5. The Perceptron Algorithm

The Perceptron Algorithm

to the problem.

16:37 / 16:37 1.25x

We talked about linear separation,
when the set of linear classifiers
is sufficient to separate the training set.
We did that by means of examples.
And we defined a learning algorithm
for the set of linear classifiers
that can take a training set as an input.
If that training set is linearly separable,
then the perceptron algorithm succeeds in finding a solution
to the problem.

End of transcript. Skip to the start.

Video

Download video file

Transcripts

Download SubRip (.srt) file
Download Text (.txt) file

Perceptron Concept Questions 1

0/1 point (graded)
Remember that the Perceptron Algorithm (without offset) is stated as the following:

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

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

☑ θ_0

✖

Submit You have used 2 of 2 attempts

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 $h(x)$ 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

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

对于那个样本，updating以后一定是对了的

- ○ 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,

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

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

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

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 or stays the same 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.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 1 Linear Classifiers and Generalizations (2 weeks):Lecture 2. Linear Classifier and Perceptron / 5. The Perceptron Algorithm