―― JARIP 学会賞受賞記念論文 ――

深層学習の近似能力について

楠岡成雄 *

2020 年 9 月 4 日投稿

1　はじめに

機械学習には、対象や目的に応じて多くの手法が存在する。その代表的なものとして、教師あり学習、教師なし学習、強化学習がある。どの場合も関数近似、則ち未知の関数を何らかの関数で近似するということが必要となる。

機械学習を厳密に定義することは大変難しいが、大雑把に言って、集合 $\tilde{\mathcal{X}}$ を定義域とし、集合 $\tilde{\mathcal{Y}}$ を値域とする写像 $\tilde{f}:\tilde{\mathcal{X}}\to\tilde{\mathcal{Y}}$ に対して、それを近似する写像を機械的に見つける手法を指す。ただし、定義域 $\tilde{\mathcal{X}}$ 、値域 $\tilde{\mathcal{Y}}$ は未知であることが多く、集合 $\tilde{\mathcal{X}}$ を含む既知である集合 $\mathcal{X}$、集合 $\tilde{\mathcal{Y}}$ を含む既知である集合 $\mathcal{Y}$ を考え、さらに写像 $\tilde{f}:\tilde{\mathcal{X}}\to\tilde{\mathcal{Y}}$ の拡張である写像 $f:\mathcal{X}\to\mathcal{Y}$ を考え、写像 f を近似する写像を見つけるということが一般に行われる。写像 f を近似する方法として、多くの場合用いられるのが、パラメータ θ 付きの写像 $F(\cdot,\theta):\mathcal{X}\to\mathcal{Y}$ を考え、パラメータ θ を動かして、写像 f をよく近似する写像を見つけるという方法である。どのようなパラメータ付き写像のクラスを考えるかについても様々な方法が提案されているが、本論文では、深層学習と呼ばれる写像のクラスについて考えていく。特に、パラメータ数を増やすことで、どのような写像がどのような意味で、その写像のクラスで近似できるかという問題について考察する。

本論文では一般論を扱うが、例がないとわかりにくいと思うので、教師あり学習を例として考えていく。教師あり学習においては、先ず既知データとして、入力・出力のデータの組の列 $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ が与えられている。簡単のために入力 $x_k, k=1,\ldots,n,$ は N 次元ベクトル、出力 $y_k, k=1,\ldots,n,$ は M 次元ベクトルとする。

考える例として、以下の 2 つの例を考える。

例 1 (犬猫判定問題)

これは、jpeg で記述された画像データを見て、それが犬か猫かを判定するという問題である。$\tilde{\mathcal{X}}$ は犬またはネコの jpeg データの全体となるが、これを確定させせることは困難である。犬、猫に限らず考え得る可能なすべての jpeg データは $[0,1]^N$ に埋め込むことができるので $\mathcal{X}=[0,1]^N$ とすればよい（白黒の画像データであれば N は素子数とすればよい）。$\tilde{\mathcal{Y}}$ は今の場合 { 犬、猫 } であるので、犬に対して 0, 猫に対して 1 を割り当てることにすれば、$\tilde{\mathcal{Y}}=\{0,1\}$ とおけばよく、$M=1$ である。$\tilde{f}$ は $x\in\tilde{\mathcal{X}}$ が犬のデータであれば $\tilde{f}(x)=0$, 猫のデータであれば $\tilde{f}(x)=1$ という写像である。写像 f は写像 $\tilde{f}$ の拡張であれば何でもよい。

例 2 (時系列データに基づく予測問題)

$(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ が時系列データであり、将来の出力データを予測するという問題を考える。各時点 m での入力に対応する N 次元のベクトル値確率変数を X_m 出力に対応する M 次元のベクトル値確率変数を Y_m とし、$N+M$ 次元のベクトル値の確率過程 $\{(X_m,Y_m);\ m\in\mathbf{Z}\}$ が定常過程であり、$(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ が $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ の実現値と考えられる場合を考察する。

過去の入力データ $X_k,\ k=1,\ldots,n,$ から出力 Y_{n+1} を推定しようとすると、二乗損失関数の下では、

$$E[Y_{n+1}|X_n,.X_{n-1},,\ldots,X_1]=\tilde{f}_n(\{X_{n+k}\}_{k=0,-1,\ldots,-n+1})$$

*明治大学客員教授、東京大学名誉教授

がわかればよいが、これはデータ数 n に依存する関数で扱いにくい。真に求めるべき関数は

$$E[Y_{n+1}|X_k, k \leqq n] = f(\{X_{n+k}\}_{k \in \mathcal{Z}_{\leqq 0}})$$

とするべきであり、この時、求めるべき関数の定義域は $(\mathbf{R}^N)^{\mathbf{Z}_{\leqq 0}}$ の部分集合であるので、$\mathcal{X} = (\mathbf{R}^N)^{\mathbf{Z}_{\leqq 0}}$ という大きな集合となる。これを近似するパラメータ付き関数のクラスは $F(\cdot, \theta) : (\mathbf{R}^d)^{\mathbf{Z}_{\leqq 0}} \to \mathbf{R}^M,\ \theta \in \Theta,$ という形となる。

2　深層学習

深層学習の明確な定義はないが、まず基本となるのが、次に述べる順伝搬型ニューラルネットワーク (Forward Neural Network) である。以下ではこれを FNN と略称する。

FNN では集合 $\mathcal{X}, \mathcal{Y}$ は共にユークリッド空間の部分集合である。則ち、次元 $N, M \geqq 1$ が与えられており、$\mathcal{X} \subset \mathbf{R}^N$, $\mathcal{Y} \subset \mathbf{R}^M$ となる。

活性化関数 (activation function) と呼ばれる関数 $\rho : \mathbf{R} \to \mathbf{R}$ を一つ定める。この時、$N \geqq 1$ に対して写像 $\vec{\rho}_N : \mathbf{R}^N \to \mathbf{R}^N$ を

$$\vec{\rho}_N(\xi_1, \ldots, \xi_N) = (\rho(\xi_1), \ldots, \rho(\xi_N)), \qquad (\xi_1, \ldots, \xi_N) \in \mathbf{R}^N$$

で定義する。

さらに、$m \geqq 2$ および $N_1, \ldots, N_{m-1} \geqq 1$ を定め、$N_0 = N$, $N_m = M$ とおく。

今、アファイン写像の列 $W_k : \mathbf{R}^{N_k} \to \mathbf{R}^{N_{k+1}}$, $k = 0, \ldots, m-1$, を与える。ここで、$W_k : \mathbf{R}^{N_k} \to \mathbf{R}^{N_{k+1}}$ がアファイン写像であるとは、$w_{k,i,j} \in \mathbf{R}$, $i = 1, \ldots, N_{k+1}$, $j = 0, 1, \ldots, N_k$, が存在して

$$W_k(\vec{z})_i = \sum_{j=1}^{N_k} w_{k,i,j} z_j + w_{k,i,0}, \qquad i = 1, \ldots, N_{k+1},\ \vec{z} = (z_1, \ldots, z_{N_k}) \in \mathbf{R}^{N_k} \tag{1}$$

