1. Objectives

Bookmark this page

Linear Classifiers and Perceptron Algorithm

At the end of this lecture, you will be able to

- understand the concepts of Feature vectors and labels, Training set and Test set, Classifier, Training error, Test error, and the Set of classifiers
- derive the mathematical presentation of linear classifiers
- understand the intuitive and formal definition of linear separation
- use the perceptron algorithm with and without offset

Concept Review Problem: car accident prediction 1

1/1 point (graded)

In this problem, we will put ourselves in the shoes of a car insurance company. Our goal is to find out whether customers were involved in an accident on July 4th, 1998.

For 8 customers, we know the following information:

1. number of accidents the customer made in the past.
2. number of miles the customer has driven.
3. the customer's age

Also, for 5 of the customers, we know whether each of them was involved in an accident on July 4th, 1998.

If we want to learn a model in a supervised way, what is n, the number of training examples?

$n =$ 5 ✔ **Answer:** 5

Solution:

We have 5 data points with known labels.

Show answer

The insurance company recorded relevant information for all 8 customers, as illustrated in the table below.

	number of past accidents	miles customer drove so far	customer's age
customer 1	0	2710.9	21
customer 2	2	13209.2	40
customer 3	1	89001.4	32
customer 4	3	12381.1	18
customer 5	0	1893.5	24
customer 6	2	32493.5	24
customer 7	1	5443.5	30
customer 8	0	4493.5	28

What is the dimension of each feature vector?

$d =$ 3 ✔ **Answer:** 3

Solution:

Each feature vector has length 3 (columns in the table), and thus its dimension is 3.

Concept Review Problem: car accident prediction 3

1/1 point (graded)

How many feature vectors are there in the above table?

$\text{Number of Feature vectors}$ 8 ✔ **Answer:** 8

Solution:

There are 8 rows in the table.

Show answer

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Concept Review Problem: Classifier and Training Error 1

1/1 point (graded)

Assume we have training data and a classifier like the following: (where $h\left(x\right)$ denotes the value outputted by the classifier with the data point as input)

	$h(x)$	y
data 1	1	1
data 2	-1	1
data 3	1	1
data 4	1	-1
data 5	-1	-1

What is the training error?

$\varepsilon_n(h) =$ 0.4 ✔ **Answer:** 0.4

Solution:

We have 5 data points total, two of which $h(x)$ does not match y (data2 and data4). Thus $\varepsilon_n(h) = \frac{1}{5}\sum_{i=1}^{5}[[h(x_i) \neq y]] = \frac{2}{5}$

Show answer

Submit You have used 1 of 3 attempts

Question 3: What movies are in the **test set**? Select all those apply.

- [ ] movie 1
- [ ] movie 2
- [ ] movie 3
- [ ] movie 4
- [x] movie 5
- [x] movie 6
- [x] movie 7

✔

Solution:

Movies whose labels are not yet available are in the test set. Thus movies $5, 6, 7$ are in the test set. Remember that it is our end goal to predict these movies' labels.

What is the inner product of $[0,1,1]$ and $[1,1,1]$?

2

✔ **Answer:** 2

Solution:

$$0 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 = 2$$

Show answer

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Linear Classifier Practice

1/1 point (graded)

We saw in the lecture above that for a linear classifier h, $h(x;\theta) = sign(\theta \cdot x)$, i.e. the sign of the dot product of θ and x.
Now consider θ which is given by

$$\theta = (1, -1) \tag{4.1}$$

- [x] $+1$
- [ ] -1

As an aside, classifiers need not be linear. They can be of any shape!

Solution:

$(10, 10)$ belongs to the region where $h(x) = +1$.

Show answer

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

Training Error

1/1 point (graded)

Which of the following points would be classified as positive by θ? Please choose all correct answers.

- [x] $(1,-1)$
- [x] $(1,0)$
- [ ] $(0,1)$
- [ ] $(0,0)$

Solution:

$X$is positively classified by the classifier if and only if $x\cdot\theta>0$. The dot product of $(1,-1)$ with θ is positive. Also, $(1,0)\cdot\theta$ is positive. On the other hand, $(0,1)\cdot\theta$ and $(0,0)\cdot\theta$ are nonpositive. Thus the first and second points are positively classified by θ.

Show answer

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

1/1 point (graded)

Again, we have a linear classifier with θ given by

$$\theta = (1, -1) \tag{4.2}$$

and the offset, θ_0 given by $\theta_0 = -1$ Now which of the following points would be classified as positive by θ? Please choose all correct answers.

