

edX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the [Privacy Policy](#).

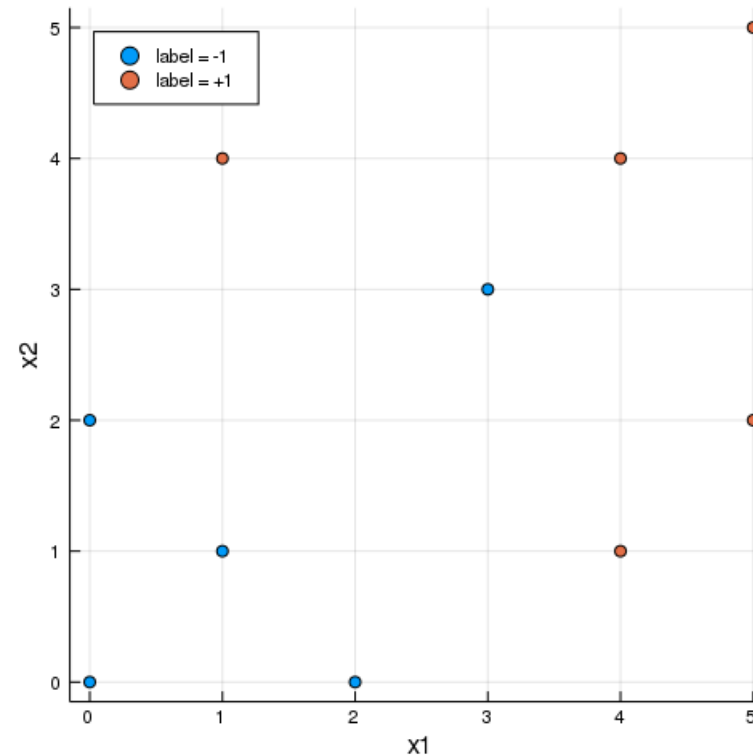


[Course](#) > [Midterm Exam \(1 week\)](#) > [Midterm Exam 1](#) > Problem 2

## Problem 2

### Problem 2. Kernel Methods

In this problem, we want to do classification over a different training dataset, as shown in plot below:



## 2. (1)

1/1 point (graded)

If we again use the linear perceptron algorithm to train the classifier, what will happen?

**Note:** In the choices below, "converge" means given a certain input, the algorithm will terminate with a fixed output within finite steps (assume  $T$  is very large: the output of the algorithm will not change as we increase  $T$ ). Otherwise we say the algorithm diverges (even for an extremely large  $T$ , the output of the algorithm will change as we increase  $T$  further).

- ☐ The algorithm always converges and we get a classifier that perfectly classifies the training dataset.
- ☐ The algorithm always converges and we get a classifier that does not perfectly classifies the training dataset.
- ☒ The algorithm will never converge. ✓
- ☐ The algorithm might converge for some initial input of  $\theta, \theta_0$  and certain sequence of the data, but will diverge otherwise. When it converges, we always get a classifier that does not perfectly classifies the training dataset.
- ☐ The algorithm might converge for some initial input of  $\theta, \theta_0$  and certain sequence of the data, but will diverge otherwise. When it converges, we always get a classifier that perfectly classifies the training dataset.

**Solution:**

The algorithm will never converge. Since this dataset is not linearly separable anymore, we will always get some mistakes at some points and we will update the parameters (which will give us more mistakes at other points, and this cycle never ends).

You have used 1 of 3 attempts

---

**i** Answers are displayed within the problem

---

2. (2)

2.0/2 points (graded)

We decide to run the kernel perceptron algorithm over this dataset using the quadratic kernel. The number of mistakes made on each point is displayed in the table below. (These points correspond to those in the plot above.)

Label	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1
Coordinates	(0,0)	(2,0)	(1,1)	(0,2)	(3,3)	(4,1)	(5,2)	(1,4)	(4,4)	(5,5)
Perceptron mistakes	1	65	11	31	72	30	0	21	4	15

Define the feature map of our quadratic kernel to be:

$$\phi(x) = [x_1^2, \sqrt{2}x_1x_2, x_2^2]^T.$$

Assume all parameters are set to zero before running the algorithm.

Based on the table, what is the output of  $\theta$  and  $\theta_0$ ?

(Enter  $\theta_0$  accurate to at least 2 decimal places.)

$\theta_0 =$

✓ Answer: -110

(Enter  $\theta$  as a vector, enclosed in square brackets, and components separated by commas, e.g. type  $[\theta, 1]$  for  $\begin{bmatrix} \theta & 1 \end{bmatrix}^T$ . Note that this sample vector input may not be of the same dimension of the answer. Enter each component accurate to at least 2 decimal places.)

$\theta =$

✓ Answer: [21.00, -22.63, 22.00]

STANDARD NOTATION

**Solution:**

$$\theta_0 = \sum_{i=1}^{10} \alpha_i y^{(i)}$$

$$\theta = \sum_{i=1}^{10} \alpha_i y^{(i)} \phi(x^{(i)})$$

Again, the answers do not depend on the order of data points used in the algorithm. (For reference, the sequence of the kernel perceptron algorithm used here is  $(3, 3), (0, 2), (4, 1), (1, 4), (0, 0), (1, 1), (5, 2), (2, 0), (5, 5), (4, 4), (5, 2)$ .)

