2. Perceptron Performance

In class we initialized the perceptron algorithm with $\theta = 0$. In this problem we will also explore other initialization choices.

2. (a)

2.0/2 points (graded)

The following table shows a data set and the number of times each point is misclassified during a run of the perceptron algorithm (**with offset** θ_0). θ and θ_0 are initialized to zero.

i	$x^{(i)}$	$y^{(i)}$	times misclassified
1	[-4, 2]	+1	1
2	[-2, 1]	+1	0
3	[-1, -1]	-1	2
4	[2, 2]	-1	1
5	[1, -2]	-1	0

Write down the state of θ and θ_0 after this run has completed (note, the algorithm may not yet have converged). Enter θ as a list $[\theta_1, \theta_2]$ and θ_0 as a single number in the following boxes.

Please enter θ:

[-4,2]

Answer: [-4, 2]

Please enter θ_0 :

-2

Answer: -2

Solution:

- Since perceptron update rule updates θ simply by adding $x^{(i)}y^{(i)}$, the resulting θ should be the summation of all mistakes.

- Additional Insight: since perceptron update rule is additively associative, doing updates in any order would lead to the same result.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

2. (b)

0/2 points (graded)

Provide one example of a different initialization of θ such that the perceptron algorithm with this initialization would not produce any mistakes during a run through the data.

$[\theta_1, \theta_2]$: [-1, -1]

Answer: See solution

θ_0: 10

Answer: See solution

Solution:

The answer (θ, θ_0) should be such that:

- $-4\theta_1 + 2\theta_2 + \theta_0 > 0$
- $-2\theta_1 + \theta_2 + \theta_0 > 0$
- $-\theta_1 - \theta_2 + \theta_0 < 0$
- $2\theta_1 + 2\theta_2 + \theta_0 < 0$
- $1\theta_1 - 2\theta_2 + \theta_0 < 0$

For instance, any line strictly inside the boundry $x_2 = x_1$ and $x_2 = x_1 + 3$ are valid solutions.

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

2. (c)

2.0/3 points (graded)

The theorem from question 1. (e) provides an upper bound on the number of steps of the Perceptron algorithm and implies that it indeed converges. In this question, we will show that the result still holds even when θ is not initialized to 0.

In other words: Given a set of training examples that are linearly separable through the origin, show that the initialization of θ does not impact the perceptron algorithm's ability to eventually converge.

To derive the bounds for convergence, we assume the following inequalities holds:

- There exists θ^* such that $\dfrac{y^{(i)}\,(\theta^* x^{(i)})}{\|\theta^*\|} \geq \gamma$ for all $i = 1, \cdots, n$ and some $\gamma > 0$
- All the examples are bounded $\|x^{(i)}\| \leq R, i = 1, \cdots, n$

If θ is initialized to 0, we can show by induction that:

$$\theta^{(k)} \cdot \frac{\theta^*}{\|\theta^*\|} \geq k\gamma$$

For instance,

$$\theta^{(k+1)} \cdot \frac{\theta^*}{\|\theta^*\|} = (\theta^{(k)} + y^{(i)} x^{(i)}) \cdot \frac{\theta^*}{\|\theta^*\|} \geq (k+1)\,\gamma$$

If we initialize θ to a general (not necessarily 0) $\theta^{(0)}$, then:

$$\theta^{(k)} \cdot \frac{\theta^*}{\|\theta^*\|} \geq a + k\gamma$$

Determine the formulation of a in terms of θ^* and $\theta^{(0)}$:

Important: Please enter θ^* as `theta^{star}` and $\theta^{(0)}$ as `theta^{0}`, and use `norm(...)` for the vector norm $\|...\|$.

$a =$ theta^{0}*theta^{star}/norm(theta^{star}) ✓

Answer: theta^{0}*theta^{star} / norm(theta^{star})

If θ is initialized to 0, we can show by induction that:

$$\|\theta^{(k)}\|^2 \leq kR^2$$

For instance,

$$\|\theta^{(k+1)}\|^2 \leq \|\theta^{(k)} + y^{(i)} x^{(i)}\|^2 \leq \|\theta^{(k)}\|^2 + R^2$$

If we initialize θ to a general (not necessarily 0) $\theta^{(0)}$, then:

$$\|\theta^{(k)}\|^2 \leq kR^2 + c^2$$

Determine the formulation of c^2 in terms of $\theta^{(0)}$:

$c^2 =$ norm(theta^{0})^2 ✔ **Answer:** norm(theta^{0})^2

From the above inequality, we can derive the inequality $\|\theta^{(k)}\| \leq c + \sqrt{k}R$ by applying the following inequality: $\sqrt{x^2+y^2} \leq \sqrt{(x+y)^2}$ if $x, y > 0$.