となることをいう。

さて、入力 $x \in \mathbf{R}^N = \mathbf{R}^{N_0}$ に対して

$$\begin{aligned} \vec{z}^{(0)} &= x, \\ \vec{u}^{(k+1)} &= W_k(\vec{z}^{(k)}) \in \mathbf{R}^{N_{k+1}}, \qquad k = 0, \ldots, m-1, \\ \vec{z}^{(k)} &= \vec{\rho}_{N_k}(u_1^{(k)}, \cdots, u_{N_k}^{(k)}) \in \mathbf{R}^{N_k}, \qquad k = 1, \ldots, m-1, \end{aligned}$$

と定め、出力 $\hat{f}(x) = \vec{u}^{(m)}$ とおく。この写像 $\hat{f} : \mathbf{R}^N \to \mathbf{R}^M$ はアファイン写像の列 $W_1, \ldots, W_{m-1}$ を一つ決めれば定まるので、$W_1, \ldots, W_{m-1}$ の関数でもある。則ち、

$$\hat{f}(x) = F(x, W_0, \ldots, W_{m-1}), \qquad x \in \mathbf{R}^N,$$

となる。アファイン写像 $W_k : \mathbf{R}^{N_k} \to \mathbf{R}^{N_{k+1}}$, $k = 1, \ldots, m-1$, は式 (1) より $w_{k,i,j} \in \mathbf{R}$, $i = 1, \ldots, N_{k+1}$, $j = 0, 1, \ldots, N_k$, で決まるので、それぞれ $(N_k + 1)N_{k+1}$ 個のパラメータを持つ。従って、活性化関数 ρ, 及び $m \geqq 2$, $N_1, \ldots, N_{m-1}$, を固定し、アファイン写像の列 $W_1, \ldots, W_{m-1}$ を変化させることを考えると、パラメータの次元は $\displaystyle\sum_{k=0}^{m-1} (N_k + 1)N_{k+1}$ となる。なお、m は層数と呼ばれ、与えた関数 $\hat{f}$ は m 層ネットワークと呼ばれる。また、$\vec{z}^{(k)}$, $\vec{u}^{(k)}$, $k = 1, \ldots, m-1$, は層 k のユニットの出力と呼ばれる。

活性化関数の例としては、

$$\rho(u) = 1_{[0,\infty)}(u) = \begin{cases} 1, & u \geqq 0, \\ 0, & u < 0, \end{cases}$$

$$\rho(u) = \frac{1}{1+e^{-u}}$$

$$\rho(u) = \tanh(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}$$

などがあるが、近年は、以下のような正規化線形関数 (rectified linear function)

$$\rho(u) = \max\{0, u\}$$

がよく用いられるようである。

FNN では N 次元データを入力とし、M 次元データを出力として閉じた関数の形をしており、犬猫判定問題のように、出力データは過去の入力データに影響されないような問題に適している。しかし、時系列データから将来予測をするといった問題の場合、出力データが過去の入力データに依存しており、どこまで過去を遡ればよいかは自明ではない。このような問題に対応するため、様々な手法が考えられているが、その一つが再帰型ニューラルネットワーク（Reccurent Neural Network）である。以下、これを RNN と略称する。ただし、RNN は FNN と違い明確に定義することは難しい（その説明は例えば 岡谷 [8] を見られたい）。ただ、RNN の特徴は途中段階の各層のユニットの出力を決めるのに、過去のデータも用いるという点にある。その例については、第 4 節で述べることにする。

3　FNN の近似能力

大雑把に言って、$\mathbf{R}^N$ 上の関数は 2 層の FNN で近似できることが知られている。これについて説明する。

単純のために、先ず $f : \mathbf{R}^N \to \mathbf{R}$ がなめらか（C^∞-級）であり、$f(x) = 0, |x| > R,$ となるような $R > 0$ が存在する場合（則ち $f \in C_0^\infty(\mathbf{R}^N)$ となる場合）を考える。f のフーリエ変換を $\hat{f} : \mathbf{R}^N \to \mathbf{R}$ とする。則ち

$$\hat{f}(\xi) = \int_{\mathbf{R}^N} \exp(-\sqrt{-1}x \cdot \xi) f(x) dx, \quad \xi \in \mathbf{R}^N$$

である。この時

$$f(x) = (\frac{1}{2\pi})^N \int_{\mathbf{R}^N} \exp(\sqrt{-1}x \cdot \xi) \hat{f}(\xi) d\xi \quad x \in \mathbf{R}^N$$

となる。よって

$$f(x) = \lim_{n\to\infty} \sum_{k_1,\ldots,k_N=-n2^n}^{n2^n} (\frac{1}{2\pi})^N 2^{-Nn} \exp(\sqrt{-1}(\sum_{i=1}^N x_i \frac{k_i}{2^n})) \hat{f}(\frac{k_1}{2^n}, \ldots, \frac{k_N}{2^n})$$

となり、この f に強い仮定をおいたので、この収束は $\{x \in \mathbf{R}^N;\ |x| \leqq R\}$ 上の一様収束である。$\sum_{i=1}^N x_i \frac{k_i}{2^n}$ は x の一次関数であるので、$\cos t, \sin t$ を t の有界な範囲で 2 層 FNN で一様に近似できれば一般の関数も 2 層 FNN で近似できることがわかる。例えば、活性化関数が正規化線形関数の場合は、折れ線近似を用いて近似できる。このようなことにより、ユークリッド空間 $\mathbf{R}^N$ 上の関数の 2 層 FNN による近似について多くの結果が知られている。([1], [2], [3], [4], [5], [6], [7])

前節で述べたように、定義域が有限次元のユークリッド空間であるような関数の場合にはフーリエ変換等の実解析的な方法で、関数が近似できることが示せるが、例えば時系列のようなものを考えたいときは、空間が大きくなりすぎて、このような方法では表現力をうまく論じることができない。

4　RNN の表現力

この節では以下のような RNN を考える（ただし、この RNN は時系列の予測問題に対して RNN に近似能力があることを示す例で、あまり推奨できるものではないことをお断りしておく）。

入力データが N 次元ベクトルで、出力データが M 次元ベクトルとする。活性化関数 $\rho : \mathbf{R} \to \mathbf{R}$ を一つ定め、$m \geqq 3,\ N_2, \ldots, N_{m-1} \geqq 1$ を定める。$N_0 = N,\ N_m = M,\ N_1 = N_2,$ とおく。さらにアファイン写像の列 $W_k : \mathbf{R}^{N_k} \to \mathbf{R}^{N_{k+1}},\ k = 0, 2, 3, \ldots, m-1,$ を与える。ここで、$k = 1$ は除かれていることに注意されたい。

さて、入力データの列 $x_r \in \mathbf{R}^{N_1},\ r = 1, 2, \ldots, n,$ に対して

$$\vec{u}_r^{(1)} = W_0 x_r \in \mathbf{R}^{N_1}, \qquad r = 1, \ldots, n,$$

$$\vec{z}_r^{(1)} = \vec{\rho}_{N_1}(\vec{u}_r^{(1)}) \in \mathbf{R}^{N_1}, \qquad r = 1, \ldots, n,$$

とおく。

$$\vec{z}_0^{(2)} = 0 \in \mathbf{R}^{N_2},$$

とし、$\vec{u}_r^{(2)},\ \vec{z}_r^{(2)},\ r = 1, 2, \ldots, n,$ を帰納的に

$$\vec{u}_r^{(2)} = \frac{1}{3}\vec{z}_r^{(1)} + \frac{1}{3}\vec{z}_{r-1}^{(2)} + \frac{1}{6}\vec{1}_{N_2} \in \mathbf{R}^{N_2}, \qquad r = 1, \ldots, n,$$

$$\vec{z}_r^{(2)} = \vec{\rho}_{N_2}(\vec{u}_r^{(2)}) \in \mathbf{R}^{N_2}, \qquad r = 1, \ldots, n,$$

