Assignment 1

Alberto Mejia
RIN: 661514960

CSCI 4100 - Machine Learning from Data

September 8, 2019

1. (100) LFD Exercise 1.3

(a) We know that $h(x) = \text{sign}(w^T(t)x(t))$. Given that x(t) is misclassified by w(t), that means $y(t) \neq \text{sign}(w^T(t)x(t))$ which means $y(t) \neq h(x)$. They will always have opposite signs.
The update rule then is $w(t + 1) = w(t) + y(t)x(t)$
There are two ways x(t) can be misclassified:

- $y(t) = -1$ thus $h(x) = 1$, that means $y(t)w^T(t)x(t) < 0$
- $y(t) = 1$ thus $h(x) = -1$, that means $y(t)w^T(t)x(t) < 0$

$\therefore y(t)w^T(t)x(t) < 0$

(b) $y(t)w^T(t+1)x(t) > y(t)w^T(t)x(t)$

$y(t)w^T(t+1)x(t) - y(t)w^T(t)x(t) > 0$	Subtract term over
$y(t)x(t)(w^T(t+1) - w^T(t)) > 0$	Factor out
$y(t)x(t)(w(t+1) - w(t))^T > 0$	Take out Transpose
$y(t)x(t)([w(t) + y(t)x(t)] - w(t))^T > 0$	Substitute in Update rule
$y(t)x(t) \times y(t)x(t)^T > 0$	Apply Transpose(y(t) is a scalar)
$y^2(t)x(t)x(t)^T > 0$	

The 1st element, x(0), is = 1 which means $x(t)x(t)^T > 0$. The square of y will always be positive thus $y^2(t)x(t)x(t)^T > 0$ so this makes sense.

We also know that
$y(t)w^T(t+1)x(t) - y(t)w^T(t)x(t) > 0$
Which means $y(t)w^T(t)x(t)$ must be smaller than $y(t)w^T(t+1)x(t)$

Vice versa, we also know that
$y(t)w^T(t+1)x(t) = y(t)x(t)(w(t) + y(t)x(t))^T = y(t)x(t)(w^T(t) + y(t)x(t)^T)$
$= y(t)x(t)w^T(t) + y(t)x(t)x(t)^T$
Since we already know what $y(t)x(t)x(t)^T > 0$, that means for the equation:
$y(t)w^T(t+1)x(t) = y(t)x(t)w^T(t) + y(t)x(t)x(t)^T$
The term $y(t)x(t)w^T(t)$, must strictly be smaller than $y(t)w^T(t+1)x(t)$ since we are always adding a positive number with $y(t)x(t)x(t)^T$.

$\therefore y(t)w^T(t+1)x(t) > y(t)w^T(t)x(t)$

(c) We can see from part (b) that the updated weight vector moves in the right direction since it moves towards a positive polarity. We can tell it is always moving in the right direction because in part (a), it does not matter if the result was negative, the results of part (b) is showing that the updated weight vectors are always strictly increasing.

2. (100) LFD Exercise 1.5

(a) Learning

(b) Design

(c) Learning

(d) Design

(e) Learning

3. (100) LFD Exercise 1.6

(a) This can be supervised learning. The data can be the users history of bought books, helping us understand what he/she likes.

(b) This can be reinforcement learning. The data can be games we played in the past versus opponents. These old games will help us learn our common mistakes and pitfalls and process to correct them in future games.

(c) This can be unsupervised learning. The data can be a set of movies and then we can watch the movies and try to determine their classification. The more movies we watch, the more we'd be able to adjust the classifications (Ex: "Oh, this one seems like a Thriller now that I've seen a lot of scary movies that I'd classify as Horror.")

(d) If the task is to learn how to play music without examples, it would be Unsupervised learning. However, if training examples are to be used, then the task can be labeled (pun intended) as Supervised learning. For example, watching a video of more experienced musicians playing, or reading articles on how to play that instrument would be learning from examples thus, Supervised Learning.

(e) This can be supervised learning. Our training examples can include the customer's past information, which can include whether the customer has any current debt and if she has paid if off. Over time, this data could include dates of when the debt was incurred and when it was payed off.