- [x] $(1, -1)$
- [ ] $(1, 0)$
- [ ] $(0, 1)$
- [ ] $(0, 0)$

✔

Solution:

$X$is positively classified by the classifier if and only if $x \cdot \theta + \theta_0 > 0$. The dot product of $(1, -1)$ with θ is 2, and adding -1 makes it still positive. However,$x \cdot \theta + \theta_0 \leq 0$ for other data points.

Given θ and θ_0, a **linear classifier** $h : X \to \{-1, 0, +1\}$ is a function that outputs $+1$ if $\theta \cdot x + \theta_0$ is positive, 0 if it is zero, and -1 if it is negative. In other words, $h(x) = \text{sign}(\theta \cdot x + \theta_0)$.

Basics 1

1/1 point (graded)

As described in the lecture above, h is a linear classifier which is defined by the boundary $\theta \cdot x = 0$ (where theta is a vector perpendicular to the plane.) The ith training data is $(x^{(i)}, y^{(i)})$, where $x^{(i)}$ is a vector and $y^{(i)}$ is a scalar quantity. If θ is a vector of the same dimension as $x^{(i)}$, what are $y^{(i)}$ and sign($\theta \cdot x^{(i)}$) respectively?

- ( ) output of the classifier h, label
- ( ) label, dimension of the feature vector
- ( ) label, distance of the point from the linear classifier
- (•) label, output of the classifier h

✔

Solution:

By definition, $y^{(i)}$ is the label of $x^{(i)}$. Also, by the definition of a linear classifier $h(x)$ =sign($\theta \cdot x^{(i)}$), the output of h is given by sign($\theta \cdot x^{(i)}$).

Show answer

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Basics 2

1/1 point (graded)

For the ith training data (x^i, y^i), what values can $y^{(i)}$ take, **conventionally** (in the context of linear classifiers)? Choose all those apply.

- [x] -1
- [x] $+1$
- [ ] 0
- [ ] $+10$

Solution:

By the convention of linear classification, because $y^{(i)}$ is a label, it can take -1 or $+1$. Note that 0 is not a possible value.

Show answer

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Basics 3

1/1 point (graded)

For the ith training data (x^{i}, y^{i}), what values can $sign\left(\theta \cdot x^{(i)}\right)$ take? Choose all those apply.

- [x] -1
- [x] $+1$
- [x] 0
- [ ] $+10$

By definition the $sign\left(\theta \cdot x^{(i)}\right)$ function can only take one of $0, -1, +1$ as its value. Remember that a linear classifier outputs one of $-1, 0, 1$.

Show answer

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

When the Product is Positive

1/1 point (graded)

When does $y^{(i)}\left(\theta \cdot x^{(i)}\right) > 0$ happen? Choose all those apply.

- [x] $y^{(i)} > 0$ and $\theta \cdot x^{(i)} > 0$
- [ ] $y^{(i)} < 0$ and $\theta \cdot x^{(i)} > 0$
- [ ] $y^{(i)} > 0$ and $\theta \cdot x^{(i)} < 0$
- [x] $y^{(i)} < 0$ and $\theta \cdot x^{(i)} < 0$

$y^{(i)}\left(\theta \cdot x^{(i)}\right) > 0$ is true if and only if $y^{(i)}$ and $\left(\theta \cdot x^{(i)}\right)$ are both positive both negative. In other words, they have the same sign.

Show answer

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Intuitive Meanings of Positive Product

1/1 point (graded)

What is the intuitive meaning of $y^{(i)}\left(\theta \cdot x^{(i)}\right) > 0$?

- (selected) x^{i} label and classified result match
- x^{i} label and classified result do not match
- x^{i} is on the boundary of the classifier
- training error is positive

$y^{(i)}\left(\theta\cdot x^{(i)}\right)>0$ is true if and only if $y^{(i)}$ and $\left(\theta\cdot x^{(i)}\right)$ are both positive both negative. In other words, they have the same sign.

Show answer

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Intuitive Meanings of Negative Product

1/1 point (graded)

What is the intuitive meaning of $y^{(i)}\left(\theta\cdot x^{(i)}\right)<0$?

- ○ x^i label and classified result match
- ● x^i label and classified result do not match
- ○ x^i is on the boundary of the classifier
- ○ training error is negative

$y^{(i)}\,(\theta \cdot x^{(i)}) < 0$ is true if and only if $y^{(i)}$ and $(\theta \cdot x^{(i)})$ have different signs.

