Homework 1

The questions below are due on Sunday September 17, 2017; 11:00:00 PM.

You are not logged in.

If you are a current student, please Log In (https://introml.mit.edu/fall17/homework/hw01?loginaction=login) for full access to this page.

1) PERCEPTRON MISTAKES

Let's apply the perceptron algorithm (through the origin) to a small training set containing three points:

i	Data Points $x^{(i)}$	Labels $y^{(i)}$
1	$[1, -1]$	1
2	$[0, 1]$	-1
3	$[-1.5, -1]$	1

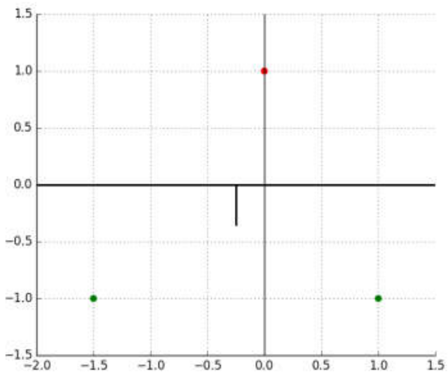
Given that the algorithm starts with $\theta^{(0)} = 0$, the first point that the algorithm sees is always a mistake. The algorithm starts with *some* data point (to be specified in the question), and then cycles through the data until it makes no further mistakes.

1.1) Take 1

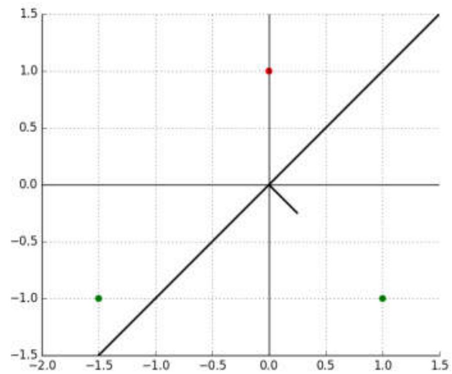
1) How many mistakes does the algorithm make until convergence if the algorithm starts with data point $x^{(1)}$?

Number of mistakes is 2

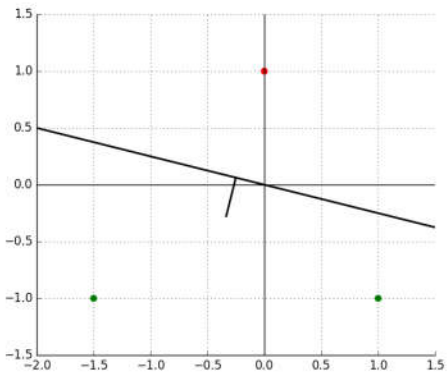
(1)



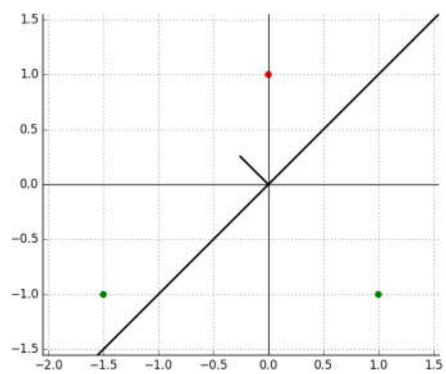
(2)



(3)



(4)



2) Which plot(s) correspond to the progression of the hyperplane as the algorithm cycles? Ignore the initial 0 weights.

Please provide the plot number(s) in the order of progression as a Python list.

```
[2,3]
```

3) How many mistakes does the algorithm make if it starts with data point $x^{(2)}$ (and then does $x^{(3)}$ and $x^{(1)}$)?

Number of mistakes is `1`

4) Which plot(s) correspond to the progression of the hyperplane as the algorithm cycles? Ignore the initial 0 weights.

Please provide the plot number(s) in the order of progression as a Python list.

```
[1]
```

1.2) Take 2

Now assume that $x^{(3)} = [-10, -1]$, with label 1.

1) How many mistakes does the algorithm make until convergence if cycling starts with data point $x^{(1)}$?

Number of mistakes is 6

2) How many mistakes if it starts with data point $x^{(2)}$?

Number of mistakes is 1

2) DUAL VIEW

The following table shows a data set and the number of times each point is misclassified during a run of the perceptron algorithm (with offset). θ is initialized to zero.

i	$x^{(i)}$	$y^{(i)}$	times misclassified
1	$[-3, 2]$	1	2
2	$[-1, 1]$	-1	4
3	$[-1, -1]$	-1	2
4	$[2, 2]$	-1	1
5	$[1, -1]$	-1	0

1) What is the post training θ?

Provide it as a python list of the form $[a, b]$.

[-2,0]

2) What is the post training θ_0?

Provide it as a number. -5

3) DECISION BOUNDARIES

3.1) AND

Consider the AND function defined over three binary variables: $f(x_1, x_2, x_3) = (x_1 \wedge x_2 \wedge x_3)$.

