

EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the [Privacy Policy](#). ×



[Unit 1 Linear Classifiers and](#)  
[Course](#) > [Generalizations \(2 weeks\)](#) > [Homework 2](#) > 3. Perceptron Updates

### 3. Perceptron Updates

In this problem, we will try to understand the convergence of perceptron algorithm and its relation to the ordering of the training samples for the following simple example.

Consider a set of  $n = d$  labeled  $d$ —dimensional feature vectors,  $\{(x^{(t)}, y^{(t)}), t = 1, \dots, d\}$  defined as follows:

$$x_i^{(t)} = \cos(\pi t) \quad \text{if } i = t \quad (3.7)$$

$$x_i^{(t)} = 0 \quad \text{otherwise,} \quad (3.8)$$

Recall the no-offset perceptron algorithm, and assume that  $\theta \cdot x = 0$  is treated as a mistake, regardless of label. Assume that in all of the following problems, we initialize  $\theta = 0$  and when we refer to the perceptron algorithm we only consider the no-offset variant of it.

## Working out Perceptron Algorithm

3/3 points (graded)

Consider the  $d = 2$  case. Let  $y^{(1)} = 1, y^{(2)} = 1$ . Assume that the feature vector  $x^{(1)}$  is presented to the perceptron algorithm before  $x^{(2)}$ .

For this particular assignment of labels, work out the perceptron algorithm until convergence.

Let  $\hat{\theta}$  be the resulting  $\theta$  value after convergence. Note that for  $d = 2$ ,  $\hat{\theta}$  would be a two-dimensional vector. Let's denote the first and second components of  $\hat{\theta}$  by  $\hat{\theta}_1$  and  $\hat{\theta}_2$  respectively.

Please enter the total number of updates made to  $\theta$  by perceptron algorithm:

✓ Answer: 2

Please enter the numerical value of  $\hat{\theta}_1$ :

✓ Answer: -1

Please enter the numerical value of  $\hat{\theta}_2$ :

✓ Answer: 1

**Solution:**

The first iteration of perceptron with data point  $(x^{(i)}, y^{(i)})$  will be a mistake due to our initialization of  $\theta^{(0)} = \bar{0}$ . The first update sets  $\theta^{(1)} = y^{(i)} x^{(i)} = x^{(i)}$ . The second iteration of perceptron with data point  $(x^{(j)}, y^{(j)})$  will also yield a mistake since

$$\theta^{(1)} \cdot x^{(j)} = x^{(i)} \cdot x^{(j)} = 0$$

. Thus, for the  $d = 2$  example, after 2 updates  $\theta^{(2)} = x^{(i)} + x^{(j)}$ . We now check whether the second pass yields mistakes

$$\begin{aligned} y^{(i)} \theta^{(2)} \cdot x^{(i)} &= y^{(i)} (y^{(i)} x^{(i)} + y^{(j)} x^{(j)}) \cdot x^{(i)} \\ &= \|x^{(i)}\|^2 + x^{(i)} \cdot x^{(j)} \\ &= \|x^{(i)}\|^2 + 0 \\ &> 0 \end{aligned}$$

so the first point is classified correctly. Similarly, the second point is also classified correctly.

Therefore, it only takes two updates to classify the  $d = 2$  dataset and  $\theta$  converges to  $\hat{\theta} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

You have used 1 of 3 attempts

**i** Answers are displayed within the problem

## General Case - Number of updates

1/1 point (graded)

Now consider any  $d > 0$ .

For the specific dataset we are considering, please choose the correct answer from the options below:

- ☒ Perceptron algorithm will make exactly  $d$  updates to  $\theta$  regardless of the order and labelings of the feature vectors ✓
- ☐ Perceptron algorithm will make at least  $d$  updates to  $\theta$  with the exact number of updates depending on the ordering of the feature vectors presented to it
- ☐ Perceptron algorithm will make at least  $d$  updates to  $\theta$  with the exact number of updates depending on the ordering of the feature vectors presented to it and their labeling
- ☐ Perceptron algorithm will make at most  $d$  updates to  $\theta$  with the exact number of updates depending on the ordering of the feature vectors presented to it and their labeling

Submit

You have used 1 of 3 attempts

## Sketching convergence

1/1 point (graded)

Consider the case with  $d = 3$ . Also assume that all the feature vectors are positively labelled. Let  $P$  denote the plane through the three points in a 3-d space whose vector representations are given by the feature vectors  $x^{(1)}, x^{(2)}, x^{(3)}$ .

