1. Perceptron Mistakes

In this problem, we will investigate the perceptron algorithm with different iteration ordering.

Consider applying the perceptron algorithm **through the origin** based on a small training set containing three points:

$x^{(1)}$ =[-1,-1],	$y^{(1)}$=1
$x^{(2)}$ =[1,0],	$y^{(2)}$=-1
$x^{(3)}$ =[-1, 1.5],	$y^{(3)}$=1

Given that the algorithm starts with $\theta^{(0)}=0$, the first point that the algorithm sees is always considered a mistake. The algorithm starts with some data point and then cycles through the data (in order) until it makes no further mistakes.

1. (a)

4.0/4 points (graded)

How many mistakes does the algorithm make until convergence if the algorithm starts with data point $x^{(1)}$? How many mistakes does the algorithm make if it starts with data point $x^{(2)}$?

Also provide the progression of the separating plane as the algorithm cycles in the following **list format**: $[[\theta_1^{(1)},\theta_2^{(1)}],\ldots,[\theta_1^{(N)},\theta_2^{(N)}]]$, where the superscript denotes different θ as the separating plane progresses. For example, if θ progress from $[0,0]$ (initialization) to $[1,2]$ to $[3,-2]$, you should enter $[[1,2],[3,-2]]$

Please enter the **number of mistakes** of Perceptron algorithm if the algorithm starts with $x^{(1)}$.

2

✔ **Answer:** 2

Please enter the **progression of the separating hyperplane (θ, in the list format described above)** of Perceptron algorithm if the algorithm starts with $x^{(1)}$.

()

[[-1,-1],[-2,0.5]]

✔ **Answer:** [[-1, -1], [-2, 0.5]]

Please enter the **number of mistakes** of Perceptron algorithm if the algorithm starts with $x^{(2)}$.

1

✔ **Answer:** 1

Please enter the **progression of the separating hyperplane (θ, in the list format described above)** of Perceptron algorithm if the algorithm starts with $x^{(2)}$.

[[-1,0]]

✔ **Answer:** [[-1, 0]]

Solution:

- If the algorithm starts with $x^{(1)}$, then it will encouter two errors: $x^{(1)}$ and $x^{(3)}$.
- Since $[0,0]\cdot x^{(1)}=0$, the result is not greater than 0.
- It would induce an update $\theta^{(1)}=\theta^{(0)}+y^{(1)}x^{(1)}=[-1,-1]$.
- Then $\theta^{(1)}\cdot x^{(2)}<0$, so there is no mistakes.

- Finally, $\theta^{(1)} \cdot x^{(3)} < 0$, which incur an error and would induce an update $\theta^{(2)} = \theta^{(1)} + y^{(3)} x^{(3)}$.

- Afterwards, there should be no mistakes.

- If the algorithm starts with $x^{(2)}$, then it will only encounter one error: $x^{(2)}$. The derivation is similar to the above analysis.

- The progression of the separating hyperplane reflects the updated hyperplane after encountering errors.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

1. (b)

1/1 point (graded)
In part (a), what are the factors that affect the number of mistakes made by the algorithm?

Note: Only choose factors that were changed in part (a), **not** all factors that can affect the number of mistakes

(Choose all that apply.)

- [x] Iteration order ✔
- [ ] Maximum margin between positive and negative data points
- [ ] Maximum norm of data points

✔

Solution:

- Only the iteration order is changed in part (a).

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

1. (c)

4.0/4 points (graded)
Now assume that $x^{(3)} = [-1, 10]$. How many mistakes does the algorithm make until convergence if cycling starts with data point $x^{(1)}$?

Also provide the progression of the separating plane as the algorithm cycles in the following **list format**: $[[\theta_1^{(1)}, \theta_2^{(1)}], \ldots, [\theta_1^{(N)}, \theta_2^{(N)}]]$, where the superscript denotes different θ as the separating plane progresses. For example, if θ progress from $[0, 0]$ (initialization) to $[1, 2]$ to $[3, -2]$, you should enter $[[1, 2], [3, -2]]$

Please enter the **number of mistakes** of Perceptron algorithm if the algorithm starts with $x^{(1)}$.

