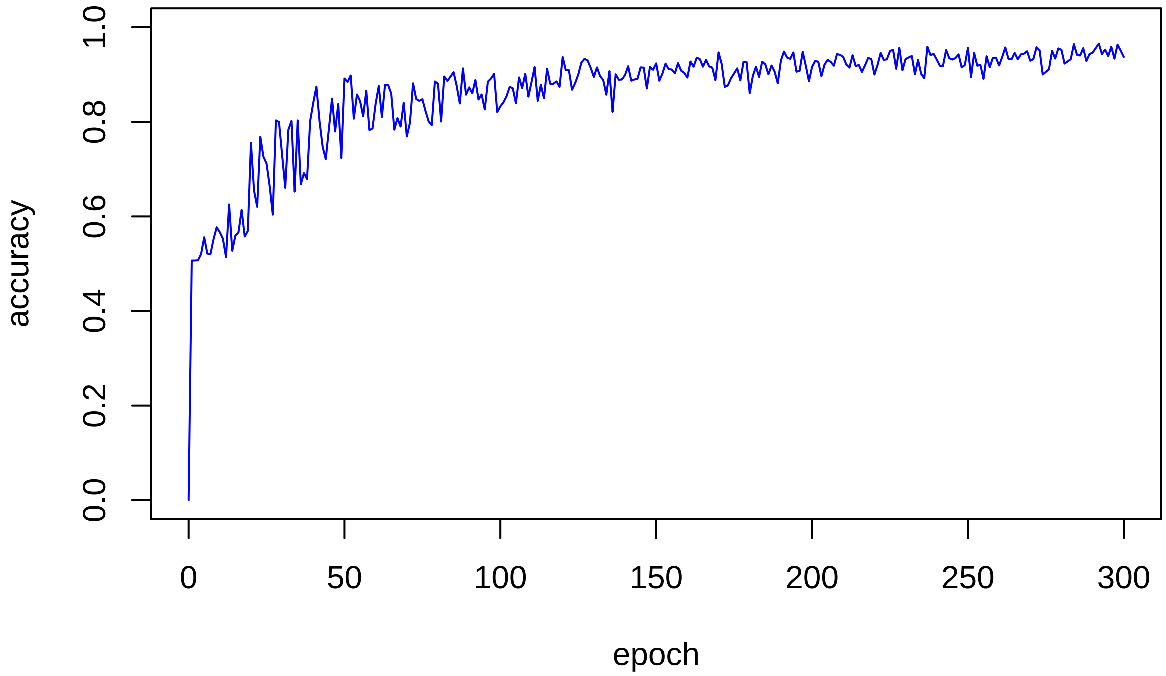
a



The accuracy does not seem too good and convergence has not achieved. We could increase the epoch. By increasing epoch, I found that accuracy increased until about epoch = 500. The final training accuracy is aroun .80.