と定める。ただし、$\vec{1}_{N_2} = (1, 1, \ldots, 1) \in \mathbf{R}^{N_2}$ である。さらに

$$\vec{u}_r^{(k+1)} = W_k(\vec{z}_r^{(k)}) \in \mathbf{R}^{N_{k+1}}, \qquad k = 2, \ldots, m-1,\ r = 1, \ldots, n,$$

$$\vec{z}_r^{(k)} = \vec{\rho}_{N_k}(\vec{u}_r^{(k)}) \in \mathbf{R}^{N_k}, \qquad k = 3, \ldots, m-1,\ r = 1, \ldots, n,$$

と定める。各 $r = 1, \ldots, n$ に対して $\vec{u}_r^{(m)}$ は $(x_1, \ldots, x_r) \in (\mathbf{R}^N)^r$ の関数であるので $\vec{u}_r^{(m)} = \hat{f}_r(x_1, \ldots, x_r)$ とおくことができる。

この時、以下のことが成立する。

定理 1 $\{X_k\}_{k \in \mathbf{Z}}$ は確率空間 $(\Omega, P, \mathcal{F})$ 上定義された $\mathbf{R}^N$-値定常過程とする。さらに、$F : (\mathbf{R}^N)^{\mathbf{Z}_{\leqq 0}} \to \mathbf{R}^N$ はボレル可測関数で

$$E[|F(\{X_k\}_{k \in Z_{\leqq 0}})|^2] < \infty$$

を満たすとする。

この時、任意の $\delta > 0$ に対して、$m = 5,\ N_2, N_3, N_4 \geqq 1,$ 及びファイン写像の列 $W_k : \mathbf{R}^{N_k} \to \mathbf{R}^{N_{k+1}},$ $k = 0, 2, 3, 4$ が存在して、

$$\overline{\lim_{n \to \infty}} E[|F(\{X_{n+k}\}_{k \in Z_{\leqq 0}} - \hat{f}_n(X_1, \ldots, X_n)|^2] < \delta$$

が成立する。

もし、定常過程 $\{X_k\}_{k \in \mathbf{Z}}$ がエルゴード的であるならば、任意の $\delta > 0$ に対して、$m = 5,\ N_2, N_3, N_4 \geqq 1,$ 及びアファイン写像の列 $W_k : \mathbf{R}^{N_k} \to \mathbf{R}^{N_{k+1}},\ k = 0, 2, 3, 4$ が存在して、確率 1 で

$$\overline{\lim_{n \to \infty}} \frac{1}{n} \sum_{\ell=1}^{n} |F(\{X_{\ell+k}\}_{k \in Z_{\leqq 0}}) - \hat{f}_\ell(X_1, \ldots, X_\ell)|^2 < \delta$$

が成立する。

以下ではこの定理を証明していく。

5　FNN と Stone-Weierstrass の定理

$Aff(\mathbf{R}^N;\mathbf{R}^M)$ は $\mathbf{R}^N$ から $\mathbf{R}^M$ へのアファイン写像全体の集合とする。則ち、$W \in Aff(\mathbf{R}^N;\mathbf{R}^M)$ であるとは W が $\mathbf{R}^N$ から $\mathbf{R}^M$ への写像であり $a_{i,j} \in \mathbf{R}$, $i = 1,\dots,M$, $j = 0,1,\dots,N$, が存在して

$$W(x_1,\dots,x_N)_i = \sum_{j=1}^{N} a_{i,j}x_j + a_{i,0}, \quad i = 1,\dots,M, \quad (x_1,\dots,x_N) \in \mathbf{R}^N$$

となることをいう。

以下は明らかであろう。

命題 2 $N_1, N_2, N_3 \geqq 1$ であり、$W_i \in Aff(\mathbf{R}^{N_i};\mathbf{R}^{N_{i+1}})$, $i = 1,2$, ならば $W_2 \circ W_1 \in Aff(\mathbf{R}^{N_1};\mathbf{R}^{N_3})$ である。

以後、連続関数 $\rho : \mathbf{R} \to \mathbf{R}$ を一つ固定する。$\vec{\rho}_N$, $N \geqq 1$, は

$$\vec{\rho}_N(x_1,\dots,x_N) = (\rho(x_1),\dots,\rho(x_N)), \quad (x_1,\dots,x_N) \in \mathbf{R}^N$$

で与えられる $\mathbf{R}^N$ から $\mathbf{R}^N$ への写像であった。

定義 3 $N, M \geqq 1$, $m \geqq 2$ とする。$f \in \mathbf{F}_{\rho,m}(\mathbf{R}^N;\mathbf{R}^M)$ であるとは $N_1,\dots,N_{m-1} \geqq 1$ 及び $W_k \in Aff(\mathbf{R}^{k-1};\mathbf{R}^k)$, $k = 1,\dots,m$, が存在して（ただし、$N_0 = N$, $N_m = M$ とする）、

$$f = W_m \circ \vec{\rho}_{N_{m-1}} \circ W_{m-1} \circ \vec{\rho}_{N_{m-2}} \circ \cdots \circ W_2 \circ \vec{\rho}_{N_1} \circ W_1$$

となることをいう。

また、$\mathbf{F}_{\rho,1}(\mathbf{R}^N;\mathbf{R}^M) = Aff(\mathbf{R}^N;\mathbf{R}^M)$ と定める。

定義より明らかに $\mathbf{F}_{\rho,m}(\mathbf{R}^N;\mathbf{R}^M)$ は ρ を活性化関数としたときの m 層 FNN で表すことのできる関数全体である。以下のことは容易に示せる。

命題 4 (1) $N_1, N_2, N_3 \geqq 1$, $m_1, m_2 \geqq 1$ とする。もし、$f_i \in \mathbf{F}_{\rho,m_i}(\mathbf{R}^{N_i};\mathbf{R}^{N_{i+1}})$, $i = 1,2$, ならば $f_2 \circ f_1 \in \mathbf{F}_{\rho,m_1+m_2-1}(\mathbf{R}^{N_1};\mathbf{R}^{N_3})$ となる。
(2) $N_i, M_i \geqq 1$, $i = 1,2$, $m \geqq 1$ とする。$f_i \in \mathbf{F}_{\rho,m}(\mathbf{R}^{N_i};\mathbf{R}^{M_i})$, $i = 1,2$, に対して写像 $f : \mathbf{R}^{N_1+N_2} \to \mathbf{R}^{M_1+M_2}$ を

$$f(x_1,\dots,x_{N_1+N_2}) = (f_1(x_1,\dots,x_{N_1}), f_2(x_{N_1+1},\dots,x_{N_1+N_2})), \quad (x_1,\dots,x_{N_1+N_2}) \in \mathbf{R}^{N_1+N_2}$$

で定めると $f \in \mathbf{F}_{\rho,m}(\mathbf{R}^{N_1+N_2};\mathbf{R}^{M_1+M_2})$ となる。
(3) $N, M, m \geqq 1$ とする。この時、$\mathbf{F}_{\rho,m}(\mathbf{R}^N;\mathbf{R}^M)$ はベクトル空間である。

