Assignment 01

Mushfika Rahman

February 08, 2022

1 Learning From Data exercises

1.1 Bayes Theorem

Let us define some events,

$P(B_2)$= probability of choosing second black ball

$P(B_1)$= probability of choosing first black ball

$P(Bag_2)$= probability of choosing second bag

$P(Bag_1)$= probability of choosing first bag

$P(B_2|B_1)$= probability of choosing second ball as black given that first ball is black

$P(B_1|Bag_1)$= probability of choosing first ball as black given that it is chosen from first bag

$P(B_1|Bag_1)$= probability of choosing first ball as black given that it is chosen from second bag

$P(B_2 \cap B_1|Bag_1)$= probability of choosing first and second ball as black given that is chosen from first bag

$P(B_2 \cap B_1|Bag_2)$= probability of choosing first and second ball as black given that is chosen from second bag

By using the Naive Bayes Theorem we get,

$$P(B_2|B_1) = \frac{P(B_1|B_2)P(B_2)}{P(B_1)} = \frac{P(B_2 \cap B_1)}{P(B_1)} \tag{1}$$

Now, probability of choosing first black ball is,

$$P(B_1) = P(B_1|Bag_1)P(Bag_1) + P(B_1|Bag_2)P(Bag_2) = \frac{2}{2} * \frac{1}{2} + \frac{1}{2} * \frac{1}{2} = \frac{3}{4} \tag{2}$$

$$P(B_2 \cap B_1) = P(B_2 \cap B_1|Bag_1)P(Bag_1) + P(B_2 \cap B_1|Bag_2)P(Bag_2) = \frac{1}{2} \tag{3}$$

From eqn (1), by using the value from (2) and (3) we find that,

$$P(B_2|B_1) = \frac{2}{3} \tag{4}$$

1.2

(a) We are given a perceptron in two dimensional $h(x) = \text{sign}\ (w^T x)$, where $w = [w_0, w_1, w_2]^T$, $x = [1, x_1, x_2]^T$. The regions $h(x) = +1$ and $h(x) = -1$ are seperated by a line. The regions are determined whether $w^T x > 0$ falls in suppoe $+1$ region otherwise in different region. The line seperating the regions are $w^T x = 0$, we can deduce further that,