We aim to find a θ such that, for any $x = [x_1, x_2, x_3]$, where $x_i \in \{0, 1\}$:

$$\theta \cdot x + \theta_0 > 0 \text{ when } f(x_1, x_2, x_3) = 1, \text{ and}$$

$$\theta \cdot x + \theta_o < 0 \text{ when } f(x_1, x_2, x_3) = 0.$$

1) For each of the combination of values of (x_1, x_2, x_3), that is, $[(0,0,0),(0,0,1),(0,1,0),(0,1,1),(1,0,0),(1,0,1),(1,1,0),(1,1,1)]$ enter the values of $f(x_1, x_2, x_3)$.

Please enter the values of $f(x_1, x_2, x_3)$ as a Python list.

```
[0, 0, 0, 0, 0, 0, 0, 1]
```

2) Assuming $\theta_0 = 0$ (no offset), enter θ as a Python list of length 3 or enter `'none'` as a Python string (with quotes) if none exists.

Enter a Python list of 3 numbers or the string `'none'`

```
'none'
```

3) Assuming θ_0 is non-zero (offset), enter a θ and θ_0 as a Python list of length 4 (θ_0 last) or enter `'none'` as a Python string (with quotes) if none exists.

Enter a Python list with 4 numbers or the string `'none'`.

```
[4, 3, 2, -8]
```

3.2) Families

You are given the following labeled data points:

- Positive examples: $[-1, 1]$ and $[1, -1]$,
- Negative examples: $[1, 1]$ and $[2, 2]$.

For each of the following parameterized families of classifiers, find the parameters of a family member that can correctly classify the above data, or think about why no such family member exists.

Is there a classifier of the following forms that can correctly classify the above data?

1) Inside or outside of an origin-centered circle with radius r

Enter a value for r or the string `'none'` if none exists.'

```
'none'
```

2) Inside or outside of a circle centered on x_0 with radius r

Enter a list with 3 entries for coordinates of x_0 and r (`[x0_1,x0_2,r]`) or the string `'none'` if none exists.' `[2, 2, 2]`

3) On one side of a line through the origin with normal θ

Enter a list with 2 entries for coordinates of θ or the string `'none'` if none exists.'

`'none'`

4) On one side of a line with normal θ and offset θ_0.

Enter a list with 3 entries for coordinates of θ and θ_0 (`[theta_1, theta_2, theta0]`)or the string `'none'` if none exists.'

`[-1, -1, 1]`

5) Which of the above are families of linear classifiers?

Enter a Python list with a subset of the numbers 1, 2, 3, 4.

`[3,4]`

4) INITIALIZATION

1) If we were to initialize the perceptron algorithm with $\theta = [1000, -1000]$, how would it effect the number of mistakes made in order to separate the data set from question 1?

○	It would have small or no effect on the number of mistakes.
○	It would significantly decrease the number of mistakes.
◉	It would significantly increase the number of mistakes.

2) Provide a value of $\theta^{(0)}$ for which running the perceptron algorithm (through origin) on the data set from question 1 returns a different result then using $\theta^{(0)} = 0$. The data set is repeated below:

i	Data Points $x^{(i)}$	Labels $y^{(i)}$
1	$[1, -1]$	1
2	$[0, 1]$	-1
3	$[-1.5, -1]$	1

Enter 2 values for θ as a Python list or the string `'none'` if none exists.'

[1000, -1000]

3) This question is a placeholder. Please ignore it.

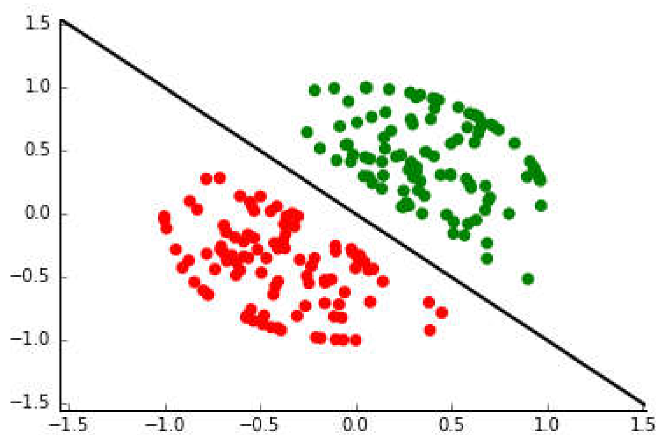
○	Bad
○	Bad
○	Bad
○	Bad
◉	Good

5) MISTAKES AND GENERALIZATION

5.1) Plots

Consider the following plots. For each one estimate plausible values of R and γ. Consider values of R in the range $[1, 10]$ and values of γ in the range $[0.01, 2]$.

