2. Perceptron Performance

Bookmark this page

Verified Track Access

Graded assessments are available to Verified Track learners. Upgrade to unlock ($300)

Homework due Sep 29, 2020 19:59 EDT Completed

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]

Please enter θ_0 :

-2

Submit You have used 1 of 3 attempts

Save

2. (b)

2/2 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]

θ_0: -2

Submit You have used 1 of 3 attempts

Save

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 $\frac{y^{(i)}\,(\theta^* x^{(i)})}{||\theta^*||} \geq \gamma$ 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}))

$\theta^0 \cdot \frac{\theta^\star}{||\theta^\star||}$

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

$||\theta^0||^2$

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`).

$k \leq$

a*r*c

$a \cdot r \cdot c$

STANDARD NOTATION

Submit You have used 2 of 3 attempts

Save

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

Submit You have used 2 of 3 attempts

Save