FNN の近似能力に関して述べていく。

以下では M は 可分な完備距離空間とし、$C(M)$ は M 上の連続関数全体の集合とする。Stone [9] により以下の定理が示されている。

定理 5 (Stone-Weierstrass) M がコンパクトと仮定する。V が $C(M)$ の部分ベクトル空間であり、以下の *2* つの仮定を満たすとする。
(1) $1 \in V$ であり、$f, g \in V$ ならば $fg \in V$.
(2) 任意の $x, y \in M$, $x \neq y$, に対して $g(x) \neq g(y)$ を満たす $g \in V$ が存在する。

この時、任意の $f \in C(M)$ に対して

$$\inf\{\sup_{x\in M} |f(x) - g(x)|;\ g \in V\} = 0$$

が成立する。

この定理の証明のアイデアは FNN の表現力の証明にも有効である。以下それを示していく。

定義 6 C を $C(M)$ の部分集合とし、$m \geqq 1$ とする。$C(M)$ の部分集合 $\mathcal{C}_{\rho,m}(C)$ を以下で定める。
$f \in \mathcal{C}_{\rho,m}(C)$ であるとは、$n \geqq 1$, $g_1, \ldots, g_n \in C$ 及び $F \in \mathbf{F}_{\rho,m}(\mathbf{R}^n; \mathbf{R})$ が存在して

$$f(x) = F(g_1(x), \ldots g_n(x)), \quad x \in M$$

が成り立つことをいう。

次の命題は明らかであろう。

命題 7 M は 可分な完備距離空間、$m \geqq 1$ とする。また、C を $C(M)$ の部分集合とする。この時、$\mathcal{C}_{\rho,m}(C)$ はベクトル空間であり、$1 \in \mathcal{C}_{\rho,m}(C)$ である。また、$f \in \mathcal{C}_{\rho,m}(C)$ ならば $\rho \circ f \in \mathcal{C}_{\rho,m+1}(C)$ となる。

以後、$\rho : \mathbf{R} \to [0,1]$ は連続関数で

$$\lim_{t \downarrow -\infty} \rho(t) = 0, \qquad \lim_{t \uparrow \infty} \rho(t) = 1$$

を満たすものとする。
(注意) $\rho(t) = \max\{t, 0\}$ の時は $\tilde{\rho}(t) = \rho(t) - \rho(t-1) = \max\{0, \min\{t, 1\}\}$ が条件を満たすので、以下の結果がこの場合も成立する。

補題 8 K_0, K_1 は $K_0 \cap K_1 = \emptyset$ を満たす M のコンパクト部分集合とする。また、C は $C(M)$ の部分集合で任意の $x \in K_0$ 及び $y \in K_1$ に対して $\varphi(x) \neq \varphi(y)$ となる $\varphi \in C$ が存在すると仮定する。この時、任意の $\delta > 0$ に対して $0 \leqq f \leqq 1$,

$$\sup_{x \in K_0} f(x) \leqq \delta, \qquad \inf_{x \in K_1} f(x) \geqq 1 - \delta,$$

となる $f \in \mathcal{C}_{\rho,4}(C)$ が存在する。

証明. 仮定より任意の $x \in K_0$, $y \in K_1$ に対して $\varphi_{x,y} \in \mathcal{C}_{\rho,1}(C)$ で

$$\varphi_{x,y}(x) \geqq 2, \qquad \varphi_{x,y}(y) \leqq -2$$

となるものが存在する。この時、任意の $y \in K_1$ に対して

$$K_0 \subset \bigcup_{x \in K_0} \{z \in M;\ \varphi_{x,y}(z) > 1\}$$

が成立する。K_0 はコンパクトであり、$\{z \in M;\ \varphi_{x,y}(z) > 1\}$ は開集合であるので、任意の $y \in K_1$ に対して $n(y) \geqq 1$, 及び $x_k(y) \in K_0$, $k = 1, \ldots, n(y)$, で

$$K_0 \subset \bigcup_{k=1}^{n(y)} \{z \in M;\ \varphi_{x_k(y),y}(z) > 1\}$$

となるものが存在する。$U(y)$, $y \in K_1$, を

$$U(y) = \bigcap_{k=1}^{n(y)} \{z \in M;\ \varphi_{x_k(y),y}(z) < -1\}$$

とおく。この時、$U(y)$ は y の 開近傍である。K_1 はコンパクトであるので $m \geqq 1$ 及び $y_\ell \in K_1$, $\ell = 1, \ldots, m$, が存在して

$$K_1 \subset \bigcup_{\ell=1}^{m} U(y_\ell)$$

となる。

$$\varliminf_{\alpha\to\infty}\inf_{z\in K_0}\sum_{k=1}^{n(y)}\rho(\alpha\varphi_{x_k(y),y}(z)) \geqq \varliminf_{\alpha\to\infty}\rho(\alpha)=1, \qquad y\in K_1,$$

であり、

$$\varlimsup_{\alpha\to\infty}\sup_{z\in U(y)}\sum_{k=1}^{n(y)}\rho(\alpha\varphi_{x_k(y),y}(z)) \leqq \varlimsup_{\alpha\to\infty}n(y)\rho(-\alpha)=0, \qquad y\in K_1$$

であることに注意する。

$g_{\alpha,\ell}\in\mathcal{C}_{\rho,2}(C)$, $\alpha>0$, $\ell=1,\dots,m$, を

$$g_{\alpha,\ell}=\sum_{k=1}^{n(y_\ell)}\rho(\alpha\varphi_{x_k(y_\ell),y_\ell}), \qquad \ell=1,\dots,m,$$

で定める。この時、上の注意より $\alpha>0$ で

$$\inf_{z\in K_0}g_{\alpha,\ell}(z)\geqq\frac{3}{4}, \qquad \ell=1,\dots,m,$$

かつ

$$\sup_{z\in U(y_\ell)}g_{\alpha,\ell}(z)\leqq\frac{1}{4} \qquad \ell=1,\dots,m,$$

となるものが存在する。この時、

$$\varliminf_{\beta\to\infty}\inf_{z\in K_1}\sum_{\ell=1}^{m}\rho(\beta(\frac{1}{2}-g_{\alpha,\ell}(z)))\geqq 1,$$

であり、

$$\varlimsup_{\beta\to\infty}\sup_{z\in K_0}\sum_{\ell=1}^{m}\rho(\beta(\frac{1}{2}-g_{\alpha,\ell}(z)))=0$$

が成立する。よって、$\beta>0$ で

$$\inf_{z\in K_1}\sum_{\ell=1}^{m}\rho(\beta(\frac{1}{2}-g_{\alpha,\ell}(z)))\geqq\frac{3}{4},$$

かつ

$$\sup_{z\in K_0}\sum_{\ell=1}^{m}\rho(\beta(\frac{1}{2}-g_{\alpha,\ell}(z)))\leqq\frac{1}{4}$$

