

卒研ゼミ 「深層学習」

6.4 Architecture Design

6.4.1 Universal Approximation Properties and Depth

6.4.2 Other Architectural Considerations

2019年12月16日？

岡崎健人

6.4 Architecture Design

アーキテクチャ：ネットワークの全体的な構造のこと

- ユニットの数、どのようにユニット同士が関係しているかなど
- ユニットは層(layers)をなし、1つ前の層の関数になっている

$$\mathbf{h}^{(i+1)} = g^{(i+1)} \left(\mathbf{W}^{(i+1)\top} \mathbf{h}^{(i)} + \mathbf{b}^{(i+1)} \right)$$

層の深さ・幅を決めることが主な検討事項

6.4 Architecture Design

アーキテクチャ：ネットワークの全体的な構造のこと

- 隠れ層ひとつでも十分まともな結果が得られる
- 層を増やすと、層のユニット数・パラメータ数を大幅に減らせるが、最適化が難しくなる
- 検証集合の誤差を実験して、理想のアーキテクチャを決めるべき

Universal Approximation Theorem

“Let $\varphi(\cdot)$ be a non-constant, bounded, and monotone-increasing continuous function. Let I_{m_0} denote the m_0 -dimensional unit hypercube $[0,1]^{m_0}$. The space of continuous functions on I_{m_0} is denoted by $C(I_{m_0})$. Then, given any function $f \in C(I_{m_0})$ and $\varepsilon > 0$, there exist an integer m_1 and sets of real constants α_i , b_i , and w_{ij} , where $i = 1, \dots, m_1$ and $j = 1, \dots, m_0$ such that we may define

$$F(x_1, \dots, x_{m_0}) = \sum_{i=1}^{m_1} \alpha_i \varphi \left(\sum_{j=1}^{m_0} w_{ij} x_j + b_i \right)$$

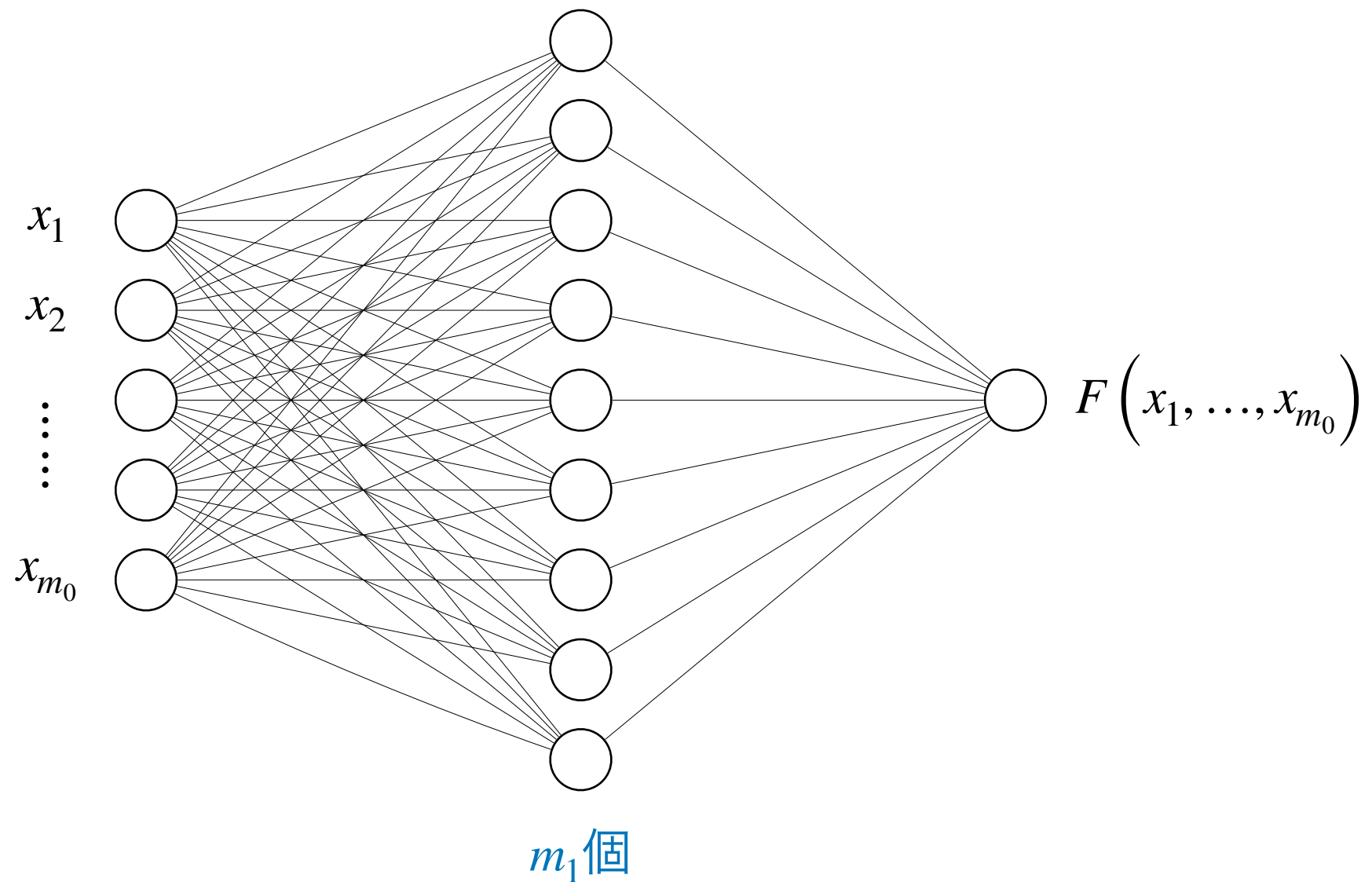
as an approximate realization of the function $f(\cdot)$; that is,

