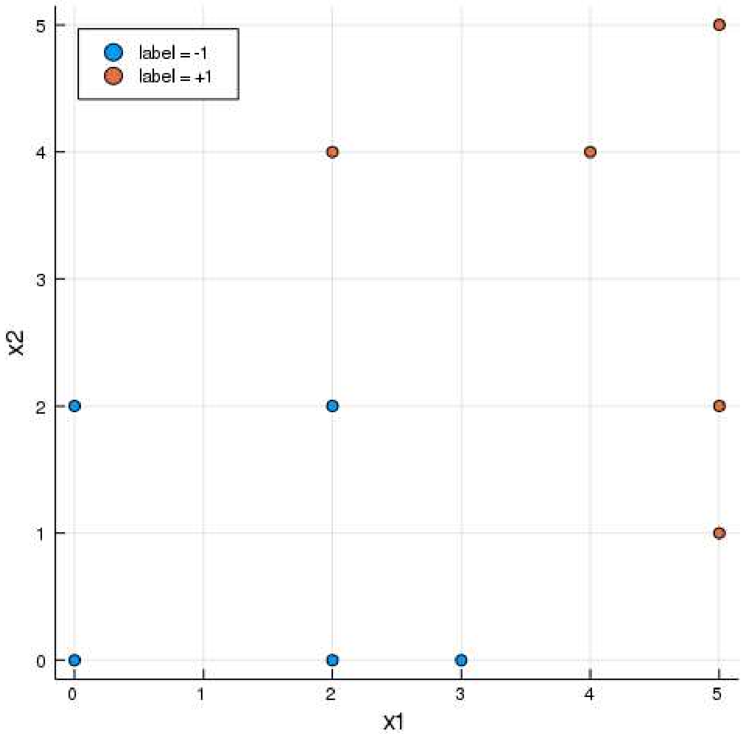
Problem 1

Midterm due Nov 10, 2020 05:29 IST *Completed*

Problem 1. Linear Classification

Consider a labeled training set shown in figure below:



1. (1)

1.0/2 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:

$\theta_0 =$ -17 ✗ **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

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

$$\theta_0 = \sum_{i=1}^{10} \alpha_i y^{(i)} = -18 \tag{7.1}$$

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)

0/1 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 points 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 (selected) ✖

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

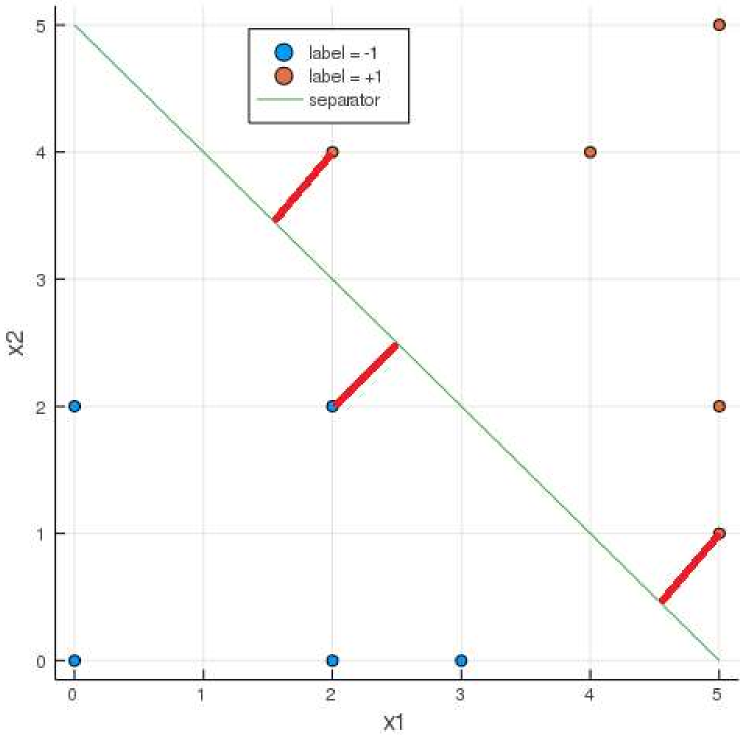
(Enter θ as a vector, enclosed in square brackets, and components separated by commas, e.g. type `[0,1]` for $[0 \quad 1]^T$.)

