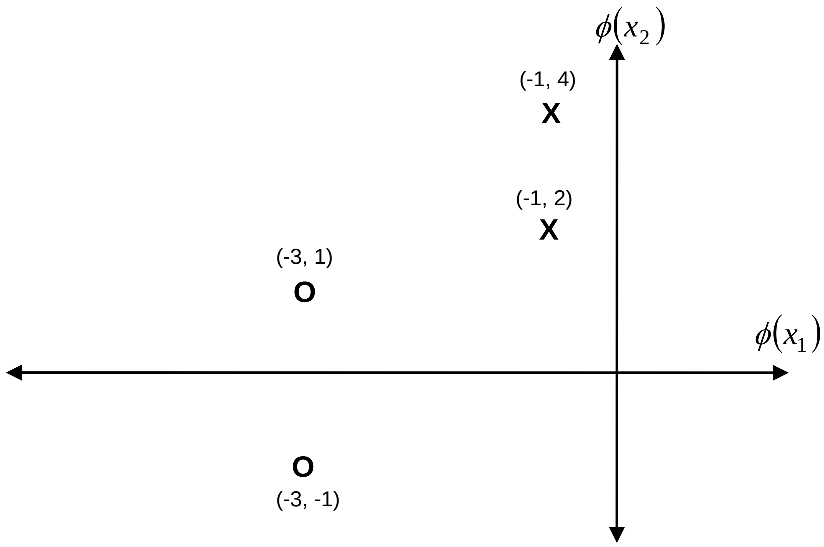
German University in Cairo
Faculty of Media Engineering and Technology
Winter 2019

CSEN 1022 – Machine Learning

Problem Set #4 Solution

Problem 1

For the data given below, apply the perceptron update rule to find the weight vector **w** of the decision boundary. Show the output of each iteration till convergence. Assume that the weight vector is initialized $\mathbf{w}^{(0)} = [3\ 1]$. Use learning rate parameter $\eta = 0.1$.

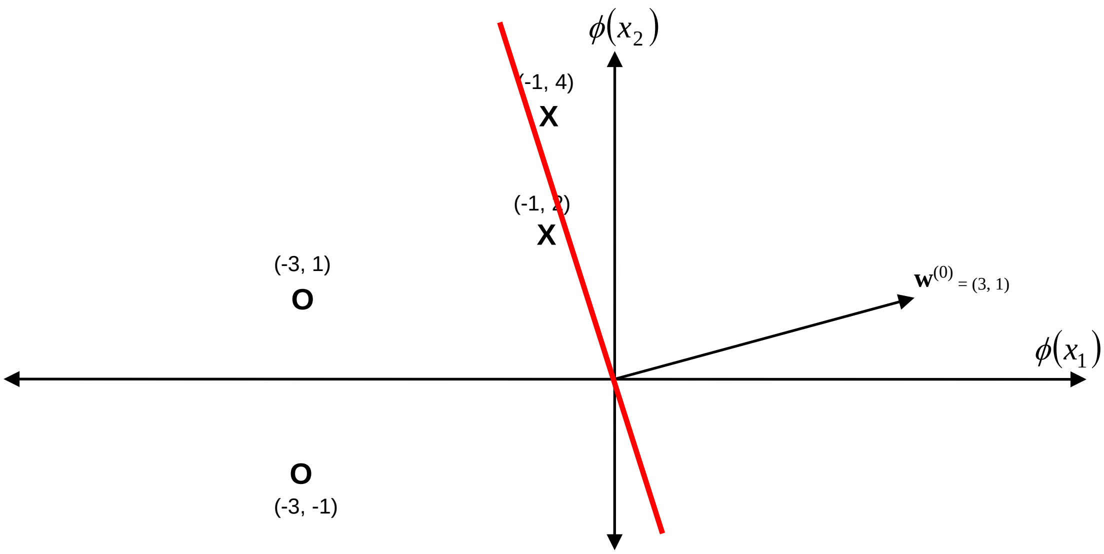


Solution:

The initial weight vector is shown below and the corresponding decision boundary $w_1x_1 + w_2x_2 = 0$ shown in red (which is perpendicular to the weight vector). For $\mathbf{w}^{(0)}$, the decision boundary is $3x_1 + x_2 = 0$. Let the target value t for points in the **X** class be 1 and for the **O** class be -1.

