



[Unit 1 Linear Classifiers and Course](#) > [Generalizations \(2 weeks\)](#) > [Homework 2](#) >
2. Passive-aggressive algorithm

Audit Access Expires May 11, 2020

You lose all access to this course, including your progress, on May 11, 2020. Upgrade by Mar 25, 2020 to get unlimited access to the course as long as it exists on the site. [Upgrade now](#)

2. Passive-aggressive algorithm

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

The passive-aggressive (PA) algorithm (without offset) responds to a labeled training example (x, y) by finding θ that minimizes

$$\frac{\lambda}{2} \|\theta - \theta^{(k)}\|^2 + \text{Loss}_h(y\theta \cdot x) \quad (3.5)$$

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. We could replace the loss function with something else (e.g., the zero-one loss). The form of the update is similar to the perceptron algorithm, i.e.,

$$\theta^{(k+1)} = \theta^{(k)} + \eta yx \quad (3.6)$$

but the real-valued step-size parameter η is no longer equal to one; it now depends on both $\theta^{(k)}$ and the training example (x, y) .

Update equation

1/1 point (graded)

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

☐ If λ is large, the step-size of the algorithm (η) would be large

☒ If λ is large, the step-size of the algorithm (η) would be small



Solution:

As λ increases, the passive-aggressive algorithm is dissuaded from updating very far from the previous θ . This is because λ serves as the weight for the regularization term (the portion that prevents overfitting: $\|\theta - \theta^{(k)}\|$) of the minimizing function.

Thus, as λ increases, we expect the step size between updates to decrease, so η should be smaller.

Submit

You have used 2 of 3 attempts

i Answers are displayed within the problem

Calculating the Step size

1/1 point (graded)

Suppose $\text{Loss}_h(y\theta^{(k+1)} \cdot x) > 0$ after the update. Express the value of η in terms of λ in this case. (Hint: you can simplify the loss function in this case).

1/lambda

**Answer:** 1/lambda $\frac{1}{\lambda}$ STANDARD NOTATION**Solution:**

When $\text{Loss}_h(y\theta^{(k+1)} \cdot x) > 0$, the hinge loss,

$$\max\{0, 1 - y\theta \cdot x\}$$

can be simply rewritten as $1 - y\theta \cdot x$. Thus, our minimizing function simplifies as follows:

$$\begin{aligned} f(\theta) &= \frac{\lambda}{2} \|\theta - \theta^{(k)}\|^2 + \text{Loss}_h(y\theta \cdot x) \\ &= \frac{\lambda}{2} \|\theta - \theta^{(k)}\|^2 + 1 - y\theta \cdot x \end{aligned}$$

We compute the minimum by setting gradient of $f(\theta)$ w.r.t. θ to 0:

$$\begin{aligned} \nabla_{\theta} f &= \lambda(\theta - \theta^{(k)}) - yx = 0 \\ (\theta - \theta^{(k)}) &= \frac{1}{\lambda} yx \\ \theta &= \theta^{(k)} + \frac{1}{\lambda} yx \end{aligned}$$

where $\frac{1}{\lambda}$ is just the step size.

Thus, $\eta = \frac{1}{\lambda}$ when $\text{Loss}_h(y\theta^{(k+1)} \cdot x) > 0$.