Let  $\hat{\theta}$  denote the value of  $\theta$  after perceptron algorithm converges for this example. Let  $v$  denote the vector connecting the origin and  $\hat{\theta}$ . Which of the following options is true regarding the vector represented by  $\hat{\theta}$ .

- ☐  $v$  is parallel to the plane  $P$
- ☐  $v$  is perpendicular to the plane  $P$  and pointing away from it
- ☒  $v$  is perpendicular to the plane  $P$  and pointing towards it ✓
- ☐  $\hat{\theta}$  lies on the plane  $P$

**Solution:**

Note that from the previous problem

$$\hat{\theta} = \sum_{i=1}^d y^{(i)} x^{(i)}$$

Evaluating the above expression, for the current example gives,

$$\hat{\theta} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}$$

The equation of the plane is given by:

$$\begin{vmatrix} x & -1 & 0 \\ y-1 & -1 & -1 \\ z & 0 & -1 \end{vmatrix} = 0$$

That is,

$$x - y + z = -1$$

Or equivalently,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix} = 1$$

Hence,  $v$  is perpendicular to  $P$  and pointing towards it.

Submit

You have used 3 of 3 attempts

**i** Answers are displayed within the problem

## Discussion

[Hide Discussion](#)

**Topic:** Unit 1 Linear Classifiers and Generalizations (2 weeks):Homework 2 / 3. Perceptron Updates

[Add a Post](#)

Show all posts ▼

by recent activity ▼

**[Staff]** [About Sketching convergence](#)

4 new\_ 20 ▼

**?** [Sketching convergence. Vector direction.](#)

6 ▼

**✓** [\[Staff\] Question 3 Part 1](#)

Any hints regarding how to turn  $\cos(\pi \cdot t)$  into a 2R vector so we can run perceptron on it? I'm a little confused by that.

12 ▼

**?** [Working out Perceptron Algorithm - why  \$x\(i\) \cdot x\(j\) = 0\$](#)

Hi guys, I am still not sure why  $x(i) \cdot x(j) = 0$  in the first qs. Probably because I can't interpret the start of the question where it says:  $x(i)(t) = \cos(\pi \cdot t)$

4 ▼

**[staff]** [Sketching convergence](#)

Hi, the first time I submitted my answer, it was write but it marked wrong. So, I changed my answer thinking I misunderstand the concept. Now, I...

1 ▼

**[Staff]** [For fetching convergence, I literally projected a graph to see lol. Image in description.](#)

2 ▼

✓ <u>Sketching convergence</u>	11
🗨 <u>first question X values</u> <u>can you help me with understanding what possible values of X are we evaluating?. I'm a bit lost here on how to run a simulation</u>	4
✓ <u>Question about sketching convergence</u>	4
✓ <u>Use of software/python till now</u> <u>Hi, I have till now not used python or any software till now in any homework or lecture exercises is that right approach or way it is supposed to b...</u>	6
✓ <u>Working out Perceptron Algorithm</u> <u>I got the correct answers by luck but don't understand why "the total number of updates" is like that</u>	7
✓ <u>x(i) dimensions</u> <u>I run into a issue when running the perceptron update. In previous exercises, such as HW 1, when theta was a 2 dimensional vector, x(i) was also...</u>	7
🗨 <u>[STAFF] Show Answer Button should appear at least in homework</u> <u>There are so many difficulties in this course and this appears in the number of posts in the discussion forum (help needed, etc). It is way too unfa...</u>	2
✓ <u>Q on "General Case - Number of updates"</u>	2

© All Rights Reserved