

Universal Approximation Theorem of Neural Network

数理科学科 3 回生 小泉 孝弥

概要

機械学習という単語を一言で説明するならば、「関数近似器」である。今回の ReMakers の機関紙では機械学習および第 3 次人工知能ブームの火付け役となった「深層学習」を数学的に定義し、深層学習の中心となる「ニューラルネットワーク」が万能関数近似器であることを示す。なお、本機関紙の前提知識は数学科 3 回生レベルの数学である。

1 人工知能・AI について

現在、人工知能 (Artificial Intelligence) の厳密な定義は完成しておらず、専門家の中でも意見が分かれている。なので、この機関紙では人工知能の定義について触れることはせず、現在の、人工知能の中心技術である「機械学習」について説明する。

1.1 機械学習・深層学習

最初に述べたとおり、機械学習 (Machine Learning) とは関数近似器である。もう少し詳細に述べれば、入力空間と呼ばれる集合 \mathcal{X} から出力空間と呼ばれる集合への良い写像 $f: \mathcal{X} \rightarrow \mathcal{Y}$ を構成するのが機械学習である。機械学習は主に、株価などのあるものの値を予想するモデルである「回帰 (Regression)」, あるものがどのクラスに属しているかを予測するモデルである「分類 (Classification)」の 2 つに分けられる。機械学習の中でも特に、あとで定義する、ニューラルネットワークというものをを用いる学習手法を深層学習 (Deep Learning) という。

1.2 深層学習と万能近似性

深層学習は以下の定理から万能関数近似器と呼ばれている。

Theorem. (万能関数近似定理)

X を集合とし $f: X \rightarrow \mathbb{R}$ を性質の良い関数とする。この時、 f を任意の精度で近似できるようなニューラルネットワークが存在する。

今回の機関紙の目標はこの定理を証明することであるが、今のままでは曖昧な単語が多すぎるため、数学の問題として扱うことができない。したがって、今からこの定理を数学的に述べなければならない。しかしながら、流れの都合上この定理を数学的に述べる前に、この定理の証明に不可欠な

「ハーン・バナッハの拡張定理」と「リースの表現定理」を証明することにする.

2 関数解析の基礎

リースの表現定理およびハーンバナッハの拡張定理は共に関数解析学 (Functional Analysis) と呼ばれる分野の定理である. したがって, この節では関数解析学の基礎事項を説明する. なお, ベクトル空間の係数体は実数体または複素数体とし, \mathbb{K} で表記する.

2.1 関数解析の基礎空間

Definition 2.1. (ノルム空間)

V をベクトル空間とする. 写像 $\|\cdot\| : V \rightarrow \mathbb{R}$ が以下の性質を満たすとする.

1. $\forall v \in V, \|v\| \geq 0$,
2. $\forall v \in V, \|v\| = 0 \iff v = 0$,
3. $\forall \alpha \in \mathbb{K}, \forall v \in V, \|\alpha v\| = |\alpha| \|v\|$ and
4. $\forall v, w \in V, \|v + w\| \leq \|v\| + \|w\|$.

この時, $\|\cdot\|$ を V のノルムといい, $(V, \|\cdot\|)$ をノルム空間 (normed space) という.

Proposition 2.2. $(V, \|\cdot\|)$ をノルム空間とする. この時, 以下で定義される写像 $d : V \times V \rightarrow \mathbb{R}$ は V 上の距離となる.

$$d(x, y) = \|x - y\|.$$

この距離を, ノルムから入る距離と呼ぶ.

Proof. ノルムの定義より明らかである. □

この定理より, ノルム空間はノルムから入る距離によって距離空間となる. 次にバナッハ空間を定義していく.

Definition 2.3. (完備)

(X, d) を距離空間とする. X の任意のコーシー列 $\{x_n\}_{n \in \mathbb{N}}$ が収束する時, (X, d) を完備距離空間 (complete metric space) または単に, 完備 (complete) という.

Definition 2.4. (バナッハ空間)

ノルム空間が距離空間として完備である時, バナッハ空間 (Banach space) という.