Show answer

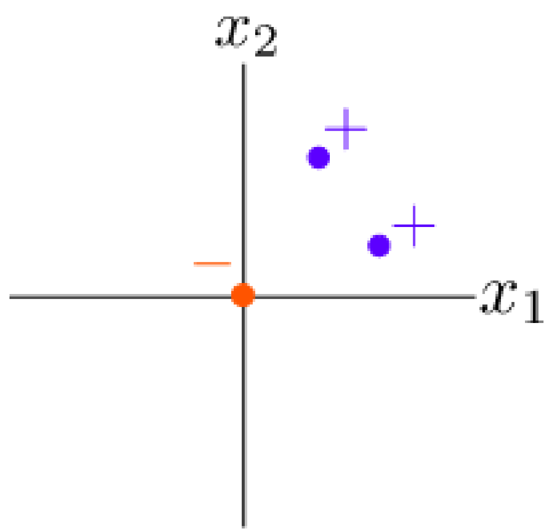
Submit You have used 1 of 1 attempt

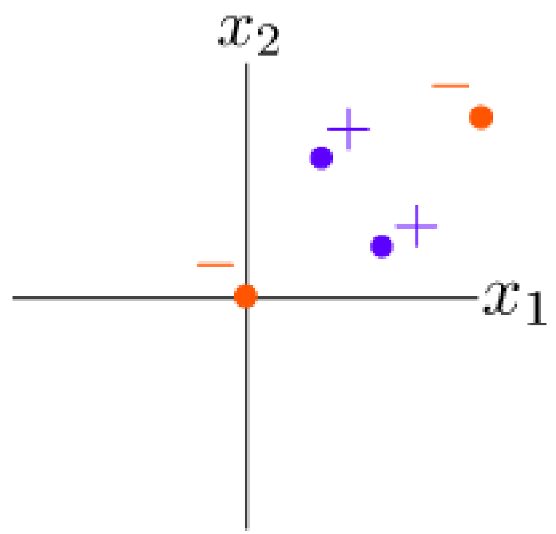
Answers are displayed within the problem

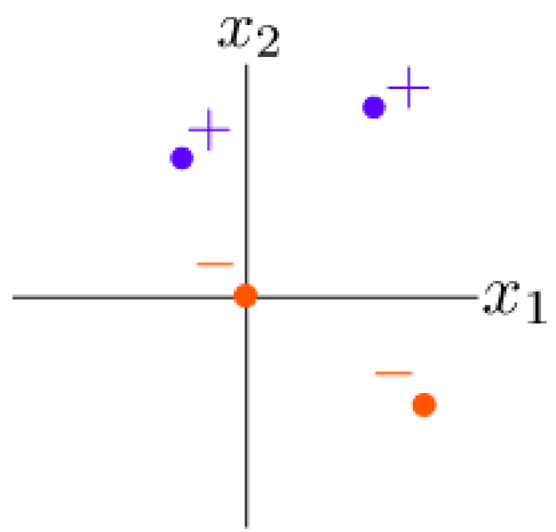
Linear Separation 1

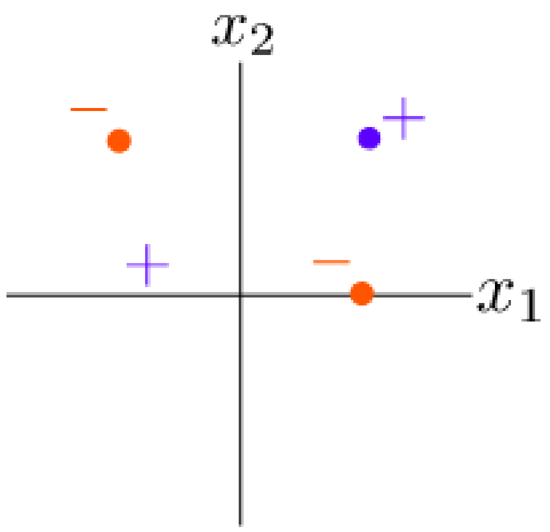
1/1 point (graded)

Of the following, which is linearly separable? Choose all those apply.

- [x] 


x_2
x_1
+
+
−
−


x_2
x_1
+
+
−
−



Solution:

Linearly separable data can be separated with $+$ labels on one side of the line and $-$ labels on the other side, by some line on the plane.

Show answer

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

A set of Training examples is illustrated in the table below, with the classified result by some linear classifier h and the label y^i. Is it linearly separable?

	$h(x^i)$	y^i
example 1	-1	-1
example 2	1	1
example 3	1	1
example 4	-1	-1
example 5	-1	-1

- (•) yes
- ( ) no

✔

Solution:

For linearly separable data, a linear classifier can perfectly separate the data. The provided classifier $h(x)$ classifies all the given points correctly.