4. (100) LFD Exercise 1.7

(a) Based on the table given, we can determine that the hypothesis that always returns '•' is what agrees with the data the most. Comparing this g with the possible choices for f ∈ {f_1, ..., f_8}, we have 1 that agrees with g on all three test points; 3 that agree with g on two test points; 3 that agree with g on one test point; 1 that agrees with g on none of the test points. Here is the breakdown:

- Agreed with g on 3 points: {f_8}
- Agreed with g on 2 points: {f_4, f_6, f_7}
- Agreed with g on 1 point: {f_2, f_3, f_5}
- Agreed with g on 0 points: {f_0}

(b) Now we use the hypothesis that always returns '○' since it is what agrees with the data the least. Comparing this g with the possible choices for f ∈ {f_1, ..., f_8}, we have 1 that agrees with g on all three test points; 3 that agree with g on two test points; 3 that agree with g on one test point; 1 that agrees with g on none of the test points. Here is the breakdown:

- Agreed with g on 3 points: {f_1}
- Agreed with g on 2 points: {f_2, f_3, f_5}
- Agreed with g on 1 point: {f_4, f_6, f_7}
- Agreed with g on 0 points: {f_8}

(c) Now our hypothesis is equal to {XOR} where XOR(x) = '•' if the number of 1's in x is odd and XOR(x) = '○' if the number of 1's in x is even. Comparing this g with the possible choices for f ∈ {f_1, ..., f_8}, we have 1 that agrees with g on all three test points; 3 that agree with g on two test points; 3 that agree with g on one test point; 1 that agrees with g on none of the test points. Here is the breakdown:

- Agreed with g on 3 points: {f_2}
- Agreed with g on 2 points: {f_1, f_4, f_6}
- Agreed with g on 1 point: {f_3, f_5, f_8}
- Agreed with g on 0 points: {f_7}

(d) We can see that all of our hypothesis contain the same performance. We will now use a hypothesis that disagrees the most with XOR where XOR(x) = '•' if the number of 1's in x is odd and XOR(x) = '○' if the number of 1's in x is even. Comparing this g with the possible choices for f ∈ {f_1, ..., f_8}, we have 1 that agrees with g on all three test points; 3 that agree with g on two test points; 3 that agree with g on one test point; 1 that agrees with g on none of the test points. Here is the breakdown:

- Agreed with g on 3 points: {f_7}
- Agreed with g on 2 points: {f_3, f_5, f_8}
- Agreed with g on 1 point: {f_1, f_4, f_6}

- Agreed with g on 0 points: $\{f_2\}$

5. (200) LFD Problem 1.1

Bayes' Theorem states that P[A and B] = $P[A|B] \cdot P[B] = P[B|A] \cdot P[A]$ or more commonly known, $P(A|B) = P(A \cap B)/P(B)$

Let B be the event the 1st ball we choose is black.
Let A be the event the 2nd ball we choose is black.

Let the bag with 2 black balls be B_1 and the bag with 1 black and 1 white ball be B_2.
The probability of choosing the B_1 is 50%, AKA 1/2. Now, the probability of choosing a black ball in B_1 is 100%, so $(1/2) \cdot (1) = 1/2$.

Now, if we were to choose B_2, the chance of pulling a black ball is 50%, with the 50% chance of choosing B_2, we get $(1/2) \cdot (1/2) = 1/4$.

Thus, P(B) is $(1/2) + (1/4) = 3/4$.

The probability that we choose 2 black balls in a row given we choose B_1 is 1
$P(A \cap B|B_1) = 1$
$P(B_1) = 1/2$

The probability that we choose 2 black balls given we choose B_2 is 0 (There is only 1 black ball).
$P(A \cap B|B_2) = 0$
$P(B_2) = 1/2$

Now that both bags are considered, the probability we get two black balls is

$$\begin{aligned} P(A \cap B) &= P(A \cap B|B_1) \cdot P(B_1) + P(A \cap B|B_2) \cdot P(B_2) \\ &= (1) \cdot (1/2) + (0) \cdot (1/2) \\ &= 1/2 \end{aligned}$$

Now, we can consider the probability that we get a 2nd black ball given that the 1st ball we picked is black. We can substitute in our values into Bayes' Theorem:

$$\begin{aligned} P(A|B) &= P(A \cap B)/P(B) \\ &= \frac{1/2}{3/4} \\ &= (1/2) \cdot (4/3) \\ &= 4/6 \\ &= 2/3 \end{aligned}$$

6. (200) LFD Problem 1.2

(a) $h(x) = \text{sign}(w^T x)$ where $w = [w_0, w_1, w_2]^T$ and $x = [1, x_1, x_2]^T$.

