# Approximation by Superpositions of a Sigmoidal Function

Kim Juhyeong

Statistical Artificial Intelligence Lab

*juhyeong@kaist.ac.kr*

September 22, 2020

# Conclusion

Finite ***superpositions(stack)*** of a fixed, univariate function that is ***discriminatory****(ex. Neural Network)* can uniformly approximate any continuous function of n real variables with support in the unit hypercube $[0,1]^n$ with arbitrary level of precision.

In detail, function with form of $\mathrm{G}(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j)$ are ***dense*** in $\mathrm{C}(\mathrm{I_n})$ providing $\sigma$ is ***continuous*** and ***discriminatory*** sigmoidal function.

# Preliminaries

- $\mathrm{I_n}$ : $[0,1]^n$.
- $\mathrm{C(I_n)}$ : Space of Continuous functions in $\mathrm{I_n}$.
- For $f \in \mathrm{C(I_n)}$, $\|f\|$ denote supremum norm of $f$.
- $\mathrm{M(I_n)}$ : Space of finite, signed regular Borel measures.

- Signed measure : Measure allowed to have negative values.
- Regular measure : Measure is regular if it is inner and outer regular.
  - Inner Regular :
    $\mu(\mathrm{A}) = \sup\{\mu(\mathrm{F}) | \mathrm{F} \subseteq \mathrm{A}, \mathrm{F}$ compact and measurable$\}$.
  - Outer Regular :
    $\mu(\mathrm{A}) = \inf\{\mu(\mathrm{G}) | \mathrm{G} \supseteq \mathrm{A}, \mathrm{G}$ open and measurable$\}$.

# Preliminaries

- Dense : A set $S \in \mathrm{X}$ is dense in $\mathrm{X}$ if, for any $\epsilon > 0$ and $x \in \mathrm{X}$, $\exists s \in \mathrm{S}$ such that $|x - s| < \epsilon$.
- Linear Functional : A map $\mathrm{T} : \mathrm{X} \to \mathrm{R}$ such that
  - $\mathrm{T}(x + y) = \mathrm{T}(x) + \mathrm{T}(y)$.
  - $\mathrm{T}(\alpha x) = \alpha \mathrm{T}(x)$.
- Bounded Linear Functional : Linear Functional such that $\exists \mathrm{M} > 0$ where $|\mathrm{T}(x)| \leq \mathrm{M}\|x\|$ for $\forall x \in \mathrm{X}$.

## Definition (sigmoidal)

We say that $\sigma$ is sigmoidal if

$$\sigma(t) \to \begin{cases} 1 \text{ as } t \to +\infty \\ 0 \text{ as } t \to -\infty \end{cases}$$

## Definition (discriminatory)

We say that $\sigma$ is discriminatory if for a measure $\mu \in \mathrm{M}(\mathrm{I_n})$.

$$\int_{\mathrm{I_n}} \sigma(y^T x + \theta) d\mu(x) = 0$$

$$\forall y \in \mathbb{R} \text{ and } \forall \theta \in \mathbb{R} \text{ implies that } \mu = 0$$

# Theorem 1

## Theorem

Let $\sigma$ be any continuous **discriminatory** function. Then finite sums of the form

$$\mathrm{G}(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j)$$

are dense in $\mathrm{C}(\mathrm{I_n})$.

Equivalently, given any $f \in \mathrm{C}(\mathrm{I_n})$ and $\epsilon > 0$,
$\exists \mathrm{G}(x)$ of the above form, such that $|\mathrm{G}(x) - f(x)| < \epsilon$ for all $x \in \mathrm{I_n}$

# Theorem 2

## Theorem

Let $\sigma$ be any continuous **sigmoidal** function. Then finite sums of the form

$$\mathrm{G}(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j)$$

are dense in $\mathrm{C}(\mathrm{I_n})$.

Equivalently, given any $f \in \mathrm{C}(\mathrm{I_n})$ and $\epsilon > 0$,
$\exists \mathrm{G}(x)$ of the above form, such that $|\mathrm{G}(x) - f(x)| < \epsilon$ for all $x \in \mathrm{I_n}$

# Theorem 3

**Theorem**

Let $\sigma$ be a continuous sigmoidal function.
Let $f$ be a decision function for any finite measurable partition of $\mathrm{I_n}$.
For any $\epsilon > 0$, there is a finite sum of the form

$$\mathrm{G}(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j)$$

and a set $\mathrm{D} \subset \mathrm{I_n}$, so that $m(\mathrm{D}) \geq 1 - \epsilon$ and $|\mathrm{G}(x) - f(x)| < \epsilon$ for $x \in \mathrm{D}$

# Theorem 4

## Theorem

Let $\sigma$ be any bounded **measurable sigmoidal** function. Then finite sums of the form

$$\mathrm{G}(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j)$$

are dense in $\mathrm{C}(\mathrm{I_n})$.

Equivalently, given any $f \in \mathrm{L}^1(\mathrm{I_n})$ and $\epsilon > 0$,
$\exists \mathrm{G}(x)$ of the above form, such that

$$\|\mathrm{G} - f\|_{\mathrm{L}^1} = \int_{\mathrm{I_n}} |\mathrm{G}(x) - f(x)| dx < \epsilon$$

# Theorem 5

## Theorem

Let $\sigma$ be a general **sigmoidal** function. Let $f$ be the decision function for any finite measurable partition of $\mathrm{I_n}$. For any $\epsilon > 0$. There is a finite sum of the form

$$\mathrm{G}(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j)$$

and a set $\mathrm{D} \subset \mathrm{I_n}$, so that $m(\mathrm{D}) \geq 1 - \epsilon$ and $|\mathrm{G}(x) - f(x)| < \epsilon$ for $x \in \mathrm{D}$