If θ is initialized to 0, we then use the fact that $1 \geq \frac{\theta^{(k)}}{\|\theta^{(k)}\|} \cdot \frac{\theta^*}{\|\theta^*\|}$ to get the upper bound $k \leq \frac{R^2}{\gamma^2}$.

In the case where we initialize θ to a general $\theta^{(0)}$, use the inequality for $\theta^{(k)} \cdot \frac{\theta^*}{\|\theta^*\|}$ above and the inequality $\|\theta^{(k)}\| \leq c + \sqrt{k}R$ to derive a bound on the number of iterations k.

Hint: Use the larger root of a quadratic equation to obtain the upper bound.

Note: Give your answer in terms of a, c, R, γ (enter the latter as `gamma`).

$k \leq$

(-(2*a*gamma-R^2) - sqrt((2*a*gamma-R^2)^2-4*gamma^2*(a^2-c^2)))/(2*gamma^2)

✖ **Answer:** (R*sqrt(R^2+4*(c-a)*gamma) + R^2 + 2*(c-a)*gamma) / (2*gamma^2)

$$\frac{-(2\cdot a\cdot\gamma-R^2)-\sqrt{(2\cdot a\cdot\gamma-R^2)^2-4\cdot\gamma^2\cdot(a^2-c^2)}}{2\cdot\gamma^2}$$

STANDARD NOTATION

Solution:

The first bound follows by recursion of $\theta^k \cdot \frac{\theta^*}{\|\theta^*\|} \geq \theta^{k-1} \cdot \frac{\theta^*}{\|\theta^*\|} + \gamma$.

The second bound follows by recursion of $\|\theta^k\|^2 \leq \|\theta^{k-1}\|^2 + R^2$.

The final bound is obtained by solving the inequality $1 \geq \theta^k \cdot \frac{\theta^*}{\|\theta^k\|\|\theta^*\|} \geq \frac{a+k\gamma}{c+\sqrt{k}R}$, i.e. $a + k\gamma - c \leq \sqrt{k}R$.

At this point, you can square both sides and solve the quadratic equation to get the upper bound.

Alternatively, solve the quadratic equation for $\sqrt{k}$ and square the answer to get the desired upper bound:

$$k \leq \frac{\left(R + \sqrt{R^2 - 4\gamma\left(a - c\right)}\right)^2}{4\gamma^2}$$

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

2. (d)

2/2 points (graded)

Since the convergence of the perceptron algorithm doesn't depend on the initialization, the end performance on the training set must be the same. Are the resulting θ's the same regardless of the initialization?

- Yes
- No

Does this necessarily imply that the performance on a test set is the same?

- Yes
- No

Solution:

- Any distinct θ that can separate the data are valid solutions, so there are infinitely many different valid correct θ in general given that the data can be separated by more than 1 line.
- Two different θ would always make different predictions for a testing data point between the two lines, so the testing performance is always different for a testing dataset that contains exactly this point.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

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

Add a Post

Show all posts by recent activity

RIght answer was marked as incorrect
Staff, can you please check my answer on the last part of 2c? I wrote the right answer but I distributed the square and it was marke... 2

Interpreting question 2.(c).3 2

[Staff] - Correct answers
Hi [Staff], can we have access to the correct answers? 2

[Staff] Please extend the deadline of hw1 and hw2 by 2 days
Please understand that we need some more time to absorb the material. The concept is a little confusing and doing all these 3 dea... 3

staff: 2.c input error
when I try to submit my answer, I have the following statement as a result: "Expected answer to be a scalar, but input is a vector". T... 3

[STAFF] 2.(c)-3 please check my answer
Could you please check my answer in 2.(c)-3, k<=....? I'm wondering whether my answer is wrong, or just input issue... Thank you. 14

[Staff] mistake on submission
Hi, I made a stupid mistake on submission. I actually got it right on 2(a) and when I moved on to 2(b) I accidently input the answer i... 9

[Staff] Can you please check my answer for 2(b)
I think I have got it right but the answer wasn't graded right. 6

[STAFF] 2C - Solve K
Staff, I believe I was able to get K right. But when entering it , it gves me wrong answer. Could you please check my anwer and advis... 2

[STAFF] 2C
Please check my answer. I dont know why is it marked wrong thank u 2

[Staff] Please don't let enter an answer if it was already accepted as correct one (2a)
Dear staff, I submitted an answer to 2(a) from 2nd attempt it was correct and graded, then I solved 2(b) and, but entered the answe... 2

Not clear with 2b
Staff, could you pls elaborate on what is required as an output in Q2b above. Don't understand this one. thanks 12 new 22

[Staff] Why is my answer in 2b wrong
The question asks for a sample value of theta and theta_0, which will not produce any mistakes. I have provided such a value. Why i... 2