5. The Kernel Perceptron Algorithm

Bookmark this page

Exercises due Oct 14, 2020 05:29 IST *Completed*

Computational Efficiency

Start of transcript. Skip to the end.

OK.

So let's see how we can do that with a simple perceptron

classifier.

So recall that perceptron utilizes the parameter vector

0.

Runs through training examples multiple times,

0:00 / 0:00 1.25x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

How the Kernel Perceptron Algorithm Works: Initalization

1/1 point (graded)

Recall that the original Perceptron Algorithm is given as the following:

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

initialize $\theta = 0$ (vector);

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

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

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

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

In the lecture, it was introduced that we can always express θ as

$$\theta = \sum_{j=1}^{n} \alpha_j y^{(j)} \phi\left(x^{(j)}\right)$$

where values of $\alpha_1, \ldots, \alpha_n$ may vary at each step of the algorithm. In other words, we can reformulate the algorithm so that we somehow initialize and update α_j's, instead of θ.

The reformulated algorithm, or **kernel perceptron** , can be given in the following form:

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

Initialize $\alpha_1, \alpha_2, \ldots, \alpha_n$ to some values;

for $t = 1, \ldots, T$

for $i = 1, \ldots, n$

if $(\text{Mistake Condition Expressed in } \alpha_j)$

Update α_i appropriately

Update α_j appropriately

Look at the initialization statement of the algorithm. Which of the following is an equivalent way to initialize $\alpha_1, \alpha_2, \ldots, \alpha_n$, if we want the same result as initializing $\theta = 0$?

- ◯ $\alpha_1 = \ldots = \alpha_n = \theta$
- ◯ $\alpha_1 = \ldots = \alpha_n = 1$
- ◉ $\alpha_1 = \ldots = \alpha_n = 0$
- ◯ $\alpha_1 = \ldots = \alpha_n = -1$

✔

Solution:

Since $\theta = \sum_{j=1}^{n} \alpha_j y^{(j)} \phi\left(x^{(j)}\right)$, setting $\alpha_j = 0$ for all j leads to $\theta = 0$.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

How the Kernel Perceptron Algorithm Works: The Update

1/1 point (graded)

As in the previous problem, our goal is to correctly reformulate the original perceptron algorithm. In other words, we want the algorithm to be about updating α_j's instead of θ.

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

  initialize $\alpha_1, \alpha_2, \ldots, \alpha_n$ to some values;
  for $t = 1, \ldots, T$
    for $i = 1, \ldots, n$
      if (Mistake Condition Expressed in α_j)
        Update α_j appropriately

Now look at the line "**Update α_j appropriately**" in the above algorithm. Remember that we express θ as

$$\theta = \sum_{j=1}^{n} \alpha_j y^{(j)} \phi\left(x^{(j)}\right)$$

Assuming that there was a mistake in classifying the ith data point i.e.

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

which of the following conditions about $\alpha_1, \ldots, \alpha_n$ is equivalent to

$$\theta = \theta + y^{(i)} \phi\left(x^{(i)}\right),$$

the update condition of the original algorithm?

- [x] $\alpha_i = \alpha_i + 1$
- [ ] $\alpha_i = \alpha_i - 1$
- [ ] $\alpha_j = \alpha_j + 1$ for all $j \in 1, \ldots, n$

✔

Solution:

Expand θ in the last equation and it turns out only α_i gets updated:

$$\alpha_i y^{(i)} \phi\left(x^{(i)}\right) + y^{(i)} \phi\left(x^{(i)}\right) = \left(\alpha_i + 1\right) y^{(i)} \phi\left(x^{(i)}\right).$$

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

How the Kernel Perceptron Algorithm Works: The Mistake Condition

0/1 point (graded)

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

initialize $\alpha_1, \alpha_2, \ldots, \alpha_n$ to some values;
for $t = 1, \ldots, T$
for $i = 1, \ldots, n$
if $(\text{Mistake Condition Expressed in } \alpha_j)$
Update α_j appropriately

Now look at the line "**Mistake Condition Expressed in** α_j" in the above algorithm. Remember that we express θ as

$$\theta = \sum_{j=1}^{n} \alpha_j y^{(j)} \phi\left(x^{(j)}\right)$$

Which of the following conditions is equivalent to $y^{(i)}\left(\theta \cdot \phi\left(x^{(i)}\right)\right) \leq 0$? Remember from the video lecture above that given feature vectors $\phi(x)$ and $\phi(x')$, we define the Kernel function K as

$$K(x, x') = \phi(x) \phi(x').$$

- [ ] $y^{(i)} \sum_{j=1}^{n} \alpha_j y^{(j)} K\left(x^j, x^i\right) \leq 0$ ✔
- [ ] $y^{(i)} \sum_{j=1}^{n} \alpha_i y^{(j)} K\left(x^j, x^i\right) \leq 0$
- [x] $y^{(i)} \sum_{j=1}^{n} \alpha_j y^{(i)} K\left(x^j, x^i\right) \leq 0$
- [ ] $y^{(i)} \sum_{j=1}^{n} \alpha_j y^{(j)} \phi\left(x^{(j)}\right) \leq 0$

✘

Solution:

Substitute θ with $\sum_{j=1}^{n} \alpha_j y^{(j)} \phi\left(x^{(j)}\right)$ in $y^{(i)}\left(\theta \cdot \phi\left(x^{(i)}\right)\right) \leq 0.$

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks):Lecture 6. Nonlinear Classification / 5. The Kernel Perceptron Algorithm

Add a Post

Show all posts by recent activity

?	What is the difference between Xi and Xj? I am trying really hard to follow up on this	3
	@staff: Question about performance benefits of kernel perceptron algorithm Please clarify if my understanding about the performance benefits of perceptron algorithm are correct. Here are the two benefits that I ...	4
	In Kernel perceptron, how do we compute theta In kernal perceptron, we are only computing alpha or mistakes we make with each training eg. How do we finally compute the theta? Do...	2
?	I am lost with the difference between (j) and (i)!!!	6
	what does x' represents?	2
?	Question on How the Kernel Perceptron Algorithm Works	4
?	kernel perceptron vs normal perceptron convergence	2
?	How the Kernel Perceptron Algorithm Works: The Mistake Condition Whats the difference between option 1 and 2? They look the same	2
	How the Kernel Perceptron Algorithm Works: The Update- index issue?	4

< Previous Next >

Affiliates
edX for Business
Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

Blog
Contact Us
Help Center
Media Kit
Donate

GET IT ON Google play
Download on the App Store