大自由度認知機構における隠れ系の統計的役割について

渡辺澄夫
東京工業大学 精密工学研究所 認知機構研究分野
〒226-8503 横浜市 緑区 長津田 4259
swatanab@pi.titech.ac.jp

1 はじめに

近年、神経系をヒントに考案された図1の型の学習装置がパターン認識や時系列予測などに利用されている。この型は外部（入出力）と直接結合していない隠れ系を持つことが特徴であるが、隠れ系を持たない型と比べて原理的な違いを有しているのだろうか。この問題を考えるとき、例えば、次のような定式化が可能である。

汎関数確率測度 μ からランダムに得られた関数 g を学習装置 f_N によって近似する。

$$g(x) \approx f_N(a,b,x) = \sum_{j=1}^{N} a_j \, \varphi(b_j \, ; \, x) \qquad (1)$$

このとき、次の二つの学習法がある。
(1) 固定隠れ系。$b = \{b_j\}$ は μ により定めておき、各 g に対しては $a = \{a_j\}$ だけを最適化する。
(2) 適応隠れ系。各 g に対して a, b の両方を最適化する。
本論では、次の問題を考える。

問題 上記2方法の平均学習精度（大自由度 $N \to \infty$ における $\|g - f_N\|$ の平均のオーダー）は μ の構造とどのような関係があるか？特に、適応隠れ系は固定隠れ系よりも優れているか？

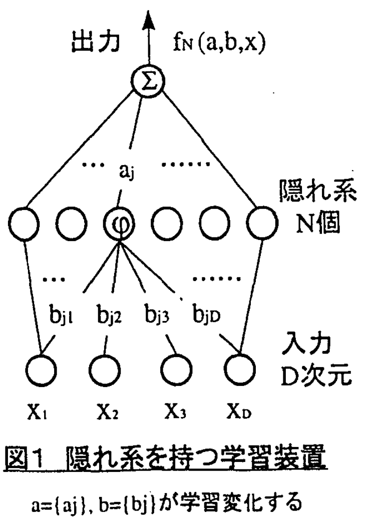

図1 隠れ系を持つ学習装置

a={aj}, b={bj}が学習変化する

2 関数空間上の確率測度と近似精度

D 次元ユークリッド空間内の集合 $[0,1)^D$ 上の複素数値を取る二乗可積分関数全体の作る空間 $H = L^2([0,1)^D)$ を考える。この空間は内積

$$(f,g) = \int_{[0,1)^D} \overline{f(x)} g(x) dx$$

によりヒルベルト空間と見なすことができる。この内積より導かれるノルムを $\|\cdot\|$ で表す（$\|f\| = (f,f)^{1/2}$）。$\{e_j(x); j = 0, \pm 1, \pm 2, \pm 3, \cdots\}$ をヒルベルト空間 $L^2([0,1)^1)$ の完全正規直交系とするとき、H の完全正規直交系を

$$\begin{aligned} n &= (n_1, n_2, \cdots, n_D) \in Z^D, \\ x &= (x_1, x_2, ..., x_D) \in [0,1)^D \end{aligned}$$

に対して（Z^D は D 次元格子の集合）、

$$e(n,x) = e_{n_1}(x_1) \, e_{n_2}(x_2) \, \cdots \, e_{n_D}(x_D)$$

によって作る。関数 g をこの直交系で展開した係数を g_b で表す。すなわち

$$g(x) = \sum_{b \in Z^D} g_b \, e(b,x), \qquad g_b = (e(b,\cdot), g)$$

本論文では、目的関数を次の関数系 $f_N(a,b,x)$ で近似する問題

$$f_N(a,b,x) = \sum_{j=1}^{N} a_j \, e(b_j, x) \quad (a_j \in \mathbf{C}, b_j \in Z^D) \qquad (2)$$

を考える。

定理 1 (Makovoz, Jones, Barron) 目的関数 $g(x)$ が条件

$$M(g) \equiv \sum_{m \in Z^D} |g_m| < \infty$$

を満たすとする。このとき、十分大きい N について

$$\inf_{a,b} \|g - f_N(a,b,\cdot)\| \leq \frac{M(g)}{N^{1/2}}$$

が成立する。

定理 2 目的関数 $g(x)$ が条件

$$L_\alpha(g) \equiv \{ \sum_{m \in Z^D} |g_m|^2 |m|^{2\alpha} \}^{1/2} < \infty$$