b



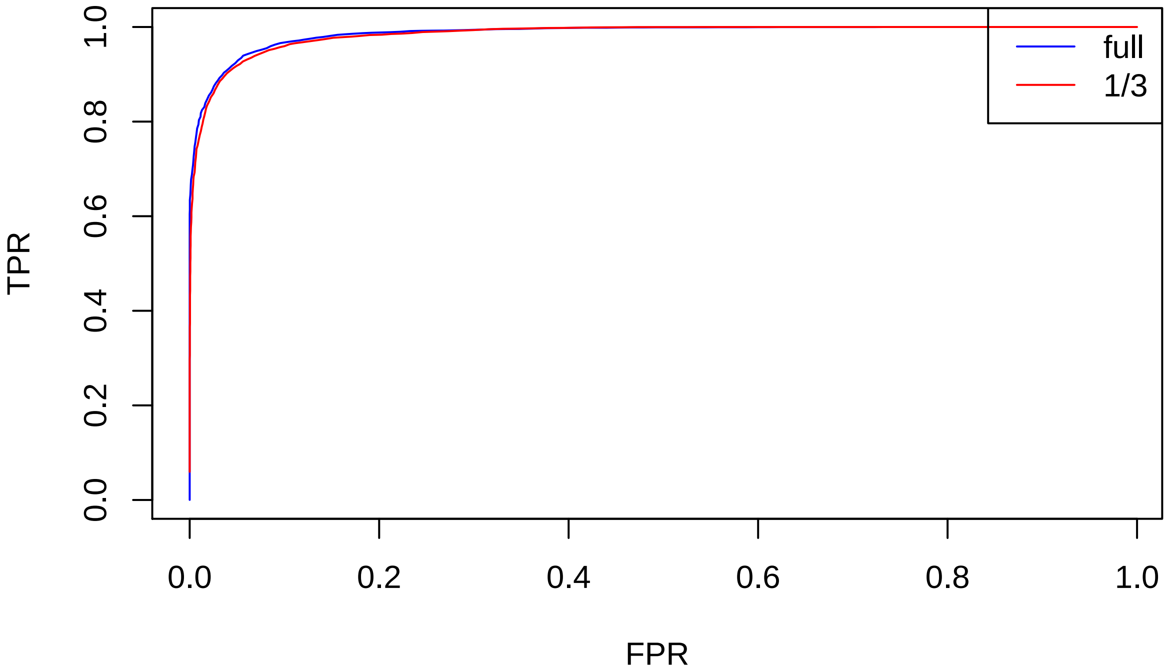
The test accuracy is much higher than training accuracy and it converges quicker. The test accuracy is around .95 after converging.

c

	Positive	Negative
Predicted Positive	857	0
Predicted Negative	125	1009

The accuracy is $\frac{857+1009}{857+0+125+1009} = 0.9372$.

d



From the ROC curve, we can see running algorithm until convergence has a slightly better performance, which means it has a better decision boundary.