AKA $h(x) = \text{sign}(w_0 + w_1 x_1 + w_2 x_2)$

We know that $h(x)$ can only be -1 or +1

If $h(x) = +1$, that means $w^T x > 0$ and if $h(x) = -1$ that means $w^T x < 0$

This means that the line that separates the two is $w^T x = 0$

Expanded out, $w^T x = 0$ is $w_0 + w_1 x_1 + w_2 x_2 = 0$

If we express this line by the equation $x_2 = a x_1 + b$

We get

$w_0 + w_1 x_1 + w_2 x_2 = 0$

$w_2 x_2 = -w_1 x_1 - w_0$

$w_2 x_2 = -1(w_1 x_1 + w_0)$

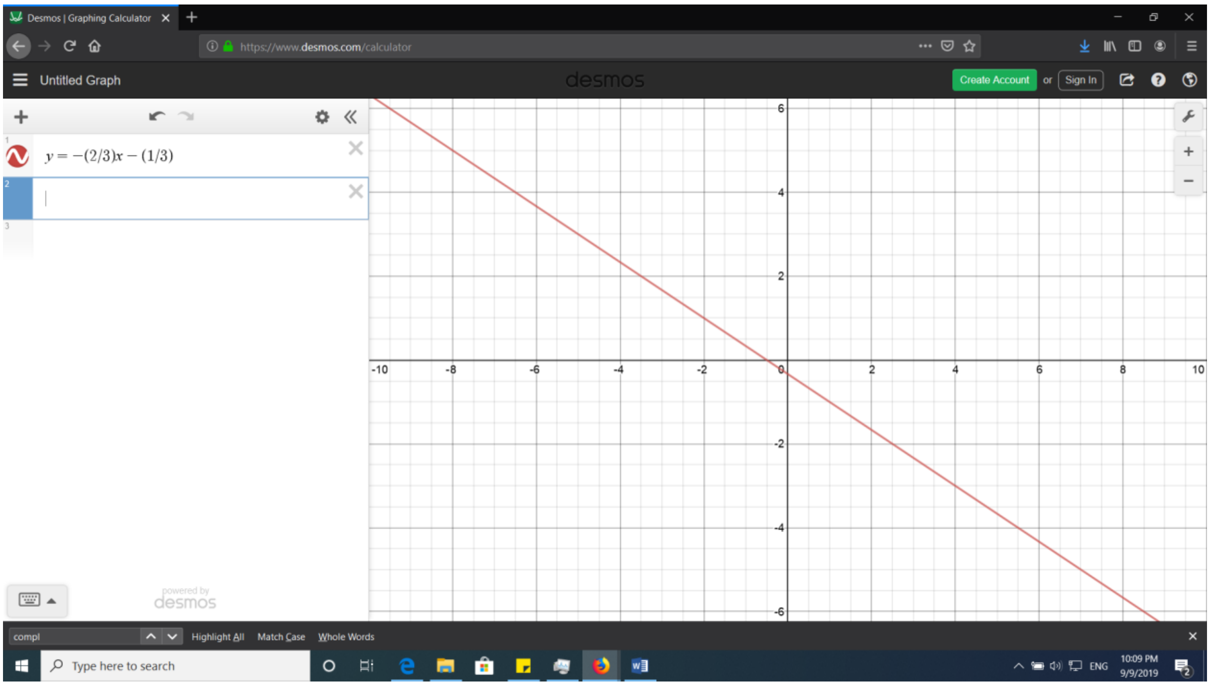
$x_2 = -1[(w_1 x_1 + w_0)]/w_2$

$x_2 = -1((w_1/w_2) \cdot x1 + w_0/w_2)$

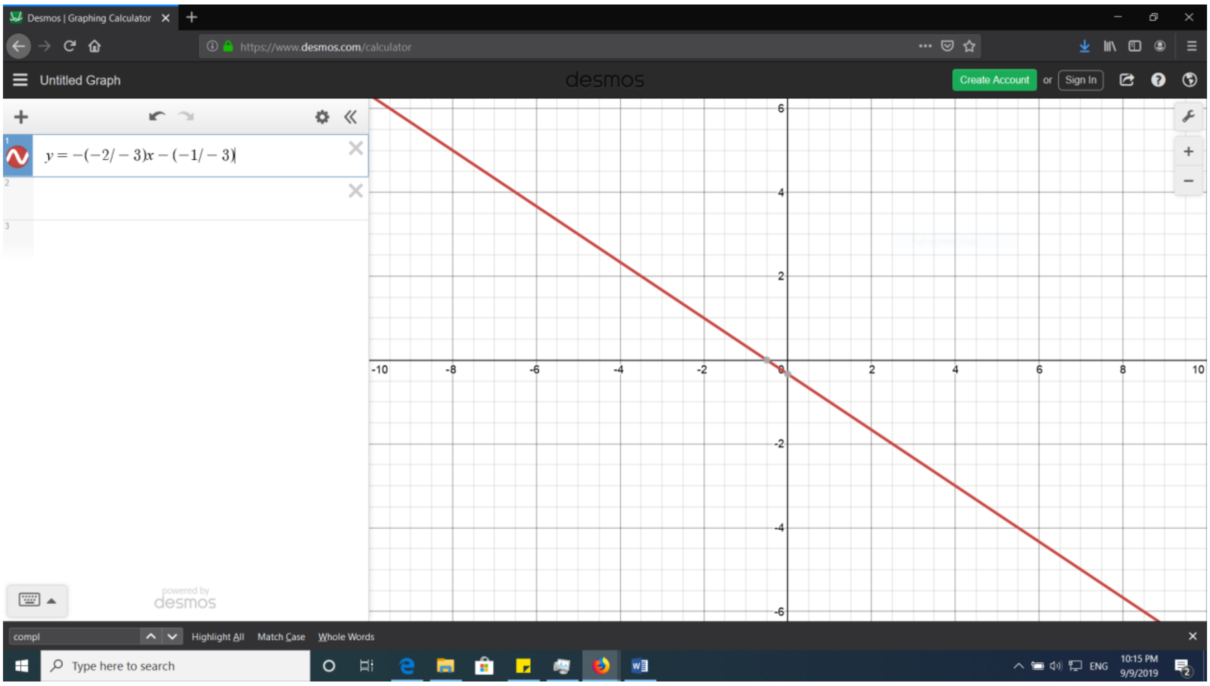
$x_2 = -(\frac{w_1}{w_2})x_1 - (\frac{w_0}{w_2})$

Thus, $a = -(\frac{w_1}{w_2})$ and $b = -(\frac{w_0}{w_2})$

(b) $x_2 = -(\frac{2}{3})x_1 - (\frac{1}{3})$ where $w = [1, 2, 3]^T$. The graph is the following:

