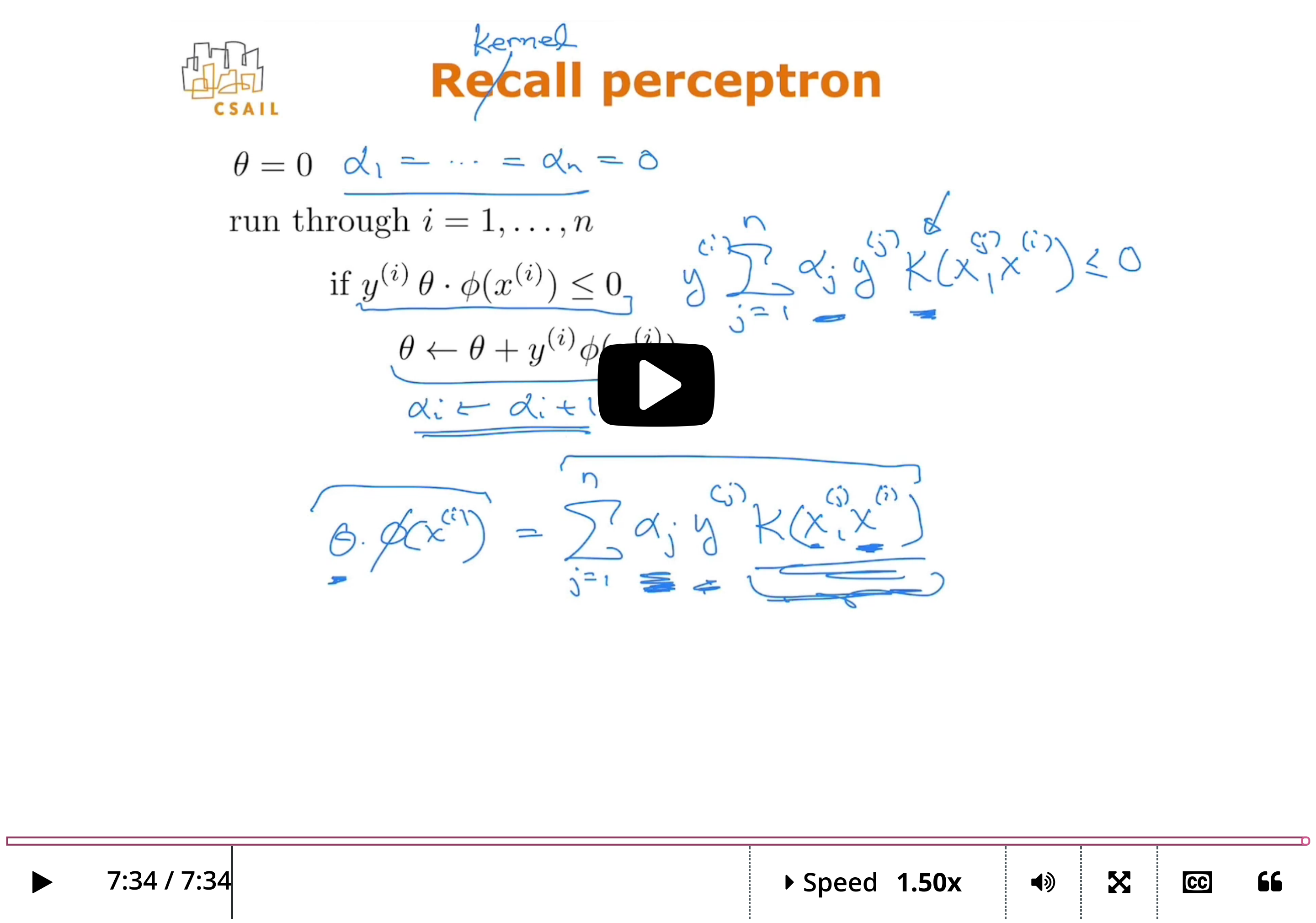
Course › Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks) › Lecture 6. Nonlinear Classification › 5. The Kernel Perceptron Algorithm

5. The Kernel Perceptron Algorithm

Computational Efficiency



Kernel
Recall perceptron

$\theta = 0$ $\alpha_1 = \cdots = \alpha_n = 0$

run through $i = 1, \ldots, n$

if $y^{(i)}\, \theta \cdot \phi(x^{(i)}) \leq 0$ $y^{(i)} \sum_{j=1}^{n} \alpha_j y^{(j)} K(x^{(j)}, x^{(i)}) \leq 0$

$\theta \leftarrow \theta + y^{(i)} \phi(x^{(i)})$

$\alpha_i \leftarrow \alpha_i + 1$

$$\theta \cdot \phi(x^{(i)}) = \sum_{j=1}^{n} \alpha_j y^{(j)} K(x^{(j)}, x^{(i)})$$

OK?

So you can think of the kernel function

here as a kind of similarity measure.

How similar the j-th example is to the i-th example.

So our predicted value here is now

how important the j-th example is.

Its label times how similar the example

we wish to make a prediction on and the j-th training example.

All right.

That is now the kernel perceptron algorithm.

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$\left(\left\{\left(x^{(i)}, y^{(i)}\right), i=1,\ldots,n\right\}, T\right)$ :
    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$\left(\left\{\left(x^{(i)}, y^{(i)}\right), i=1,\ldots,n, T\right\}\right)$
    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 $\alpha_j$ 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 $(\text{Mistake Condition Expressed in } \alpha_j)$
                Update $\alpha_j$ 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?

● $\alpha_i = \alpha_i + 1$ ✓

○ $\alpha_i = \alpha_i - 1$

○ $\alpha_j = \alpha_j + 1$ 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

1/1 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 $\alpha_j$ 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

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

● $y^{(i)} \sum_{j=1}^{n} \alpha_j y^{(j)} K\left(x^j, x^i\right) \leq 0$ ✓

○ $y^{(i)} \sum_{j=1}^{n} \alpha_i y^{(j)} K\left(x^j, x^i\right) \leq 0$

○ $y^{(i)} \sum_{j=1}^{n} \alpha_j y^{(i)} K\left(x^j, x^i\right) \leq 0$

○ $y^{(i)} \sum_{j=1}^{n} \alpha_j y^{(j)} \phi\left(x^{(j)}\right) \leq 0$

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 ▼

💬	Subscripts in alpha From question nr 1: for i=1,...,n if (Mistake Condition Expressed in qj) Update qj appropriately Two subscripts are used in the question and in the answer for alpha: "i' and 'j'. T...	1
?	Prof. Jaakkola Is it only me to whom Prof. Jaakkola's explanations don't make much sense ? It's like after watching his one lecture video, I feel like I got almost nothing. Maybe this is my Engli...	18
?	How "expressive" are kernel functions?	1
?	[STAFF] How the Kernel Perceptron Algorithm Works: The Mistake Condition First, who is x’ ? Second, in the solution, if we substitute like suggested we don’t find the solution directly. Did you miss one or many steps ? Thank you	1
?	α (alpha)	9
?	How the Kernel Perceptron Algorithm Works: The Update	3
?	How the Kernel Perceptron Algorithm Works: The Mistake Condition: none of the answers is a condition	3
💬	Here is my view of how kernel perceptron works	1
💬	For staff. Resembling symbols for students with dyslexia As a student with dyslexia, resembling symbols in mathematical terms have been a problem. Please take that into account.	4
💬	[Staff] Second video is the same as first one Second video is the same as first one	2

Learn About Verified Certificates