Chapter 1

About the theorem

1.0.1 Why does it not contradict the Weierstrass approximation theorem?

Theorem 1. *(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 2. The set of polynomial functions $\mathcal{R}^n[x]$ is dense in the space of continuous functions on a compact set $K \subset \mathbb{R}^n$, $\mathcal{C}(K)$. So any continuous function on a compact set can be approximated arbitrarly well by a polynomial.

The theorem states that: if Σ_n is dense in $\mathcal{C}(\mathbb{R}^n)$ then σ is not an algebraic polynomial. But why this statment does not contradict the Weierstrass approximation theorem ? This does not work because σ has degree fixed k, then any element in the set Σ_n has degree at most k. Hence, the set Σ_n is a finite vector space and can not be dense in $\mathcal{C}(\mathbb{R}^n)$. Not all contiunous functions can be apparoximated with a polynimial of degree fixed, for example: (comment per afegir : per exemple una funcio que sigui continua que no es pugui approximar per un polinomi de com a molt grau k , una k tingui grau mes gran que k polinomi de k+1 ??)

1.0.2 Previous results

The activation functions that were reported thus far in the literature.

Theorem 3. *(Hornik Theorem 1). Standard multilayer feedforward networks with a bounded and nonconstant activation function can approximate any function in $L^p(\mu)$ arbitrary well, given a sufficiently large number of hidden units.*

Theorem 4. *(Hornik Theorem 2) Standard multilayer feedforward networks with a continuous, bounded and nonconstant activation function can approximate any continuous function on X arbitrarily well (with respect to the uniform distance) given a sufficiently large number of hidden units.*

The theorem generalizes Hornik's Theorem 2 by establishing necessary and sufficient conditions for universal approximation. Note that the theorem merely requires "nonpolynomiality" in the activation function. Unlike Hornik's result, the activation functions do not need to be continuous or smooth. This has an important biological interpretation because the activation functions of real neurons may well be discontinuous or even non-elementary.

1.0.3 Results

Definition 5. The set $L^p(\mu)$ contains all mesurable functions f such that:

$$\|f\|_{L^p}(\mu) = \Big(\int_{R^n} |f(x)|^p d\mu(x) \Big)^{1/p} < \infty$$

Proposition 6. Let μ be a non-negative finite measure on $\mathbb{R}$ with compact support, absolutely continous with respect to Lebesgue measure. Then Σ_n is dense in $L_p(\mu)$, $1 \leq p < \infty$, if and only of, σ is not a polynomial.

Proposition 7. If $\sigma \in M$ is not a polynomial (a.e) then,

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

is dense in $\mathcal{C}(\mathbb{R}^n)$ for some $\mathcal{A} \subset \mathbb{R}^n$ if and only if there does not exist a nontrivival polynomial vanishing on $\mathcal{A}$.

Remark 1. The theorem only requires for the activation function to be nonpolynomial, we dont need continuity on sigma. For example, let σ be continuous with a jump discontinuity at 0 such that:

$$\lim_{x \to 0^-} \sigma(x) = 0 \qquad \lim_{x \to 0^+} \sigma(x) = 1$$

Given $f \in \mathcal{C}(\mathbb{R})$ and $K \subset \mathbb{R}$ compact, letting $w \to 0$ in $\sigma(wx)$ the function

$$h(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x > 0 \end{cases}$$

$h \in \overline{\Sigma_1}$.

Linear combinations of h and its translates can uniformly approximate any continuous function on any finite interval (and thus an), compact subset of R).