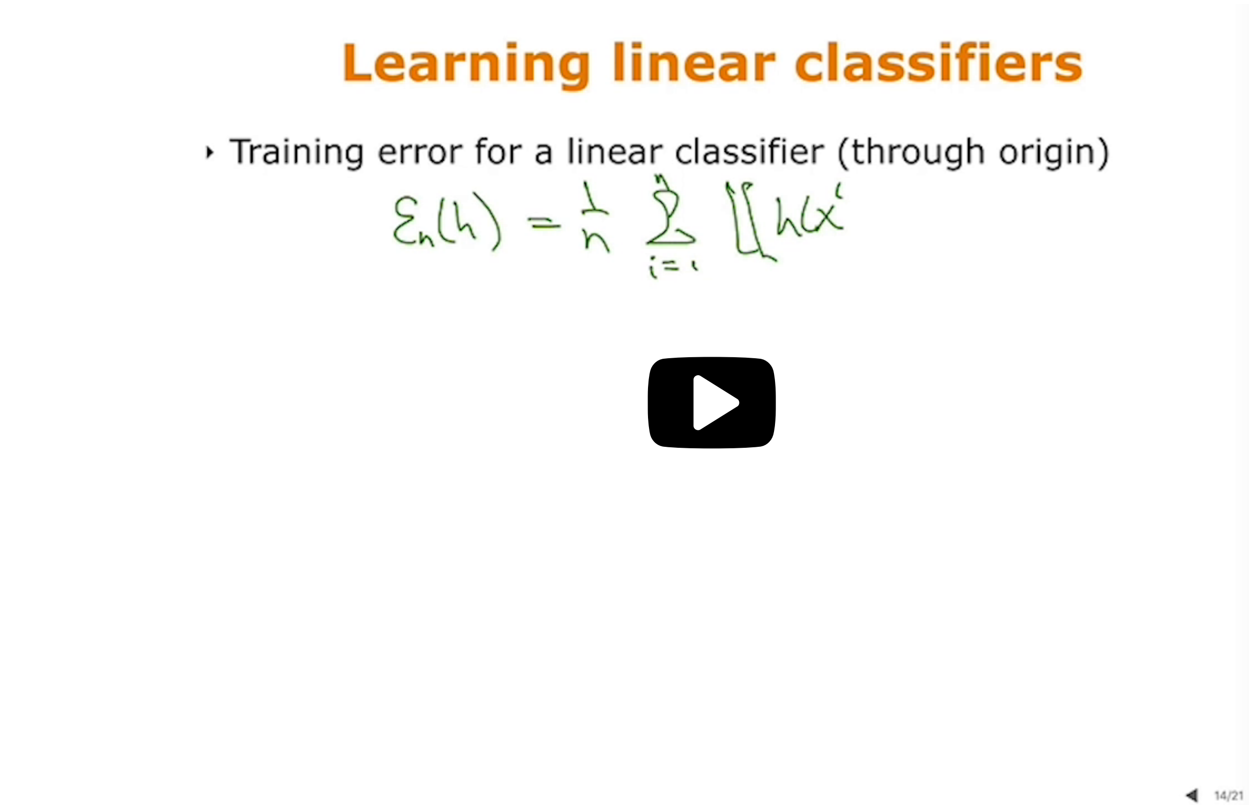
5. The Perceptron Algorithm

The Perceptron Algorithm



0:00 / 0:00 Speed 1.50x

Start of transcript. Skip to the end.

All right.

We defined earlier, training error

for any classifier as a fraction of training samples

that are misclassified--

so in terms of whether or not the classifier

applied to their training example,

whether it disagrees with a given label for that example.

All right.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Perceptron Concept Questions 1

1/1 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 ✔
- [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"

Submit You have used 2 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

- [x] 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,

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

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.

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 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 2 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

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

Add a Post

Show all posts ▼ by recent activity ▼

- Mistake on the algorithm? (restart from first observation after theta update) 2
  In the described algorithm, if there is an update of theta on the observation n, it then continues to the observation n+1. Based on the questions I believe that after each u...
- Perceptron Update 1 2
- [Staff] [Typo] Perceptron Update 2 5
  Community TA
- R & Coffee - The Perceptron Algorithm 7
- Analysis of training set 5
  How to find the optimal T to use in the perceptron algorithm for a given training set? Is there a (theoretical) way to check the separability of a training set?
- Solution to perceptron update 1 and 2 possibly not well worded 3
- Perceptron Algo 2
  Why the double square brackets?

Geometric Interpretation -- Projection or Something Else?
I remember my calc a lot better than linear algebra. When he adds the second point, a second vector is discovered between the first line of orthogonality and the second ... 2

2nd step on the perceptron
why do you consider the 1st observation y(i) = 1 and not -1? thanks! 5

[Staff] Perception Update 2 grading is wrong
Hi, My answer was correct but graded wrong by the system. --- deleted; BB --- [1]: https://edxuploads.s3.amazonaws.com/1561151064347887.png 3

Proof of the convergence of the perceptron
This set of notes has a proof of the convergence of the perceptron algorithm after a bounded number of steps): https://www.cse.iitb.ac.in/~shivaram/teaching/old/cs344+... 3

Transforming yi((θ+yix)·x+θ0+yi) for later iterations
So I understand that in the beginning θ and θ0 are null and therefore yi((θ+yix)·x+θ0+yi) = (yi)^2*x^2+(yi)^2 But how does it hold up for the further iterations? Isn't it possi... 5

[Staff]Q2
Is the question about error on ith point immediately after ith point has been processed by the algo or is it after all T passes ? 4

Is this the same as the perceptron in NN? 2

Learn About Verified Certificates