となるものが存在する。$g\in\mathcal{C}_{\rho,3}(C)$ を

$$g=\sum_{\ell=1}^{m}\rho(\beta(\frac{1}{2}-g_{\alpha,\ell}))$$

で定める。この時、

$$\varliminf_{\gamma\to\infty}\inf_{z\in K_1}\rho(\gamma(\frac{1}{2}-g(z))=1,$$

かつ

$$\varlimsup_{\gamma\to\infty}\sup_{z\in K_0}\rho(\gamma(\frac{1}{2}-g(z))=0$$

となる。

$\rho(\gamma(\frac{1}{2}-g))\in\mathcal{C}_{\rho,4}(C)$ であり、$0\leqq\rho(\gamma(\frac{1}{2}-g))\leqq 1$ であるので、命題の主張を得る。

∎

この補題より、先ず次のことを得る。

定理 9 M はコンパクト距離空間であり、$f \in C(M)$ とする。$C(M)$ の部分集合 C が以下の条件を満たすとする。

$x, y \in M$ であり、$f(x) \neq f(y)$ であるならば、$g \in C$ で $g(x) \neq g(y)$ を満たすものが存在する。

この時、

$$\inf\{\sup_{x \in M} |f(x) - g(x)|;\ g \in \mathcal{C}_{\rho,4}(C)\} = 0$$

が成立する。

証明. Step 1. すべての $x \in M$ に対して $0 < f(x) \leqq 1$ となる場合。

$n \geqq 1$ とし、M 上の関数 f_n を

$$f_n(x) = \sum_{k=1}^{n} \frac{k}{n} 1_{((k-1)/n, k/n]}(f(x))$$

で与える。ここで、1_A は $x \in A$ ならば $1_A(x) = 1$, $x \notin A$ ならば $1_A(x) = 0$ で与えられる関数である。

この時、

$$f_n(x) - \frac{1}{n} \leqq f(x) \leqq f_n(x), \qquad x \in M$$

となることが分かる。また、

$$f_n(x) = \sum_{k=1}^{n} (\sum_{\ell=1}^{k} \frac{1}{n} 1_{((k-1)/n, k/n]}(f(x))) = \sum_{\ell=1}^{n} \frac{1}{n} 1_{((\ell-1)/n, 1]}(f(x))$$

であり

$$f_n(x) - \frac{1}{n} \leqq \sum_{\ell=2}^{n} \frac{1}{n} 1_{((\ell-1)/n, 1]}(f(x))$$

となることがわかる。

補題 8 より $h_{n,k} \in \mathcal{C}_{\rho,4}(C)$, $k = 2, \ldots, n$, で $0 \leqq h_{n,k} \leqq 1$,

$$h_{n,k}(x) \leqq (\frac{1}{n})^2 \qquad x \in f^{-1}([0, (k-1)/n]),$$

$$h_{n,k}(x) \geqq 1 - (\frac{1}{n})^2, \qquad x \in f^{-1}([k/n, 1])$$

を満たすものが存在する。

$$g_n(x) = \sum_{k=2}^{n} \frac{1}{n} h_{n,k}(x),$$

とおくと、$g \in \mathcal{C}_{\rho,4}(C)$ であり、

$$f(x) - \frac{1}{n} \leqq f_n(x) - \frac{1}{n} \leqq (1 - (\frac{1}{n})^2)^{-1} g_n(x) + \frac{1}{n} \qquad x \in M,$$

かつ

$$f(x) + \frac{1}{n} \geqq f_n(x) \geqq g_n(x) - \frac{2}{n}, \qquad x \in M$$

となることがわかる。

$$g_n(x) \geqq (1 - (\frac{1}{n})^2)(f(x) - \frac{2}{n}) \geqq f(x) - \frac{2}{n}, \qquad x \in M$$

となるので

$$|f(x) - g_n(x)| \leqq \frac{2}{n}, \qquad x \in M$$

を得る。これより主張を得る。

Step 2. 一般の場合

$f \in C(M)$ とすると、M はコンパクトなので

$$a = \min f > -\infty, \qquad b = \min f < \infty$$

となる。$\tilde{f} : M \to \mathbf{R}$ を

$$\tilde{f}(x) = \frac{1}{b-a+1}(f(x) - a + 1), \qquad x \in M$$

で定めると、$\tilde{f} \in C(M)$ かつ $0 < \tilde{f} \leqq 1$ となる。$f = (b-a+1)\tilde{f} + (b-a+1)(1-a)$ であるので、Step 1 より定理の主張を得る。

∎

上記定理の帰結として以下を得る。

系 10 M はコンパクト距離空間 とする。$C(M)$ の部分集合 C が以下の条件を満たすとする。

(条件) $x, y \in M$, $x \neq y$ であるならば、$g \in C$ で $g(x) \neq g(y)$ を満たすものが存在する。

この時、

$$\inf\{\sup_{x \in M} |f(x) - g(x)|;\ g \in \mathcal{C}_{\rho,4}(C)\} = 0$$

がすべての $f \in C(M)$ に対して成立する。

定理 11 M は可分完備距離空間、$\mathcal{B}(M)$ を M 上のボレル加法族とし、ν は $(M, \mathcal{B}(M))$ 上の確率測度とする。また、$C(M)$ の部分集合 C が以下の条件を満たすとする。

(条件) $x, y \in M$, $x \neq y$ であるならば、$g \in C$ で $g(x) \neq g(y)$ を満たすものが存在する。

この時、以下が成立する。

(1) 任意の $B \in \mathcal{B}(M)$ に対して

$$\inf\{\int_M |1_B(x) - g(x)|\nu(dx);\ g \in \mathcal{C}_{\rho,4}(C),\ 0 \leqq g \leqq 1\} = 0$$

が成り立つ。

(2) ボレル可測関数 $f : M \to \mathbf{R}$ が $0 \leqq f \leqq 1$ を満たすならば

$$\inf\{\int_M |f(x) - g(x)|\nu(dx);\ g \in \mathcal{C}_{\rho,4}(C),\ 0 \leqq g \leqq 1\} = 0$$

が成り立つ。

(3) 任意のボレル可測関数 $f : M \to \mathbf{R}$ に対して

$$\inf\{\int_M \min\{|f(x) - g(x)|, 1\}\nu(dx);\ g \in \mathcal{C}_{\rho,4}(C)\} = 0$$

が成り立つ。

(4) $p \in [1, \infty)$ とする。ボレル可測関数 $f : M \to \mathbf{R}$ が

$$\int_M |f(x)|^p \nu(dx) < \infty$$

を満たすならば、

$$\inf\{\int_M |f(x) - g(x)|^p \nu(dx);\ g \in \mathcal{C}_{\rho,4}(C)\} = 0$$

が成り立つ。

証明. (1) $B \in \mathcal{B}(M)$ とする。また、$\varepsilon \in (0,1)$ とする。M は完備距離空間であるので M のコンパクト集合 K_0, K_1 で $K_0 \subset M \setminus B$, $K_1 \subset B$, $\nu(M \setminus (K_0 \cup K_1)) < \varepsilon$ となるものが存在する。この時、補題 8 より

$\varphi \in \mathcal{C}_{\rho,4}(C)$ で $0 \leqq \varphi \leqq 1$, $x \in K_0$ に対して $\varphi(x) \leqq \varepsilon$ であり、$x \in K_1$ に対して $\varphi(x) \geqq 1-\varepsilon$ となるものが存在する。この時、

$$\begin{aligned}
&\int_M |1_B(x) - \varphi(x)|\nu(dx) \\
&= \int_{K_0} |1_B(x) - \varphi(x)|\nu(dx) + \int_{K_1} |1_B(x) - \varphi(x)|\nu(dx) + \int_{M\setminus(K_0\cup K_1)} |1_B(x) - \varphi(x)|\nu(dx) \\
&\leqq \varepsilon\nu(K_0) + \varepsilon\nu(K_1) + \nu(M \setminus (K_0 \cup K_1)) \leqq 2\varepsilon
\end{aligned}$$

