CSI 5325 Assignment 1

Sadia Nasrin Tisha

8th February 2022

1 Online exercises

2 Problem 1.1

According to Bayes' Rule, if the first ball is black then the probability of picking second ball black is

$P(Black2|Black1) = \frac{P(Black1|Black2)P(Black2)}{P(Black1)} = \frac{P(Black2 \cap Black1)}{P(Black1)}$

Here, Bag1 has two black balls and Bag2 has one black and one red ball. So the probability of picking one bag is:
$P(Bag1) = 1/2$
$P(Bag2) = 1/2$

So, the probability of picking black ball first is:
$P(Black1) = P(Black1|Bag1)P(Bag1) + P(Black1|Bag2)P(Bag2)$
$=2/2 \times 1/2 + 1/2 \times 1/2$
$=1/2 + 1/4$
$=3/4$

$P(Black2 \cap Black1) = P((Black2 \cap Black1)|Bag1)P(Bag1) + P((Black2 \cap Black1)|Bag2)P(Bag2)$
$=2/2 \times 1/1 \times 1/2 + 1/2 \times 0/1 \times 1/2 = 1/2$

So, $P(Black2|Black1) = \frac{P(Black2 \cap Black1)}{P(Black1)} = \frac{1/2}{3/4} = 2/3$

Hence, the probability that second ball is also black is 2/3.

3 Problem 1.2

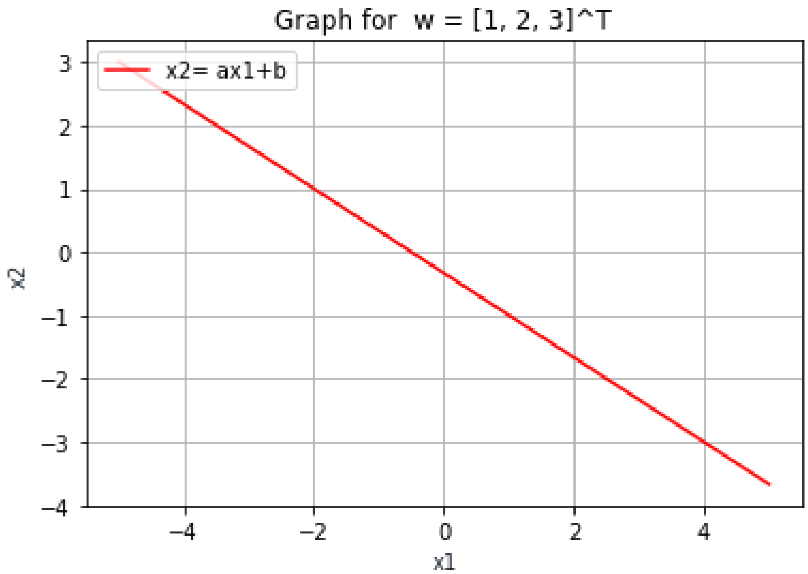
Given that, $h(x) = sign(w^T x)$
Here,
$w = [w_0, w_1, w_2]^T$
$x = [1, x_1, x_2]^T$

1. Consider $h(x) = +1$ (resp.-1), implies that $w^T x > 0$. So the equation of line, separating two region is,
   $w^T x = 0$
   $\Rightarrow w_0 + w_1 x_1 + w_2 x_2 = 0$

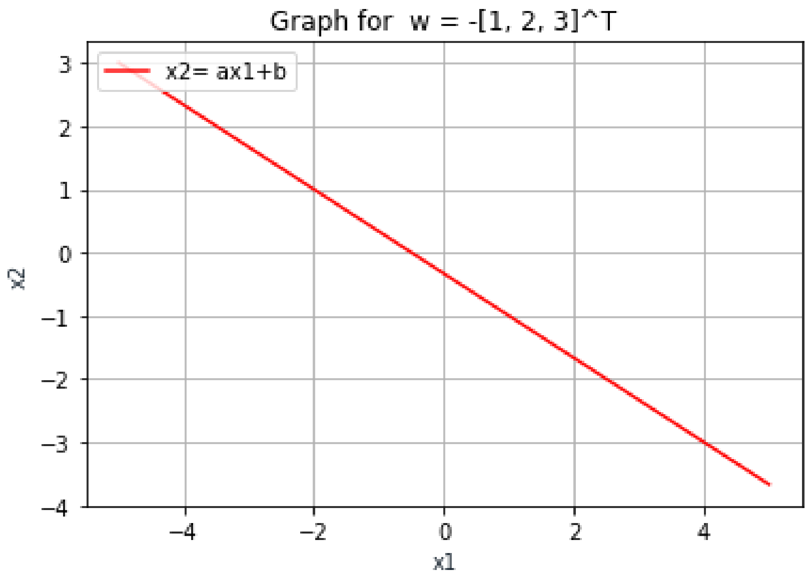
$\Rightarrow \mathrm{w}_0 + w_1 x_1 = -w_2 x_2$
If this line is separated by the given equation $x_2 = ax_1 + b$, then the slope, $a = -\frac{w_1}{w_2}$ and intercept, $b = -\frac{w_0}{w_2}$.

