



MITx 6.86x

**Machine Learning with Python-From Linear Models to Deep Learning**[Course](#)[Progress](#)[Dates](#)[Discussion](#)[Resources](#)[Course](#) / [Unit 1. Linear Classifiers and Generalizations \(2 weeks\)](#) / [Homework 1](#)[< Previous](#)

## 5. Passive-aggressive algorithm

[Bookmark this page](#)

Homework due Feb 22, 2023 08:59 -03 Past due

In this problem, we will try to understand the loss in Passive-Aggressive (PA) Perceptron

The passive-aggressive (PA) algorithm (without offset) responds to a labeled training example  $(x, y)$  by updating the parameters  $\theta$  that minimizes

$$\frac{\lambda}{2} \|\theta - \theta^{(k)}\|^2 + \text{Loss}_h(y\theta \cdot x)$$

where  $\theta^{(k)}$  is the current setting of the parameters prior to encountering  $(x, y)$  and

$$\text{Loss}_h(y\theta \cdot x) = \max\{0, 1 - y\theta \cdot x\}$$

is the hinge loss. (If we wished, we could replace the loss function with something else; this would give a different algorithm.) The above determines  $\theta^{(k+1)}$  in terms of  $\theta^{(k)}$ . Your exercise is to write  $\theta^{(k+1)}$  more concretely in terms of those latter parameters — i.e., to write an update equation. The form of the update is similar to the perceptron algorithm, i.e.,

$$\theta^{(k+1)} = \theta^{(k)} + \eta yx$$

but the real-valued step-size parameter  $\eta$  is no longer equal to one; it now depends on the training example  $(x, y)$ .

## Update equation

1 point possible (graded)

Consider minimizing the above defined loss function with the hinge loss component. What happens to the step size at large values of  $\lambda$ ? Please choose one from the options below.

☐ If  $\lambda$  is large, the step-size of the algorithm ( $\eta$ ) would be large

☐ If  $\lambda$  is large, the step-size of the algorithm ( $\eta$ ) would be small

Submit

You have used 0 of 3 attempts

## Loss functions and decision boundaries

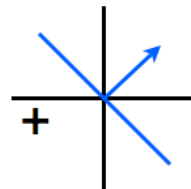
0.0/1.0 point (graded)

Consider minimizing the above defined loss function and the setting of our decision boundary. We ran our PA algorithm on the next data point in our sequence - a positively-labeled vector  $\mathbf{x}$ . We plotted the results of our algorithm after the update, by trying out a few different loss functions and  $\eta$  as follows:

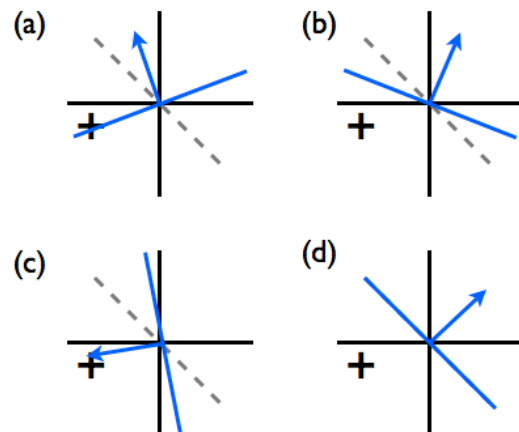
1. hinge loss and a large  $\eta$
2. hinge loss and a small  $\eta$
3. 0-1 loss and a large  $\eta$
4. 0-1 loss and a small  $\eta$

Each of the options below provides a matching between the 4 variations above with a decision boundary plotted in a-d below. Please choose the options that match them correctly. Note that the dashed lines correspond to the previous decision boundary, and the solid blue lines correspond to the new decision boundary; also, note that these are just sketches which ignore any changes to the magnitude of the weights.

setting before update:

















possible settings after update:



(a)  ▼

(b)  ▼

	<u>Last question answer explanation:</u> For 1-b, Loss can be minimised by moving theta vector towards example. Does the example mean the positive example?
	<u>still confused about theta, theta_k and theta_{k+1}</u> I understand that theta is the output we want to get. theta_k is the current setting of parameters, and we update it to theta_{k+1}.
	<u>Unclear if we should ditch influence of x, y in optimal parameter eta</u> It is unclear if we should ditch influence of x, y in optimal parameter eta.
	<u>Hinge Loss notation -- subscript h</u> The hinge loss function is always written as "Loss_h", and I've been wondering what "h" refers to. It just dawned on me.
	<u>dot product notation</u> Hi, how can I write the dot product of the vectors x and y? he is not recognizing dot(x,y). Should I use the notation x \cdot y?
	<u>Loss function and decision boundaries</u> How do we determine the loss function and decision boundaries based on the graphs?
	<u>Where can I learn all that?</u> Can we at least get a list of resources where we could study all this? I can accept a gap, but this is all a compendium of things.
	<u>How to calculate step size question</u> lambda is in the objective function while the step size is in the updating of theta function. How do I link them?
	<u>please read these highlights first</u> difference between hinge loss and 0-1 loss <a href="https://discussions.edx.org/course-v1:MITx+6.86x+1T2023/topics/0-1-loss">https://discussions.edx.org/course-v1:MITx+6.86x+1T2023/topics/0-1-loss</a>  <u>Community TA</u>
	<u>Theta vs Theta_k?</u> This might be a stupid question, but I am having trouble wrapping my mind around the distinction between theta and theta_k.
	<u>What is the "zero-one loss"?</u> Hi everyone. If I've been following this class attentively, we haven't defined what the "zero-one loss" is, right?
	<u>differentiating a norm</u> Getting a little confused with differentiating norm vars. ie is norm(nyx)^2 differentiated with respect to theta
	<u>Relationship between magnitude of hinge loss and step size</u>

# edX

[About](#)

[Affiliates](#)

[edX for Business](#)

[Open edX](#)

[Careers](#)

[News](#)

# Connect

[Blog](#)

[Contact Us](#)

[Help Center](#)

[Security](#)

[Media Kit](#)



© 2023 edX LLC. All rights reserved.

深圳市恒宇博科技有限公司 [粤ICP备17044299号-2](#)