CSEN 1022 – Machine Learning

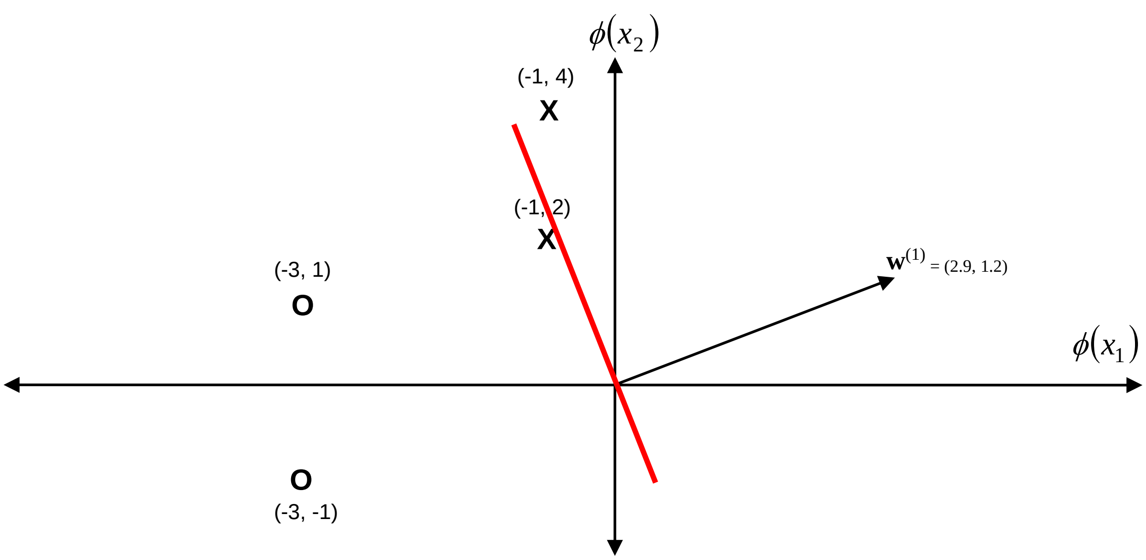
Problem Set #4 Solution



For this choice of weight vector, the point (-1, 2) is miss-classified

For iteration 1: $\mathbf{w}^{(1)} = \mathbf{w}^{(0)} + 0.1\phi(\mathbf{x}_n)t_n$ where $\mathbf{x}_n$ in this case is the miss-classified point (-1, 2) where $t_n = 1$

$$\mathbf{w}^{(1)} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} + 0.1\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2.9 \\ 1.2 \end{bmatrix}$$