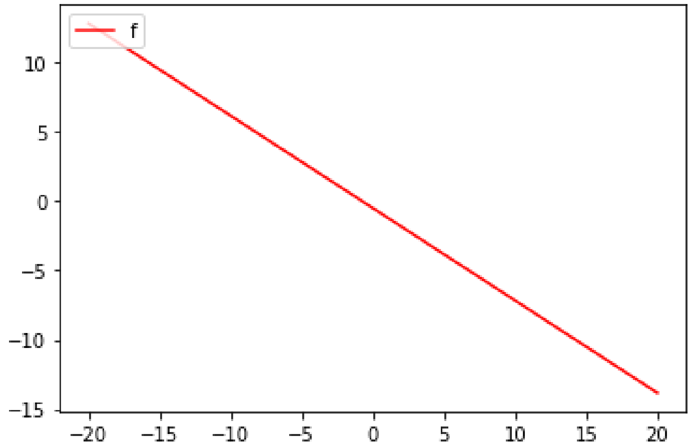
$$w_0 + w_1 x_1 + w_2 x_2 = 0 \tag{5}$$

We are given line equation as,

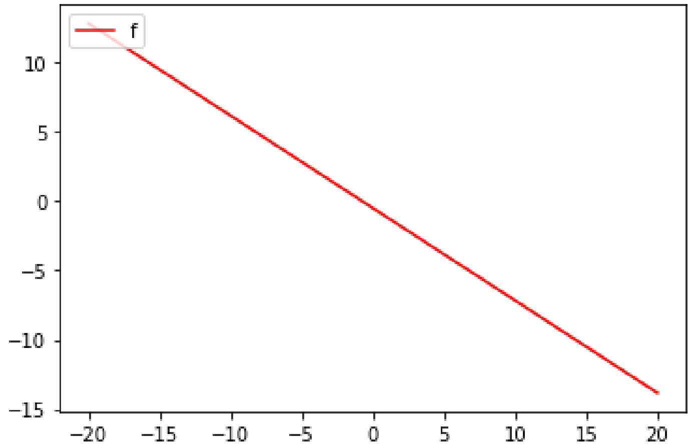
$$x_2 = a x_1 + b \tag{6}$$

So the slope a $=-\frac{w_1}{w_2}$ and intercept b $= -\frac{w_0}{w_2}$

(b) For $w = [1, 2, 3]^T$, we see the line



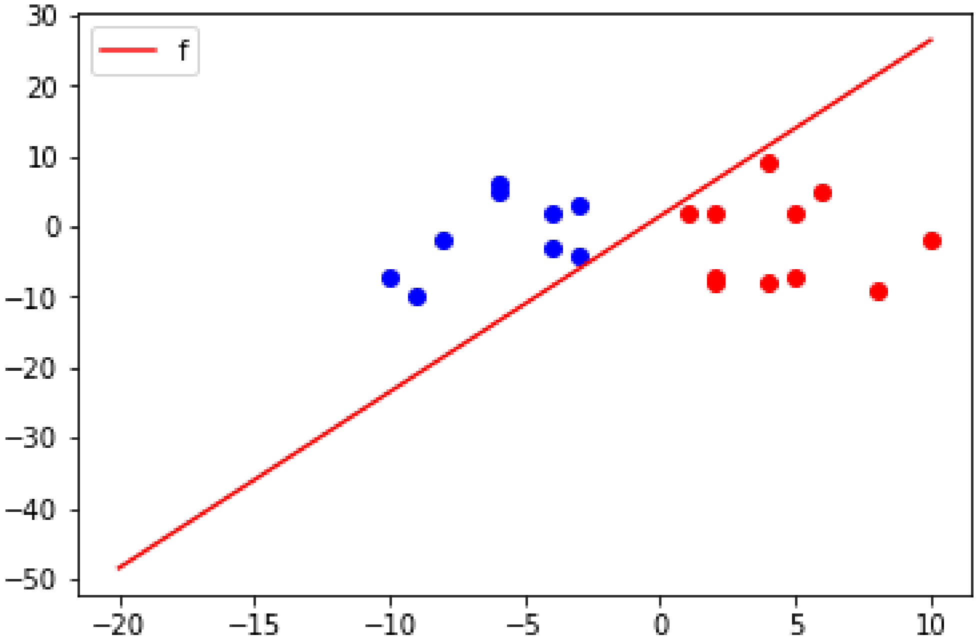
For $w = -[1, 2, 3]^T$, we see the line



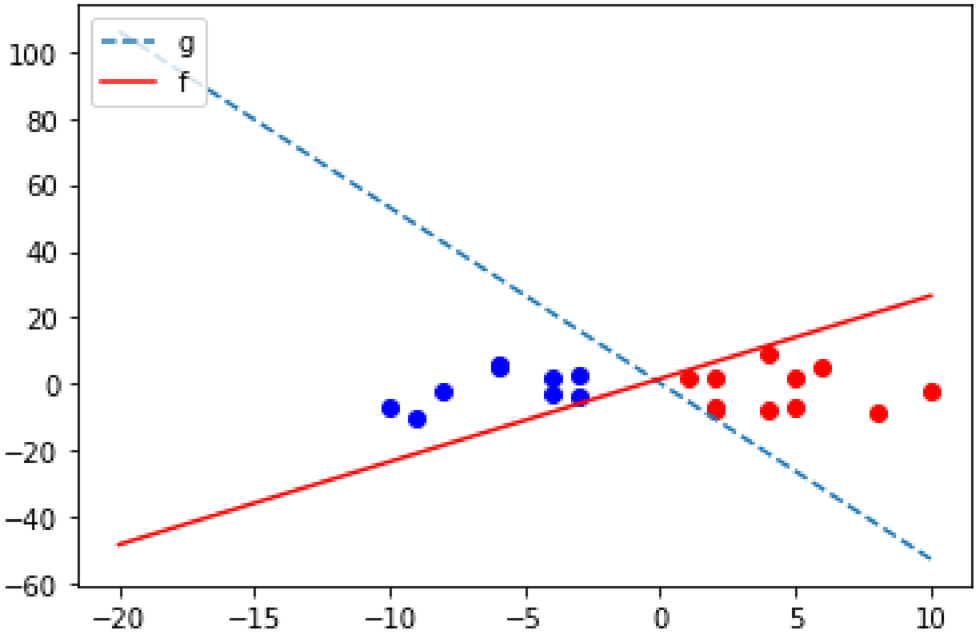
The above lines are identical to each other. The proportion to the co-efficient are equal that was they are identical. However, they signs of the proportions are opposite to each other. That means the regions they will indicate will be opposite. That impiles that, suppose the region $+1$ is on the right side of the line 1 and for line 2 it would be the left side.