となる。よって (1) の主張が示された。

(2) を示す。$f : M \to \mathbf{R}$ は可測関数で、$0 \leqq f \leqq 1$ とする。$B_{n,k} \in \mathcal{B}(M)$, $n \geqq 1$, $k = 1,\ldots,n-1$, を

$$B_{n,k} = \{x \in M;\ \frac{k}{n} < f(x) \leqq 1\}$$

とおくと $(\ell-1)/n < f(x) \leqq \ell/n$, $\ell = 1,\ldots,$ の時、

$$\frac{1}{n}\sum_{k=1}^{n-1} 1_{B_{n,k}}(x) = (\ell-1)/n$$

となるので

$$\sup_{x\in M} |f(x) - \frac{1}{n}\sum_{k=1}^{n-1} 1_{B_{n,k}}(x)| \leqq \frac{1}{n}, \qquad n \geqq 1$$

となる。よって、主張 (1) より主張 (2) を得る。

主張 (3),(4) は主張 (2) より容易にわかる。

∎

6　RNN の近似可能性

前節と同様に M を可分な完備距離空間とする。また、$\rho : \mathbf{R} \to [0,1]$ は連続関数で $0 \leqq \rho \leqq 1$, $\lim_{t\downarrow-\infty} \rho(t) = 0$, $\lim_{t\uparrow\infty} \rho(t) = 1$ を満たすものとする。さらに、この節では $\rho|_{(0,1)} : (0,1) \to \mathbf{R}$ は C^1-級で $c_0 \in (0,1)$ が存在して

$$0 < c_0 \leqq \rho'(t) \leqq 1, \qquad t \in (0,1)$$

が成り立つと仮定する。

$\hat{M} = M^{\mathbf{Z}_{\leqq 0}} = \{\{x_n\}_{n\in\mathbf{Z}_{\leqq 0}};\ x_n \in M\}$ とおく。この時、$\hat{M}$ も可分な完備距離空間となる。$\psi \in C(M)$ に対して $U^{(m)}_{\rho,i,n}(\cdot,\psi) : \hat{M} \to \mathbf{R}$, $i = 1,2$, $n,m \in \mathbf{Z}_{\leqq 0}$, を以下で定める。

まず、$\vec{x} = \{x_n\}_{n\in\mathbf{Z}_{\leqq 0}} \in \hat{M}$ に対して

$$U^{(m)}_{1,n}(\vec{x},\psi) = \begin{cases} 0, & n \leqq m-1 \text{ の時} \\ \rho(\psi(x_n)), & n \geqq m \text{ の時} \end{cases}$$

で定める。さらに、$n \leqq m$ の時、$U^{(m)}_{\rho,2,n}(\vec{x};\psi) = 0$, $\vec{x} \in \hat{M}$, と定め、$n = m+1, m+2, \ldots, 0$, に対しては帰納的に

$$U^{(m)}_{\rho,2,n}(\vec{x};\psi) = \rho(\frac{1}{3}U^{(m)}_{\rho,1,n}(\vec{x},\psi) + \frac{1}{3}U^{(m)}_{\rho,2,n-1}(\vec{x},\psi) + \frac{1}{6})$$

で定める。

この時以下が成り立つ。

命題 12 任意の $\psi \in C(M)$ に対して

$$U_{\rho,i,n}(\vec{x},\psi) = \lim_{m\downarrow -\infty} U^{(m)}_{\rho,i,n}(\vec{x},\psi), \qquad i=1,2,\ \vec{x}\in \hat{M}$$

が存在する。さらに、

$$|U_{\rho,2,n}(\vec{x},\psi) - U^{(m)}_{\rho,2,n}(\vec{x},\psi)| \leqq 3^{m-n}, \qquad \vec{x}\in \hat{M},\ n,m\in \mathbf{Z}_{\leqq 0}$$

が成り立つ。

また、$U_{\rho,i,n}(\cdot,\psi):\hat{M}\to \mathbf{R}$, $i=1,2$, $n\in \mathbf{Z}_{\leqq 0}$, は連続であり、$\vec{x}=\{x_n\}_{n\in \mathbf{Z}_{\leqq 0}}\in \hat{M}$ に対して

$$U_{\rho,1,n}(\vec{x},\psi) = \rho(\psi(x_n)),$$

$$U_{\rho,2,n}(\vec{x};\psi) = \rho(\frac{1}{3}U_{\rho,1,n}(\vec{x},\psi) + \frac{1}{3}U_{\rho,2,n-1}(\vec{x},\psi) + \frac{1}{6})$$

が成り立つ。

証明. 明らかに、$U_{\rho,1,n}(\vec{x},\psi)=\rho(\psi(x_n))$, $\vec{x}\in\hat{M}$, が成立する。よって、$U_{\rho,1,n}(\cdot,\psi):\hat{M}\to\mathbf{R}$, $n\in\mathbf{Z}_{\leqq 0}$, は $\hat{M}$ 上連続である。

$0\leqq U^{(m)}_{\rho,i,n}(\vec{x},\psi)\leqq 1$, $\vec{x}\in\hat{M}$, $i=1,2$, $n,m\in\mathbf{Z}_{\leqq 0}$ であり、$0\leqq \rho‘(t)\leqq 1$, $t\in(0,1)$, であるので、$n,m,r\in\mathbf{Z}$ が $r\leqq m\leqq n-1$ を満たすならば

$$\begin{aligned}
&|U^{(m)}_{\rho,2,n}(\vec{x},\psi) - U^{(r)}_{\rho,2,n}(\vec{x},\psi)| \\
&\leqq |\rho(\frac{1}{3}\rho(\psi(x_n)) + \frac{1}{3}U^{(m)}_{\rho,2,n-1}(\vec{x}) + \frac{1}{6}) - \rho(\frac{1}{3}\rho(\psi(x_n)) + \frac{1}{3}U^{(r)}_{\rho,2,n-1}(\vec{x}) + \frac{1}{6})| \\
&\leqq \frac{1}{3}|U^{(m)}_{\rho,2,n-1}(\vec{x},\psi) - U^{(r)}_{\rho,2,n-1}(\vec{x},\psi)|
\end{aligned}$$

が成立する。

$$|U^{(m)}_{\rho,2,m}(\vec{x},\psi) - U^{(r)}_{\rho,2,m}(\vec{x},\psi)| \leqq 1$$

であるので、$n,m,r\in\mathbf{Z}_{\leqq 0}$ が $r\leqq m\leqq n$ を満たすならば

$$|U^{(m)}_{\rho,2,n}(\vec{x},\psi) - U^{(r)}_{\rho,2,n}(\vec{x},\psi)| \leqq 3^{m-n}, \qquad \vec{x}\in\hat{M}$$

