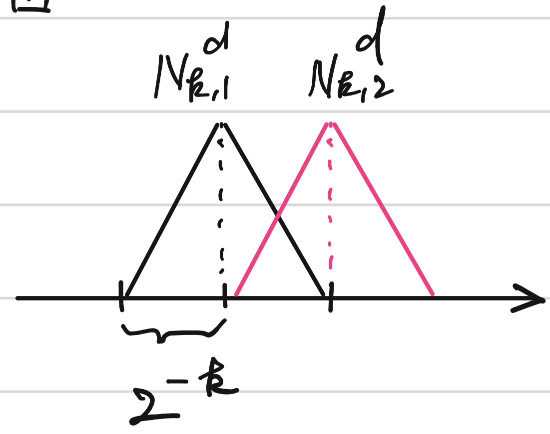
Besov空間の近似理論

B-spline基底: $N_m(x) = (\underbrace{N * \cdots * N}_{m+1\,回})(x)$

$$M^d_{k,j}(x) = \prod_{i=1}^{d} N_m(2^k x_i - j_i)$$



$(k \in \mathbb{N},\ j_i \in \mathbb{N})$

Lem (DeVore & Popov (1988))

$0 < p \le \infty$, $0 < q \le \infty$, $0 < s < \min(m, m-1+\frac{1}{p})$ とする.

$$f \in B^s_{p,q} \iff f = \sum_{k=0}^{\infty} p_k(f)$$

ただし,

$$\begin{cases} p_k(f) = \sum_{j \in J(k)} \alpha_{k,j} M_{k,j} \\ J(k) = \{ j \in \mathbb{Z}^d \mid -m \le j_i \le 2^k \} \end{cases}$$

かつ,

$$\left(\sum_{k=0}^{\infty} \left(2^{sk} \| p_k \|_{L^p} \right)^q \right)^{\frac{1}{q}} < \infty$$

が成り立つ.

実は, $\|f\|_{B^s_{p,q}} \simeq \left(\sum_{k=0}^{\infty} (2^{sk} \|p_k\|_{L^p})^q \right)^{\frac{1}{q}}$ も成り立つ.

$\|p_k\|_{L^p} \simeq \left(2^{-kd} \sum_{j \in J(k)} |\alpha_{k,j}|^p \right)^{\frac{1}{p}}$ も成り立つ.

Thm

先のLemと同じ条件かつ

$$s > d\left(\frac{1}{p} - \frac{1}{r}\right)_+$$

なら、任意の $N \gg 1$ に対し、

$$\|f - f_N\|_{L^r} \lesssim N^{-\frac{s}{d}} \|f\|_{B^s_{p,q}}$$

が以下の形をした f_N に対して成り立つ:

$$f_N(x) = \sum_{k=0}^{K} \sum_{j \in J(k)} \alpha_{k,j} M^d_{k,j}(x) + \sum_{k=K+1}^{K^*} \sum_{i=1}^{n_k} \alpha_{k,j(i)} M^d_{k,j(i)}(x)$$

ただし

$(j(i))_{i=1}^{n_k} \subset J(k)$,

$K = \lceil C_1 \log(N)/d \rceil$

$K^* = \lceil \log(\lambda N) \nu^{-1} \rceil + K + 1$

$n_k = \lceil \lambda N 2^{-\nu(k-N)} \rceil \quad (k = K+1, \dots, K^*)$

for

$\delta = d\left(\frac{1}{p} - \frac{1}{r}\right)_+$ かつ $\nu = (s-\delta)/2\delta$,

$C_1, \nu > 0$ は定数。

//

※ ・ $\frac{1}{p} \leq \frac{1}{r}$ (pが大きい) なら、$K = K^*$ と考えて良い。

つまり、低周波成分から順に取って来れば良い。

・ $\frac{1}{p} > \frac{1}{r}$ (pが小さい) なら、$K < k \leq K^*$ なる k においては

基底を置く"場所"も工夫する必要がある。

$\Longrightarrow$ スパース性

高周波成分が必要な場所に適応的に配置する。

(略証) $p \geq r$ の場合のみ考える.

$$R_K(f) = \sum_{k=0}^{K} p_k$$ とおく. (上位K個の周波数分だけ取って来る.)

Lemma 5.3 of Dũng (2011) より,

$$\|f - R_K(f)\|_{L^r}^r \lesssim \sum_{k>K} \|p_k\|_{L^r}^r$$

が成り立つ.

$p > r$ かつ $\Omega = [0,1]^d$ より, $\|p_k\|_{L^r} \leq \|p_k\|_{L^p}$ である.

(∵ Jensenの不等式)

よって,

$$\|f - R_K(f)\|_{L^r}^r \lesssim \sum_{k>K} \|p_k\|_{L^p}^r$$

である.

