Problem 3

Midterm due Nov 9, 2020 18:59 EST

Stochastic gradient descent (SGD) is a simple but widely applicable optimization technique. For example, we can use it to train a Support Vector Machine. The objective function in this case is given by:

$$J(\theta) \quad = \quad \left[\frac{1}{n}\sum_{i=1}^{n} \text{Loss}_h\left(y^{(i)}\theta\cdot x^{(i)}\right)\right] + \frac{\lambda}{2}\|\theta\|^2$$

where $\text{Loss}_h(z) = \max\{0, 1-z\}$ is the hinge loss function, $(x^{(i)}, y^{(i)})$ with for $i = 1, \ldots n$ are the training examples, with $y^{(i)} \in \{1, -1\}$ being the label for the vector $x^{(i)}$.

For simplicity, we ignore the offset parameter θ_0 in all problems on this page.

3. (1)

3 points possible (graded, results hidden)

The stochastic gradient update rule involves the gradient $\nabla_\theta \text{Loss}_h\left(y^{(i)}\theta\cdot x^{(i)}\right)$ of $\text{Loss}_h\left(y^{(i)}\theta\cdot x^{(i)}\right)$ with respect to θ.

*Hint:*Recall that for a k-dimensional vector $\theta = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_k \end{bmatrix}^T$, the gradient of $f(\theta)$ w.r.t. θ is $\nabla_\theta f(\theta) = \begin{bmatrix} \frac{\partial f}{\partial \theta_1} & \frac{\partial f}{\partial \theta_2} & \cdots & \frac{\partial f}{\partial \theta_k} \end{bmatrix}^T$.)

Find $\nabla_\theta \mathrm{Loss}_h(y\theta \cdot x)$ in terms of x.

(Enter `lambda` for λ, `y` for y and `x` for the vector x. Use `*` for multiplication between scalars and vectors, or for dot products between vectors. Use `0` for the zero vector.)

For $y\theta \cdot x \leq 1$:

$\nabla_\theta \mathrm{Loss}_h(y\theta \cdot x) =$

For $y\theta \cdot x > 1$:

$\nabla_\theta \mathrm{Loss}_h(y\theta \cdot x) =$

Let θ be the current parameters. What is the stochastic gradient update rule, where $\eta > 0$ is the learning rate? (Choose all that apply.)

$\theta \rightarrow$

- [ ] $\theta + \eta \nabla_\theta \left[\text{Loss}_h \left(y^{(i)}\theta \cdot x^{(i)}\right)\right] + \eta\lambda\theta$ for random $x^{(i)}$ with label $y^{(i)}$
- [ ] $\theta - \eta \nabla_\theta \left[\text{Loss}_h \left(y^{(i)}\theta \cdot x^{(i)}\right)\right] - \eta\lambda\theta$ for random $x^{(i)}$ with label $y^{(i)}$
- [ ] $\theta + \eta \nabla_\theta \left[\text{Loss}_h \left(y^{(i)}\theta \cdot x^{(i)}\right)\right] + \eta \nabla_\theta \left[\frac{\lambda}{2}\|\theta\|^2\right]$ for random $x^{(i)}$ with label $y^{(i)}$
- [ ] $\theta - \eta \nabla_\theta \left[\text{Loss}_h \left(y^{(i)}\theta \cdot x^{(i)}\right)\right] - \eta \nabla_\theta \left[\frac{\lambda}{2}\|\theta\|^2\right]$ for random $x^{(i)}$ with label $y^{(i)}$
- [ ] $\theta + \eta \sum_{i=1}^{n} \nabla_\theta \left[\text{Loss}_h \left(y^{(i)}\theta \cdot x^{(i)}\right)\right] + \eta \nabla_\theta \left[\frac{\lambda}{2}\|\theta\|^2\right]$
- [ ] $\theta - \eta \displaystyle\sum_{i=1}^{n} \nabla_\theta \left[\text{Loss}_h \left(y^{(i)}\theta \cdot x^{(i)}\right)\right] - \eta \nabla_\theta \left[\frac{\lambda}{2}\|\theta\|^2\right]$

Grader is correct: The grader behaves as intended in this problem. If you get an input error, please check your answers carefully. You will also need to complete all parts of the question before the submit button will be un-grayed.