2. For case, $w = [1,2,3]^T$ and $w = -[1,2,3]^T$ the individual plot is generated by python.

For,$w = [1,2,3]^T$ the graph is:



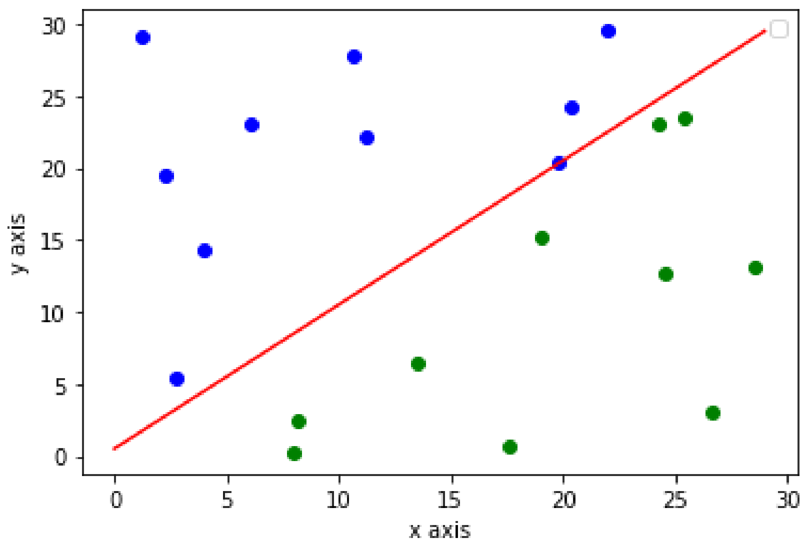
For,$w = -[1,2,3]^T$ the graph is:



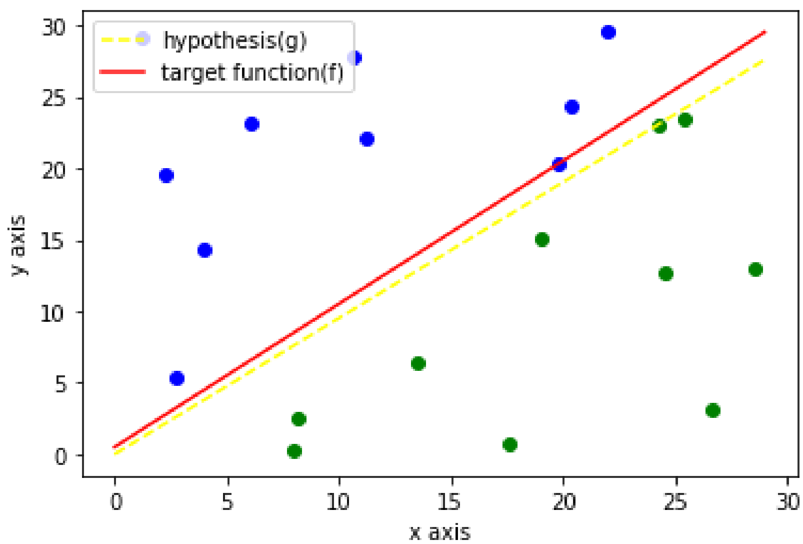
From the above two graph of $h(x) = +1$ and $h(x) = +1$, we can see both of them are identical but the regions are opposite. For the plot of $w = [1,2,3]^T$, the positive region is above the line and negative region is below the line. For plot of $w = [1,2,3]^T$, it is opposite. The positive region is below the line and negative region is above the line

4 Problem 1.4

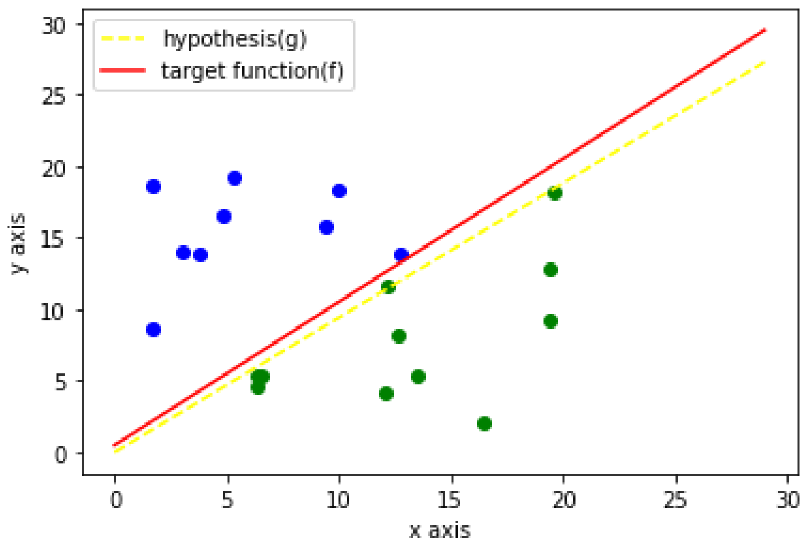
1. The linearly separable data set of size 20. Here the target function f on a plane is indicated as red mark.