(i) $q \leq r$ なら,

$$\begin{aligned}
\|f - R_K(f)\|_{L^r}^q &\lesssim \left(\sum_{k>K} \|p_k\|_{L^p}^r \right)^{\overset{\leq 1}{\frac{q}{r}}} \\
&\lesssim \sum_{k>K} \|p_k\|_{L^p}^q \\
&\lesssim \sum_{k>K} \left(\|p_k\|_{L^p} \cdot 2^{sk} \right)^q \cdot \overbrace{2^{-skq}}^{\leq N^{-\frac{s}{d} q}} \\
&\lesssim \|f\|_{B^s_{p,q}}^q \cdot N^{-\frac{s}{d} \cdot q} \qquad (\because \text{先のLemma})
\end{aligned}$$

//

(ii) $q > r$ なら、

$$\|f - R_K(f)\|_{L^r}^r \lesssim \sum_{k>K} \left(2^{sk}\|p_k\|_{L^p}\right)^r \cdot 2^{-skr}$$

$$\lesssim \sum_{k>K} \left(2^{sk}\|p_k\|_{L^p}\right)^r \cdot 2^{-s(k-K)r} \cdot \underbrace{2^{sKr}}_{\simeq N^{-\frac{s}{d}\cdot r}}$$

$$\le N^{-\frac{s}{d}\cdot r} \sum_{k>K} 2^{-s(k-K)r} \left(2^{sk}\|p_k\|_{L^p}\right)^{q\cdot\frac{r}{q}}$$

$$\lesssim N^{-\frac{s}{d}\cdot r} \left(\underbrace{\sum_{k>K} 2^{-s(k-K)r\cdot(1-\frac{r}{q})^{-1}}}_{<\infty}\right)^{1-\frac{r}{q}} \times \left(\sum_{k>K} \left(2^{sk}\|p_k\|_{L^p}\right)^q\right)^{\frac{r}{q}}$$

($\because$ Hölderの不等式)

$$\lesssim \left(N^{-\frac{s}{d}} \cdot \|f\|_{B^s_{p,q}}\right)^r$$

//

($p < r$ の状況は省略: Dũng (2011) を参照)

あとは、$M^d_{k,j}$ を深層NNで近似すれば良い。

Prop.

幅 $W_0 = O(dm)$, 深さ $L_0 = O(\log(\varepsilon^{-1}) \cdot c_{d,m})$

パラメータ数 $= O(L_0 W_0^2)$, 各パラメータの絶対値 $= O(m^m)$

なるDNN $\hat{\phi}$ が存在して、

$$\|M^d_{0,0} - \hat{\phi}\|_{L^\infty} \le \varepsilon.$$

//

あとは、$M^d_{\bar{\alpha}, j}$ を深層NNで近似すれば良い。

Prop.

幅 $W_0 = O(dm)$, 深さ $L_0 = O(\log(\varepsilon^{-1}) \cdot c_{d,m})$

パラメータ数 $= O(L_0 W_0^2)$, 各パラメータの絶対値 $= O(m^m)$

なるDNN $\hat{\phi}$ が存在して、

$$\| M^d_{0,0} - \hat{\phi} \|_{L^\infty} \leq \varepsilon .$$

//

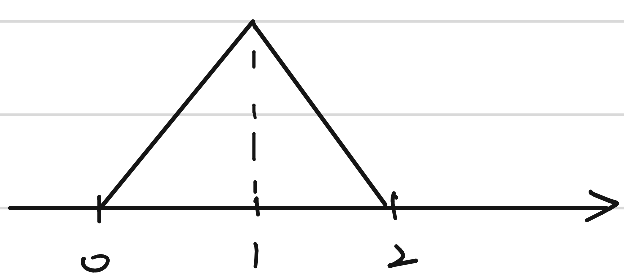
(略証)

$$N_m(x) = \frac{1}{m!} \sum_{j=0}^{m+1} (-1)^j \binom{m+1}{j} (x-j)_+^m$$

になることが知られている。((4.28) of Mhaskar & Micchelli (1992))

($m=1$ でチェック:

$$N_1(x) = x_+ - 2(x-1)_+ + (x-2)_+$$



← 合っている。)

- $(x-j)_+$ は ReLU を用いたニューロン1つで表現できる
- $(\cdot)^m$ は ReLU DNN を用いて効率的に近似できる。($\log\left(\frac{1}{\varepsilon}\right)$ くらいの深さ)

$\Longrightarrow$ $(x-j)_+^m$ は作れる。 $\Longrightarrow$ これらの線形結合をとる。

← これもNNで実現可

$\Longrightarrow$ $N_m(x)$ を作れる.

$\Longrightarrow$ $M_{0,0}^d(x) = \prod_{j=1}^{d} N_m(x_j)$ も NN で表現可.

※実際は $M_{0,0}^d$ のサポートの外 ($x \notin [0, m+1]^d$) で 0 になるように NN を構成できる.

◦ 線形推定量の Minimax リスクの下限.