が成り立つ。

これより主張を得る。

∎

命題 13 C_0 は $C(M)$ の部分集合で、任意の異なる $x,y\in M$ に対して $\varphi(x)\neq\varphi(y)$ を満たす $\varphi\in C_0$ が存在すると仮定する。この時、任意の異なる $\vec{x},\vec{y}\in\hat{M}$ に対して

$$U_{\rho,2,0}(\hat{x},a\varphi+b) \neq U_{\rho,2,0}(\hat{y},a\varphi+b)$$

となる $a,b\in\mathbf{R}$ および $\varphi\in C_0$ が存在する。

証明. $\vec{x}=\{x_n\}_{n\in\mathbf{Z}_{0\leqq 0}}, \vec{y}=\{y_n\}_{n\in\mathbf{Z}_{\leqq 0}}\in\hat{M}$ であり、$\vec{x}\neq\vec{y}$ とする、$m=\sup\{n\in\mathbf{Z}_{\leqq 0};\ x_n\neq y_n\}$ とおく。この時、 $x_n=y_n$, $n>m$, であるので、任意の $\psi\in C(M)$ に対して $U_{\rho,1,n}(\vec{x},\psi)=U_{\rho,1,n}(\vec{y},\psi)$, $n>m$, となる。$x_m\neq y_m$ であるので $\varphi\in C_0$ で $\varphi(x_m)\neq\varphi(y_m)$ となるものが存在する。

$$b_0 = \frac{1}{2}(\varphi(x_m)+\varphi(y_m)), \qquad \varepsilon = \mathrm{sign}(\varphi(x_m)-\varphi(y_m)),$$

とおくと

$$\lim_{t\to\infty}\rho(t\varepsilon(\varphi(x_m)-b_0))=1,\ \text{及び}\ \ \lim_{t\to\infty}\rho(t\varepsilon(\varphi(y_m)-b_0))=0$$

となることがわかる。従って

$$0\leqq\rho(a\varphi(y_m)+b)<1/8<7/8<\rho(a\varphi(x_m)+b)\leqq 1$$

となる $a\neq 0,\, b\in\mathbf{R}$ が存在する。よって、

$$0\leqq U_{\rho,1,m}(\vec{y};a\varphi+b)<1/8<7/8<U_{\rho,1,m}(\vec{x};a\varphi+b)\leqq 1$$

となる。仮定より $0<s\leqq t<1$ ならば

$$\rho(t)-\rho(s)=\int_s^t\rho'(u)du\in[c_0(t-s),(t-s)]$$

であるので

$$c_0|t-s|\leqq|\rho(t)-\rho(s)|\leqq|t-s|,\qquad s,t\in(0,1)$$

かつ

$$|\rho(t)-\rho(s)|\leqq|t-s|\leqq\frac{2}{3},\qquad s,t\in[\frac{1}{6},\frac{5}{6}]$$

となることがわかる。

任意の $n\in\mathbf{Z}_{\leqq 0},\,\psi\in C(M),$ に対して

$$|U_{\rho,2,n}(\vec{x};a\varphi+b)-U_{\rho,2,n}(\vec{y};a\varphi+b)|$$

$$=|\rho(\frac{1}{3}U_{\rho,1,n}(\vec{x},\psi)+\frac{1}{3}U_{\rho,2,n-1}(\vec{x},\psi)+\frac{1}{6})-\rho(\frac{1}{3}U_{\rho,1,n}(\vec{y},\psi)+\frac{1}{3}U_{\rho,2,n-1}(\vec{y},\psi)+\frac{1}{6})|$$

であるので、

$$|U_{\rho,2,n}(\vec{x};\psi)-U_{\rho,2,n}(\vec{y};\psi)|\geqq\frac{c_0}{3}|(U_{\rho,1,n}(\vec{x},\psi)-U_{\rho,1,n}(\vec{y},\psi))+(U_{\rho,2,n-1}(\vec{x},\psi)-U_{\rho,2,n-1}(\vec{y},\psi))|,$$

かつ

$$|U_{\rho,2,n}(\vec{x};\psi)-U_{\rho,2,n}(\vec{y};\psi)|\leqq\frac{2}{3}$$

となることがわかる。よって、

$$|U_{\rho,2,0}(\vec{x};a\varphi+b)-U_{\rho,2,0}(\vec{y};a\varphi+b)|$$

$$\geqq(\frac{c_0}{3})^{-m}|(U_{\rho,1,m}(\vec{x},a\varphi+b)-U_{\rho,1,m}(\vec{y},a\varphi+b))+(U_{\rho,2,m-1}(\vec{x},a\varphi+b)-U_{\rho,2,m-1}(\vec{y},a\varphi+b))|$$

$$\geqq(\frac{c_0}{3})^{-m}(\frac{3}{4}-\frac{2}{3})>0$$

となる。

よって命題が示された。

∎

上記の命題、及び定理 11 より以下のことがわかる。

定理 14 V_0 は $C(M)$ の部分ベクトル空間で以下の仮定を満たすとする。
(i) $1\in V_0$.
(ii) 任意の異なる $x,y\in M$ に対して $\varphi(x)\neq\varphi(y)$ を満たす $\varphi\in V_0$ が存在する。

今、$C(\hat{M})$ の部分集合 C を

$$C=\{U_{\rho,2,0}(\cdot,\psi);\ \psi\in V_0\}$$

で定める。この時、任意の $(\hat{M},\mathcal{B}(\hat{M}))$ 上の確率測度 ν 及び任意のボレル可測関数 $f:\hat{M}\to\mathbf{R}$ に対して以下が成立する。
(1) $\inf\{\int_{\hat{M}} \min\{|f(\vec{x})-g(\vec{x})|,1\}\nu(d\vec{x});\ g\in\mathcal{C}_{\rho,4}(C)\}=0$ が成り立つ。
(2) $p\in[1,\infty)$ とする。f が

$$\int_{\hat{M}}|f(\vec{x})|^p\nu(d\vec{x})<\infty$$

を満たすならば、

$$\inf\{\int_{\hat{M}} |f(\vec{x})-g(\vec{x})|^p\nu(d\vec{x});\ g\in\mathcal{C}_{\rho,4}(C)\}=0$$

が成り立つ。

系 15 V_0 は $C(M)$ の部分ベクトル空間で以下の仮定を満たすとする。
(i) $1\in V_0$.
(ii) 任意の異なる $x,y\in M$ に対して $\varphi(x)\neq\varphi(y)$ を満たす $\varphi\in V_0$ が存在する。
　今、$C(\hat{M})$ の部分集合 C を

$$C=\{U_{\rho,2,n}(\cdot,\psi);\ \psi\in V_0\}$$

で定める。この時、任意の $(\hat{M},\mathcal{B}(\hat{M}))$ 上の確率測度 ν 及び

$$\int_{\hat{M}}|f(\vec{x})|^2\nu(d\vec{x})<\infty$$

を満たすボレル可測関数 $f:\hat{M}\to\mathbf{R}$ に対して以下が成立する。

　任意の $\delta>0$ に対して $N\geqq 1$, $\psi_i\in V_0$, $i=1,\ldots,N$, 及び $F\in\mathbf{F}_{\rho,4}(\mathbf{R}^N;\mathbf{R})$ が存在して

$$\overline{\lim_{m\to\infty}}\int_{\hat{M}} |f(\vec{x})-F(U^{(m)}_{\rho,2,0}(\vec{x},\psi_1),\ldots,U^{(m)}_{\rho,2,0}(\vec{x},\psi_N))|^2\nu(d\vec{x})<\delta$$