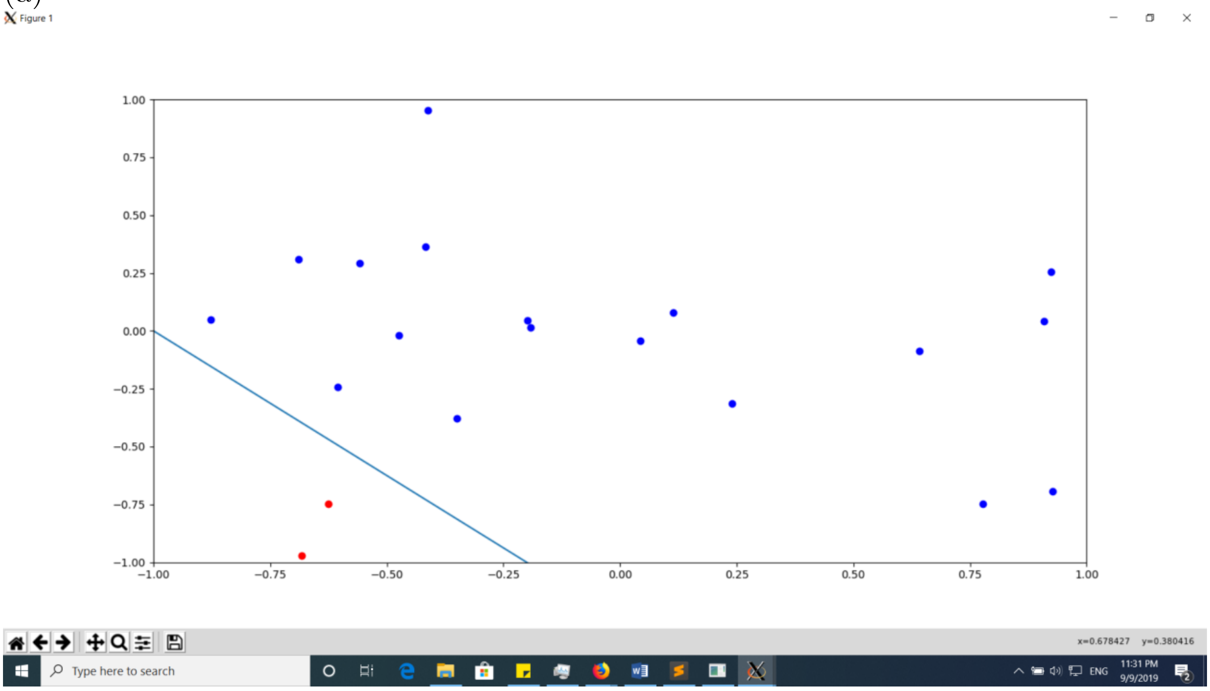


$x_2 = -(\frac{-2}{-3})x_1 - (\frac{-1}{-3})$ where $w = -[1, 2, 3]^T$. The graph is the following:



7. (200) LFD Problem 1.4 (a – e)

(a)



(b)

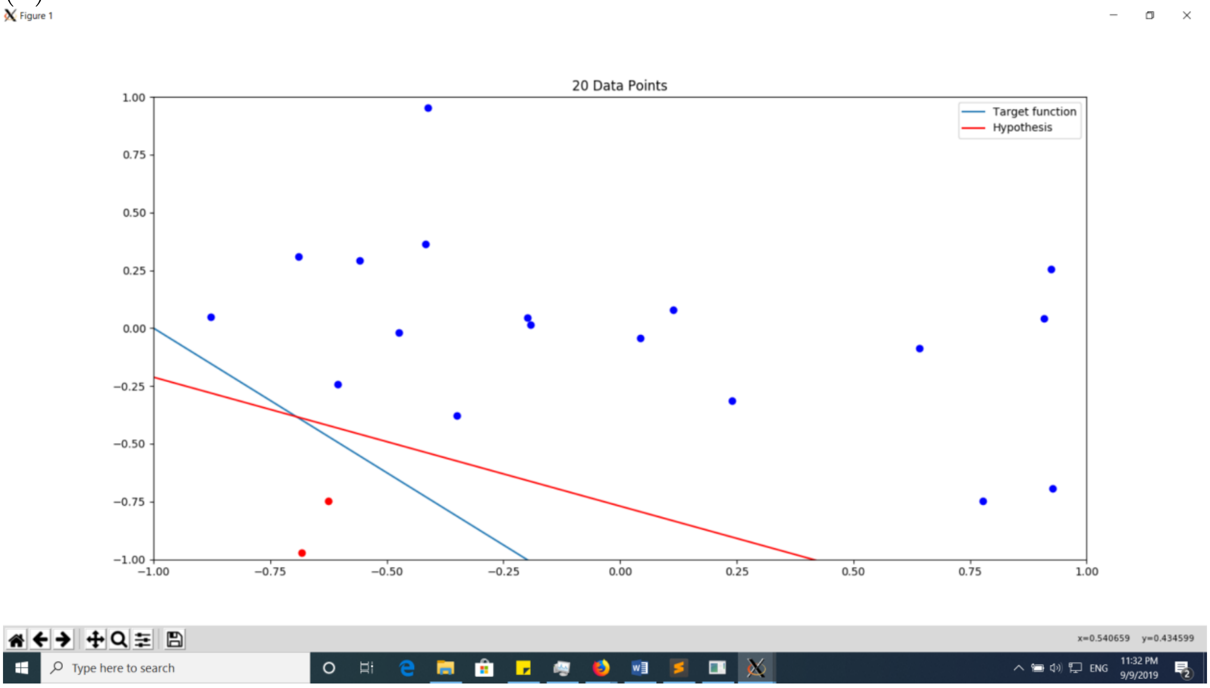


Figure 1
20 Data Points
Target function
Hypothesis
x=0.540659 y=0.434599
Type here to search
11:32 PM
9/9/2019

(c)

