Universidade de São Paulo
Brasil

11/11/2020

28^o USP International Symposium of Undergraduate Research

Study of Singularities in Deep Neural Networks

Student: Alan Gonelli Miranda (Bolsista CNPq INCTMat/IMPA)

ICMC Coordinator : Raimundo N. Araújo dos Santos (ICMC-USP-BR)

ICMC Partner: Luis Gustavo Nonato (ICMC-USP-BR)

i-PRoBe Lab / MSU Coordinator : Arun Ross (MSU-EUA)

1. Introduction

11/11/2020

❑ 1.1 Project Motivation

- Use tools of Pure and Applied Mathematics to create mathematical models that allow determining and classifying singularities in neural networks.
- These singularities exist due to problems in training data, training algorithm, network architecture, algorithm parameters, etc.
- There are applications in various areas of knowledge such as robotics, computer graphics, image processing, bioinformatics and dynamic systems.

Singularities can result in 'adversarial inputs" that destabilize network output and can also cause instabilities in the training process.

1. Introduction

11/11/2020

❑ 1.2 Group composition

1. **Brazil**

Raimundo Araújo dos Santos (SMA-ICMC/USP);

- *Gustavo Nonato* (SME-ICMC/USP);

- *Alan Gonelli Miranda* (USP-São Carlos/CNPq-INCTMat).

2. **United States**

- *Arun Ross* (Director i-PRoBe Lab/MSU);
- Members of the Research Lab.

This project contributes to the SPRINT/FAPESP Project, process 2019/08939-0, the result of a collaboration between ICMC/USP and Michigan State University (MSU).

2. Objectives

11/11/2020

❑ *2.1 Main Proposal*

- Analyze artificial neural networks and pattern classification problems, such as object detection, through linear algebra mathematical tools, probabilities and applied mathematics techniques.
- Propose mathematical models to study their behavior through "singular" phenomena: adverse examples and training problems.
- Forecast expected results through known input arguments.

3. Application Examples

11/11/2020

3.1 Project Contextualization

- A deep neural network is reminiscent of the complex neural structure of the human brain.
- Consists of input nodes (neurons) connected from a graph with different layers of nodes preceding an output layer.
- Our investigation considers the problem of adverse cases **[1]**: characterized by small imperceptible disturbances in input data that may result in unwanted output.
- In this sense, these examples can be used to reveal singularities in a trained network.

3. Application Examples

11/11/2020

- A neural network consists of: one input layer, multiple intermediate layers, and one output layer, according to **Figure 1** below:

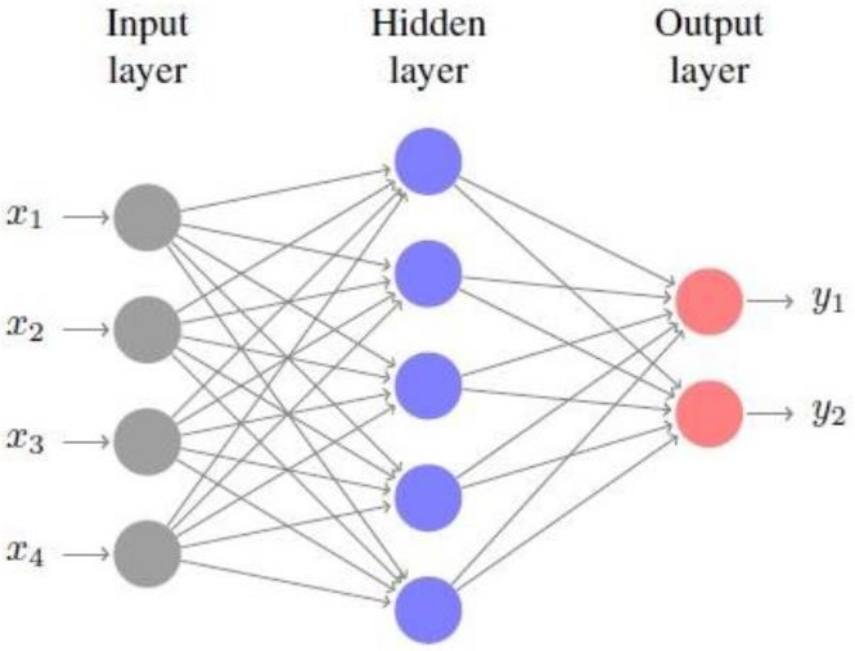

Figure 1: Illustration of a neural network with 4 inputs, 5 intermediate layers and 2 outputs

Fonte: S. Soatto; R. Giryes; J. Bruna; R. Vidal. Mathematics of Deep Learning **[2]**

3. Application Examples

11/11/2020

❑ 3.2 Deep Neural Network: Processes

- The first layer receives the input data, the middle tier applies some nonlinear functions to learn the characteristics of the data, and the output layer collects the responses from the intermediate layers to calculate a prediction rate.

