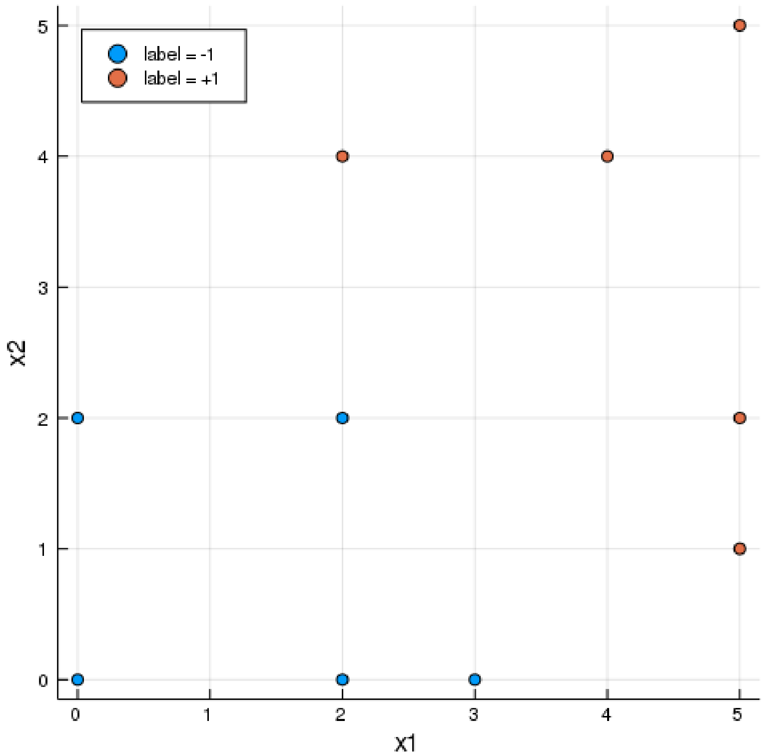
Problem 1

Problem 1. Linear Classification

Consider a labeled training set shown in figure below:



1. (1)

2.0/2 points (graded)

We initialize the parameters to all zero values and run the **linear perceptron algorithm** through these points in a particular order until convergence. The number of mistakes made on each point are shown in the table below. (These points correspond to the data point in the plot above)

Label	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1

Coordinates	(0,0)	(2,0)	(3,0)	(0,2)	(2,2)	(5,1)	(5,2)	(2,4)	(4,4)	(5,5)
Perceptron mistakes	1	9	10	5	9	11	0	3	1	1

Note: You should be able to arrive at the answer without programming.

What is the resulting offset parameter θ_0?

Enter the numerical value for θ_0:

$\theta_0 =$ -18 **Answer:** -18

What is the resulting parameter θ?

(Enter θ as a vector, e.g. type `[0,1]` if $\theta = [0 \quad 1]^T$.)

$\theta =$ [4,4] **Answer:** [4,4]

STANDARD NOTATION

Correction note: July 30 17:00UTC In an earlier version, the note "You should be able to arrive at the answer without programming are not present."

Solution:

Let α_i be the number of mistakes that perceptron makes on the point $x^{(i)}$ with label $y^{(i)}$. The resulting offset parameter is

$$\theta_0 = \sum_{i=1}^{10} \alpha_i y^{(i)} = -18 \tag{6.1}$$

The resulting parameter θ is

$$\theta = \sum_{i=1}^{10} \alpha_i y^{(i)} x^{(i)} = [4 \quad 4]^T.$$

Note that the answer does not depend on the order of data points used in the algorithm. (For reference, the sequence in the perceptron algorithm used here is $(4,4)$, $(0,0)$, $(2,0)$, $(3,0)$, $(5,5)$, $(2,4)$, $(0,2)$, $(2,2)$, $(5,1)$, $(5,2)$.)

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

1. (2)

1/1 point (graded)

Setup as above: We initialize the parameters to all zero values and run the **linear perceptron algorithm** through these points in a particular order until convergence. The number of mistakes made on each point are shown in the table below. (These points correspond to the data point in the plot above.)

Label	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1
Coordinates	(0,0)	(2,0)	(3,0)	(0,2)	(2,2)	(5,1)	(5,2)	(2,4)	(4,4)	(5,5)
Perceptron mistakes	1	9	10	5	9	11	0	3	1	1

The mistakes that the algorithm makes often depend on the order in which the points were considered. Could the point $(5, 2)$ labeled $+1$ have been the first one considered?

- ○ yes
- ◉ no ✔
- ○ depends

Correction Note July 29 15:00UTC: An earlier version of the exam does not include the clarification titled “Setup as above".

Solution:

When perceptron is initialized to all zeros, the first point considered is always a mistake. Since no mistakes were made on the point $(5, 2)$ labeled $+1$, it could not have been the first point considered.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

1. (3)

2.0/2 points (graded)

Suppose that we now find the linear separator that **maximizes** the margin instead of running the perceptron algorithm.

What are the parameters θ_0 and θ corresponding to the **maximum margin separator**?

(Enter θ_0 accurate to at least 3 decimal places.)

$\theta_0 =$ -5 ✔ **Answer:** -5

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

$\theta =$ [1,1] ✔ **Answer:** [1,1]

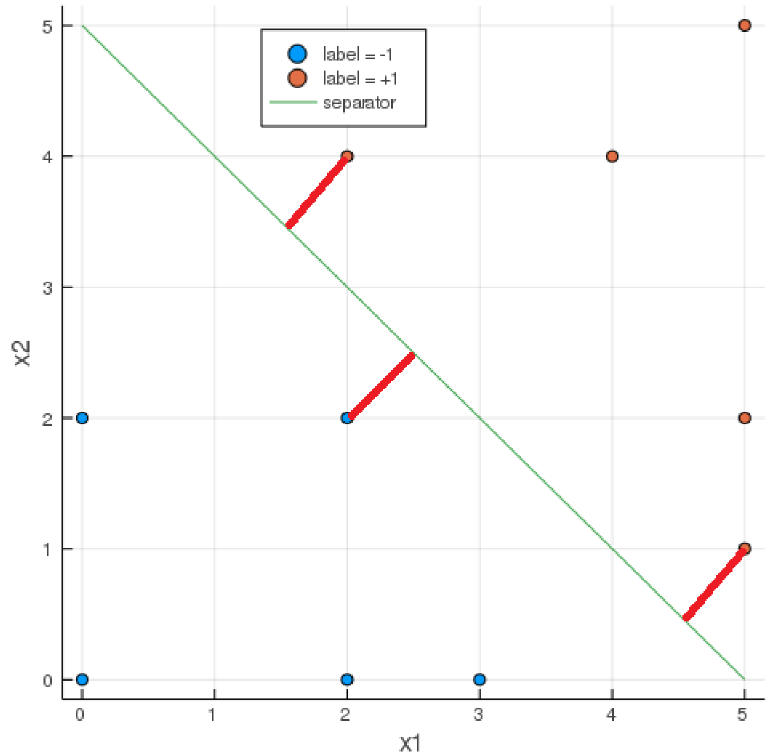
STANDARD NOTATION

Solution:

The margin of a separator is the minimal distance between the separator and any point in the dataset. The equation of the line that maximizes the margin on the given points is $x_1 + x_2 - 5 = 0$. The parameters corresponding to the maximum margin separator are:

$$\theta = [1,1]^T \text{ and } \theta_0 = -5 \quad (6.2)$$

In the following plot, we show the maximum linear separater with the data points. We observe points $(2,2)$, $(2,4)$, $(5,1)$ are support vectors that has margin $1/\sqrt{2}$ (the red line segments denote the margins). Note we have intentionally arranged the data points so you should be able to derive this unique answer by inspection.



Discussion:

Although it is not required by the problem, we feel it might be helpful to also discuss how to find “maximum margin separater" generally. In fact, this is known as the hard margin SVM, i.e., to find the linear hyperplane with the maximum margin, such that all points are classified correctly. In particular, if we omit the bias term θ_0, the goal is to maximize $1/||\theta||$ (or minimize θ^2, equivalently), given $y^{(i)}\left(\theta \cdot x^{(i)}\right) \geq 1$ for all i. This optmization problem can be then written as the following quadratic programming formulation:

$$\begin{aligned} \min_{\theta} \quad & \theta^2 \\ \text{s.t.} \quad & y^{(i)}\left(\theta \cdot x^{(i)}\right) \geq 1, \forall\, i \end{aligned}$$

Note this is different from the commonly seen soft margin SVM, where you can allow some classification errors (that's why it's called "soft" instead of "hard"). In comparison, the soft margin SVM corresponds to the following optmization problem:

$$\begin{aligned} \min_{\theta, \xi \geq 0} \quad & \theta^2 + C \sum_i \xi_i \\ \text{s.t.} \quad & y^{(i)}\left(\theta \cdot x^{(i)}\right) \geq 1 - \xi_i, \forall\, i \end{aligned}$$

If you use SVM program directly from machine learning toolkits such as scikit-learn, it is by default the soft-margin SVM (in scikit-learn the default C equals to 1); and some regularizer term is also likely to be there by default on the objective. If you want to numerically use the soft SVM program to approximately produce a hard SVM result, you need to be very careful on the parameters ($C \leftarrow \infty$, $\lambda \leftarrow 0$). Then you would get something very close within the numerical tolerance.

