ニューラルネット普遍近似定理

ニューラルネットの普遍近似定理の主張を述べてください．ニューラルネットの定義は隠れ1層の全結合ネットとします．

ニューラルネットの普遍近似定理は、隠れ層が1層のみの全結合ネットワークであっても、十分な数の隠れユニットを用いれば、コンパクト集合上の任意の連続関数を任意の精度で一様に近似できるという主張です。

具体的には、以下のように表現されます：

- **対象関数:** コンパクト集合 $K \subset \mathbb{R}^n$ 上の任意の連続関数 $f : K \to \mathbb{R}$ に対して、
- **近似精度:** 任意の正の数 $\varepsilon > 0$ が与えられたとき、
- **近似の存在:** 適切な重みやバイアス、そして一般にはシグモイドなどの非線形かつ連続な活性化関数を用いることで、1層の隠れ層を持つ全結合ニューラルネットワーク g を構成でき、全ての $x \in K$ に対して

  $$|f(x) - g(x)| < \varepsilon$$

  が成り立つようにできるというものです。

この定理は、理論的にニューラルネットワークが非常に強力な近似能力を持つことを示しており、実際の応用においても幅広い関数近似やパターン認識の基盤となっています。

証明を与えてください

以下に、代表的な証明のアイデア（スケッチ）を示します。以下の証明は Cybenko の手法に基づくもので、隠れ層1層の全結合ネットワークでの普遍近似性を示す標準的なアプローチです。

定理の再表現

コンパクト集合 $K \subset \mathbb{R}^n$ 上の任意の連続関数 $f \in C(K)$ と任意の正数 $\varepsilon > 0$ に対して、適当なパラメータ $N \in \mathbb{N}$、重み $w_j \in \mathbb{R}^n$、バイアス $b_j \in \mathbb{R}$、および係数 $a_j \in \mathbb{R}$ ($j = 1, \dots, N$) を選べば、

$$G(x) = \sum_{j=1}^{N} a_j \, \sigma(w_j^T x + b_j)$$

が K 上一様に f を近似し、

$$\sup_{x\in K}\big|f(x)-G(x)\big|<\varepsilon$$

となる、という主張です。

証明のスケッチ

1. **関数系の定義と稠密性の帰結**

   単一隠れ層ニューラルネットワークで表現可能な関数全体の集合を

   $$\mathcal{F}=\left\{G(x)=\sum_{j=1}^{N}a_j\,\sigma(w_j^Tx+b_j)\;\middle|\;N\in\mathbb{N},\,a_j\in\mathbb{R},\,w_j\in\mathbb{R}^n,\,b_j\in\mathbb{R}\right\}$$

   とします。普遍近似定理は「$\mathcal{F}$ の閉包が $C(K)$ 全体になる」、すなわち

   $$\overline{\mathcal{F}}=C(K)$$

   であることを主張しています。

2. **背理法による議論**

   仮に $\overline{\mathcal{F}}\neq C(K)$ であるとすると、Hahn–Banach の定理より、
   $C(K)$ の非零な連続線形汎関数 $L:C(K)\to\mathbb{R}$ が存在して、
   $\mathcal{F}$ 上ゼロ、すなわち

   $$L(g)=0\quad\text{for all }g\in\mathcal{F}$$

   となります。

3. **Riesz 表現定理の適用**

   $C(K)$ はコンパクト集合上の連続関数空間であり、Riesz 表現定理より、
   ある有限な符号付きBorel測度 $\mu\neq0$ が存在して

   $$L(h)=\int_K h(x)\,d\mu(x)\quad\text{for every }h\in C(K)$$

   と書けます。従って、$\mathcal{F}$ の任意の関数 G について

   $$\int_K\sigma(w^Tx+b)\,d\mu(x)=0\quad\text{for all }w\in\mathbb{R}^n,\,b\in\mathbb{R}$$

が成り立ちます。

4. **シグモイド関数の性質と「discriminatory」性**

   シグモイド関数 σ は、一般に以下の性質を持ちます：

   - $\lim_{t \to -\infty} \sigma(t) = 0$
   - $\lim_{t \to \infty} \sigma(t) = 1$

   これを用いて、固定された w に対し、バイアス b を変化させたときの関数

   $$\varphi(b) = \int_K \sigma(w^T x + b)\, d\mu(x)$$

   を考えると、$\varphi(b) \equiv 0$ でなければなりません。
   特に $b \to \infty$ の極限では $\sigma(w^T x + b) \to 1$ となるため、

   $$\lim_{b \to \infty} \varphi(b) = \int_K 1\, d\mu(x) = \mu(K) = 0.$$

   また、適切な補助的議論（例えば、微分可能性やFourier変換を用いる方法）を通じて、
   任意の w に対して $\int_K \sigma(w^T x + b)\, d\mu(x) = 0$ から、
   測度 μ の「Fourier 変換」または「ラプラス変換」が零関数となり、
   その結果として μ が零測度でなければならないことが示されます（$\mu = 0$）。

5. **矛盾の成立と結論**

   しかし、Riesz 表現定理で得られた μ は非零と仮定していたため、
   $\mu = 0$ となるのは矛盾です。
   この矛盾により、初めの仮定（$\overline{\mathcal{F}} \neq C(K)$）が誤りであると結論付けられ、
   よって $\mathcal{F}$ は $C(K)$ に一様ノルムで稠密である、すなわち
   単一隠れ層の全結合ニューラルネットワークは任意の連続関数を任意の精度で近似可能であることが示されます。

補足

- この証明では、シグモイド関数が「discriminatory（識別的）」であるという性質が重要です。
  すなわち、

  $$\int_K \sigma(w^T x + b)\, d\mu(x) = 0 \quad \text{for all } w \in \mathbb{R}^n,\, b \in \mathbb{R}$$

であるならば、μ はゼロ測度でなければならないという性質です。

- 証明の詳細としては、Fourier変換やラプラス変換を用いた解析的議論を組み入れることで、より厳密に「μ が零である」ことを導出します。

以上が、隠れ層1層の全結合ニューラルネットワークに対する普遍近似定理の証明の概要です。