STANDARD NOTATION

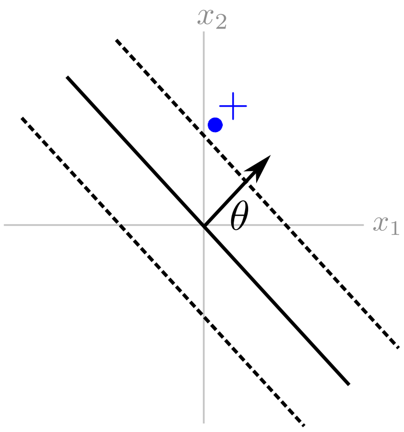
Submit

You have used 0 of 3 attempts

3. (2)

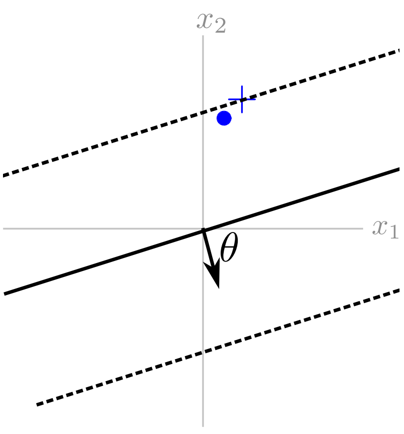
1 point possible (graded, results hidden)

Suppose the current parameter θ is as in the figure below:

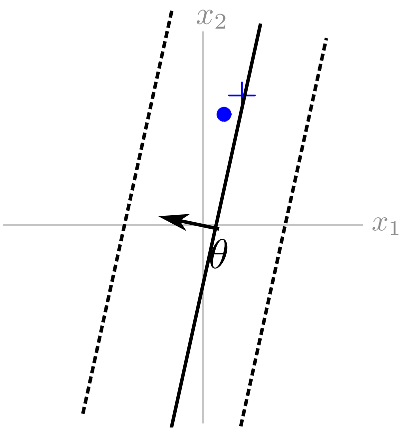


Here, θ is in the direction of the arrow, the solid line represents the classifer defined by $\theta,$ and the dotted lines represent the positive and negative margin boundaries.

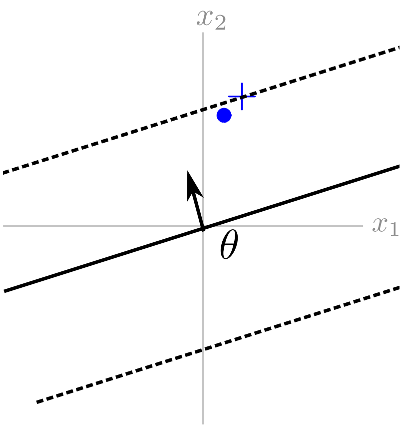
For large η (i.e. η close to 1) $0.5 < \eta\lambda < 1$, which of the following figure corresponds to a single SGD update made in response to the point labeled '+' above?