Submit

You have used 2 of 3 attempts

**i** Answers are displayed within the problem

2. (3)

1/1 point (graded)

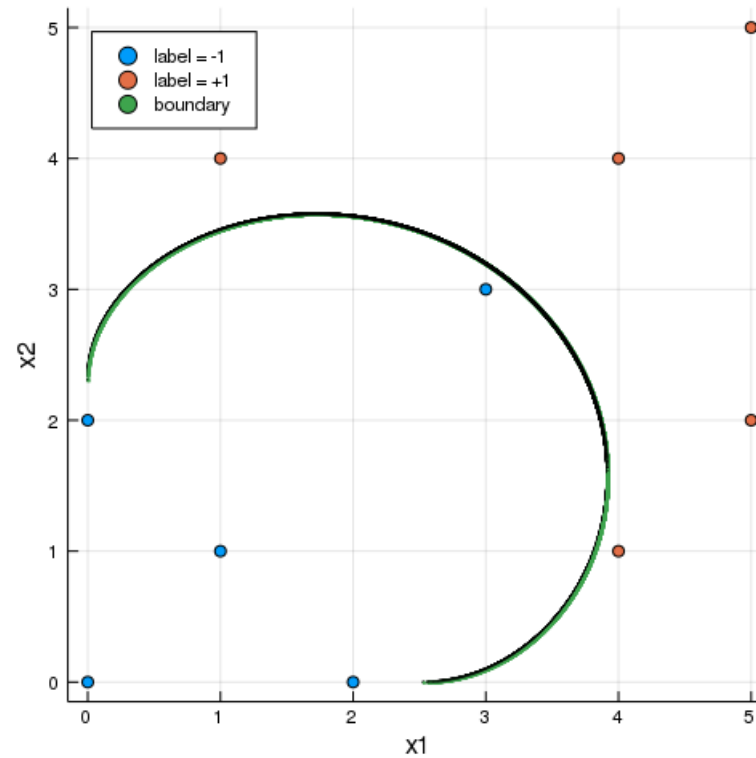
Based on the calculation of  $\theta$  and  $\theta_0$ , does the decision boundary  $\theta^T \phi(x) + \theta_0 = 0$  correctly classify all the points in the training dataset?

☒ Yes ✓

☐ No

**Solution:**

Check that  $y^{(i)} (\theta^T \phi(x^{(i)}) + \theta_0) \geq 0$  for all  $i = 1, 2, \dots, 10$ . The boundary looks like the following in the  $(x_1, x_2)$ -space.



You have used 2 of 3 attempts

**i** Answers are displayed within the problem

2. (4)

1/1 point (graded)

Recall for  $x = [x_1 \ x_2]^T$ 

$$\phi(x) = [x_1^2, \sqrt{2}x_1x_2, x_2^2]^T.$$

Define the kernel function

$$K(x, x') = \phi(x)^T \phi(x').$$

Write  $K(x, x')$  as a function of the dot product  $x \cdot x'$ . To answer, let  $z = x \cdot x'$ , and enter  $K(x, x')$  in terms of  $z$ . $K(x, x') =$ 

z^2

✓ Answer: z^2

z^2

STANDARD NOTATION

**Solution:**

Given

$$\phi(x) = [x_1^2 \ \sqrt{2}x_1x_2 \ x_2^2]^T,$$

we have



$$\begin{aligned}
 \phi(x)^T \phi(x') &= (x_1 x'_1)^2 + 2(x_1 x'_1)(x_2 x'_2) + (x_2 x'_2)^2 \\
 &= (x_1 x'_1 + x_2 x'_2)^2 \\
 &= z^2 \quad \text{where } z = x \cdot x' = x_1 x'_1 + x_2 x'_2.
 \end{aligned}$$

Submit

You have used 1 of 3 attempts

**i** Answers are displayed within the problem

## Error and Bug Reports/Technical Issues

[Hide Discussion](#)

**Topic:** Midterm Exam (1 week):Midterm Exam 1 / Problem 2

[Add a Post](#)

Show all posts ▼

by recent activity ▼

- |  |     |
|--|-----|
| <p><b>[Staff]</b> <u>My answer is right, although it is marked as wrong</u><br/> <u>If you could check the first part of 2.2. My answer is right, but I do not get a green tick from grader. Thank you</u></p> | 3 ▼ |
| <p><b>?</b> <u>Are we using programing here</u><br/> <u>I felt confused, it didn't mention that we are able to get the answer without programming.</u></p>   | 6 ▼ |
| <p><b>?</b> <u>[STAFF] clarification on 2.(1)</u></p>  | 3 ▼ |
| <p><b>---</b> <u>edited ---</u><br/> <u>edited ---</u></p>   | 6 ▼ |

✓	<a href="#">How to plot the classifier like the solution in 2-(3)?</a>	6
✓	<a href="#">[STAFF] Problem 2.4 Kernel Function</a>	3
💬	<a href="#">alpha i</a> Can someone explain what is the alpha i parameter in the first problem? I did not get it	2
?	<a href="#">Clarification 2. (2)</a> Is the question asking for theta and theta_0 in the original space or the space of the feature map?	6
?	<a href="#">--- edited ---</a> <a href="#">--- edited ---</a>	5
?	<a href="#">Clarification question 2(4)</a> When you ask "enter $K(x, x')$ in terms of $z$ ", should I understand "in terms of $z_1$ and $z_2$ "? Remove if too revealing, Thanks, J.	2
?	<a href="#">Q2 part 2</a> How does the grader correct this? Won't there be infinitely many thetas that will work, i.e., correctly come up with a classifier? Thanks	1
?	<a href="#">Question about convergence</a> What is the output of the algorithm? The parameters or the predictions?	2
💬	<a href="#">Can I use in answers <math>\text{sqrt}()</math> or should calculate the result</a> Afraid to make mistake when round the answer.	2
?	<a href="#">[STAFF] Partial Credit Application</a>	2

© All Rights Reserved