6 ✔ **Answer:** 6

Please enter the **progression of the separating hyperplane (θ, in a list format described above)** of Perceptron algorithm if the algorithm starts with $x^{(1)}$.

[[-1,-1],[-2,9],[-3,8],[-4,7],[✔ **Answer:** [[-1, -1],[-2, 9],[-3,8],[-4,7],[-5,6],[-6,5]]

Please enter the **number of mistakes** of Perceptron algorithm if the algorithm starts with $x^{(2)}$.

1 ✔ **Answer:** 1

Please enter the **progression of the separating hyperplane (θ, in the list format described above)** of Perceptron algorithm if the algorithm starts with $x^{(2)}$.

[[-1,0]]

✔ Answer: [[-1, 0]]

Solution:

- The derivation is similar to part (a).
- If the algorithm starts with $x^{(1)}$, then it will encounter one error for $x^{(1)}$, one error for $x^{(3)}$, and then 4 errors for $x^{(1)}$.
- If the algorithm starts with $x^{(2)}$, then it will only encounter one error: $x^{(2)}$.
- The progression of the separating hyperplane reflects the updated hyperplane after encountering errors.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

1. (d)

1/1 point (graded)

For a fixed iteration order, what are the factors that affect the number of mistakes made by the algorithm between part (a) and part (c)?

Note: Only choose factors that were changed between part (a) and part (c), **not** all factors that can affect the number of mistakes

(Choose all that apply.)

- [ ] Iteration order
- [ ] Maximum margin between positive and negative data points
- [x] Maximum norm of data points ✔

✔

Solution:

- The maximum norm of data points cause part (c) to make more mistakes.

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

1. (e) (Optional)

0 points possible (ungraded)

In 1962, Novikoff has proven the following theorem.

Assume:

- 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 $\|x^{(i)}\| \leq R, i = 1, \cdots, n$

Then the number k of updates made by the perceptron algorithm is bounded by $\frac{R^2}{\gamma^2}$.

(Note that the first condition implies that the data is linearly separable)

For proof, refer to theorem 1 of this paper. Based on this theorem, what are the factors that constitute the bound on the number of mistakes made by the algorithm?

(Choose all that apply.)

- [ ] Iteration order
- [x] Maximum margin between positive and negative data points ✔
- [x] Maximum norm of data points ✔
- [ ] Average norm of data points

✔

Solution:

- Iteration order would affect relative convergence speed, but does not constitute the bound in this theorem.
- The maximum margin and the maximum norm of data points consitute the bounds in the maximum number of mistakes.
- We can always scale an easy dataset to achieve the same average norm of data points with the same number of mistakes.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

1. (f) (Optional)

0 points possible (ungraded)

Now we want to establish an adversarial procedure to maximize the number of mistakes the perceptron algorithm makes. What are possible solutions? You should consider a general dataset instead of part (a) and part (c). (Choose all that apply.)

- [x] Exhaustic search the worst ordering ✔
- [x] Dynamic Programming the worst ordering ✔
- [x] Greedily select the data point with the maximum norm

✖

Solution:

- There are only finitely possible iteration orders, since the algorithm can converge in finite adjustments. As a direct result, exhaustic search can always find the worst ordering.
- Given any prior mistakes made by the algorithm, we know that the maximum number of mistakes should be prior mistakes plus maximum future mistakes. Hence, optimal substructure exist and dynamic programming can be applied.
- Choosing the data point with maximum norm does not maximimze the number of mistakes. For instance, take example from part (c) and adjust $x^{(2)}$ to be $[100000, 0]$ to get a counter example. norm大不代表 " 难分割 "

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

Discussion

Show Discussion

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