を満たすとする。この時、b_j $(j=1,2,...,)$ を $|b_j|$ の小さな順に定めると（g と無関係に原点に近い方から順序づけると）、十分大きい N について

$$\inf_a \|g - f_N(a,b,\cdot)\| \leq \frac{L_\alpha(g)}{N^{\alpha/D}}$$

が成立する。また $L_\alpha(g) < \infty$ を満たす $g(x)$ であって、任意の $\epsilon > 0$ について

$$\inf_{a,b} \|g - f_N(a,b,\cdot)\| \geq \frac{L_\alpha(g)}{N^{(\alpha/D)+\epsilon}}$$

を満たす $g(x)$ が存在する。

定理 1 と 2 の主張する所は、関数近似オーダーが基礎空間の次元 D に影響されるかどうかについては、対象関数 $g(x)$ の性質次第だということである。さて、情報環境を取り扱うために、関数空間 $H = L^2([0,1)^D)$ 上の確率測度 μ を考える。μ はどのような関数が現れやすいかを表現するものである。以下、μ のタイプと関数近似法との関係を調べよう。μ による汎関数 $F(g)$ $(g \in H)$ の平均を $E_g\{F(g)\}$ と書くことにする。次の二つの近似効率を考える。

$$Fix(N) = \inf_b E_g\{\inf_a \|g - f_N(a,b,\cdot)\|\}$$

$$Adap(N) = E_g\{\inf_b \inf_a \|g - f_N(a,b,\cdot)\|\}$$

$Fix(N)$ は固定隠れ系による近似精度を、$Adap(N)$ は適応隠れ系による近似精度をそれぞれ表している。定義からすぐに次のことが導かれる。

(1) $Fix(N) \geq Adap(N)$。

(2) $E_g\{\|g\|^2\} < \infty$ ならば、

$N \to \infty$ のとき $Fix(N) \to 0, Adap(N) \to 0.$

定義 1 確率測度 μ が独立等方的でオーダー α を持つとは、$\{g_n(n \in Z^D)\}$ が独立であり、$|g_n|$ の分布関数を $F_n(x)$ とするとき、ある絶対連続な分布関数 $G(x)$ が存在して次の 2 条件を満たすことである。
(1) ある定数 $\lambda > 0$ が存在して、十分大きい x について

$$G(x) \geq 1 - \exp(-\lambda x)$$

(2) ある定数 $A, B > 0$ が存在して

$$G(\frac{(1+|n|^\alpha)x}{A}) \leq F_n(x) \leq G(\frac{(1+|n|^\alpha)x}{B})$$

定理 3 確率測度 μ を独立等方的な測度としオーダー α を持つとする。$E_g\{\|g\|^2\} < \infty$ を仮定する。このとき、$2\alpha - D > 0$ であり、近似のオーダーについて次が成立する。

$$\begin{aligned} Fix(N) &= \frac{Const.}{N^{(2\alpha-D)/2D}} \\ Adap(N) &\geq \frac{Const.}{N^{(2\alpha-D)/2D}(\log N)^{(2\alpha-D)/2\alpha}} \end{aligned}$$

例 1（Wiener 型の測度） 次の確率測度を考える。

$$\frac{d\mu}{dg} \propto \exp(-(g,(1+\Delta^{2\alpha})g))$$

フーリエ展開を考えると、複素係数 $\{\hat{g}_n\}$ は独立であり、定義 1 を満たすことが確かめられる。従って定理 3 より、適応的な近似系選択はそれほど有効でないことがわかる。

定理 4 ある定数 $\alpha, A > 0$ が存在して

$$E_g\{|g_n|\} \leq \frac{A}{1+|n|^{D+\alpha}}$$

が成り立つとする。このとき

$$Adap(N) \leq \frac{Const.}{N^{1/2}}$$

例 2（競合型の相互作用） $\{Z^D\}$ を次のように分解する。

$$\begin{aligned} Z^D &= \bigcup_{r=0}^{\infty} S_r \\ S_r &= \{n \in Z^D; r \leq |n| < r+1\} \end{aligned}$$

また測度 μ を次のように構成する。

(1) $\{g_n; n \in S_r\}$ と $\{g_n; n \in S_{r'}\}$ は $r \neq r'$ であれば独立。
(2) 各 $\{g_n; n \in S_r\}$ では、等確率でひとつの g_n が選ばれて、選ばれた $|g_n|$ は平均 $\frac{A}{1+r^{1+\epsilon}}$ $(\epsilon > 0)$ を持つ確率変数 ξ_r に一致する。残りの g_n は 0。