1.4

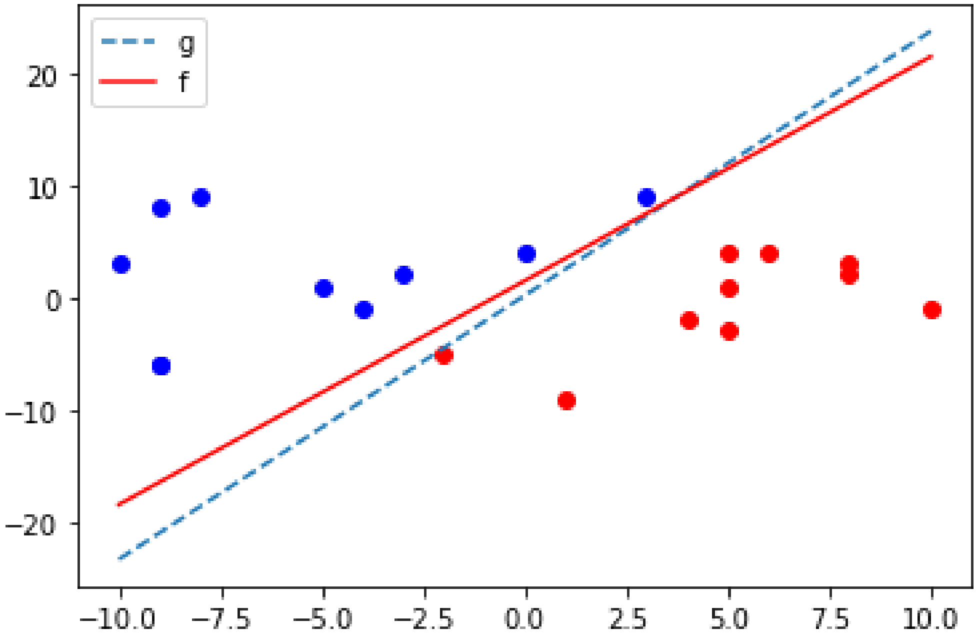
(a) Linear seperable line for a dataset size 20 according to Excercise 1.4. The target function f is in red color. The red points are +1 labels and blue points are 0 points.



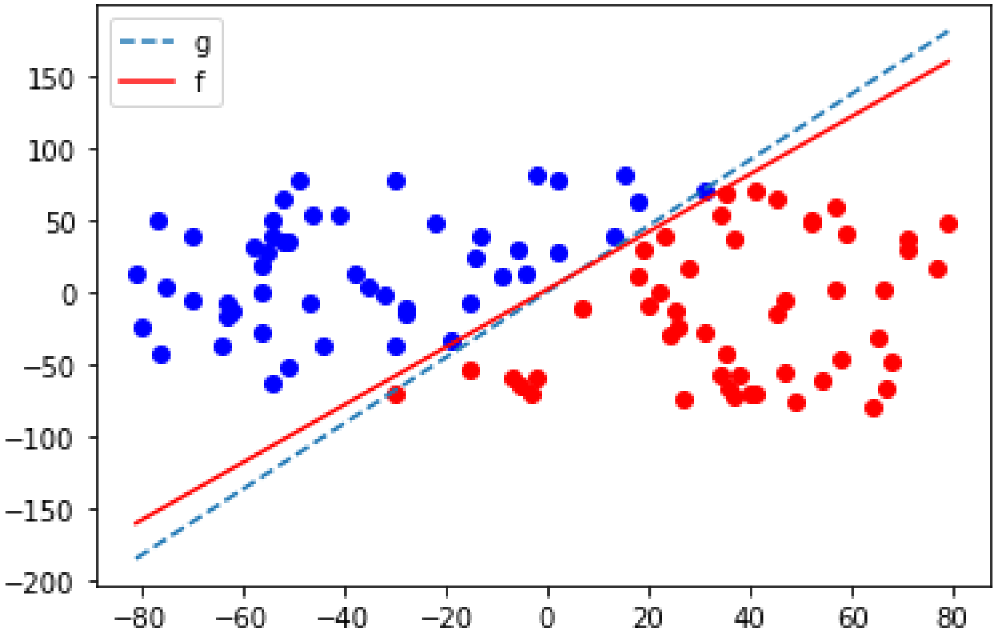
(b) The number of updates required before converging was 9. The hypothesis is different than the target function, the lines intersect each other. Here the hypothesis was able to correctly classify two regions.