$$\left| F(x_1, \dots, x_{m_0}) - f(x_1, \dots, x_{m_0}) \right| < \varepsilon$$

for all x_1, x_2, \dots, x_{m_0} that lie in the input space.”

Simon Harkin, ‘*Neural Networks and Learning Machines*’, Third Edition, p. 167. Pearson Education Inc., 2008.

$\varphi(x) = \sigma(x), \tanh(x)$ など



$$\left| F(x_1, \dots, x_{m_0}) - f(x_1, \dots, x_{m_0}) \right| < \varepsilon \text{ となる整数 } m_1 \text{ が存在する}$$

(存在定理)

UATは改良されている。現在では…

- f はボレル可測関数であればよい（閉じていて、有界な \mathbb{R}^n の部分集合上の関数）
- φ はReLUを含む多くの関数について証明されている
- 微分についてもUATは証明されている(Horkin *et al.* 1990)

- 表現できても学習できるとはかぎらない (👉???)
 - ▶ 目的関数に対応するパラメータを見つけられない
 - ▶ 過剰適合が理由で間違った関数を選択してしまう
- 「ノーフリーランチ定理」 (普遍的に優れた機械学習アルゴリズムは存在しない) を思い出す
- **与えられた目的関数に対して、任意の精度で近似できるMLPが存在する**としか主張していない

MLPが存在するのはわかった。でもつまるところ、隠れユニット数 m_1 をいくつにすれば良いのか？

- 隠れ層ひとつの場合、**指数関数的な数の隠れユニット**が必要になる (Barron, 1993)
 - ▶ n 次元の二値ベクトルに対する二値関数を表すのに、隠れユニットは $O(2^n)$ くらい必要

万能近似定理のまとめ

- 隠れ層ひとつであっても、十分に近似できるようなMLPはある
- しかし隠れ層のユニットを非現実的に多くしなきゃいけない可能性もある
- 正しい学習・汎化に失敗する可能性もある
- 隠れ層を多くする（深くする）ことでこれらの問題を解決できる