この測度によって

$$E_g\{|g_n|\} = \frac{1}{|n|^{D-1}} \cdot \frac{A}{|n|^{1+\epsilon}} = \frac{A}{|n|^{D+\epsilon}}$$

従って、定理 4 より

$$Adap(N) \leq \frac{Const.}{N^{1/2}}$$

がわかる。一方、

$$
\begin{aligned}
Fix(N) &= E_g\{(\sum_{|n|>N^{1/D}} |g_n|^2)^{1/2}\} \\
&\geq \frac{1}{N^{(1+2\epsilon)/2D}}
\end{aligned}
$$

従って、適応隠れ系が有効である。

定理 5 ある $\alpha \geq 0, \epsilon > 0$ が存在して、

$$E_g\{L_\alpha(g)^2\} < \infty$$

であり、かつ μ が条件

$$\mu(\{g \in H; 0 < |g_n| < \frac{\epsilon}{1+|n|^\alpha} \ (n \in Z^D)\}) = 0$$

を満たすとする。このとき、

$$Adap(N) \leq \frac{Const.}{N^{1/2}}$$

例 3 展開係数 g_n が離散値 $0, \pm\epsilon, \pm 2\epsilon, \pm 3\epsilon, ...,$ を取る場合を考えよう。これは例えば、得られる情報が、フーリエ係数で表され量子化されている場合に相当する。また μ として次のものを考える。

(1) $\{g_n (n \in Z^D)\}$ は独立。
(2) $|g_n| = k\epsilon \ (k > 0)$ となる確率は

$$Pr\{|g_n| = k\epsilon\} = \frac{1}{(1+|n|)^\delta}$$

このとき、

$$Fix(N) \geq \frac{1}{N^{(\delta-D)/D}}$$

であり、一方、定理 5 において、$\alpha = 0$ の場合を考えれば、(2) の条件から $|n| \geq 1$ のとき

$$E_g\{|g_n|^2\} = \sum_{k=1}^{\infty} \frac{k^2\epsilon^2}{|n|^{\delta k}} \approx \frac{\epsilon^2}{|n|^\delta}$$

従って $E_g\{L_0(g)^2\} < \infty$ が得られて定理 5 の仮定が満たされるので、

$$Adap(N) = \frac{Const.}{N^{1/2}}$$

となる。すなわち、この場合は $D < \delta < 2D$ であれば、適応隠れ系選択が固定の場合に比べて有効である。一般に展開係数が離散値を取る場合には、適応隠れ系が有効であると考えて良い。

3 考察

定理 2 のように、基礎空間の次元 D が大きくなると関数近似精度が指数のオーダーで悪化することは、人工知能の研究者により「次元の呪い」と呼ばれ、高次元空間上の関数を学習によって獲得する場合の原理的な困難と考えられてきた。これに対して、定理 1 の結果は「次元の呪いが階層モデルにより解決される」ように見えるため、適応隠れ系を持つモデルの有用性としてしばしば言及されている。

しかしながら、本論で示したように、適応隠れ系を持つモデルが有効であるかどうかは、そのモデルが置かれた情報環境の構造に本質的に依存している。Wiener 型の測度は場の理論や確率積分では中心的な役割を果たし、汎関数測度の代表的な例であるが、そのような測度で記述される環境では隠れ系が適応的であっても学習精度はほとんど向上しない。自由場に近い環境では適応的な認知機構は積極的な意義を有しないのである。

一方、エネルギーが有限で強い相互作用を持つような環境（エネルギー有限のため競合型になりやすい）では、隠れ系が適応的な学習装置は固定的な装置よりも遥かに有効である。今後、学習装置が利用される様々な情報環境（画像、音声、人間など）が実際にどのような確率測度に従っているかを推定し、実世界における認知機構の構成への基礎としたい。

4 結論

汎関数測度を用いて、隠れ系を持つ学習装置の精度解析を行い、情報環境と近似精度との関係を考察した。階層型の学習装置は、自由場に近い環境では積極的な意義を持たないが、強い競合的な相互作用のある環境では単層型の装置よりも有効である。

5 謝辞

本研究は文部省科研費 09680362 により行われた。

文献

本論の定理などの証明は次をご覧下さい。

[1] S. Watanabe, *Proc. of Int. Conf. on Approximation Theory*, 1998.

[2] 渡辺澄夫, 信学技報, Vol. NC98, No. 3, 1998.

なお適応近似系は統計モデルとして非正則であり、統計的推定でも固定近似系と異なる性質を持っている。

[3] S.Watanabe, *Proc. of Int. Conf. on Information Science*, Vol.2, pp.149-152, 1997.