Finite Gaussian Neurons

A Defense Against Adversarial Attacks?

Felix Grezes

Graduate Center
City University of New York

Thesis Proposal Fall 2020

THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK

Table of Contents

THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK

Abstract

I introduce the Finite Gaussian Neuron, a novel neural network architecture.
My works aims to:

- make it easy to convert existing models to the FGN architecture
- while preserving the existing model's behavior on real data
- and offering resistance against some adversarial attacks.

THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK

Introduction

THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK

Related Work

THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK

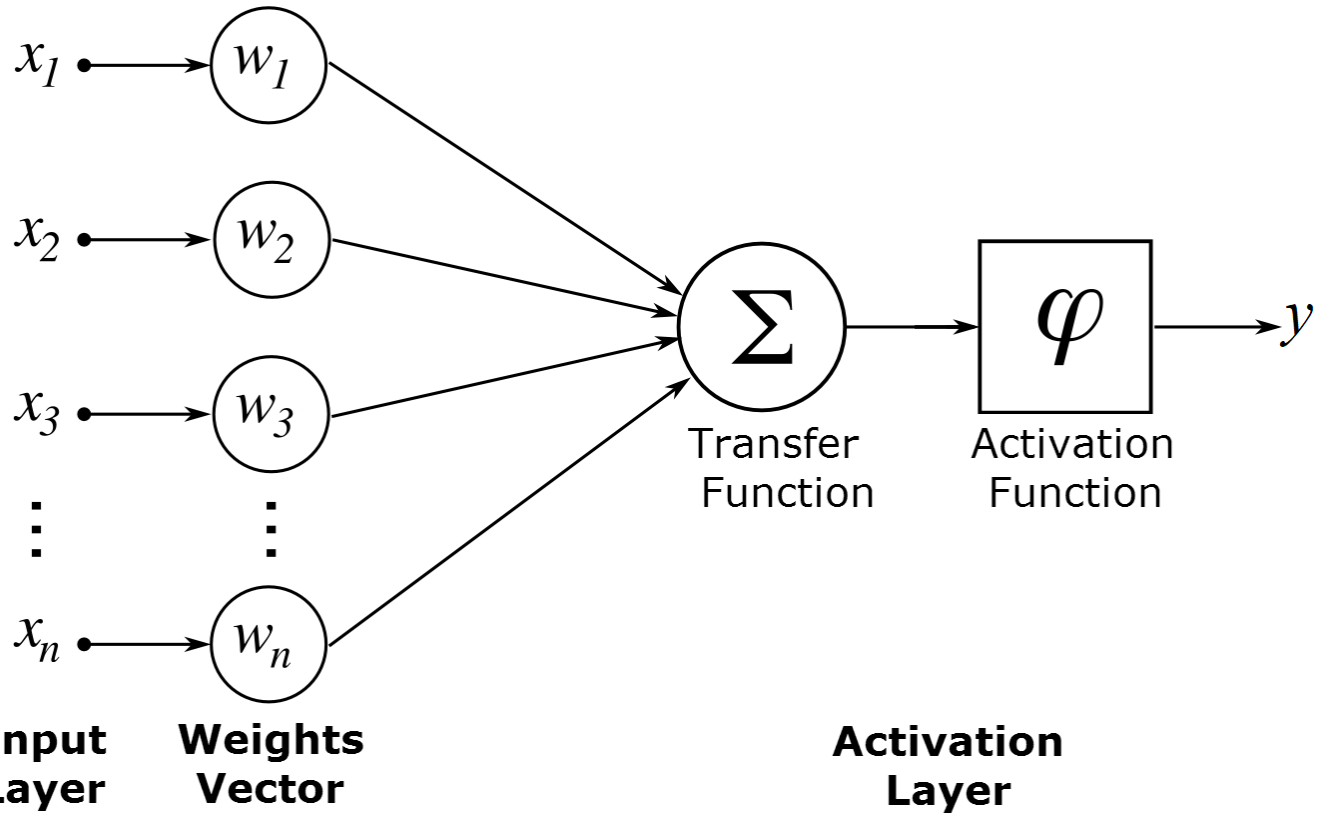
The Classic Neuron

Neuron output:

$$y = \varphi(l)$$

Linear component:

$$l = \sum_i x_i w_i$$



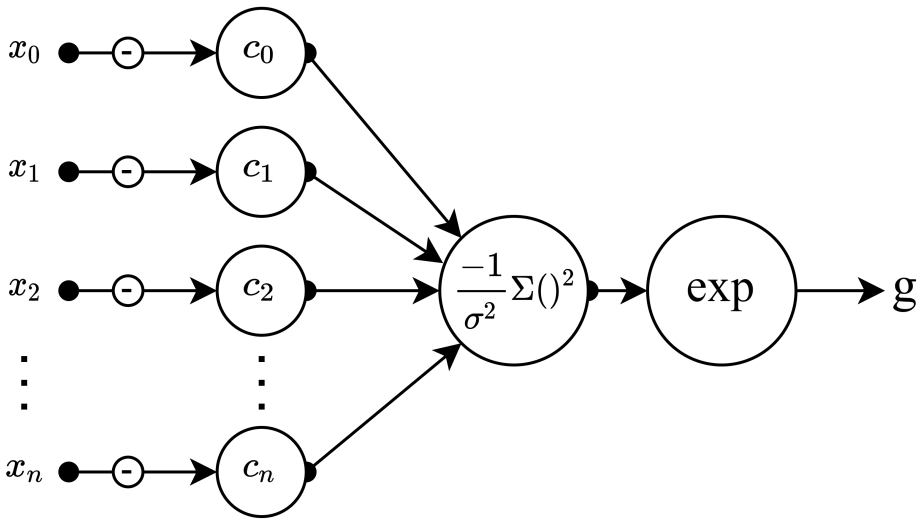
The Finite Gaussian Neuron

Neuron output:

$$y = \varphi(\sum_i x_i w_i) * g$$

Gaussian component:

$$g = exp\left(\frac{-1}{\sigma^2} * \sum_i (x_i - c_i)^2\right)$$



THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK

2D Illustration

THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK