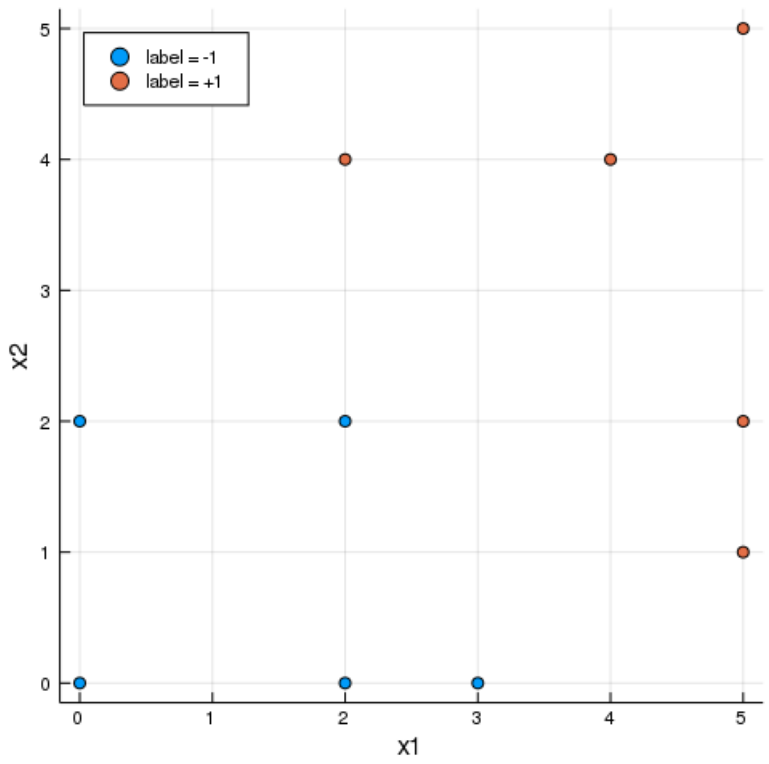
Problem 1

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 $\theta = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$.)

$\theta =$ [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

$$\theta = \sum_{i=1}^{10} \alpha_i y^{(i)} x^{(i)} = \begin{bmatrix} 4 & 4 \end{bmatrix}^T.$$

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 4 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)

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.)

这里我忘记了sklearn的SVM有一个regularization term，
需要把C设定的很大，才能有一个近似解。

$\theta_0 =$ -1.50554144 ✖ **Answer:** -5

所以后面的都错了。
这道题的要求是，不能有一点hinge loss，也就是每个点至少都在margin boundary上面。
本来SVM可以达到这个要求，但是sklearn的SVM有一个regularization term，把θ控制的很小。

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

$\theta =$ [0.33559688, 0.35592169] ✖ **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:

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

Submit You have used 1 of 4 attempts

Answers are displayed within the problem

1. (4)

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.)

2.0442012322326675 ✖ **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)

0/1 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: 2.7487090847583184 ✖ **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:

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

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

1. (6)

0/1 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: 5.515291696489003 ✖ **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 2 of 3 attempts

Answers are displayed within the problem

Error and Bug Reports/Technical Issues

Show Discussion

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