Submit

You have used 2 of 3 attempts

i Answers are displayed within the problem

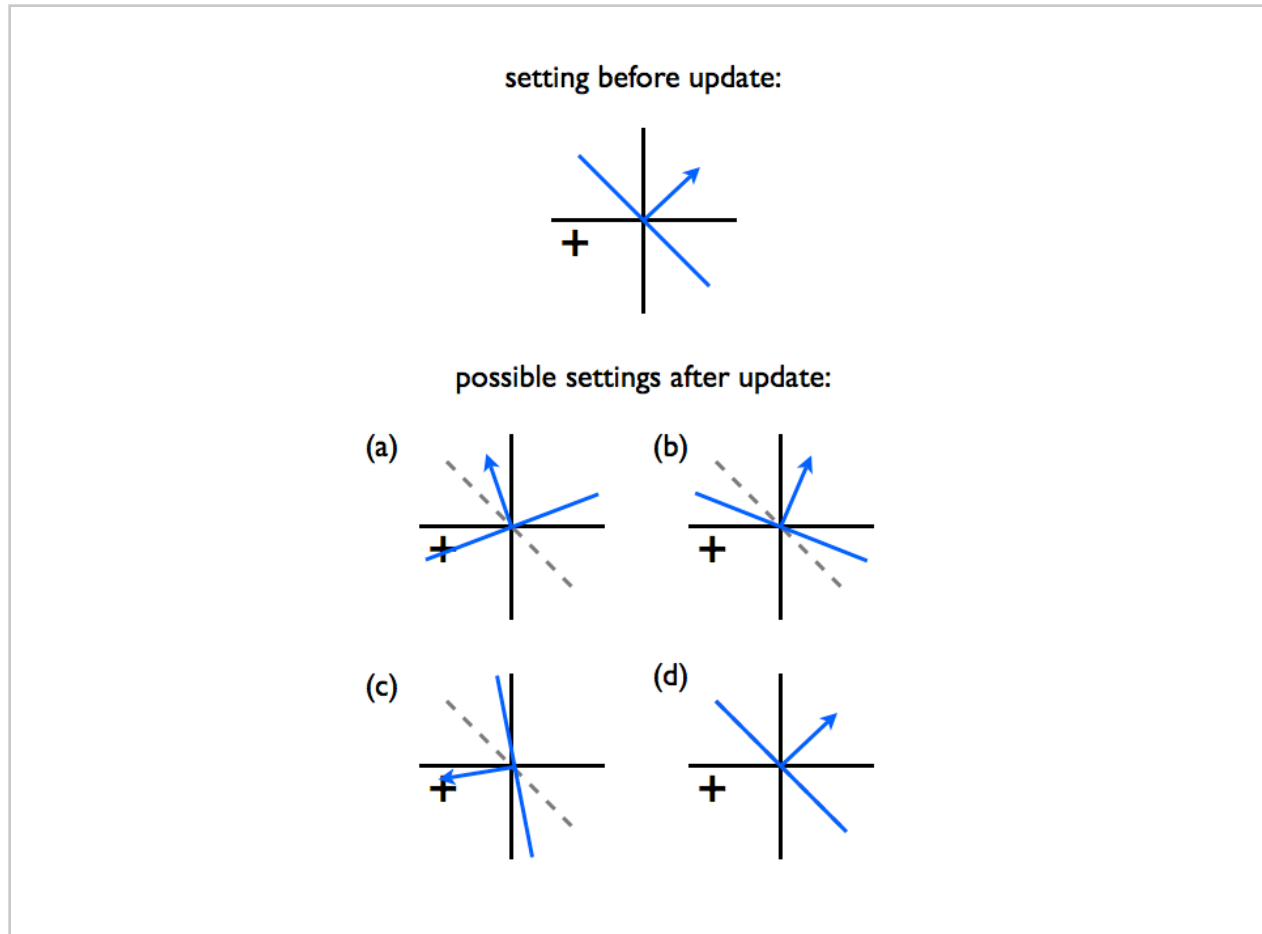
Loss functions and decision boundaries

1.0/1.0 point (graded)

Consider minimizing the above defined loss function and the setting of our decision boundary plotted below. We ran our PA algorithm on the next data point in our sequence - a positively-labeled vector (indicated with a $+$). We plotted the results of our algorithm after the update, by trying out a few different variations of loss function and λ as follows:

1. hinge loss and a large λ
2. hinge loss and a small λ
3. 0-1 loss and a large λ
4. 0-1 loss and a small λ

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 dotted 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 θ .



- (a) **✓ Answer:** 0-1 loss, small lambda
- (b) **✓ Answer:** Hinge loss, large lambda
- (c) **✓ Answer:** Hinge loss, small lambda
- (d) **✓ Answer:** 0-1 loss, large lambda

Solution:

- 1 - b hinge loss and a large λ . For hinge loss, loss can be improved by moving the θ vector towards the example. However, since λ is large, the change in θ term dominates, so θ only rotates slightly.
- 2 - c hinge loss and a small λ . For hinge loss, loss can be improved by moving the θ vector towards the example. Since λ is small, the loss term dominates, so the loss is minimized when the example is correctly classified. The difference

between hinge loss and 0-1 loss with small λ is that hinge loss will correctly classify the example with a margin from the boundary. This is because hinge loss is minimized when $y\theta \cdot x \geq 1$ while 0-1 is when $y\theta \cdot x \geq 0$.

- 3 - d 0-1 loss and a large λ . Since the loss function is zero-one, it can only take on two values whether its classified correctly or not. With λ large, the change in θ term dominates. In order to minimize the change in θ , θ simply stays the same after the update and does not improve the loss at all.
- 4 - a 0-1 loss and a small λ . Since λ is small, the zero-one loss term dominates. Since the loss function can only improve when it is classified correctly, the example must be on the positive side. Because all correctly classified examples show the same loss, the secondary goal of minimizing the change in θ will cause the point to be immediately on the positive side of the decision boundary.

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 / 2. Passive-aggressive algorithm

[Add a Post](#)

Show all posts

by recent activity

- | | |
|---|---|
| ? Help understanding "minimizing the above defined loss function with the hinge loss component" | 5 |
| ? Loss functions and decision boundaries - Question about 0-1 loss function
The gradient of the 0-1 loss function is 0 except when X=0. If we replace the loss function in e... | 5 |
| 💬 [staff]: Calculating the Step size - completely don't understand what to there 6 new | 9 |
| ? Misleading statement in 2b | 2 |

?	[STAFF] Please Advise Question Loss functions and decision boundaries	6
	Greetings, in the lecture notes, we are given mainly that the margin boundary and decision b...	
?	zero-one loss function	1 new_
	hi how will one differentiate the zero-one loss function to calculate theta increment? Regards...	
💬	Hint Loss Functions and Decision Boundaries	1
	This video really helped me visualise the new vector positions due to algorithm updates. http...	
💬	question 1	3
	Seriously? The question has 2 possible options, but we're allowed 3 attempts ;-)	
💬	Found these two resources very insightful	2
	https://www.bonaccorso.eu/2017/10/06/ml-algorithms-addendum-passive-aggressive-algorit...	
✓	What is k indexing in equation (3.6)?	2
💬	Loss functions and decision boundaries	3
with a decision boundary, plotted in **a-d** below.... Could somebody, please, tell me w...	
💬	Some useful text!	5
	Hello everyone! How are you? In this text you will find some useful information for this sectio...	
💬	hint	1
	answering this question requires a bit of search for "online passive aggressive algorithms"	

Learn About Verified Certificates

© All Rights Reserved