Week-9, Graded

Question-1

Statement

Assume that Perceptron algorithm is applied to a data set in which the maximum of the lengths of the data points is 4 and the value of margin (γ) of the optimal separator is 1. If the algorithm has made 10 mistakes at some point of the execution of the algorithm, which of the following can be valid squared length(s) of the weight vector obtained in the 11th iteration?

Options

(a)

90

(b)

150

(c)

190

Answer

(c)

Solution

The number of mistakes after t updates to the weight vector is:

$$\gamma^2 t^t \leq ||w^t||^2 \leq tR^2$$

For $t = 11$, we have:

$$121 \leq ||w^{11}||^2 \leq 11 \times 16$$

From this, we get:

$$121 \leq ||w^{11}||^2 \leq 176$$

From the options given, we see that $||w^{11}||^2 = 150$ is valid.

Question-2

Statement

Consider the following data set:

f_1	f_2	y
-1	-1	-1
0	1	+1
1	0	+1
1	1	+1

If Perceptron algorithm is applied on this data set with the weight vector initialized to [0, 0], how many times the weight vector will be updated during the training process?

Options

(a)

0

(b)

1

(c)

2

(d)

3

Answer

(b)

Solution

$w_0 = [0\ 0]$

$\mathbf{I_1}$:

The predictions for each of the four data points x_1, x_2, x_3, x_4 as per $w^T x$ will be +1, +1, +1, +1.

The mistake occurs for x_1

Hence, $w_1 = w_0 + x_1 y_1$

Resulting into $w_1 = [1\ 1]$

$\mathbf{I_2}$:

The predictions for each of the four data points x_1, x_2, x_3, x_4 as per $w^T x$ will be -1, +1, +1, +1, which are correct.

Hence, the weight update happened once.

Question-3

Statement

Consider the following data set with three data points:

$$\left(\begin{bmatrix}2\\2\end{bmatrix}, +1\right), \left(\begin{bmatrix}2\\-2\end{bmatrix}, +1\right), \left(\begin{bmatrix}-2\\1\end{bmatrix}, -1\right)$$

If the Perceptron algorithm is applied to this data with the initial weight vector w^0 to be a zero vector, what will be the outcome?

Options

(a)

The algorithm will converge with $w = \begin{bmatrix}-2\\1\end{bmatrix}$

(b)

The algorithm will converge with $w = \begin{bmatrix}2\\1\end{bmatrix}$

(c)

The algorithm will converge with $w = \begin{bmatrix}2\\-1\end{bmatrix}$

(d)

The algorithm will never converge.

Answer

(c)

Solution

$w_0 = [0\ 0]$

$\mathbf{I_1}$:

The predictions for each of the three data points x_1, x_2, x_3 as per $w^T x$ will be +1, +1, +1

The mistake occurs for x_3

Hence, $w_1 = w_0 + x_3 y_3$

Resulting into $w_1 = [2\ \ -1]$

$\mathbf{I_2}$:

