Chapter 1

Multilayer Feedforward Networks

1.1 Function Approximation

In this paper we take $C(\mathbb{R}^n)$ to be the family of "real world" functions that one may wish to approximate. If we can show that a given set of functions F is dense in $C(\mathbb{R}^n)$, we can conclude that for every continuous function $g \in C(\mathbb{R}^n)$ and each compact set $K \subset \mathbb{R}^n$, there is a function $f \in F$ such that f is a good approximation to g on K.

Definition 1. A *metric* (or *distance*) on a set X is a function $d : X \times X \to \mathbb{R}$ such that forall $s, t \in X$ the following properties are satisfied:

1. $d(s,t) \geq 0$ and $d(s,t) = 0$ if and only if $s = t$.
2. $d(s,t) = d(t,s)$.
3. $d(s,t) \leq d(s,u) + d(u,t) \quad$ (*triangular inequality*).

A *metric space* is a pair (X, d), where X is a set and d is a distance in X.

If we take X to be a set of functions, the metric $d(f,g)$ will enable us to measure the distance between functions $f, g \in X$.

Definition 2. Let f, g be real-valued functions with compact support. We define the *convolution* of f with g as

$$(f * g)(x) = \int f(x-t)g(t)\,dt$$

1.2 Lebesgue measure

An essential part of measure theory is the calculation of lengths, areas, volumes, etc. We are already familiar with the Riemann integral, and now we aim to introduce a more general integral, the Lebesgue integral, that comes from the Lebesgue measure.

Definition 3. A box in $\mathbb{R}^d$ is a set of the form

$$Q = [a_1, b_1] \times ... \times [a_d, b_d] = \prod_{i=1}^{d} [a_i, b_i]$$

The volume of the box is

$$vol(Q) = (b_1, a_1)...(b_d - a_d) = \prod_{i=1}^{d} (b_i - a_i)$$

The *exterior measure* (or outer measure) of a set $E \subseteq \mathbb{R}^d$ is

$$|E|^* = \inf\{\sum_k vol(Q_k)\}$$

where the infimum is taken over all finite or countable collection of boxes Q_k such that $E \subseteq \cup_k Q_k$

Definition 4. A set $E \subseteq \mathbb{R}^n$ is *Lebesgue mesurable* (or mesurable) if $\forall \epsilon > 0$, there exist U open set such that $E \subseteq U$ and $|U \setminus E|^* < \epsilon$

Definition 5. A function u defined almost everywhere on a measurable set $\Omega \in \mathbb{R}^n$ is said to be *essentially bounded* on Ω if $|u(x)|$ is bounded almost everywhere on Ω. We denote $u \in L^\infty(\Omega)$ with the norm

$$\|u\|_{L^\infty(\Omega)} = inf(\lambda|\{x : |u(x)| \geq \lambda\} = 0) = ess \sup_{x \in \Omega} |u(x)|$$

We have that $L^\infty(\mathbb{R})$ is the space of essentially bounded functions.

Examples and counterexamples of functions essentially bounded.

- $f : \Omega \to$

Definition 6. A function u defined almost everywhere on a domain Ω (a domain is an open set in $\mathbb{R}^n$) is said to be *locally essentially bounded* on Ω if for every compact set $K \subset \Omega$, $u \in L^\infty(K)$. We denote $u \in L^\infty_{loc}(K)$

Definition 7. We say that a set of functions $F \subset L^\infty_{loc}(\mathbb{R})$ is *dense* in $C(\mathbb{R}^n)$ if for every function $g \in C(\mathbb{R}^n)$ and for every compact $K \subset \mathbb{R}^n$, there exist a sequence of functions $f_j \in F$ such that

$$\lim_{j \to \infty} \|g - f_j\|_{L^\infty(K)} = 0.$$

1.3 Results

Definition 8. Let $\mathcal{M}$ denote the set of functions which are in $L^\infty_{loc}(\mathbb{R})$ and have the following property. The closure of the set of points of discontinuity of any function in $\mathcal{M}$ is of zero Lebesgue measure.

Proposition 9. (This implies that) for any $\sigma \in \mathcal{M}$, interval $[a, b]$. and $\delta > 0$, there exists a finite number of open intervals, the union of which we denote by U, of measure δ, such that σ is uniformly continuous on $[a, b]/U$.

Definition 10. $\mathcal{C}^\infty_0$ functions $\mathcal{C}^\infty$ with compact support.

1.4 Multilayer Feedforward Network

intro??? Multilayer feedforward networks (MFNs) are a type of artificial neural network that consist of several layers of interconnected nodes, with each node taking input from the previous layer and producing output for the next layer. The general architecture of a multilayer feedforward network, MFN, consist of: input layer: n-input units, one/more hidden layers : intermediate processing units, output layer: m output-units. dibuix???

Definition 11. (Multilayer feedforward networks) The function that a MFN compute is:

$$f(x) = \sum_{j=1}^{k} \beta_j \cdot \sigma(w_j \cdot x - \theta_j)$$

where $x \in \mathbb{R}^n$ is the input vector, $k \in \mathbb{N}$ is the number of processing units in the hidden layer, $w_j \in \mathbb{R}^n$ is the weight vector that connects the input to processing unit j in the hidden layer, $\sigma : \mathbb{R} \to \mathbb{R}$ is an activation function applied element-wise to the vector $w_j^T x - \theta_j$, where $\theta_j \in \mathbb{R}$ is the threshold (or bias) associated with processing unit j in the hidden layer, and $\beta_j \in \mathbb{R}$ is the weight that connects processing unit j in the hidden layer to the output of the network.

Let N_w be the family of all functions implied by the network's architecture. If we can show that N_w is dense in $C(\mathbb{R}^n)$, we can conclude that for every continuous function $g \in C(\mathbb{R}^n)$ and each compact set $K \subset \mathbb{R}^n$, there is a function $f \in N_w$ such that f is a good approximation to g on K.

Under which necessary and sufficient conditions on σ will the family of networks N_w be capable of approximating to any desired accuracy any given continuous function?

1.5 Theorem

Theorem 12. *Let $\sigma \in M$. Set*

$$\Sigma_n = span\{\sigma(w \cdot x + \theta) : w \in \mathbb{R}^n, \theta \in \mathbb{R}\}$$

Then Σ_n is dense in $\mathcal{C}(\mathbb{R}^n)$ if and only if σ is not an algebraic polynomial.

1.5.1 Why does not contradict the Weierstrass approximation theorem?

Theorem 13. *(Weierstrass approximation theorem). Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then, there exists polynomials $p_n \in \mathcal{R}[x]$ such that the sequence (p_n) converge uniformly to f on $[a,b]$.*

Corollary 14. The set of polynomial functions $\mathcal{R}^n[x]$ is dense in the space of continuous functions on a closed and bounded interval $\mathcal{C}([a,b]^n)$.

Com que σ té grau fix "k", aleshores $\sigma(wx + \theta)$ grau com a molt k. Per tant el conjunt σ_n és un espai vectorial finit i no pot ser dens.