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

1. (4)

1.0/1 point (graded)

What is the value of the margin attained?

(Enter an exact answer or decimal accurate to at least 2 decimal places.)

1.414 ✔ **Answer:** 1/sqrt(2)

Grading note: Both reasonable answers, i.e. $1/\sqrt{2}$ and $2/\sqrt{2},$ are accepted.

Solution:

The support vectors (points closest to the max-margin separator) are $(2,2)\,,(2,4)$ and $(5,1)$. The distance between any one of these points and the separator is $\frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}}$. Alternatively, we know the margin is $\frac{1}{\|\theta\|} = \frac{1}{\sqrt{2}}$.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

1. (5)

1/1 point (graded)

Using the parameters θ_0 and θ corresponding to the **maximum margin separator**, what is the sum of Hinge losses evaluated on each example?

Sum of hinge losses: 0 **Answer:** 0

Correction Note (July 31 15:00 UTC): An earlier version does not include "Using the parameters θ_0 and θ corresponding to the **maximum margin separator**" in the problem statement.

Solution:

Since the points are linearly separated, the hinge loss is 0. Alternatively, the sum of the hinge losses can be calculated by:

$$\sum_{i=1}^{10} \max\{0, 1 - y^{(i)} \left(\theta \cdot x^{(i)} + \theta_0\right)\} = 0. \tag{6.3}$$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

1. (6)

1/1 point (graded)

Suppose we modify the maximum margin solution a bit and divide both θ and θ_0 by 2. What is the sum of Hinge losses evaluated on each example for this new separator?

Sum of hinge losses: 1.5 **Answer:** 1.5

Solution:

The sum of the hinge losses for the new parameters is:

$$\sum_{i=1}^{10} \max\{0, 1 - y^{(i)}\,(\frac{1}{2}\theta \cdot x^{(i)} + \frac{1}{2}\theta_0)\} = 1.5$$

We can also find the hinge loss visually. Since both θ and θ_0 are scaled by the same constant (0.5), the decision boundary stays the same, but the margin is twice what it was before. The points that were right on the margin [i.e. $(2,2)$,$(2,4)$ and $(5,1)$] will now have a loss of 1/2 each, and all other points has loss of 0, resulting in a total loss of 1.5.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Error and Bug Reports/Technical Issues

Hide Discussion

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

Add a Post

Show all posts ▼ by recent activity ▼

[Staff] Solutions Are Too Concise 6
The answers are way too concise and don't contribute much to the learning. Can we have more detailed answers. I am particularly interested in ...

Mideterm 1.(3) & 1.(4) (memo) 1

? [I answered Q 1.3 correctly](#)
Dears, I sent an email with no reply. I answered question 1.3 correctly. I did submit the correct answer. My mark is not showing this. Actually, i w... 2

? [STAFF] Cannot see question 1.(3) at all
I don't see question 1(3) at all. Instead, I see a message "Could not format HTML for problem. Contact course staff in the discussion forum for ass... 3

? [STAFF] Is it possible to award partial credit for correct solutions to 4, 5, 6 based on response for question 3?
Since all questions 4, 5, 6 all hinge on a correct response for 3, would it be possible to award credit for correct solutions to 4, 5, 6 based on the re... 4

[Staff] Can someone show the entire solution to question 1.3, the maximum margin separator?
I am seeing bits and pieces of info being thrown around but in the interests of learning, I would appreciate it if someone can show the solution fr... 7

I now understand Q1.3 (how to get the maximum margin separator) 3

Q1: There is no reasonable way to figure out what training order did you use.
[Edited to remove exam content. Please email 686exam@mit.edu for further questions.] 6

? [Staff] Can you please evaluate my approach and answers for 1.(3) to 1.(6)
For questions 1.(3) to 1.(6) as there was no mention of any specific algorithm/approach to be used For finding maximum margin separator, I hav... 1

1.3 why is [4,4], -20 wrong
I submitted [4,4] für Theta and -20 for Theta_0 which is the same Graph as with using [1,1] -5; Why is my answer wrong? 5

? [Staff] Midterm Question 1.5 and 1.6
In Midterm question 1.5, wouldn't the hinge loss for each of points (2,4), (2,2), and (5,1) would be 1-sqrt(2)/2 so the sum be 3-3*sqrt(2)/2, since th... 5

Question 1.(3): How to obtain the maximum margin separator?
Given that the solution only states the answer, could somebody elaborate on how to find the maximum margin separator in a formal way withou... 8

? [Staff] Solution of first problem
The solution to first problem is [4,4] and -18 which is the equation theta * x + theta_0 = 0, where theta = [4,4] and theta_0 = -18 so why is [1,1], -4.... 9

[STAFF] The following code gives the right count of perceptron mistakes but not the values of theta. Anyone can explain why?