次にヒルベルト空間を定義する.

Definition 2.5. (内積空間)

V をベクトル空間とする. $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{K}$ が以下の性質を満たすとする.

1. $\forall v \in V, \langle \cdot, v \rangle, v \rangle \geq 0,$
2. $\forall v \in V, \langle \cdot, v \rangle, v \rangle = 0 \iff v = 0,$
3. $\forall v, w \in V, \langle \cdot, v \rangle, w \rangle = \overline{\langle \cdot, w \rangle, v \rangle},$
4. $\forall u, v, w \in V, \langle \cdot, u \rangle + v, w \rangle = \langle \cdot, u \rangle, w \rangle + \langle \cdot, v \rangle, w \rangle$ and
5. $\forall v, w \in V, \forall \alpha \in \mathbb{K}, \langle \cdot, \alpha v \rangle, w \rangle = \alpha \langle \cdot, v \rangle, w \rangle.$

この時 $\langle \cdot, \cdot \rangle, \cdot \rangle$ を V の内積 (inner) といい, 組 $(V, \langle \cdot, \cdot \rangle, \cdot \rangle)$ を内積空間 (inner space) という.

Proposition 2.6. $(V, \langle \cdot, \cdot \rangle)$ を内積空間とする. この時, 以下で定義される写像 $\| \cdot \| : V \rightarrow \mathbb{R}$ はノルムとなる.

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Proof. 内積の定義より明らか. □

これにより, 内積空間はノルム空間となる.

Definition 2.7. (ヒルベルト空間)

内積空間がバナッハ空間であるとき, ヒルベルト空間 (Hilbert space) という.

2.2 有界線形作用素

Definition 2.8. (線形作用素)

V, W をベクトル空間とし $f : V \rightarrow W$ が

1. $\forall v, w \in V, f(v + w) = f(v) + f(w),$
2. $\forall \alpha \in \mathbb{K}, \forall v \in V, f(\alpha v) = \alpha f(v).$

を満たす時, f を線形作用素 (linear operator) という.

Definition 2.9. (有界)

$(V_1, \| \cdot \|_1), (V_2, \| \cdot \|_2)$ をノルム空間とし f を V_1 から V_2 への写像とする. f が,

$$\exists c \geq 0 \text{ s.t. } \forall x \in V_1, \|f(x)\|_2 \leq c\|x\|_1$$

を満たす時, f は有界 (bounded) であるという.

Definition 2.10. (有界線形作用素)

有界であり線形である写像を有界線形作用素 (bounded linear operator) という.

Theorem 2.11. $(V_1, \|\cdot\|_1), (V_2, \|\cdot\|_2)$ をノルム空間とし, $T : V_1 \rightarrow V_2$ を有界線形作用素とする. また, 集合 C を以下で定義する.

$$C = \{c \geq 0 \mid \forall x \in V_1, \|f(x)\|_2 \leq c\|x\|_1\}.$$

この時, 以下のことが成立する.

1. C には最小値が存在する.
2. C の最小値を $\|T\|$ とすれば, 以下の式が成り立つ.

$$\|T\| = \sup_{\|x\|_1 \leq 1} \frac{\|T(x)\|_2}{\|x\|_1} = \sup_{\|x\|_1 = 1} \frac{\|T(x)\|_2}{\|x\|_1}.$$

3 ハーン・バナッハの拡張定理

4 リースの表現定理

5 機械学習の基礎

この節では機械学習の数学的な定式化および, 機械学習の具体例を紹介する. \mathcal{X} と \mathcal{Y} を集合とする.

5.1 機械学習の基礎空間

Definition 5.1. (仮説空間, 仮説)

\mathcal{X} から \mathcal{Y} へのなんらかの条件を満たす写像の集まりのことを仮説空間といい \mathcal{H} と表記する. すなわち,

$$\mathcal{H} = \{f : \mathcal{X} \rightarrow \mathcal{Y} \mid f \text{ が満たす条件} \}$$

である. (条件の具体例は後に述べる). 仮説空間 \mathcal{H} の元のことを仮説と呼ぶ. また, この時の \mathcal{X} を入力空間, \mathcal{Y} を出力空間と呼ぶ.