Perceptron Concept Questions 1

1/1 point (graded)

Remember that the Perceptron Algorithm (without offset) is stated as the following:

Perceptron $\left(\{(x^{(i)}, y^{(i)}), i = 1, \ldots, n\}, T\right)$:

initialize $\theta = 0$(vector);

for $t = 1, \ldots, T$ do

for $i = 1, \ldots, n$ do

if $y^{(i)}\left(\theta \cdot x^{(i)}\right) \leq 0$ then

update $\theta = \theta + y^{(i)}x^{(i)}$

What does the Perceptron algorithm take as inputs among the following? Choose all those apply.

- [x] Training set
- [x] T - the number of times the algorithm iterates through the whole training set
- [ ] Test set
- [ ] θ
- [ ] θ_0

Solution:

The perceptron algorithm takes T and the training set as input, and aims to learn the optimal "θ","θ_0"

Show answer

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Perceptron Update 1

1/1 point (graded)

Now consider the Perceptron algorithm with Offset. Whenever there is a "mistake" (or equivalently, whenever $y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right) \leq 0$ i.e. when the label y^i and $h\left(x\right)$ do not match), perceptron updates

$$\theta \text{ with } \theta + y^{(i)} x^{(i)}$$

and

$$\theta_0 \text{ with } \theta_0 + y^{(i)}.$$

More formally, the Perceptron Algorithm with Offset is defined as follows:

Perceptron $\left(\left\{\left(x^{(i)}, y^{(i)}\right), i=1, \ldots, n\right\}, T\right)$:

initialize $\theta=0$(vector); $\theta_0=0$(scalar)

for $t=1, \ldots, T$ do

for $i=1, \ldots, n$ do

if $y^{(i)}\left(\theta \cdot x^{(i)}+\theta_0\right) \leq 0$ then

update $\theta=\theta+y^{(i)} x^{(i)}$

update $\theta_0=\theta_0+y^{(i)}$

In the next set of problems, we will try to understand why such an update is a reasonable one.

When a mistake is spotted, do the updated values of θ and θ_0 provide a better prediction? In other words, is

$$y^{(i)}\left(\left(\theta+y^{(i)} x^{(i)}\right) \cdot x^{(i)}+\theta_0+y^{(i)}\right)$$

always greater than or equal to

$$y^{(i)}\left(\theta \cdot x^{(i)}+\theta_0\right)$$

- ( ) Yes, because $\theta + y^{(i)} x^{(i)}$ is always larger than θ
- (•) Yes, because $\left(y^{(i)}\right)^2 \|x^{(i)}\|^2 + \left(y^{(i)}\right)^2 \geq 0$
- ( ) No, because $\left(y^{(i)}\right)^2 \|x^{(i)}\|^2 - \left(y^{i}\right)^2 \leq 0$
- ( ) No, because $\theta + y^{(i)} x^{(i)}$ is always larger than θ

✔

Solution:

Comparing the two terms,

$$y^{(i)}\left(\left(\theta + y^{(i)} x^{(i)}\right) \cdot x^{(i)} + \theta_0 + y^{(i)}\right) - y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right) = \left(y^{(i)}\right)^2 \|x^{(i)}\|^2 + \left(y^{(i)}\right)^2 = \left(y^{(i)}\right)^2 \left(\|x^{(i)}\|^2 + 1\right)) > 0$$

the first is always greater than the latter. Considering that our goal is to minimize the training error, the update always makes the training error decrease, which is desirable.

Show answer

Submit You have used 1 of 2 attempts

Perceptron Update 2

0 points possible (ungraded)

For a given example i, we defined the training error as 1 if $y^{(i)}\left(\theta \cdot x^{(i)}+\theta_0\right) \leq 0$, and 0 otherwise:

$$\varepsilon_i\left(\theta, \theta_0\right)=\left[\left[y^{(i)}\left(\theta \cdot x^{(i)}+\theta_0\right) \leq 0\right]\right]$$

Say we have a linear classifier given by θ, θ_0. After the perceptron update using example i, the training error $\varepsilon_i\left(\theta, \theta_0\right)$ for that example can (select all those apply):

- [ ] Increase
- [x] Stay the same
- [x] Decrease

Solution:

From the previous problem, we saw that $y^i\left(\theta \cdot x+\theta_0\right)$ increases after the perceptron update. Thus $\left[\left[y^i\left(\theta \cdot x+\theta_0\right) \leq 0\right]\right]$ becomes zero or stays 1.