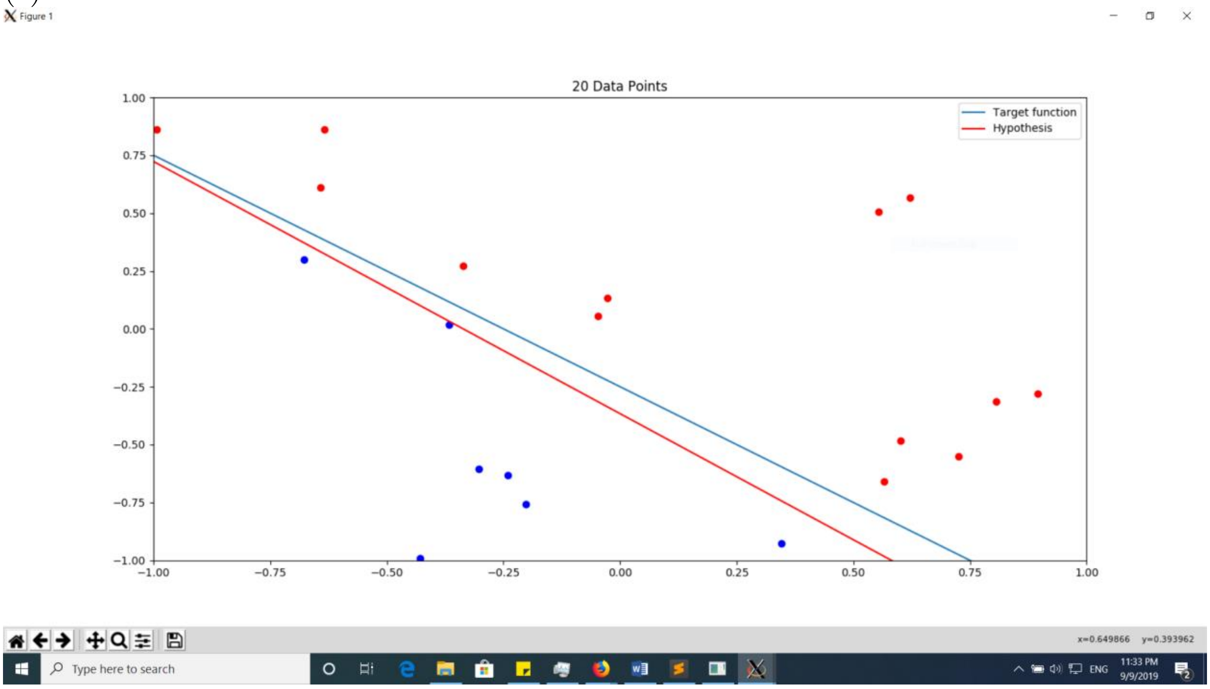


Figure 1
20 Data Points
Target function
Hypothesis
x=0.649866 y=0.393962
Type here to search
11:33 PM
9/9/2019

(d)