となる。

証明. 定理 14 より任意の $\delta>0$ に対して $N\geqq 1$, $\psi_i\in V_0$, $i=1,\ldots,N$, 及び $F\in\mathbf{F}_{\rho,4}(\mathbf{R}^N;\mathbf{R})$ が存在して

$$\overline{\lim_{m\to\infty}}\int_{\hat{M}} |f(\vec{x})-F(U_{\rho,2,0}(\vec{x},\psi_1)\ldots,U_{\rho,2,0}(\vec{x},\psi_N))|^2\nu(d\vec{x})<\delta$$

が成り立つ。仮定より $F:\mathbf{R}^N\to\mathbf{R}$ は有界な連続関数であるので、命題 12 より

$$\lim_{m\to\infty}\int_{\hat{M}} |F(U_{\rho,2,0}(\vec{x},\psi_1)\ldots,U_{\rho,2,0}(\vec{x},\psi_N))-F(U^{(m)}_{\rho,2,0}(\vec{x},\psi_1)\ldots,U^{(m)}_{\rho,2,0}(\vec{x},\psi_N))|^2\nu(d\vec{x})=0$$

となるので、主張を得る。 ∎

系 16 V_0 は $C(M)$ の部分ベクトル空間で以下の仮定を満たすとする。
(i) $1\in V_0$.
(ii) 任意の異なる $x,y\in M$ に対して $\varphi(x)\neq\varphi(y)$ を満たす $\varphi\in V_0$ が存在する。
　この時、$C(\hat{M})$ の部分集合 C を

$$C=\{U_{\rho,2,n}(\cdot,\psi);\ \psi\in V_0\}$$

で定める。
　今、 $\{X_k\}_{k\in\mathbf{Z}}$ は確率空間 $(\Omega,P,\mathcal{F})$ 上定義された M-値定常過程とし、ボレル可測関数 $f:M^{\mathbf{Z}_{\leqq 0}}\to\mathbf{R}$ は

$$E[|f(\{X_k\}_{k\in\mathbf{Z}_{\leqq 0}})|^2]<\infty$$

を満たすものとする。この時、任意の $\delta > 0$ に対して $N \geqq 1$, $\psi_i \in V_0$, $i = 1, \ldots, N$, 及び $F \in \mathbf{F}_{\rho,4}(\mathbf{R}^N; \mathbf{R})$ が存在して

$$\overline{\lim_{m\to\infty}} E[|f(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}) - F(U^{(m)}_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_1), \ldots, U^{(m)}_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_N))|^2] < \delta$$

となる。

さらに、もし定常過程 $\{X_k\}_{k\in\mathbf{Z}}$ がエルゴード的であるならば任意の $\delta > 0$ に対して $N \geqq 1$, $\psi_i \in V_0$, $i = 1, \ldots, N$, 及び $F \in \mathbf{F}_{\rho,4}(\mathbf{R}^N; \mathbf{R})$ が存在して、確率 1 で

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{m=1}^{n} |f(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}) - F(U^{(m)}_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_1), \ldots, U^{(m)}_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_N))|^2 < \delta$$

となる。

証明. $\{X_k\}_{k\in\mathbf{Z}_{\leqq 0}}$ の $\hat{M} = M^{\mathbf{Z}_{\leqq 0}}$ 上に与える確率法則を ν とする。$\{X_k\}_{k\in\mathbf{Z}}$ は定常過程であるので、任意の $N \geqq 1$, $\psi_i \in V_0$, $i = 1, \ldots, N$, 及び $F \in \mathbf{F}_{\rho,4}(\mathbf{R}^N; \mathbf{R})$ に対して

$$\begin{aligned} &E[|f(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}) - F(U_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_1), \ldots, U_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_N))|^2] \\ &= \int_{\hat{M}} |f(\vec{x}) - F(U_{\rho,2,0}(\vec{x}, \psi_1), \ldots, U_{\rho,2,0}(\vec{x}, \psi_N))|^2 \nu(d\vec{x}) \end{aligned}$$

かつ

$$\begin{aligned} &E[|f(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}})] - F(U^{(m)}_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_1), \ldots, U^{(m)}_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq}0}, \psi_N))|^2] \\ &= \int_{\hat{M}} |f(\vec{x}) - F(U^{(m)}_{\rho,2,0}(\vec{x}, \psi_1), \ldots, U^{(m)}_{\rho,2,0}(\vec{x}, \psi_N))|^2 \nu(d\vec{x}) \end{aligned}$$

となる。これと系 15 より前半の主張を得る。

もし、$\{X_k\}_{k\in\mathbf{Z}}$ がエルゴード的であれば、確率 1 で

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} |f(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}) - F(U_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_1), \ldots, U_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_N))|^2 \\ &= \int_{\hat{M}} |f(\vec{x}) - F(U_{\rho,2,0}(\vec{x}, \psi_1), \ldots, U_{\rho,2,0}(\vec{x}, \psi_N))|^2 \nu(d\vec{x}) \end{aligned}$$

が成立する。

$F : \mathbf{R}^N \to \mathbf{R}$ は連続であるので、$[0, 1]^N$ 上一様連続である。よって、命題 12 より

$$\begin{aligned} &\frac{1}{n} \sum_{m=1}^{n} |F(U_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_1), \ldots, U_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_N)) \\ &\qquad -F(U^{(m)}_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_1), \ldots, U^{(m)}_{\rho,2,0}(\{X_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_N))|^2 \\ &\qquad \to 0, \quad n \to \infty \end{aligned}$$

であるので、後半の主張を得る。 ∎

さて、定理 1 を証明しよう。

M がユークリッド空間であり、V_0 がそのユークリッド空間上のアファイン関数全体とする。$N \geqq 1$, $\psi_i \in V_0$, $i = 1, \ldots, N$, 及び $F \in \mathbf{F}_{\rho,4}(\mathbf{R}^N; \mathbf{R})$ に対して

$$F(U^{(m)}_{\rho,2,0}(\{x_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_1), \ldots, U^{(m)}_{\rho,2,0}(\{x_{m+k}\}_{k\in\mathbf{Z}_{\leqq 0}}, \psi_N))$$

は $x_1, \ldots, x_m$ の関数であり、第 4 節で与えた RNN となる。このことと系 16 より、定理の主張が示される。

参考文献

[1] Cotter, N.E., The Stone-Weierstrass theorem and its application to neural networks. IEEE Transactions on Neural Networks 1(1990), 290-295.

[2] Cybenko, G., Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals and Systems 2(1989), 304-314.

[3] Funahashi, K., On the approximate realization of continuous mappings by neural networks. Neural Networks, 2(1989), 183-192.

[4] Hornik, K., M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators. Neural Networks 2(1989), 55-67.

[5] Hornik, K., Approximation capabilities of multilayer feedforward networks. Neural Networks 4(1991), 251-257.

[6] Ito, Y., Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory. Neural Networks 4(1991), 385-394.

[7] Kreinovich, V.Y., Arbitrary nonlinearity is sufficient to represent all functions by neural networks: a theorem. Neural Networks 4(1991), 381-383.

[8] 岡谷貴之, 深層学習 (機械学習プロフェッショナルシリーズ). 講談社，2015

[9] Stone, M. H. , The generalized Weierstrass approximation theorem. Mathematics Magazine 21(1948) 167-184, 237-254.