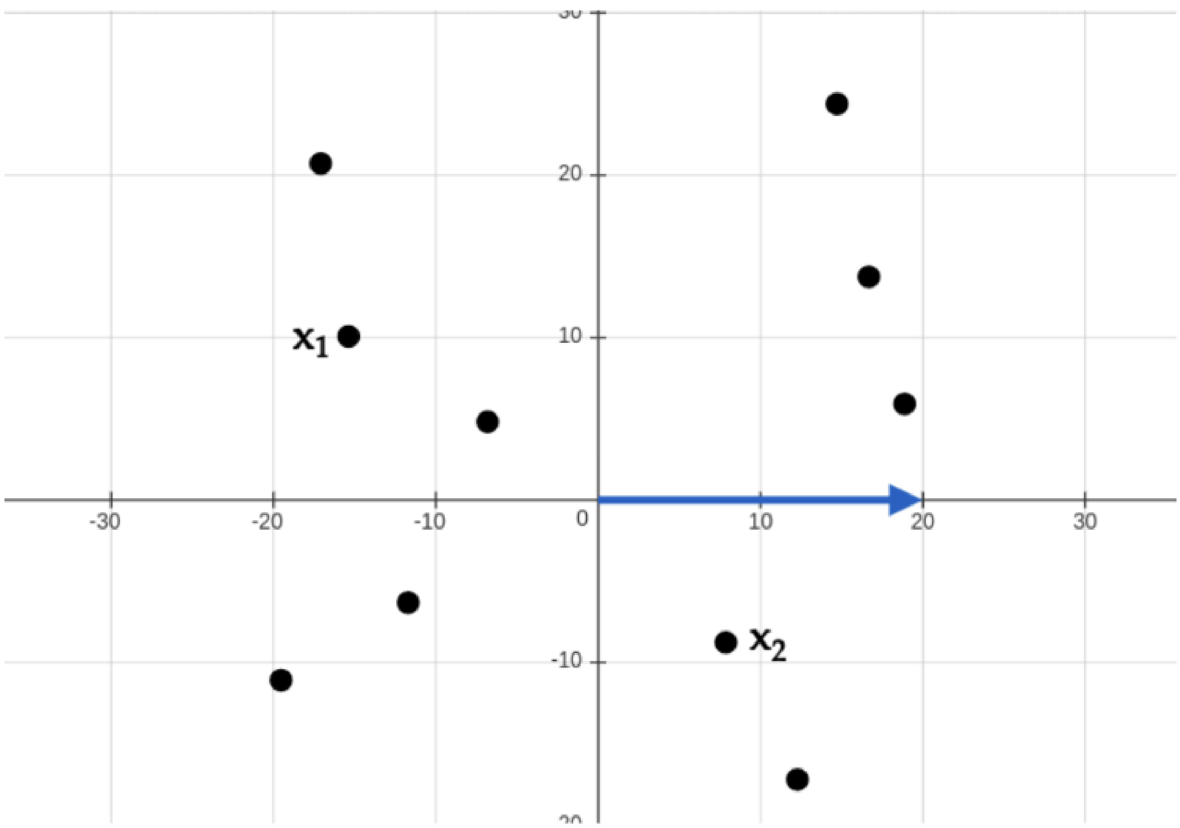
The predictions for each of the three data points x_1, x_2, x_3 as per $w^T x$ will be +1, +1, -1, which are correct.

Hence, the algorithm will converge with $w = [2 \ -1]$

Question-4

Statement

Consider ten data points as shown in the following image:



The blue line represents the weight vector. As per this weight vector, the Perceptron algorithm will predict which classes for the data points x_1 and x_2?

Options

(a)

$x_1 : +1, x_2 : -1$

(b)

$x_1 : -1, x_2 : +1$

(c)

$x_1 : +1, x_2 : +1$

(d)

$x_1 : -1, x_2 : -1$

Answer

(b)

Solution

The decision boundary will be perpendicular to w. For the data points on the right side of it, $w^T x$ will be greater than equal to zero, on the LHS, it will be less than zero.

Accordingly, RHS data points (and hence x_2) will be predicted as +1.

And, LHS data points (and hence x_1) will be predicted as -1.

Question-5

Statement

In the previous question, if the weight vector is multiplied by -1, which classes will be predicted by the Perceptron for the data points x_1 and x_2?

Options

(a)

$x_1 : +1, x_2 : -1$

(b)

$x_1 : -1, x_2 : +1$

(c)

$x_1 : +1, x_2 : +1$

(d)

$x_1 : -1, x_2 : -1$

Answer

(a)

Solution

If the weight vector is multiplied by -1, then for the data points on the RHS, $w^T x$ will be less than 0 and on the LHS, it will be greater than equal to zero.

Accordingly, LHS data points (and hence x_1) will be predicted as +1.

And, RHS data points (and hence x_2) will be predicted as -1.

Question-6

Statement

Given a data set with R = 4, γ = 2, what is the maximum number of mistakes that Perceptron algorithm can make on the data?

Options

(a)

2

(b)

4

(c)

8

(d)

16

Answer

(b)

Solution

The maximum number of mistakes is given by R^2/γ^2

Which is $\frac{4^2}{2^2} = 4$

Question-7

Statement

If the scores (i.e, $w^T x$ values) for some data points are -4, 3, 1, 2, -6 respectively, what will be the probabilities returned for these points by Logistic Regression?

Options

(a)

0.25, 0.1875, 0.0625, 0.125, 0.375

(b)

-1, 1, 1, 1, -1

(c)

0.017, 0.95, 0.73, 0.88, 0.002

(d)

0, 1, 1, 1, 0

Answer

(c)