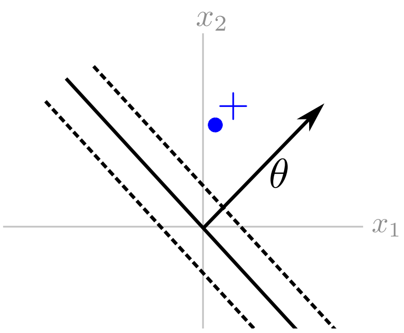
x_2
x_1
θ



x_2
x_1
θ



x_2
x_1
θ



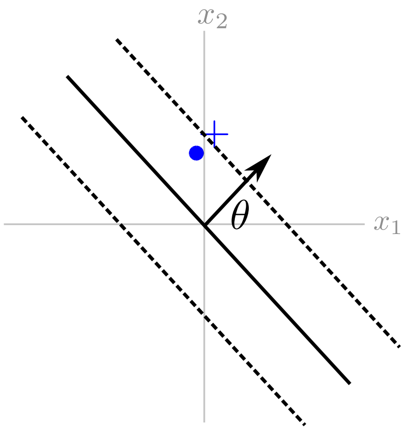
x_2
x_1
θ

Submit You have used 0 of 3 attempts

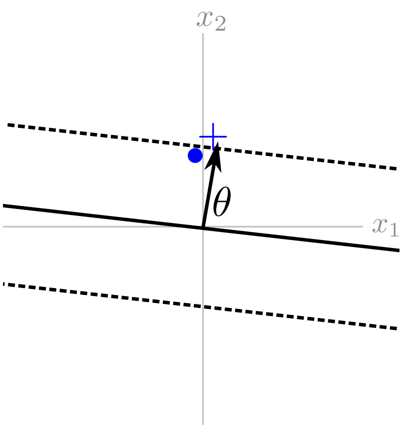
3. (3)

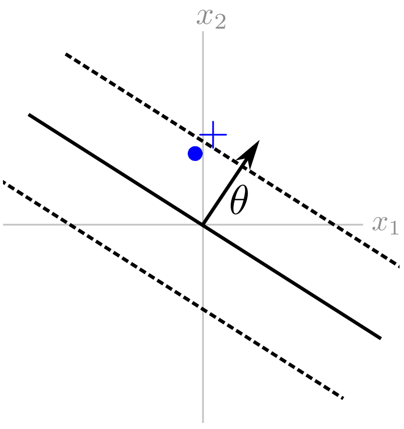
1 point possible (graded, results hidden)

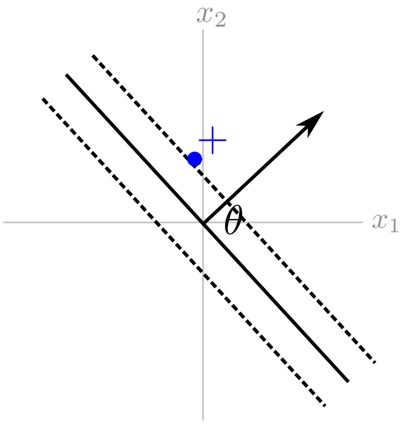
Again for large η (i.e. η close to 1) and $0.5 < \eta\lambda < 1$, but now we perform a single SGD update made in response to a different point labeled '+', shown below:

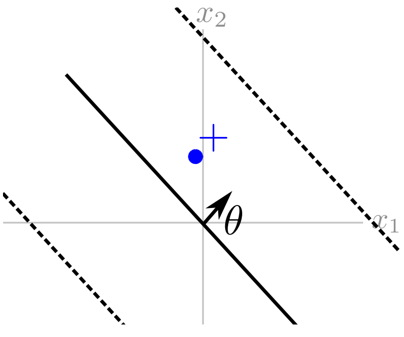


which of the following figure corresponds to a single SGD update made in response to the point labeled '+' above?


x_2
θ
x_1


x_2
θ
x_1


x_2
θ
x_1


x_2
θ
x_1

Submit You have used 0 of 3 attempts

Error and Bug Reports/Technical Issues

Hide Discussion

Topic: Midterm Exam (1 week):Midterm Exam 1 / Problem 3

Add a Post

Show all posts | by recent activity

- [staff] 3. (1). Clarification needed. — 1
  The problem statement mentions SGD. The question 3.(1) mentions stochastic gradient, without r...
- Use 'xxx' as zero vector. — 3
  Question 1, asks us to input 'something similar to phi' as zero vector. What should we actually typ...
- Clarification with figures — 1
  Could you please clarify : In both the figures there is a single point denoted with a dot? The "+" is j...
- [STAFF] Isn't the update arrow the wrong way round in Q3.(1)? — 1
- Question 3(1) — 3
  I find the phrasing of the question a bit confusing. Should the answer to "Find ∇θLossh(yθ·x) in te...
- Invalid Input: \'lambda\' not permitted in answer as a variable — 4
  Why I am getting the error while Submitting the answer Invalid Input: \'lambda\' not permitted in ...