An Introduction To Neural Networks

Adithya Nair

The Journal Society, Amrita Vishwa Vidyapeetham

November 14, 2024

An Introduction

- The Journal Society is a group of people coming together to discuss mathematic and scientific ideas in a rigorous fashion.
- We also have sessions happening in parallel on Real Analysis, following Terence Tao's textbooks on Analysis.

What *are* neural networks?

A neural network is a function. It takes an input, and maps it to some output.

$$f : \mathbb{R}^m \to \mathbb{R}^n$$

An Example

Let's say we had M greyscale images of the digits $0 \cdots 9$ on a square. How do you make a computer recognize each digit.

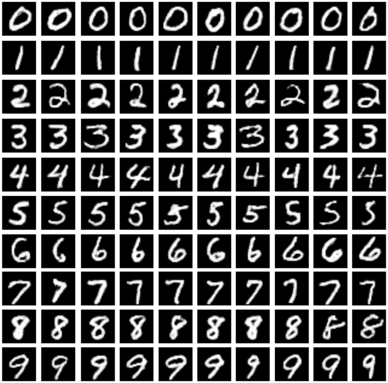

Figure: Each digit drawn out clearly

An Example

1. Each image is a set of p pixels, or a vector $\vec{v} = (v_1, v_2, \cdots, v_p)$. Each element v_i tells us how 'dark' or 'light' that pixel is. 0 is black, and 1 is white.

An Example

1. Each image is a set of p pixels, or a vector $\vec{v} = (v_1, v_2, \cdots, v_p)$. Each element v_i tells us how 'dark' or 'light' that pixel is. 0 is black, and 1 is white.
2. We have M vectors $\vec{v}$ in p -dimensional space or $\mathbb{R}^p$

An Example

1. Each image is a set of p pixels, or a vector $\vec{v} = (v_1, v_2, \cdots, v_p)$. Each element v_i tells us how 'dark' or 'light' that pixel is. 0 is black, and 1 is white.
2. We have M vectors $\vec{v}$ in p -dimensional space or $\mathbb{R}^p$
3. We do have *a* function f that can map these specific vectors v to an output in the discrete set $\{0, \cdots, 9\}$.

The Problem

The function that maps these vectors cannot do this for **new** vectors.

The Problem

The function that maps these vectors cannot do this for **new** vectors.
We need a good “rule”.

Universality

Neural networks are capable of **universal approximation**.

The Big Picture

Neural networks are compositions of simple functions.

$$F(v) = F_L(F_{L-1} \cdots F_2 F_1(\vec{v})))$$

The reason they work so well, is that we are finding or computing the right parameters such that the output of this function matches the expected output as much as possible.

The Big Picture

Neural networks are compositions of simple functions.

$$F(v) = F_L(F_{L-1} \cdots F_2 F_1(\vec{v})))$$

The reason they work so well, is that we are finding or computing the right parameters such that the output of this function matches the expected output as much as possible. These parameters perform operations such as shifting, inverting and scaling functions that allows them to match real-world outputs in effective ways.

The Big Picture

Neural networks are compositions of simple functions.

$$F(v) = F_L(F_{L-1} \cdots F_2 F_1(\vec{v})))$$

The reason they work so well, is that we are finding or computing the right parameters such that the output of this function matches the expected output as much as possible.
These parameters perform operations such as shifting, inverting and scaling functions that allows them to match real-world outputs in effective ways.
What are these 'simple' functions?

Linear Functions

$$F(v) : \mathbb{R}^p \rightarrow \mathbb{R}^{10}$$

Definition - Linear Functions Or Linear Maps

A linear function, in linear algebra terms, means a mapping between two vector spaces $V \rightarrow W$ which preserves the operations of vector addition and scalar multiplication.
The function f is such that,

1. $f(u + v) = f(u) + f(v)$
2. $f(cu) = cf(u)$

They Don't Work

Linear functions don't work well.

They Don't Work

Linear functions don't work well. The input-output mapping of real world problems are rarely linear.

Continuous Piecewise Linear Functions

Definition

A function $f(x)$ is said to be piecewise continuous on an interval $[a, b]$ if it is defined and continuous except possibly at a finite number of points $a \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq b$
Furthermore, at each point of discontinuity, we require that the left and right hand limits exists.

$$f(x_k^-) = \lim_{x \to x_k^-} f(x); f(x_k^+) = \lim_{x \to x_k^+} f(x)$$

At the ends of the domain, the left hand is ignored at a and the right hand limit is ignored at b.

An Example Of Such Functions

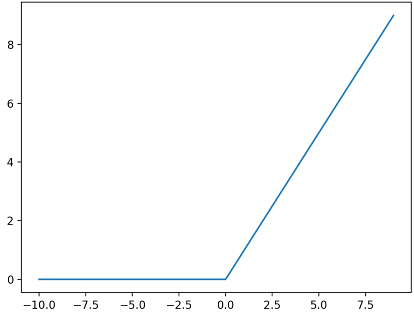

Figure: The RELU (REctifier Linear Unit) function

$$f(x) = \max(x, 0)$$

Why Do They Work?

"Linear for simplicity, continuous to model an unknown but reasonable rule and piecewise to achieve the non-linearity that is an absolute requirement for real images and data." — Gilbert Strang, Linear Algebra And Learning From Data (2019)

Affine Functions

Definition

An affine function is a function with a linear transformation as well as a translation.

$$F(\vec{v}) = A\vec{v} + b$$

A Small Introduction To Partial Derivatives

Definition

The derivative of a function dependent on multiple variables with respect to one of the variables, while the others are kept constant. For a function $f(x, y, \cdots)$,

$$\frac{\partial f}{\partial x} = f'_x$$

Gradient Operator

The gradient operator is a linear operator that helps calculate the total derivative.

$$\text{where, } \nabla = \frac{\partial()}{\partial x_1}\hat{e_1} + \frac{\partial()}{\partial x_2}\hat{e_2} \ldots \frac{\partial()}{\partial x_n}\hat{e_n}$$

$$\nabla f = \frac{\partial f}{\partial x_1}\hat{e_1} + \frac{\partial f}{\partial x_2}\hat{e_2} + \cdots + \frac{\partial f}{\partial x_n}\hat{e_n}$$