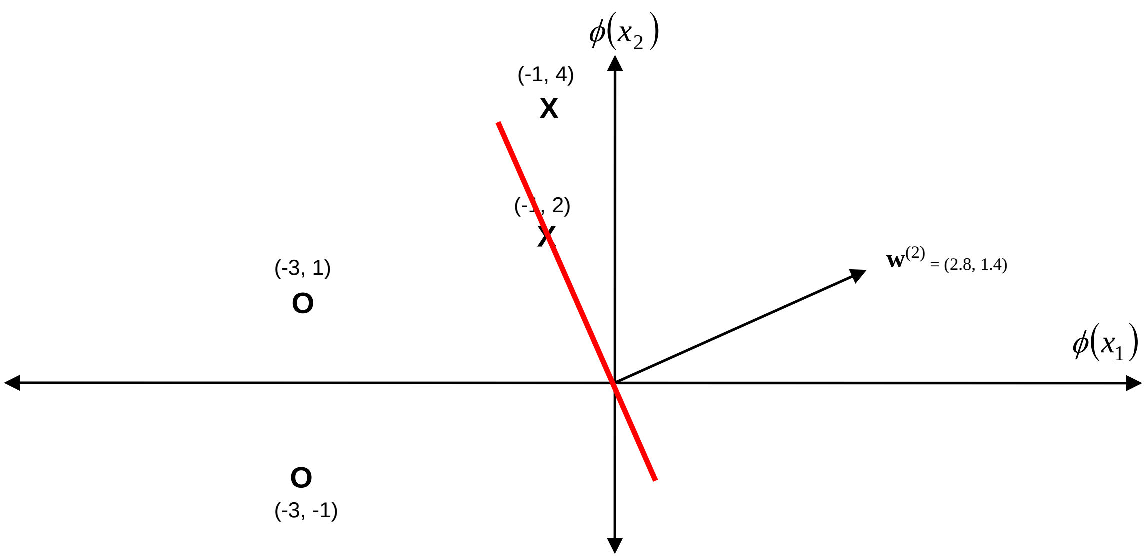


CSEN 1022 – Machine Learning

Problem Set #4 Solution

For iteration 2: $\mathbf{w}^{(2)} = \mathbf{w}^{(1)} + 0.1\phi(\mathbf{x}_n)t_n$ where $\mathbf{x}_n$ in this case is the miss-classified point (-1, 2) where $t_n = 1$

$$\mathbf{w}^{(2)} = \begin{bmatrix} 2.9 \\ 1.2 \end{bmatrix} + 0.1\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2.8 \\ 1.4 \end{bmatrix}$$

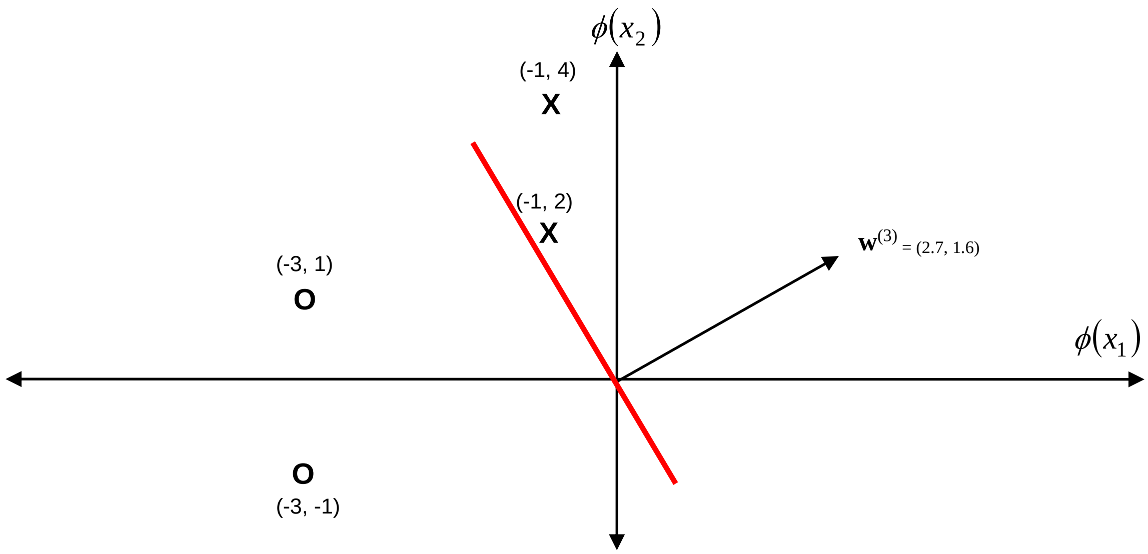


For iteration 3: $\mathbf{w}^{(3)} = \mathbf{w}^{(2)} + 0.1\phi(\mathbf{x}_n)t_n$ where $\mathbf{x}_n$ in this case is the miss-classified point (-1, 2) where $t_n = 1$

$$\mathbf{w}^{(3)} = \begin{bmatrix} 2.8 \\ 1.4 \end{bmatrix} + 0.1\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2.7 \\ 1.6 \end{bmatrix}$$

CSEN 1022 – Machine Learning

Problem Set #4 Solution



Now all points are classified correctly, so the algorithm stops.

CSEN 1022 – Machine Learning

Problem Set #4 Solution

Problem 2

i- Consider the dataset shown in Figure 1 for the weight and height of two classes representing male and female persons. The initial random weight vector of a perceptron results in the decision boundary (straight line) shown in the figure. **Explain** the steps of the perceptron training algorithm that will be taken to correctly classify the data where the weight update rule of the perceptron is given by $\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} + \eta \mathbf{x}_n t_n$. Do not solve numerically. Just explain the steps.

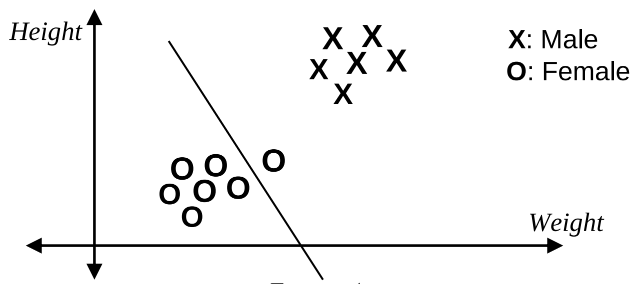


Figure 1

Solution:

The perceptron training algorithm will follow the steps below to correctly classify the data:

1. It will loop over the data points and stop when a misclassified point ($\mathbf{x}_n$) is found. In the above example, the perceptron will stop at the point **O** that is misclassified in class **X**.
2. It will update the weight vector using the update rule: $\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} + \eta \mathbf{x}_n t_n$ where $\mathbf{x}_n$ is the misclassified point found in step 1.
3. It will loop again over the data points. If a misclassified point is found (the same point detected in step 1 or any other point), it will go back to step 2. If no misclassified points are found, the perceptron will stop.

CSEN 1022 – Machine Learning

Problem Set #4 Solution

ii- Outlier points of one class are points that belong to the class but their values are significantly different from the majority of the class. For example, in the weight and height dataset, basketball players represent outliers compared to normal people as their height and weight are much higher than normal people. If the weight and height of two male basketball players is added to the data as shown in Figure 2, **explain** what effect this will have on the decision boundary obtained using the perceptron.

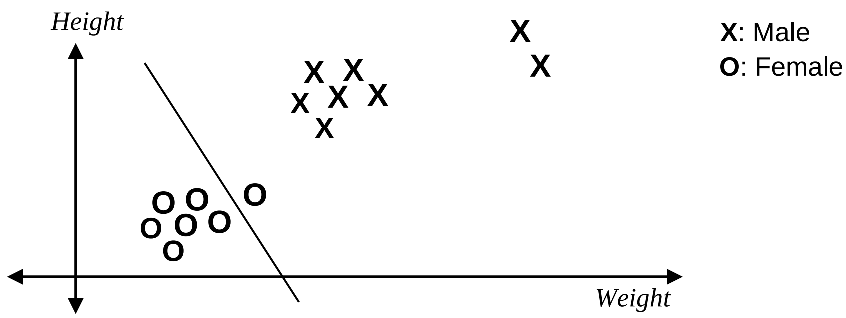


Figure 2

Solution:

The presence of outliers will not affect the decision boundary obtained by the perceptron. In the above example, the outliers are already classified correctly, so the steps showed in (i) will not change. Even if, in any other case, the outliers are misclassified, the perceptron will be able to classify them correctly. In other words, **the perceptron is robust to outliers.**

CSEN 1022 – Machine Learning

Problem Set #4 Solution

iii- If the weight and height of two female basketball players is added to the dataset as shown in Figure 3, **explain** what effect this will have on the decision boundary obtained using the perceptron.

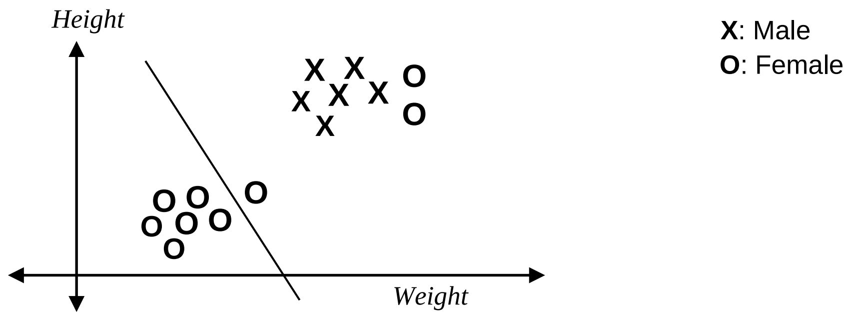


Figure 3

Solution:

The addition of the new data points made the data nonlinearly separable. Given that the perceptron is a linear classifier, it will not be able to classify nonlinearly separable data. It will keep repeating the steps explained in (i) given that it will always find a misclassified point. In this case, the perceptron will never converge.