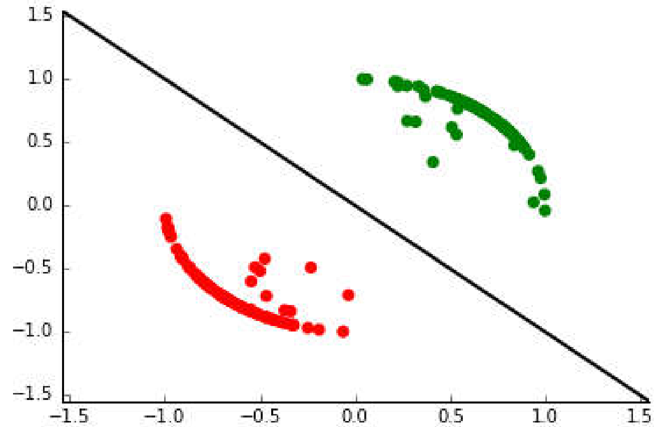
(1)



Enter a Python list with 2 floats, a value of R and a value of gamma

```
[0.1, 2.1]
```

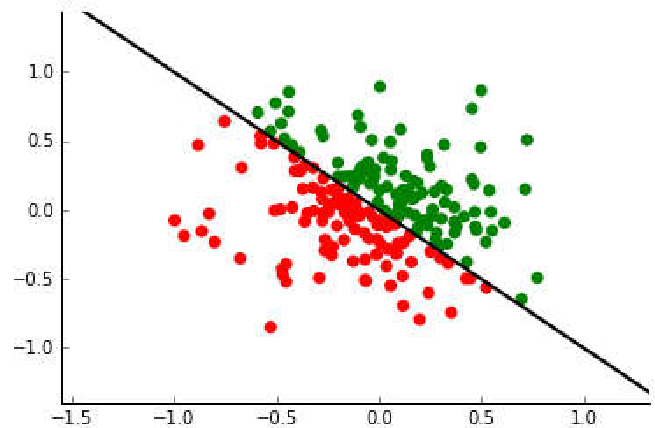
(2)



Enter a Python list with 2 floats, a value of R and a value of gamma

```
[0.4, 2.1]
```

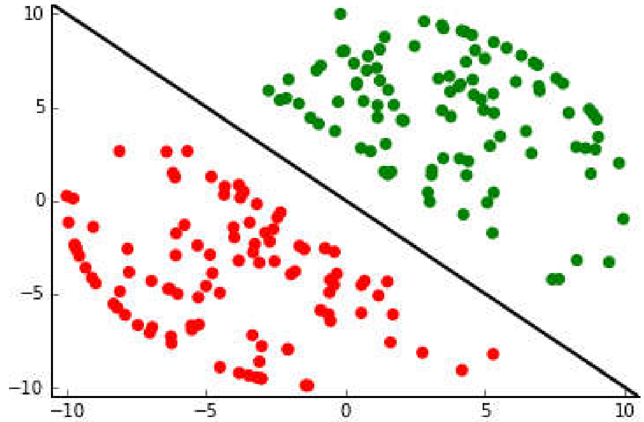
(3)



Enter a Python list with 2 floats, a value of R and a value of gamma

```
[0.01, 2.1]
```

(4)



Enter a Python list with 2 floats, a value of R and a value of gamma

[0.2, 10]

5.2) Mistake Bound

1) What is an upper bound on mistakes when $R = 1, d = 4, n = 1000$ for each of the following values of γ?

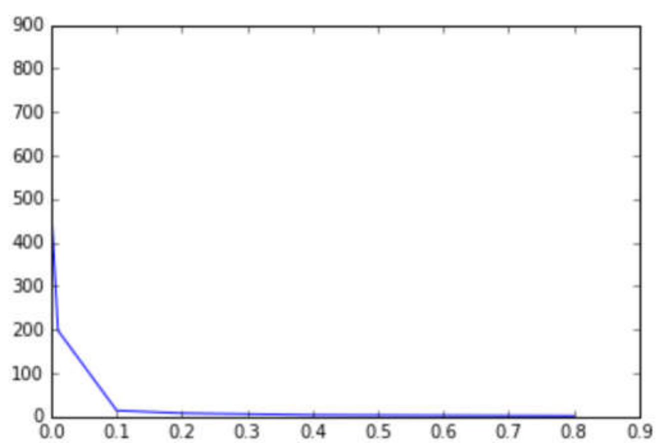
```
(.00001, .0001, .001, .01, .1, .2, .4)
```

Enter a Python list with 7 floats.

[9999999999.999996, 100000000.0, 1000000.0, 10000.

2) Here is a plot, for $R = 1, d = 4, n = 1000$ of the actual numbers of mistakes made by the perceptron on one particular run, as a function of γ.

The actual numbers of mistakes (on y axis) are: [862, 414, 446, 198, 14, 8, 4].



Are the data consistent with the theory? Yes.

6) SEPARATION

Write a Python procedure that takes as input a dataset that is:

- not linearly separable without an offset
- linearly separable with an offset

and returns a new data set such that running perceptron without offset on this new data set should enable us to find a separator for the original data set. This transformation should be the same independent of the data values; in particular, it should not need to know the separator.

The input data set is specified by a d by n data array and a 1 by n labels array. The output of the procedure should be a tuple of a data array and a label array, but possibly with a dimension different from d.

```
import numpy as np

#row of 1s on the original data, that accounts for the theta_o.
#output: (d+1,n)
def new_data(data, labels):
    n = data.shape[1]
    return np.vstack((np.array([[1]*n]), data)), labels

```