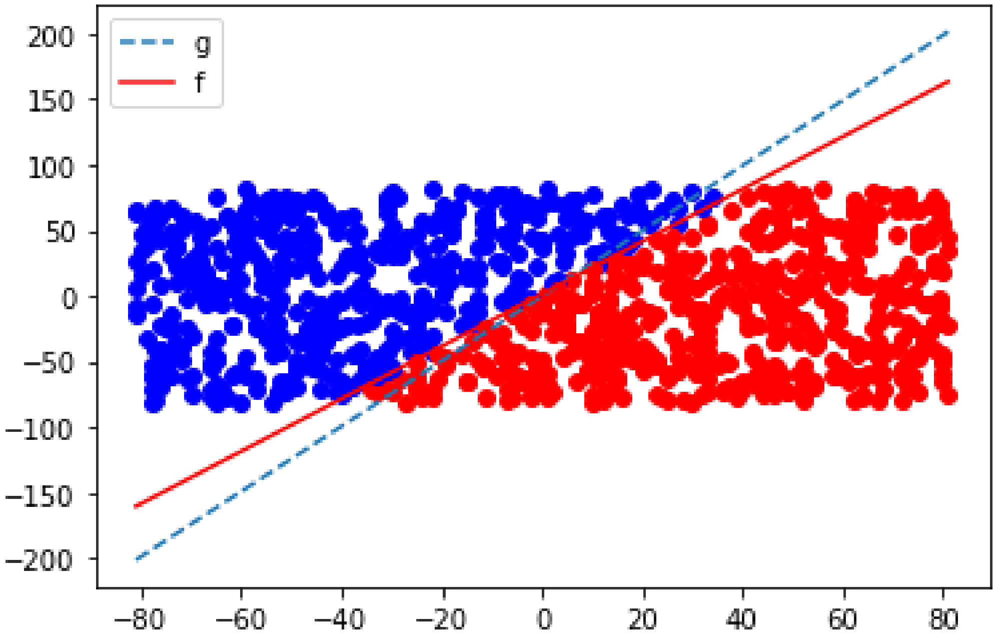
(c) The number of updates required before converging was 19. In here the hypothesis g is almost identical with the target function. It was able to classify all data points into two regions. The hypothesis is different than (b) because it had different set of data points.



(d) The number of updates required before converging was 43. The hypothesis g was almost identical with the target function. It was able to classify all data points into two regions. In (b) we can observe that all the point's visibility clearly into two regions, in here some points were just on the line for the blue class labels.

