${\sf DeepLearning}$

Abdelrahman Khaled

I he

rerceptroi

Perceptron Convergence Perceptron In Regression

Learning

Concepts

Programming

Multi-Layer

Non-Linearly Separable Problems

Solving the X

eferences

Deep Learning Lecture 2: The Perceptron (cont)

Abdelrahman Khaled

Machine Learning Research Cluster German University in Cairo

April 8, 2019

Outline

DeepLearning

Abdelrahmar Khaled

The Perceptro

Perceptron Convergence Perceptron In Regression Problems

Machine Learning Concepts Batch Training

Programming Task

Multi-Layer Perceptron Non-Linearly

Non-Linearly Separable Problems Solving the XOI problem

References

- 1 The Perceptron
 - Perceptron Convergence
 - Perceptron In Regression Problems
- 2 Machine Learning Concepts
 - Batch Training
- 3 Programming Task
- 4 Multi-Layer Perceptron
 - Non-Linearly Separable Problems
 - Solving the XOR problem
- 5 References

Recall: Perceptron Update Rule

DeepLearning

Abdelrahmai Khaled

Perceptron
Convergence
Perceptron In
Regression
Problems

Learning Concepts

Programming

Multi-Layer

Non-Linearly Separable Problems Solving the XO

References

Previously we used this equation to calculate the error and update the weights:

$$\Delta w_i = (t_i - y_i)x_i$$

We also said that for that specific purpose, proportionality matters a lot more than magnitude, so in that sense the below equation is equivalent (sufficient is a better term here).

$$\Delta w_i = t_i x_i$$

Perceptron Convergence Theorem

DeepLearning

Abdelrahmai Khaled

The Perceptron Convergen

Perceptron Convergence Perceptron In Regression Problems

Learning Concepts
Batch Trainin

Programming Task

Multi-Laye Perceptron

Non-Linearly Separable Problems Solving the XOF problem

References

Theorem

Given an untrained perceptron with an initial set of weights w and dataset X with n datapoints.

Assuming:

- There exist an optimal set of weights w^* such that $\|w^*\| = 1$
- There exist constants $\gamma > 0$ and R > 0.
- That for all i = 1 ... n, $||x_i|| \le R(x_i \in X)$

Such that
$$t_i(x_i \cdot w^*) \geq \gamma$$

Then the perceptron algorithm makes at most $\frac{R^2}{\gamma^2}$ errors before being fully trained.

DeepLearning

Abdelrahma Khaled

The Percentro

Perceptron

Convergence Perceptron In Regression Problems

Machine Learning Concepts

Concepts

Batch Training

Programming Task

Multi-Laye Perceptron

Separable Problems Solving the XO

Reference

Proof.

Proof Idea: If we can bound $||w^k||$ we can use those bounds to bind k

Let w^k denote the weight vector after the algorithm makes its kth error.

Let $w^0 = 0$

DeepLearning

Abdelrahmai Khaled

The Perceptron

Perceptron Convergence Perceptron In Regression Problems

Machine Learning Concepts

Programming Task

Multi-Laye Perceptron

Separable Problems Solving the XC

References

Proof.

Proof Idea: If we can bound $||w^k||$ we can use those bounds to bind k

Let w^k denote the weight vector after the algorithm makes its kth error.

Let $w^0 = 0$

■ Take the dot product between w^k and w^*

DeepLearning

Abdelrahma Khaled

The
Perceptron
Perceptron
Convergence
Perceptron In
Regression
Problems

Machine Learning Concepts Batch Trainin

Programming Task

Multi-Laye Perceptror

Non-Linearly Separable Problems Solving the XO

Doforoncoc

Proof.

Proof Idea: If we can bound $||w^k||$ we can use those bounds to bind k

Let w^k denote the weight vector after the algorithm makes its kth error.

Let $w^0 = 0$

- Take the dot product between w^k and w^*
- Since $||w^*|| = 1$, therfore

$$w^k \cdot w^* \le \|w^k\| \tag{1}$$

DeepLearning

Perceptron Convergence

Proof.

Assuming the kth error occurs on example i, we have:

$$w^{k} \cdot w^{*} = (w^{k-1} + t_{i}x_{i}) \cdot w^{*}$$
$$= w^{k-1} \cdot w^{*} + t_{i}x_{i} \cdot w^{*}$$
$$\geq w^{k-1} \cdot w^{*} + \gamma$$

DeepLearning

Abdelrahmai Khaled

The

^oerceptro

Perceptron Convergence Perceptron In Regression Problems

Learning Concepts

Programming

Multi-Laye

Non-Linearly Separable Problems Solving the XO

References

Proof.

