



MITx 6.86x

Machine Learning with Python-From Linear Models to Deep Learning[Course](#)[Progress](#)[Dates](#)[Discussion](#)[Resources](#)[Home](#) [Course](#) / [Unit 1. Linear Classifiers and Generalizations \(2 weeks\)](#) / [Homework 1](#)[← Previous](#)

1. Perceptron Mistakes

[Bookmark this page](#)

Homework due Feb 22, 2023 08:59 -03 Past due

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

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

$$\mathbf{x}^{(1)} = [-1, -1],$$

$$\mathbf{x}^{(2)} = [1, 0],$$

$$\mathbf{x}^{(3)} = [-1, 1.5],$$

Given that the algorithm starts with $\theta^{(0)} = \mathbf{0}$, the first point that the algorithm sees is 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)

1/4 points (graded)

How many mistakes does the algorithm make until convergence if the algorithm starts with $\theta^{(0)} = \mathbf{0}$?
How many mistakes does the algorithm make if it starts with data point $\mathbf{x}^{(2)}$?

Also provide the progression of the separating plane as the algorithm cycles in the following format: $[[\theta_1^{(1)}, \theta_2^{(1)}], \dots, [\theta_1^{(N)}, \theta_2^{(N)}]]$, 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 $[[0, 0], [1, 2], [3, -2]]$.

Please enter the **number of mistakes** of Perceptron algorithm if the algorithm starts with $\theta^{(0)} = \mathbf{0}$.



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

()



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

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

(Choose all that apply.)

☒ Iteration order

☐ Maximum margin between positive and negative data points

☒ Maximum norm of data points



Submit

You have used 3 of 3 attempts

1. (c)

4 points possible (graded)

Now assume that \mathbf{x}_1 . How many mistakes does the algorithm make until it starts with data point \mathbf{x}_1 ?

Also provide the progression of the separating plane as the algorithm cycles in the following form: $\mathbf{w}^0, \mathbf{w}^1, \dots, \mathbf{w}^k$, where the superscript denotes different iterations as the separating plane progresses.

For example, if \mathbf{w} progress from \mathbf{w}^0 (initialization) to \mathbf{w}^1 to \mathbf{w}^2 , you should enter $\mathbf{w}^0, \mathbf{w}^1, \mathbf{w}^2$.

Please enter the **number of mistakes** of Perceptron algorithm if the algorithm starts with \mathbf{x}_1 .

Please enter the **progression of the separating hyperplane** ($\mathbf{w}^0, \mathbf{w}^1, \dots, \mathbf{w}^k$), in a list format describing the progression of the separating hyperplane as the algorithm cycles if the algorithm starts with \mathbf{x}_1 .

Please enter the **number of mistakes** of Perceptron algorithm if the algorithm starts with \mathbf{x}_1 .

Please enter the **progression of the separating hyperplane** ($\mathbf{w}^0, \mathbf{w}^1, \dots, \mathbf{w}^k$), in the list format describing the progression of the separating hyperplane as the algorithm cycles if the algorithm starts with \mathbf{x}_1 .

number of mistakes

(Choose all that apply.)

☐ Iteration order

☐ Maximum margin between positive and negative data points

☐ Maximum norm of data points

Submit

You have used 0 of 3 attempts

1. (e) (Optional)

0 points possible (ungraded)

In 1962, Novikoff has proven the following theorem.

Assume:

- There exists γ such that $\gamma \cdot \sum_{i=1}^n \mathbf{x}_i \cdot \mathbf{w} \geq 1$ for all \mathbf{w} and some n
- All the examples are bounded

Then the number of updates made by the perceptron algorithm is bounded by $\frac{1}{\gamma^2}$.

(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 bound on the number of mistakes made by the algorithm?

(Choose all that apply.)

☐ Iteration order

☐ Maximum margin between positive and negative data points

☐ Maximum norm of data points

☐ A function of the data points

☐ Dynamic Programming the worst ordering☐ Greedily select the data point with the maximum norm**Submit**

You have used 0 of 3 attempts

< Previous

Next >

Discussion

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

edX

[About](#)[Affiliates](#)[edX for Business](#)[Open edX](#)[Careers](#)[News](#)

Legal

[Terms of Service & Honor Code](#)[Privacy Policy](#)[Accessibility Policy](#)[Trademark Policy](#)[Sitemap](#)[Cookie Policy](#)[Do Not Sell My Personal Information](#)

Connect

[Blog](#)[Contact Us](#)[Help Center](#)