✕

edX

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

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

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

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

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

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.,

$$\theta^{(k+1)} = \theta^{(k)} + \eta\, y x \tag{3.6}$$

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 1 of 3 attempts

ℹ Answers are displayed within the problem

Calculating the Step size

1/1 point (graded)

Suppose $\mathrm{Loss}_h \left(y\theta^{(k+1)} \cdot x \right) > 0$ 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 $\mathrm{Loss}_h \left(y\theta^{(k+1)} \cdot x \right) > 0$, 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:

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

$$= \frac{\lambda}{2}\left\|\theta - \theta^{(k)}\right\|^2 + 1 - y\theta \cdot x$$

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

$$\nabla_\theta f = \lambda(\theta - \theta^{(k)}) - yx = 0$$

$$(\theta - \theta^{(k)}) = \frac{1}{\lambda}yx$$

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

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 1 of 3 attempts

ⓘ   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:
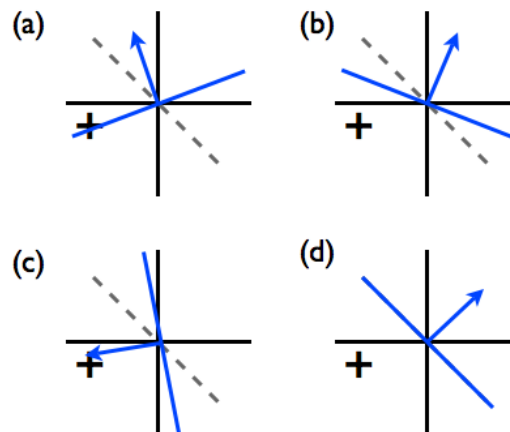
1. hinge loss and a large λ

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.

setting before update:
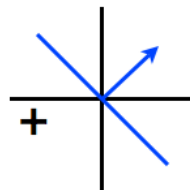
possible settings after update:

(a) [0-1 loss, small lambda   ▼]   ✔ **Answer:** 0-1 loss, small lambda

(b) [Hinge loss, large lambda ▼]   ✔ **Answer:** Hinge loss, large lambda

(c) [Hinge loss, small lambda ▼]   ✔ **Answer:** Hinge loss, small lambda

(d) [0-1 loss, large lambda ▼] ✔ **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 2 of 3 attempts

ⓘ Answers are displayed within the problem

Discussion

Topic: Unit 1 Linear Classifiers and Generalizations (2 weeks):Homework 2 / 2. Passive-aggressive algorithm

Hide Discussion

Add a Post

Show all posts ▼	by recent activity ▼

☑ **Calculating the Step size: Right answer, wrong sign?** — **12**

☑ **Dont quite understand the question 3** — 2 new_ **5**
Have trouble to identify the answers. Any hint?

☑ **[STAFF] Solution to Calculating the Step size problem** — **3**
Hello, Now that the deadline for this work has passed, is it possible to have the solution for the "Calculating the Step size" problem? Tks

☑ **Still missing basic understanding of the process** — **10**

☑ **Calculating the Step size: Where can be mistake in logic?** — **2**
Greetings, I try to solve the Step size problem in the following way. I know (3.5). I know (3.6). In the task, I'm given that loss_$h(y*theta^{(k+1)}*x) > 0$...

☑ **Loss Function and decision boundaries -- Request for feedback** — 5 new_ **10**

☑ **[last part] How the different losses would affect the update of theta?** — 3 new_ **5**

💬 **Homework Questions** — **15**
Hi, I don't know if this happens to anyone else, but I am completely lost in this homework. I've gotten the right answers mainly by guessing and i...

☑ **[Staff] Clarification for Problem 2 Passive-Agressive Algorithm** — **3**

💬 **Differentiating a square norm** — **4**
I was wondering when I differentiate w.r.t to a variable within a norm, what is the explanation that I can remove the norm around the variable? I...

☑ [Staff] Passive-aggressive algorithm for non-zero offset   3

☑ [Staff] Grader not accepting answers for Loss functions and decision boundaries   2

I can neither submit nor save answers for the Loss functions and decision boundaries question. Is there an issue with the grader?

☑ Calculating the Step size   17

☑ [Question 3] is Hinge Loss and 0-1 Loss Equivalent   2

Excuse my ignorance, is Hinge Loss and 0-1 Loss equivalent…. or I miss/misinterpret some part of it? Thank you