e

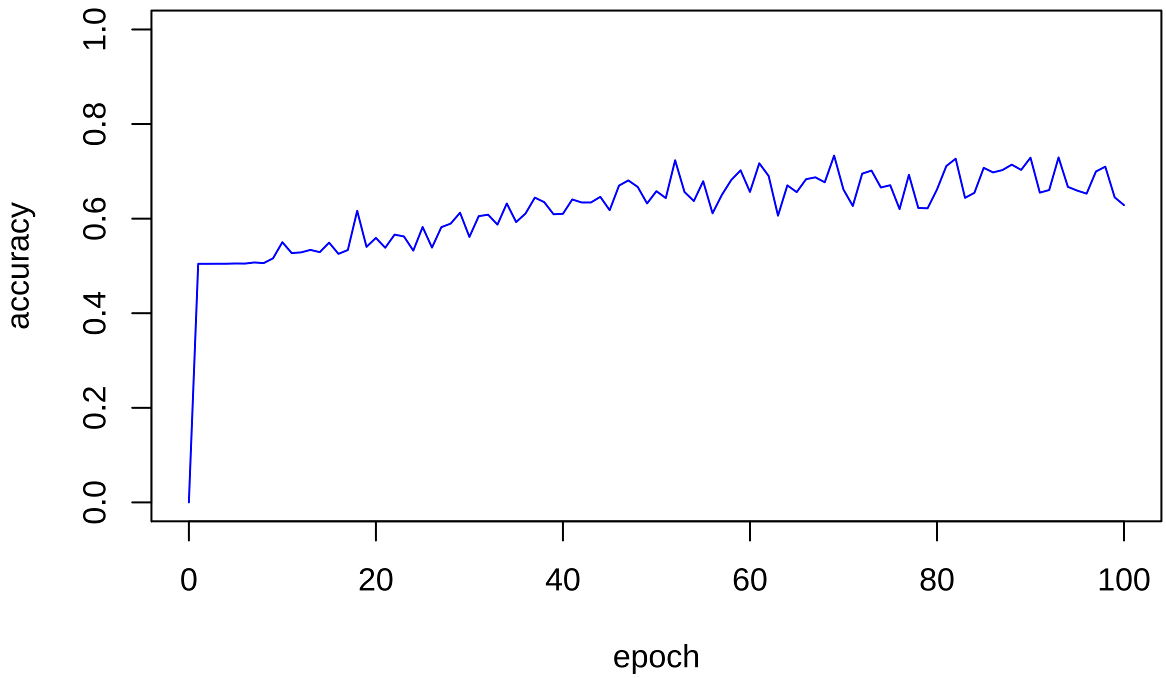
```
## [1] 0.9866415
```

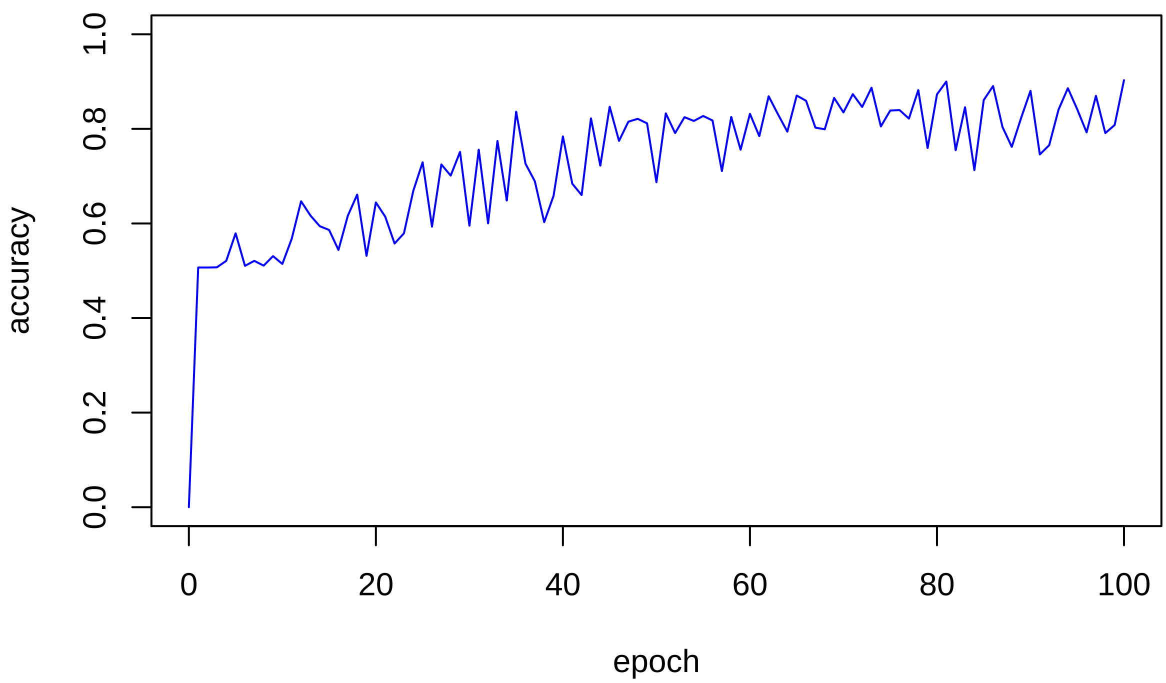
```
## [1] 0.9844665
```

The AUC of w^* is greater than the AUC of w'.

2

a





	Positive	Negative
Predicted Positive	789	0
Predicted Negative	193	1009

The test accuracy is $\frac{TP+TN}{TP+TN+FN+FP} = \frac{789+1009}{789+193+0+1009} = 0.9031$.

b

Based on my experiments, $\eta = 0.1$ seems to be the best and $\eta = 0.5$ also performs very close. The best way to decide which value to use for η is create a range of η and calculate some sort of loss function and find a value to optimize the loss function.