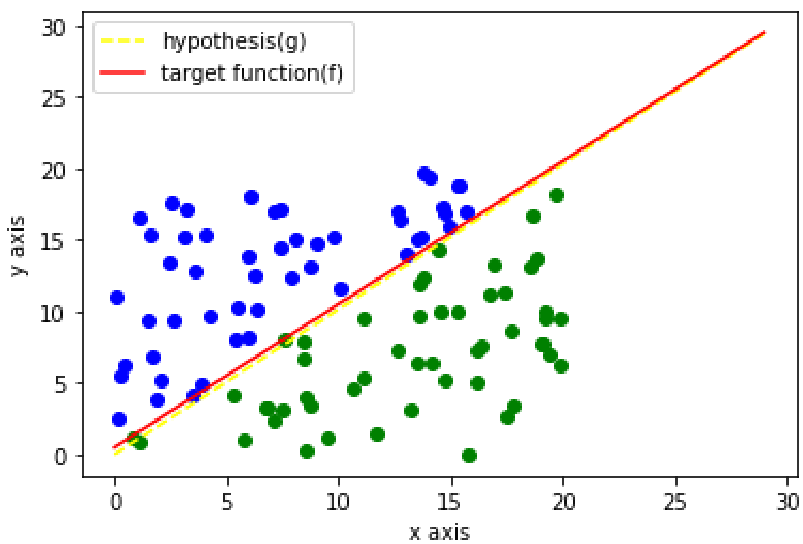
2. The linearly separable data set of size 20. Here the target function (f) on a plane is indicated as red mark and final hypothesis (g) is indicated as yellow line which is generated by perceptron learning algorithm. Here, from the figure, we can see that g is very close to f but not identical. But both of the line linearly separate data set of size 20 and the number of update value for converge is 2104.



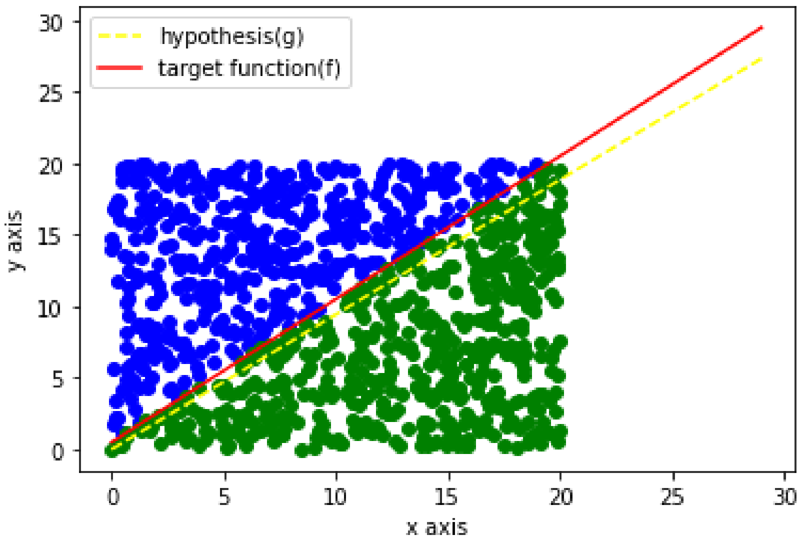
3. Here the data set of size 20 is randomly generated with different random value of 20 where in (2), it was 30 and also changed the seed value. After generating the data size and applying perceptron learning, we can see that g is very close to f but not identical. But both of the line linearly separate data set of size 20 and the number of update value for converge is 1996 which is lower than(2).