$\theta =$ [1,1] **Answer:** [1,1]

STANDARD NOTATION

Solution:

A maximum margin separator is in particular a decision boundary whose distance to the closest data point(s) are maximal. In this problem, the data points are arranged so that the maximum margin separator can be found by inspection:



The black line is the maximum linear separater with the data points.
The points $(2,2), (2,4), (5,1)$ are support vectors.
The margin, i.e. the length of the red segments, is $1/\sqrt{2}$.

In general, there are infinitely many equations that describes the same hyperplane, or line, in this case. However, the parameters θ and θ_0 of the maximum margin separator satisifes one more constraint: the positive and /negative margins are defined by $\theta \cdot x + \theta_0 = \pm 1$. This gives the margin, i.e. the distance between either margin boundary and the separator, to be $\frac{1}{||\theta||}$.

Hence, in this problem, we need to choose θ such that $||\theta|| = \sqrt{2}$, where $1/\sqrt{2}$ is the distance between the separator and the closest data point(s). The parameters corresponding to the maximum margin separator are:

$$\theta = [1,1]^T \text{ and } \theta_0 = -5 \tag{7.2}$$

Remark:

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:

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

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:

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

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.)

Grading note: Both reasonable answers wil be accepted. In case the definition of the margin is not clear, we have accepted both the distance between the separator and the margin, and the distance between the 2 margin boundaries, as correct answers.

sqrt(2)

Answer: 1/sqrt(2)

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 $\frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}}$.
By definition, this distance equals the margin of a maximum margin separator, because the support vector(s) lie on the margin boundaries in this case.

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

Solution:

Since the points are linearly separated, the hinge loss is 0.
You can also verify by computing the sum of the hinge losses explicitly by:

$$\sum_{i=1}^{10} \max\{0, 1 - y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right)\} = 0. \tag{7.3}$$

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:

We can find the hinge loss visually. Since both θ and θ_0 are scaled by the same constant $1/2$, 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.
The sum of the hinge losses for the new parameters is:

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

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] 1.(3) Answer differs -1 — 3
- ? [STAFF] My problem is not graded — 5
  see title
- ? Does anyone have a worked solution to 1.3 using the quadratic programming formulation? — 3
  I got this question correct but only after much research and learning how to employ the Lagrang...
- ? [Staff] My answer to Question 1.1 and 1.3 is correct , i want to raise dispute here — 2
  [Staff] My answer to Question 1.1 and 1.3 is correct , just that i have approached in another way ,...
- Please, next time help us with minor disabilities. — 4
  Folks, When in a timed exam. Please use +1 and -1 labels right next to each point. I'm colorblind ,...
- [Staff] Can't the values in the first answer be scaled and still be equivalent?

?	Why is the answer [Theta_zero=-18] and [Theta=4,4] correct while [Theta_zero=-4.5] and [Theta=...	4
	Poorly designed "maximum margin separator" questions I don't believe we have discussed how to find a maximum margin separator specifically, the disc...	2
	My answer is poorly qualified In question 1(3) my answer was theta = [2, 2] and theta_0 = -10 which corresponds to the line 2x ...	3
	Question 1: What were your answers?	11
	[Staff] problem about question 1[1] I ran my programing, got total number of mistakes same as the table show, but the number of m...	36
	[STAFF] A little confusion in my "Progress" section	9
	[Staff] Correct Answers are not visible Earlier it shows that answers are visible on 10th Nov 2020 at 7:45 am IST. Is this the issue with m...	3
?	"Perceptron mistakes": What does it mean? When running the perceptron algorithm and encountering a "mistake" you have to update your t...	2