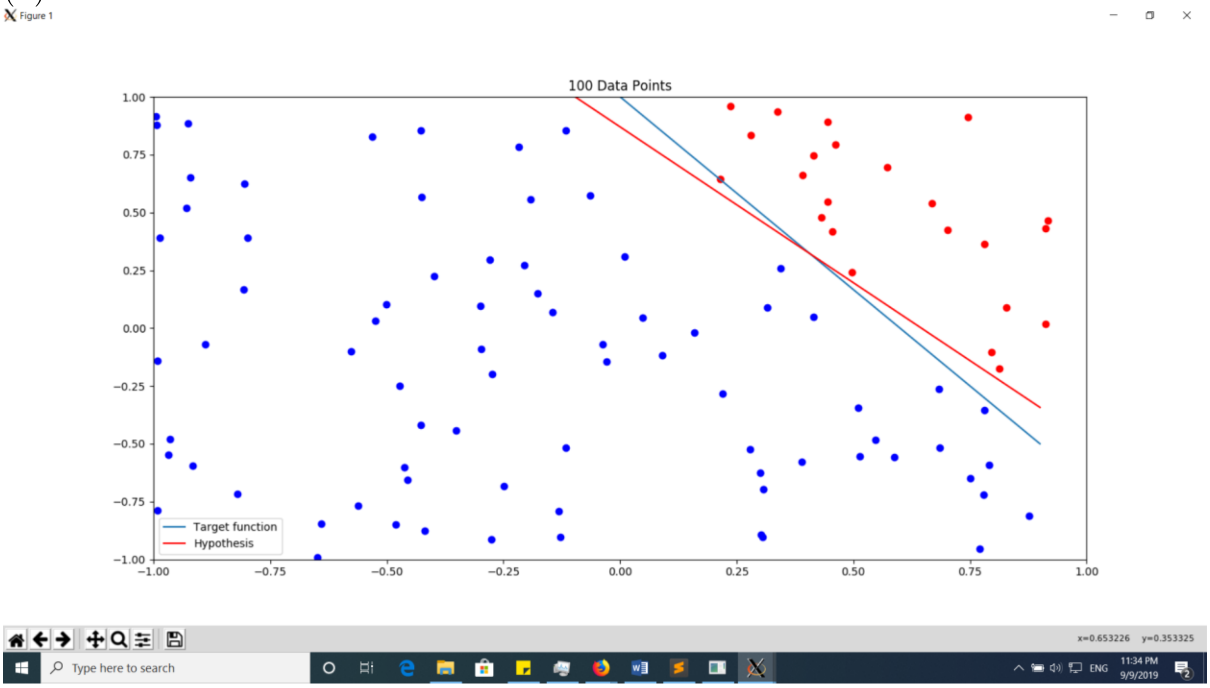


Figure 1
100 Data Points
Target function
Hypothesis
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
x=0.653226 y=0.353325
Type here to search
ENG
11:34 PM
9/9/2019

(e)

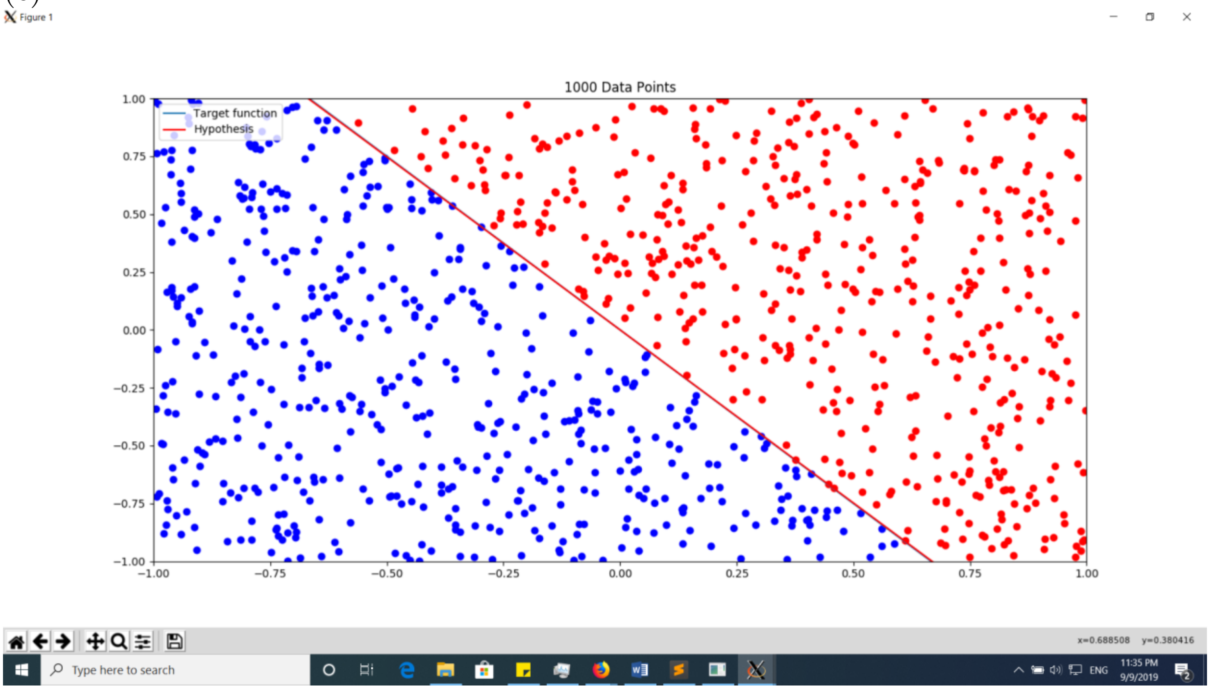


Figure 1
1000 Data Points
Target function
Hypothesis
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
x=0.688508 y=0.380416
Type here to search
ENG
11:35 PM
9/9/2019