Definition 5.2. (データ)

\mathcal{X} と \mathcal{Y} の直積集合 $\mathcal{X} \times \mathcal{Y}$ の有限部分集合のことをデータといい D で表す.

Definition 5.3. (損失関数)

\mathcal{H} を仮説空間とする. \mathcal{H} から \mathbb{R} への写像 $\mathcal{L} : \mathcal{H} \rightarrow \mathbb{R}$ を損失関数 (Loss function) と呼ぶ.

Definition 5.4. (機械学習空間)

D をデータ, \mathcal{H} を仮説空間, \mathcal{L} を損失関数とする. この時 5 つ組 $(\mathcal{X}, \mathcal{Y}, D, \mathcal{H}, \mathcal{L})$ を機械学習空間 (Machine Learning sapce) という.

Definition 5.5. (学習, 最適仮説)

\mathcal{H} を仮設空間, $\mathcal{L} : \mathcal{H} \rightarrow \mathbb{R}$ を損失関数とする.

Definition 5.6. (機械学習)

機械学習空間 $(\mathcal{X}, \mathcal{Y}, D, \mathcal{H}, \mathcal{L})$ 上での学習を機械学習という.

5.2 機械学習の具体例

前節では, 機械学習の抽象的な枠組みを紹介したが, この説では機械学習空間の具体例を述べる.

5.2.1 単回帰分析

最初に機械学習の最も基本的なモデルである単回帰分析を紹介する.

Example 5.7. (単回帰分析)

機械学習空間 $(\mathcal{X}, \mathcal{Y}, D, \mathcal{H}, \mathcal{L})$ を以下のように定義する.

$\mathcal{X} = \mathbb{R}, \mathcal{Y} = \mathbb{R},$

$$\mathcal{H} = \{f : \mathcal{X} \rightarrow \mathcal{Y} \mid f(x) = wx, w \in \mathbb{R}\},$$

$$\mathcal{L}(f) = \sum_{i=1}^N |f(x_i) - y_i|.$$

この機械学習空間 $(\mathcal{X}, \mathcal{Y}, D, \mathcal{H}, \mathcal{L})$ 上での学習を単回帰分析という.

単回帰分析の最適仮説は

6 ニューラルネットワークの万能近似定理

いよいよ最終目標であるニューラルネットワークの万能近似性を示す.

6.1 ニューラルネットワーク, 深層学習

7 Generative Adversarial Networks

最後に, ニューラルネットワークの応用として GAN を紹介する.

7.1 GAN の定式化

$(\Omega, \mathcal{F}, \mathbb{P})$ を完備な確率空間とする. \mathcal{X}, \mathcal{Y} を空でない集合とする.

$$\begin{aligned}\mathcal{H}_1 &= \{G: \mathcal{Z} \rightarrow \mathcal{X} \mid G \text{ はニューラルネット} \}, \\ \mathcal{H}_2 &= \{D: \mathcal{X} \rightarrow [0, 1] \mid D \text{ はニューラルネット} \}\end{aligned}$$

ここで、 \mathcal{Z} は潜在空間と呼ばれる、なんらかの確率分布に従う確率変数の realization の集合である. すなわち,

$$\mathcal{Z} := \{z \in \mathbb{R}^d \mid \exists X_{\mathbf{R.V.}}: \Omega \rightarrow \mathbb{R}^d, \exists \omega \in \Omega \text{ s.t. } X(\omega) = z\}.$$

参考文献

- [1] <https://tutorials.chainer.org/ja/index.html>
- [2] https://github.com/Runnriru/machinelearning_text
- [3] カーネル法入門—正定値カーネルによるデータ解析・福水健次・2010
- [4] 統計的学習理論・金森 敬文・2015
- [5] 関数解析 POD 版・前田 周一郎・2007