Solution

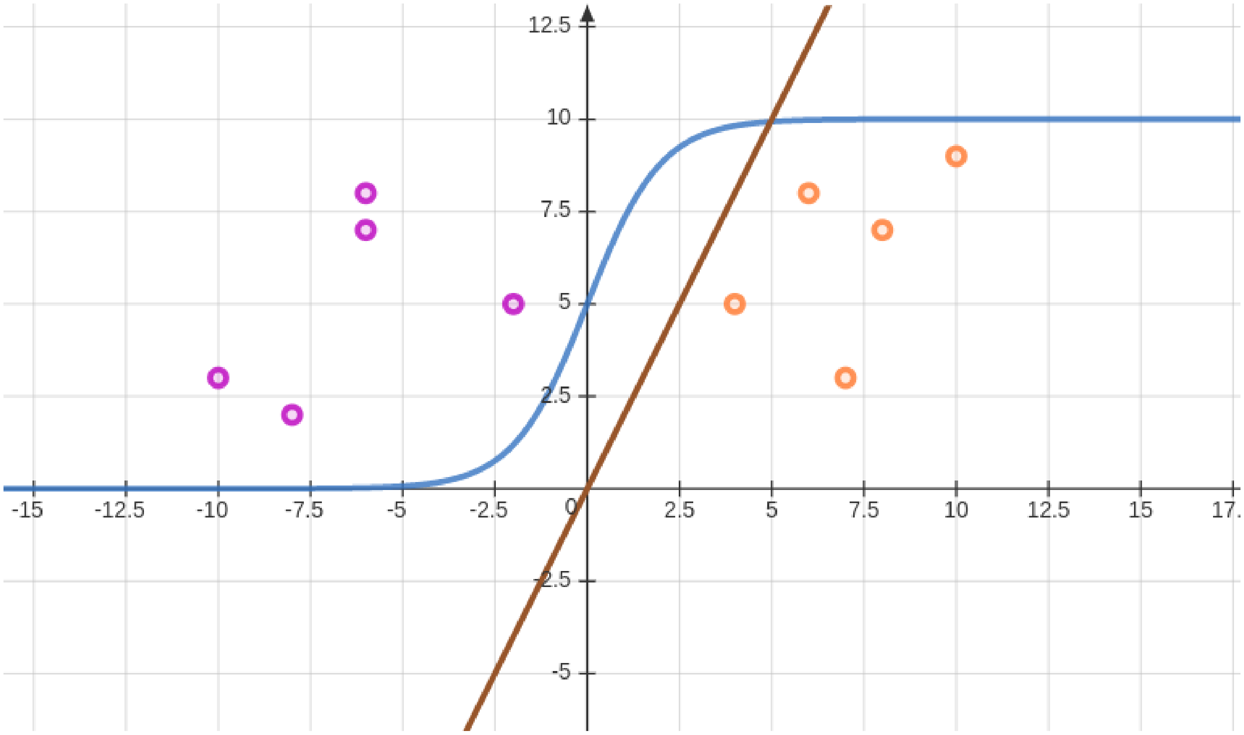
Put the values -4, 3, 1, 2, -6 in the formula $g(z) = \dfrac{1}{1+e^{-z}}$ one by one.

Ex: $g(-4) = \dfrac{1}{1+e^{4}} = \dfrac{1}{1+54.59} = \dfrac{1}{55.59} = 0.017$

Question-8

Statement

Which of the lines (blue or brown) in the following image may represent the decision boundary of Logistic Regression?



Options

(a)

Blue line

(b)

Brown line

(c)

Both

(d)

None of these

Answer

(b)

Solution

The decision boundary in logistic regression is always linear (Brown line).

It's just that the values obtained from a linear combination are reduced to values between 0 and 1 by the sigmoid function (blue line).

Question-9

Statement

Consider a data set $x_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$, $x_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$, $x_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $x_4 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$. Let the corresponding class labels be $y_1 = y_2 = y_4 = -1$ and $y_3 = +1$.

Assume you try to find the w using the Perceptron algorithm. You decide to cycle through points in the order $\{x_4, x_3, x_2, x_1\}$ repeatedly until you find a linear separator. How many mistakes does your algorithm make and what is the linear separator your algorithm outputs?

Options

(a)

3, [1, 0]

(b)

2, [1, 1]

(c)

3, [-1, 1]

(d)

4, [-1, 0]

Answer

(b)

Solution

When initial weight vector is not given, we will take zero vector as initial weight vector.

$w^0 = [0\ 0]$

The order in which we have to traverse the data points is given.

We start with x_4.

$w^T x_4$ predicts +1 class, which is a mistake.

Hence, $w = w + x_4 y_4$ gives $w = [0\ 1]$

$w^T x_3$ predicts +1 which is correct.

$w^T x_2$ predicts +1 which is a mistake.

Hence $w = w + x_2 y_2$ gives $w = [1\ 1]$

$w^T x_1$ predicts -1 which is correct.

Once again we check x_4, x_3, x_2, x_1 in that order, and they all are predicted correctly.

Hence final $w = [1\ 1]$ and w had been updated twice.