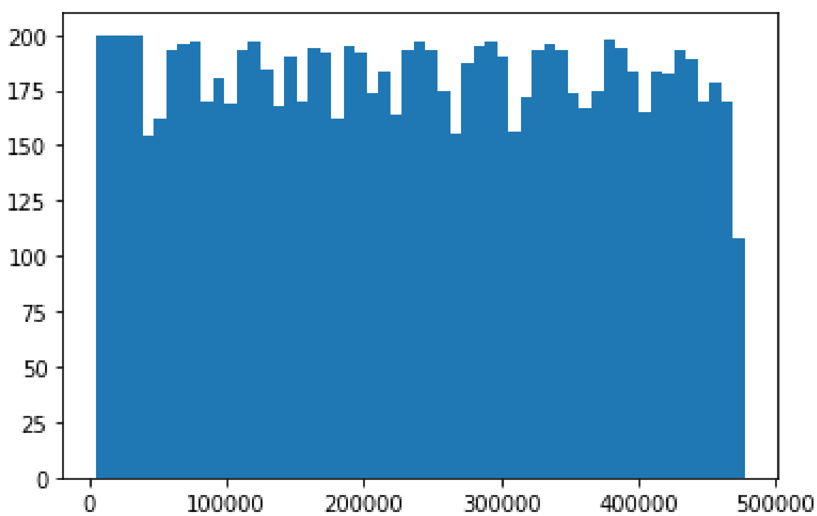
4. For the linearly separable data set of size 100. Here the target function (f) on a plane is indicated as red mark and final hypothesis (g) is indicated as yellow line which is generated by perceptron learning algorithm. Here, from the figure, we can see that g is very close to f and almost identical where both of the line linearly separate data set of size 100. Here, same number of iteration and random values are applied for both 20 and 100 sample. So in compare with 20 sample data, the perceptron algorithm works well for 100 sample and the number of update value for converge is 10398.



5. For the linearly separable data set of size 1000. Here the target function (f) on a plane is indicated as red mark and final hypothesis (g) is indicated as yellow line which is generated by perceptron learning algorithm. Here, from the figure, we can see that g is very close to f and not identical. Here, same number of iteration and random values are applied for both 20, 100, and 1000 sample. So in compare with 1000 sample data, the perceptron algorithm works slightly well for 20 and 100 sample data. In the figure we can also notice that some green data points are above the hypothesis g line. So so data point doesnot linearly separate for g. In case of target function f it shows that it linearly separate 1000 samples. Here the number of update value for converge is 107072.



6. For generate rendom dataset in a linearly separable data set of size 1000 with $x_n \in R^{10}$ instead of R^2 , there need total 441392 updates to converge which is a huge number.
7. Below histogram plot shows the number of updates for 100 experiments that the algorithm takes to converge. Here we can see the update value starts from 4295 and ends at 464358 and the updates are slightly uniform.



8. Here, N defines the data points and d defines dimension. The greater data point we get, the g become more approximate to f and also the run time increases. In case of dimension(d), the greater dimension also cause greater runtime.

5 Problem 1.6

1. Here, given that, probability of red marbel is μ.
   Probability of getting no red marbles from a sample of 10 independently drawn marbles, $P(v = 0)=$ $(1 - \mu)^{10}$

When, $\mu = 0.05$, $P(v = 0)$= $(1 - 0.05)^{10}$= 0.5987
When, $\mu = 0.5$, $P(v = 0)$= $(1 - 0.5)^{10}$= 9.7656×10^{-4}
When, $\mu = 0.8$, $P(v = 0)$= $(1 - 0.8)^{10}$= 1.024×10^{-7}

2. Here, given that, probability of red marbel is μ.
   Probability of having at least one sample ($v = 0$) from a sample of 1000 independently drawn marbles,
   P(at least one sample have $v = 0$)
   $= 1 - (no\ sample\ have\ v = 0)$
   $=1 - (1 - atleast\ one\ sample\ have\ v = 0)^{1000}$
   $=1 - (1 - (1 - \mu)^{10})^{1000}$

   When, $\mu = 0.05$, P(at least one sample have $v = 0$)= $1 - (1 - (1 - 0.05)^{10})^{1000}$= 1
   When, $\mu = 0.5$, P(at least one sample have $v = 0$)= $1 - (1 - (1 - 0.5)^{10})^{1000}$= 0.623576
   When, $\mu = 0.8$, P(at least one sample have $v = 0$)= $1 - (1 - (1 - 0.8)^{10})^{1000}$= 1.0239×10^{-4}

3. According to (2),
   Probability of having at least one sample ($v = 0$) from a sample of 1000000 independently drawn marbles,
   P(at least one sample have $v = 0$) $=1 - (1 - (1 - \mu)^{10})^{1000000}$

   When, $\mu = 0.05$, P(at least one sample have $v = 0$)= $1 - (1 - (1 - 0.05)^{10})^{1000000}$= 1
   When, $\mu = 0.5$, P(at least one sample have $v = 0$)= $1 - (1 - (1 - 0.5)^{10})^{1000000}$= 1
   When, $\mu = 0.8$, P(at least one sample have $v = 0$)= $1 - (1 - (1 - 0.8)^{10})^{1000000}$= 0.09733