❑ 3.3 Adversarial Example

- A neural network can be considered as applying the input domain to a label class:

$$f\colon \mathbb{R}^d \rightarrow \{1,2,\ldots\ldots,c\}, c\colon number\ of\ classes$$

- Consider an element x (benign example) such that $f(x) = y$. An adversarial example x' can be determined from a vector η such that for'$x' = x + \eta$, an incorrect classification of the adversarial example occurs: $f(x') \neq y$.

3. Application Examples

- An adversarial case is shown as shown in **Figure 2** below:

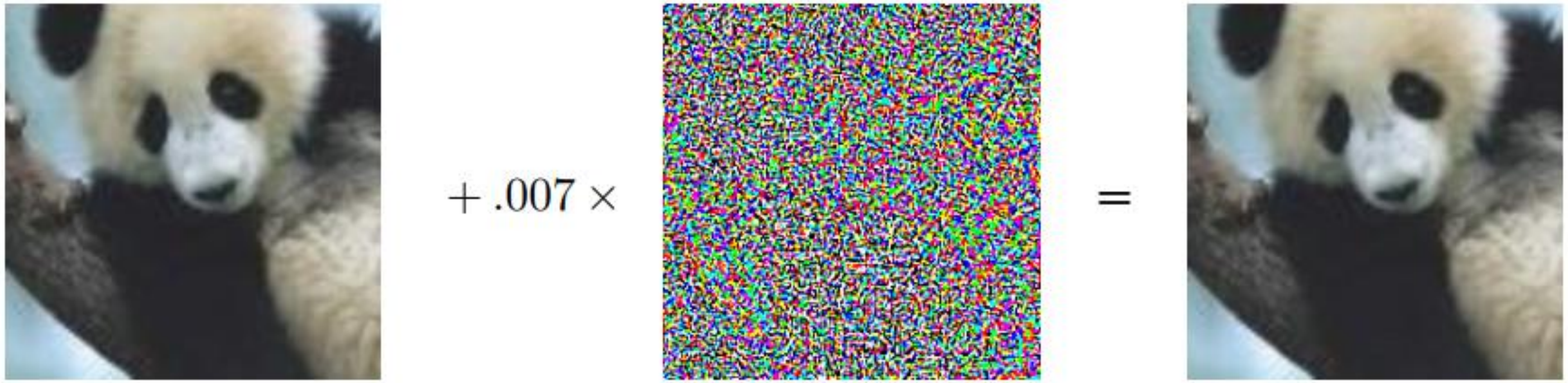

Figure 2: Illustration of adversarial sample generation using a disturbance vector weighted by a scalar value of 0.007. The "benign sample" (left) was originally classified as panda with 57.7% confidence, but the adversarial sample (right) was classified as gibbon with 99.3% confidence.

Fonte: I. J. Goodfellow, J. Shlens and C. Szegedy, "Explaining and Harnessing Adversarial Examples”, in International Conference on Learning Representations, 2015 **[3]**

4. Methods and Procedures

11/11/2020

❑ 4.1 Study Methodology

1. Reading articles and books specialized in the area of machine learning for deep neural networks and Theory of Singularities;
2. Discussion of scientific articles proposed by the coordinators of the ICMC and MSU;
3. Holding (virtual) seminars supervised by the advisors and members of the research group;
4. Develop mathematical and computational models to detect and study failures in convolutional neural networks (CNNs) and deep neural networks (DNNS).

5. Top Topics

11/11/2020

- **5.1 Computational Area**
- Adversarial Examples**;**
- Convolutional Neural Network;
- Decision Frontier.

- **5.2 Mathematical Area**
- Stability of a function or application **[4]**;
- Topology and Sets;
- Singularities: study of singular points associated with applications and sets **[4]**.

6. Results

11/11/2020

❑ **6.1 Expected Results**

- At the end of this scientific initiation project it is expected to determine :

1. Fundamental mathematical concepts for the identification of some types of singulars in deep neural networks;
2. Understanding the most common phenomena responsible for deep network failures;
3. Mathematical modeling of a convolutional neural network that allows identifying possible singularities;
4. Establishment of optimization techniques through singularity theory tools.

7. References

11/11/2020

- **[1]** S.Soatto, R. Giryes, J. Bruna, R. Vidal. Mathematics of Deep Learning, https://arxiv.org/abs/1712.04741, December 13, 2017.

- **[2]** I. J. Goodfellow, J. Shlens and C. Szegedy, "Explaining and Harnessing Adversarial Examples,“ in International Conference on Learning Representations, 2015

- **[3]** I. Goodfellow, Y. Bengio and A. Courville. Deep Learning, MIT Press, Cambridge Massachusetts, 2016.

- **[4]** Yung-Chen Lu. Singularity Theory and an Introduction to Catastrophe Theory, Springer-Verlag New York, 1976.