KU LEUVEN

FACULTEIT INGENIEURSWETENSCHAPPEN

# Distributed Adversarial Attacks

Sander Prenen

Thesis voorgedragen tot het behalen
van de graad van Master of Science
in de ingenieurswetenschappen:
computerwetenschappen, hoofdoptie
Artificiële intelligentie

**Promotoren:**
Prof. dr. ir. W. Joosen
Dr. ir. D. Preuveneers

**Assessoren:**
Ir. W. Eetveel
W. Eetrest

**Begeleiders:**
V. Rimmer
I. Tsingenopoulos

Academiejaar 2021 – 2022

# Preface

*Sander Prenen*

# List of Abbreviations

**AI** Artificial Intelligence. 3

**ANN** Artificial Neural Network. 3

**ReLU** Rectified Linear Unit. 3, 4

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

# Chapter 2

# Background

## 2.1 Neural networks

Ever since the invention of computer systems, it has always been a goal of scientists and engineers to create Artificial Intelligence (AI). Current state of the art approaches are mimicking the human brain, more specifically the neurons inside the brain. Already in the fifties, Rosenblatt introduced his perceptron [1]. The perceptron is a single neuron able to learn linearly separable patterns. It does so by finding a hyperplane that separates the two classes. This hyperplane is called the decision surface or decision boundary. The concept of linear separability is explained in Figure 2.1 in two dimensions.

Unfortunately not all patterns are linearly separable. To overcome this problem, the neurons can be layered, creating an Artificial Neural Network (ANN) in the process. Layering neurons sequentially is essentially a linear combination of neurons. This in itself does not create non-linear decision surfaces. Non-linear activation functions are added for the ANN to be able to learn more complex decision boundaries. Some commonly used activation functions are Rectified Linear Unit (ReLU) [2], Heaviside step function and softmax (or sigmoid when used on scalars). In Figure 2.2 the plots of the activation functions can be found.

## 2.2 Adversarial attacks

## 2.3 Particle swarm optimization

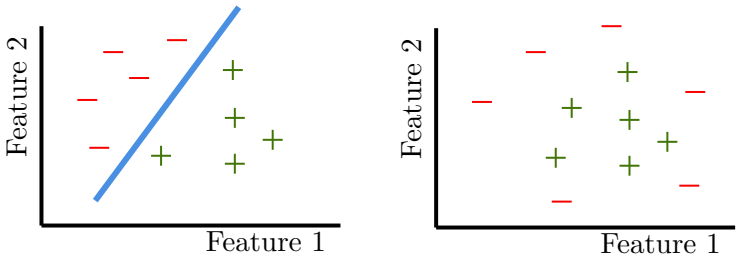


Figure 2.1: Linearly separable classes on the left and non-linearly separable classes on the right. Two classes are linearly separable if there exists a hyperplane for which all examples of one class are on the same side of this hyperplane, whilst all examples of the other class are on the other side of the hyperplane. In two dimensions, the hyperplane is a straight line.

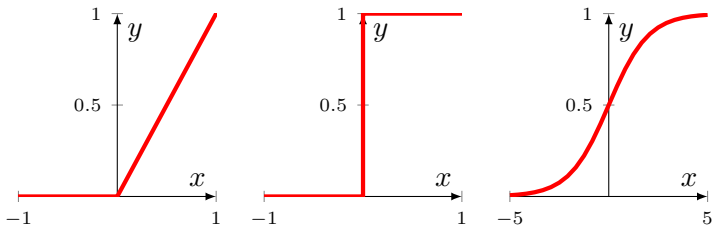


Figure 2.2: Plots of different activation functions. From left to right: ReLU, Heaviside step and sigmoid.

# Bibliography

[1] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. 65(6):386–408.

[2] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, page 807–814, Madison, WI, USA, 2010. Omnipress.