Assuming the kth error occurs on example i, we have:

$$w^{k} \cdot w^{*} = (w^{k-1} + t_{i}x_{i}) \cdot w^{*}$$
$$= w^{k-1} \cdot w^{*} + t_{i}x_{i} \cdot w^{*}$$
$$\geq w^{k-1} \cdot w^{*} + \gamma$$

It follow by induction that:

$$w^k \cdot w^* \ge k\gamma \tag{2}$$

DeepLearning

Perceptron Convergence

Proof.

Using equations 1 and 2 we can conclude the lower bound for $||w^k||$ is:

$$||w^k|| \ge k\gamma \tag{3}$$

This is the lower bound.

DeepLearning

Perceptron Convergence

Proof.

• Consider the square of $||w^k||$

$$||w^{k}||^{2} = ||w^{k-1} + t_{i}x_{i}||^{2}$$

$$= ||w^{k-1}||^{2} + ||t_{i}x_{i}||^{2} + 2t_{i}x_{i} \cdot w^{k-1}$$

$$\leq ||w^{k-1}||^{2} + R^{2}$$

DeepLearning

Abdelrahmai Khaled

The

erceptro'

Perceptron Convergence Perceptron In Regression Problems

Learning Concepts

Concepts Batch Trainin

Programming Task

Multi-Layer Perceptron

Non-Linearly Separable Problems Solving the XO

References

Proof.

■ Consider the square of $||w^k||$

$$||w^{k}||^{2} = ||w^{k-1} + t_{i}x_{i}||^{2}$$

$$= ||w^{k-1}||^{2} + ||t_{i}x_{i}||^{2} + 2t_{i}x_{i} \cdot w^{k-1}$$

$$\leq ||w^{k-1}||^{2} + R^{2}$$

It follows by induction that:

$$\|w^k\|^2 \le kR^2 \tag{4}$$

This is the upper bound.

 ${\sf DeepLearning}$

Abdelrahmai Khaled

The

Perceptro

Perceptron Convergence Perceptron In Regression Problems

Machine Learning Concepts

Programming Task

Multi-Laye Perceptron

Separable Problems Solving the XO

References

Proof.

Combining both inequalities 3 and 4 we can derive the following inequality:

$$k^2 R^2 \le \|w^k\|^2 \le kR^2$$

DeepLearning

Abdelrahmai Khaled

The

Perceptron Convergence Perceptron In Regression Problems

Machine Learning Concepts

Batch Training

Programming Task

Multi-Laye Perceptron

Separable Problems Solving the XO

References

Proof.

Combining both inequalities 3 and 4 we can derive the following inequality:

$$k^2 R^2 \le \|w^k\|^2 \le kR^2$$

Finally, we can simplify it to get:

$$k \le \frac{R^2}{\gamma^2}$$

Regression Problems VS. Classification Problems

DeepLearning

Abdelrahman Khaled

The
Perceptron
Perceptron
Convergence
Perceptron In
Regression
Problems

Learning
Concepts
Batch Trainin

Programming Task

Perceptron
Non-Linearly
Separable
Problems

References

In a classification problem datapoints in the given data set should fall into different **discrete** groups.

We used a perceptron to find a line separates the data.

In a regression problem the data points in the data set have no obvious group, but have a **continuous** real number value. We will use a perceptron to find a line that **best-fits** the data.

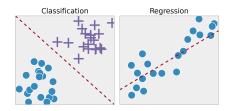


Figure: Image Source

Perceptron Regression

DeepLearning

Abdelrahma Khaled

I he
Perceptron
Perceptron
Convergence
Perceptron In
Regression
Problems

Learning Concepts

Batch Trainin

Programming Task

Multi-Layer Perceptron

Non-Linearly Separable Problems Solving the XO

Reference

- We want to find the line that best fits the data.
- We can try to do that by minimizing the distance between each data point and the current line until convergence.
- A popular and effective way of doing this is Least Square Optimization.
 - 1 Calculate the squared distance between all data points and the line (perpendicular distance).
 - 2 Sum all the distnaces
 - 3 Minimize!

