





MITx 6.86x

**Machine Learning with Python-From Linear Models to Deep Learning**[Course](#)[Progress](#)[Dates](#)[Discussion](#)[Resources](#)[Course](#) / [Unit 2. Nonlinear Classification, Linear regression,...](#) / [Lecture 6. No](#)[< Previous](#)

## 5. The Kernel Perceptron Algorithm

[Bookmark this page](#)

Exercises due Mar 8, 2023 08:59 -03 Completed

**Computational Efficiency****Video**
 [Download video file](#)
**Transcripts**
 [Download SubRip \(.srt\) file](#)
 [Download Text \(.txt\) file](#)
**How the Kernel Perceptron Algorithm Works: Initialization**

1/1 point (graded)

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

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

initialize  $\theta = \mathbf{0}$  (vector);

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

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

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

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

In the lecture, it was introduced that we can always express  $\theta$  as

Look at the initialization statement of the algorithm. Which of the following is an equivalent statement, if we want the same result as initializing  $\mathbf{w}$ ?



Submit

You have used 1 of 1 attempt

## 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  $\mathbf{w}$ 's instead of  $\mathbf{b}$ .

### Kernel Perceptron

```

initialize  $\mathbf{w}$  to some values;
for each data point  $\mathbf{x}_i$ 
    for each class  $c$ 
        if  $\text{classify}(\mathbf{x}_i, c) \neq \text{true}$ 
            Update  $\mathbf{w}$  appropriately

```

Now look at the line "**Update  $\mathbf{w}$  appropriately**" in the above algorithm. Remember that  $\mathbf{w}$  is a vector of size  $d$ .

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

**Submit**

You have used 1 of 1 attempt

## How the Kernel Perceptron Algorithm Works: The Mistake Condition

1/1 point (graded)

### Kernel Perceptron

```

initialize  $w$  and  $b$  to some values;
for each  $x_i$  in  $S$ 
    for each  $x_j$  in  $S$ 
        if  $\langle x_i, x_j \rangle < 0$ 
            Update  $w$  and  $b$  appropriately

```

Now look at the line "**Mistake Condition Expressed in**  $\langle x_i, x_j \rangle < 0$ " in the above algorithm. Remember that  $\langle x_i, x_j \rangle$  is the dot product of  $x_i$  and  $x_j$ .

Which of the following conditions is equivalent to  $\langle x_i, x_j \rangle < 0$ ? Remember from Lecture 5 that  $\langle x_i, x_j \rangle$  is the dot product of  $x_i$  and  $x_j$  above that given feature vectors  $x_i$  and  $x_j$ , we define the Kernel function  $K(x_i, x_j)$  as



&lt; Previous

Next &gt;



💬 The first three options for the last question are the same?  
Should be a typo?

💬 Deadline of Lecture 6 is March 8 , but why I cant submit the answer today? Today is 7March.  
Deadline of Lecture 6 is March 8 , but why I cant submit the answer today? Today is 7March.

💬 confusion about the iteration  
I thought that there are lots of point which is unclear about this video clip, although I have already spent so

? Thinking of kernel as a "similarity measure"  
At 6:58, prof. Jaakkola says we can think of the kernel as a kind of similarity measure. My questions are: 1. W

💬 Matrix representation mistake condition

? I have a question to clarify based on this video.  
So to clarify, does j means the number of updates and i means the number of x which is the number of featur

## Legal

[Terms of Service & Honor Code](#)

[Privacy Policy](#)

[Accessibility Policy](#)

[Trademark Policy](#)

[Sitemap](#)

[Cookie Policy](#)

[Do Not Sell My Personal Information](#)

## Connect

[Blog](#)

[Contact Us](#)

[Help Center](#)

[Security](#)

[Media Kit](#)

