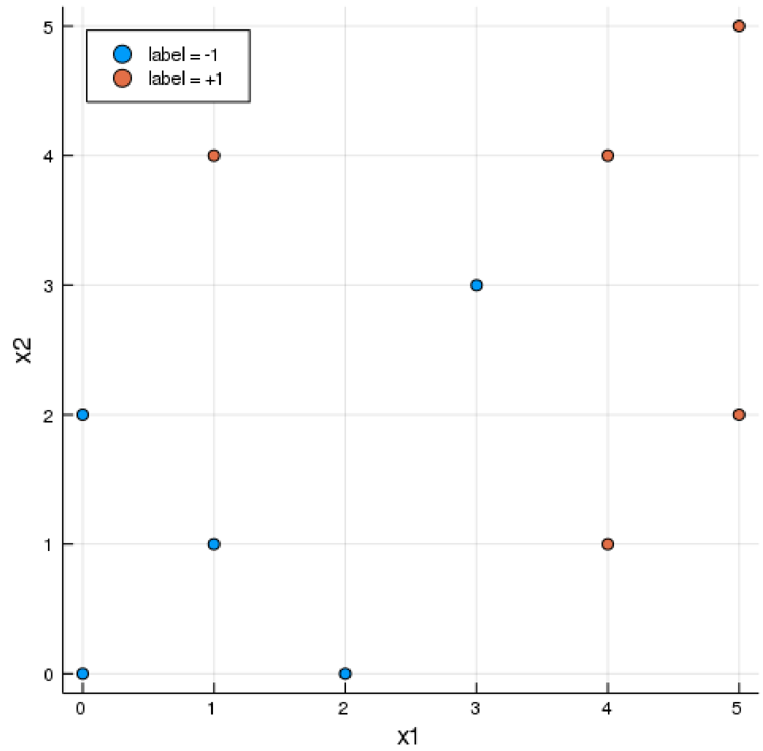
Problem 2

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 θ, θ_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 θ, θ_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).

Submit You have used 1 of 3 attempts

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 θ and θ_0?

(Enter θ_0 accurate to at least 2 decimal places.)

$\theta_0 =$ -110 ✔ **Answer:** -110

(Enter θ as a vector, enclosed in square brackets, and components separated by commas, e.g. type `[0,1]` for $[0 \quad 1]^T$. 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.)

$\theta =$ [21, -22.627417, 22] ✔ **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 1 of 3 attempts

Answers are displayed within the problem

2. (3)

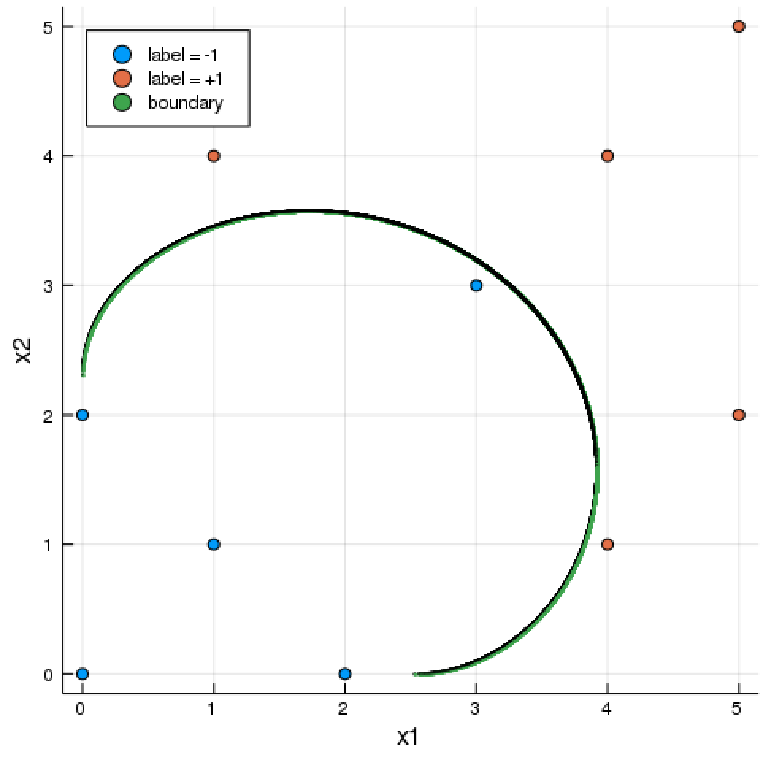
1/1 point (graded)

Based on the calculation of θ and θ_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 $y^{(i)}\left(\theta^T\phi\left(x^{(i)}\right)+\theta_0\right)\geq 0$ for all $i=1,2,\ldots,10$. The boundary looks like the following in the (x_1,x_2)-space.



Submit You have used 1 of 3 attempts

Answers are displayed within the problem

2. (4)

1/1 point (graded)

Recall for $x=\begin{bmatrix}x_1 & x_2\end{bmatrix}^T$

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

Define the kernel function

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

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

$K\left(x,x'\right)=$ z^2 ✔ Answer: z^2

z^2

STANDARD NOTATION

Solution:

Given

$$\phi\left(x\right)=\begin{bmatrix}x_1^2 & \sqrt{2}x_1x_2 & x_2^2\end{bmatrix}^T,$$

we have

$$\begin{aligned}\phi(x)^T\phi\left(x'\right) &= \left(x_1x_1'\right)^2+2\left(x_1x_1'\right)\left(x_2x_2'\right)+\left(x_2x_2'\right)^2\\ &= \left(x_1x_1'+x_2x_2'\right)^2\\ &= z^2 \qquad \text{where } z=x\cdot x' \;=\; x_1x_1'+x_2x_2'.\end{aligned}$$

Submit You have used 2 of 3 attempts