CSEN 1022 – Machine Learning

Problem Set #4 Solution

Problem 3

Consider the data about conditions for playing tennis given below:

Day	Outlook	Temperature	Humidity	Play Tennis?
1	Sunny	Hot	High	No
2	Cloudy	Hot	High	**Yes**
3	Rain	Mild	High	No
4	Rain	Cool	Normal	No
5	Cloudy	Cool	Normal	**Yes**
6	Sunny	Mild	High	**Yes**
7	Sunny	Cool	Normal	**Yes**
8	Rain	Mild	Normal	No
9	Sunny	Mild	Normal	**Yes**
10	Cloudy	Mild	High	**Yes**

Using Naïve Bayes classifier, predict if one should play tennis on a Sunny, Hot with Normal humidity day. Note that since the data is discrete, you can use the frequentist statistics to compute the needed probabilities.

Solution:

Since the goal is to classify a sunny, hot with normal humidity day as good day to play tennis or not, we first define two classes C_1 and C_2, corresponding to Play = Yes and Play = No, respectively. To classify the given day with attributes $\mathbf{x}$, we need to compute $p(C_1 | \mathbf{x})$:

p(Play = Yes | Outlook = Sunny, Temperature = Hot, Humidity = Normal)

and $p(C_2 | \mathbf{x})$:

p(Play = No | Outlook = Sunny, Temperature = Hot, Humidity = Normal)

and find which conditional probability is larger. If the first one is larger, then our prediction is Play = Yes. If the second one is larger, then our prediction is Play = No. Note that $\mathbf{x}$ here is 3 dimensional corresponding to Outlook, Temperature and Humidity.

CSEN 1022 – Machine Learning

Problem Set #4 Solution

Since $p(C_1|\mathbf{x}) = \dfrac{p(\mathbf{x}|C_1)p(C_1)}{p(\mathbf{x})}$

We need to compute $p(\mathbf{x} \mid C_1)$ = p(Outlook = Sunny, Temperature = Hot, Humidity = Normal | Play = Yes). Using the Naïve Bayes assumption which assumes that the dimensions of the input data (the attributes of the day) are independent, we can re-write $p(\mathbf{x} \mid C_1)$ as $p(\mathbf{x}|C_1) = \prod_{i=1}^{D} p(x_i|C_1)$

p(Outlook = Sunny | Play = Yes) p(Temperature = Hot | Play = Yes) p(Humidity = Normal | Play = Yes)

Similarly, $p(\mathbf{x} \mid C_2)$ can be re-written as $p(\mathbf{x}|C_2) = \prod_{i=1}^{D} p(x_i|C_2)$

p(Outlook = Sunny | Play = No) p(Temperature = Hot | Play = No) p(Humidity = Normal | Play = No)

From the available data in the table and using frequentist statistics:

p(Outlook = Sunny | Play = Yes) = 3/6

p(Temperature = Hot | Play = Yes) = 1/6

p(Humidity = Normal | Play = Yes) = 3/6

p(Play = Yes) = 6/10

p(Outlook = Sunny | Play = No) = 1/4

p(Temperature = Hot | Play = No) = 1/4

p(Humidity = Normal | Play = No) = 2/4

p(Play = No) = 4/10

Therefore,

$p(\mathbf{x} \mid C_1)$ = p(Outlook = Sunny, Temperature = Hot, Humidity = Normal | Play = Yes) = (3/6) x (1/6) x (3/6)

And

$p(\mathbf{x} \mid C_2)$ = p(Outlook = Sunny, Temperature = Hot, Humidity = Normal | Play = No) = (1/4) x (1/4) x (2/4)

CSEN 1022 – Machine Learning

Problem Set #4 Solution

$$\frac{p(C_1|\mathbf{x})}{p(C_2|\mathbf{x})} = \frac{p(\mathbf{x}|Play = Yes)p(Play = Yes)}{p(\mathbf{x}|Play = No)p(Play = No)} = \frac{0.025}{0.0125} = 2$$

Since $\frac{p(C_1|\mathbf{x})}{p(C_2|\mathbf{x})} > 1$ then $\mathbf{x} \in C_1$. Therefore, one should play tennis in a sunny, hot with normal humidity day.