Batch Training

DeepLearning

Abdelrahmar Khaled

The Perceptro

Perceptron Convergence Perceptron In Regression Problems

Learning Concepts Batch Training

Programming Task

Multi-Layer Perceptron

Separable Problems Solving the XOI

References

- The data set is split into many smaller (often disjoint) subsets.
- Datapoints in one subset are used to calculate a cumulative error which is then averaged.
- That average error is used to update the weights normally.

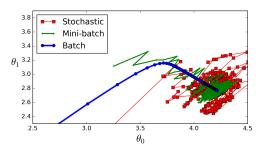


Figure: Image Source

Programming Task

DeepLearning

Abdelrahma Khaled

The
Perceptron
Perceptron
Convergence
Perceptron In
Regression
Problems

Machine Learning Concepts Batch Trainin

Programming Task

Perceptron

Non-Linearly
Separable
Problems

Solving the XOF

References

Use what you already know and implement the following:

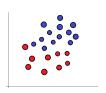
- A Perceptron Regressor
 - I Find (or create) a data set with an approximately linear behavior.
 - 2 Implement a perceptron that can solve a regression problem
 - 3 Train your algorithm multiple times on that dataset and record your results.
- Batch Training VS. Online Training
 - 1 Tweak the Perceptron that we already made so that it can be trained using batch training.
 - **2** Change the batch size, log the results and compare them.

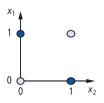
Non-Linearly Separable Problem

DeepLearning

Non-Linearly Separable Problems

A non-linearly separable problem is one where an assortment of lines or a curve would better classify the data than just a single line. A simple example of which is the XOR problem shown on the right.





The Multi-Layer Perceptron

 ${\sf DeepLearning}$

Abdelrahma Khaled

The

Perceptron Convergence Perceptron In Regression

Learning Concepts

Programming Task

Multi-Laye Perceptron

Separable Problems Solving the XOR problem

References

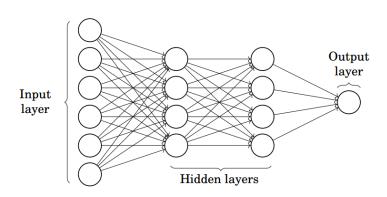


Figure: Image Source

We couldn't solve the XOR problem before, maybe we can now?

Solving the XOR Problem

${\sf DeepLearning}$

Abdelrahman Khaled

The

Perceptro

Perceptron
Convergence
Perceptron I

Regression Problems

Machin

Concepts

Batch Trainin

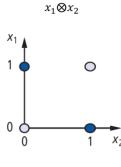
Programming Task

Multi-Layer Perceptron

Non-Linearly Separable Problems

Solving the XOR problem

ferences



$$(x_1 \land \neg x_2) \lor (\neg x_1 \land x_2)$$
$$x_1 x_2' + x_1' x_2$$

Solving the XOR Problem

DeepLearning

Abdelrahman Khaled

The

Perceptro

Perceptron In

Regression Problems

Learning

Ratch Trainin

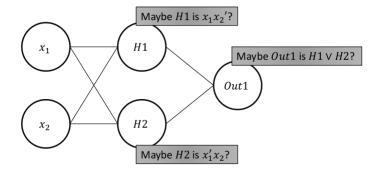
Programming Task

Multi-Laye Perceptron

> Non-Linearly Separable Problems

Solving the XOR problem

forences



Next Time

${\sf DeepLearning}$

Abdelrahman Khaled

The

Perceptro

Perceptron Convergence Perceptron In Regression Problems

Machine Learning

Concepts

Batch Training

Programming Task

Multi-Laye

Separable Problems Solving the XOR problem

D-f----

1 Multi-Layer Perceptron Continued.

Overfitting.

3 Testing and Validating our model.

References

DeepLearning

Abdelrahma Khaled

The
Perceptron
Perceptron
Convergence
Perceptron In
Regression
Problems

Machine Learning Concepts Batch Training

Programming Task

Perceptron
Non-Linearly
Separable
Problems

Separable Problems Solving the XO

References

- Goodfellow, Bengio, Courville. Deep Learning.
 Massachusetts: MIT Press, 2017. Chapter 6
- Chapman, Hall. Machine Learning: An Algorithmic Perspective. CRC Press, 2014. Chapter 4
- Non-academic resources (YouTube)
 - 3 blue 1 brown. **Neural Networks**. link
 - The Coding Train. 10: Neural Networks The Nature of Code. link