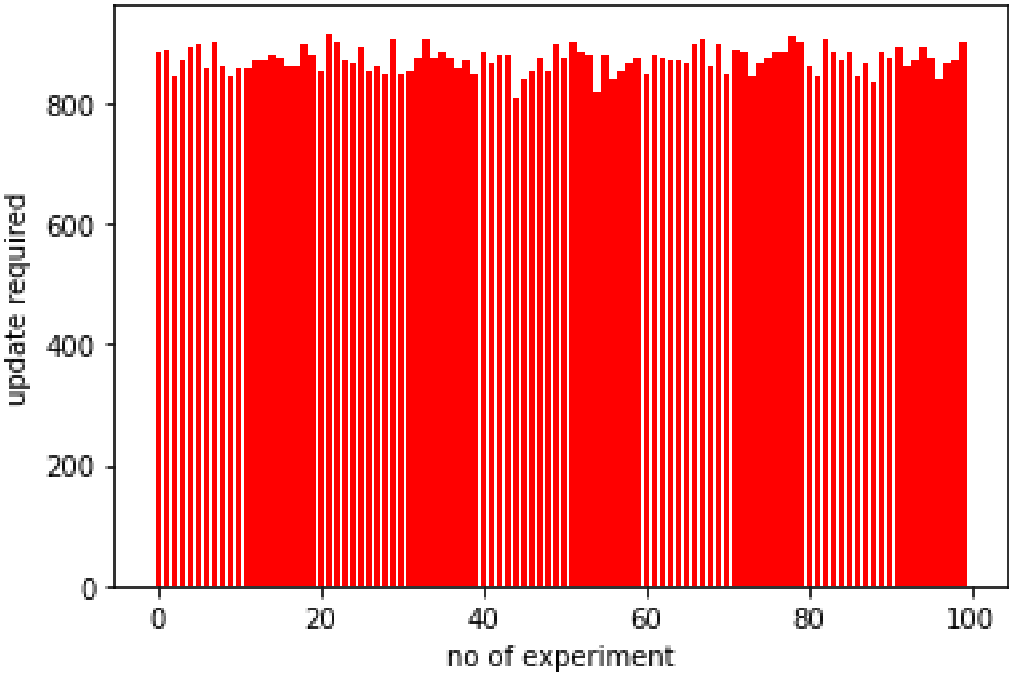


(e) The number of updates required before converging was 523. The hypothesis was very close to the target function. It misclassified some data where in (b) it was able to classify all the points perfectly.



(f) The number of updates required before converging was 492.

(g) It ran through 100 experiments. It took a long time to converge. The minimum step required to converge was around 807, and the maximum step required was 915.



(h) Let's define N and d. N is the number of samples and d is the dimensions in each sample. When the dimension was 2 and the sample size was small for 20 the algorithm converged quickly and it was able to classify the datapoints correctly. When the sample size grew it required more time converge and when the size of the sample was 1000 it misclassified some datapoints.The reason for that could be the data points were very dense. When the dimension increased the algorithm took longer time to converge. According to my view, with small sample size and dimension of 2 the algorithm showed faster convergence.The hypothesis function was very close to the target function also with increased sample size.

1.6

(a) If we define, P(E^c) as probability of an event occurring then P(E^c) as event not occurring. We know that,

$$P(E^c) = 1 - P(E) \tag{7}$$

For one sample which consists of 10 marbles we can write,

$$P(v = 0) = (1 - \mu)^{10} \tag{8}$$

By using above equation, for $\mu = 0.05$, P(v=0) = 0.5987

By using above equation, for $\mu = 0.5$, P(v=0) = $9.7656X10^{-4}$

By using above equation, for $\mu = 0.8$, P(v=0) = $1.024X10^{-7}$

(b) We know that, P(Atleast one sample has v=0) = 1- P(all the samples has no event where v=0)

= 1 - $\prod_{i=1}^{1000}[1 - P(v_i = 0)]$

= 1 - $\prod_{i=1}^{1000}[1 - (1 - \mu)^{10}]$

= 1 - [1- $(1 - \mu)^{10}]^{1000}$

By using above equation, for $\mu = 0.05$, P(Atleast one sample has v=0)= 1 - [1- $(1 - 0.005)^{10}]^{1000}$

By using above equation, for $\mu = 0.5$, P(Atleast one sample has v=0) = 1 - [1- $(1 - 0.05)^{10}]^{1000}$

By using above equation, for $\mu = 0.8$, P(Atleast one sample has v=0) = 1 - [1- $(1 - 0.8)^{10}]^{1000}$

(c) By repeating everything in (b) for 1000000 samples.

We can write, P(Atleast one sample has v=0) = 1 - [1- $(1 - \mu)^{10}]^{1000000}$

By using above equation, for $\mu = 0.05$, P(Atleast one sample has v=0)= 1 - [1- $(1 - 0.05)^{10}]^{1000000}$

By using above equation, for $\mu = 0.5$, P(Atleast one sample has v=0) = 1 - [1- $(1 - 0.5)^{10}]^{1000000}$

By using above equation, for $\mu = 0.8$, P(Atleast one sample has v=0) = 1 - [1- $(1 - 0.8)^{10}]^{1000000}$