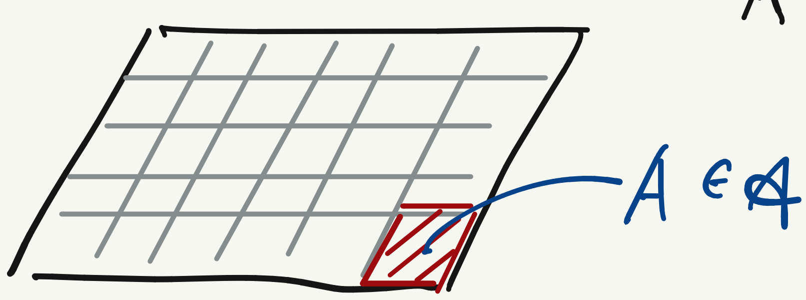
$\Omega = [0,1]^d$

$\mathcal{F}$: $\Omega \to \mathbb{R}$ なる関数の集合（真の関数の集合）

$$\boxed{R^* := \inf_{\hat{f}:\text{線形}} \sup_{f^\circ \in \mathcal{F}} E_{D^n}\left[\|\hat{f} - f^\circ\|^2_{L^2(P_X)}\right]}$$

$\mathcal{A}$: Ω の等分割 s.t. $|\mathcal{A}| = 2^K$

$\forall A \in \mathcal{A}, \ \mu(A) = 2^{-K}$.



Condition A

- $\exists r_1, r_2 > 0$ s.t. $n^{-r_1} \le 2^{-K} \le 2^{r_2}$

n個のデータがほぼ等分割されてる

- 事象 $\mathcal{E}$ s.t. $\begin{cases} \cdot \ \left|\{x_i \mid x_i \in A \ (i=1,\dots,n)\}\right| \le C' \dfrac{n}{2^K} \\ \cdot \ P(\mathcal{E}) \ge 1 + o(1) \end{cases}$ ← 事象の確率は十分高い

Condition B : $\Delta > 0$ は以下をみたす:

- $\exists F > 0$ s.t. $\forall A \in \mathcal{A}$ に対し $g_A \in \mathcal{F}$ が存在し.

$$g_A(x) \ge \frac{1}{2}\Delta F$$

- $\exists K', C'' > 0$ s.t. $\frac{1}{n}\sum_{i=1}^{n} g(x_i)^2 \le C'' \Delta^2 2^{-K'}$

$(\forall g \in \mathcal{F})$

Condition A, B のもと、ある定数 C が存在し、
以下の少なくともどちらかが成り立つ：

$$\frac{F^2}{4CC''} \frac{2^{k'}}{n} \leq R^*$$

$$\frac{F^3}{32} \cdot \Delta^2 2^{-k} \leq R^*$$

---

① k' について、Condition A, B より、

$$\frac{1}{4} F^2 \cdot \Delta^2 \cdot 2^{-k} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)^2 \leq C'' \Delta^2 2^{-k'}$$

$\Longrightarrow \quad k' \lesssim k$ である。

$k' = k$ が一番良い。

② ($k' = k$ とすると)

$$\min\left(\frac{2^k}{n}, \Delta^2 2^{-k}\right) \lesssim R^*$$

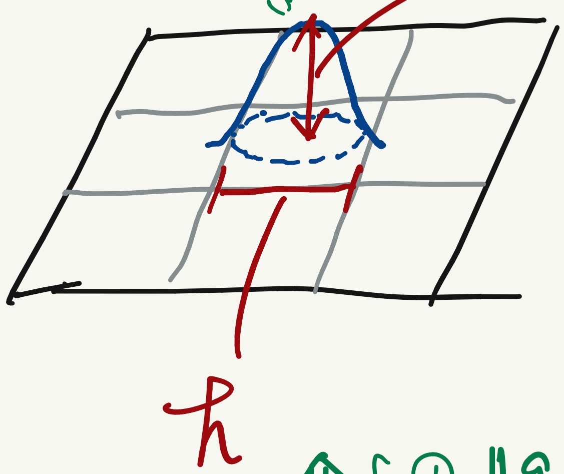
なので、左辺を最大にする k を持って来たい。

ポイント：2^k（分割の数）と Δ（高さ）のトレードオフ

典型例： Besov 空間 ($B^{\beta}_{p,q}([0,1]^d)$)

$g = \Delta \cdot M^{d}_{k,j}(x)$

$\Delta = h^{\beta - \frac{d}{p}}$



基底の場所の数は h^{-d} 個.

$$\underline{2^{k} = h^{-d}}$$

$\Downarrow$

$\frac{2^{k}}{n} = \Delta^{2} 2^{-k}$ は

$$\boxed{\frac{h^{-d}}{n} = h^{2(\beta - \frac{d}{p})} \cdot h^{d} \lesssim R^{*}}$$

$\Downarrow$

$$h = n^{-\frac{1}{2(\beta - \frac{d}{p} + d)}}$$

$\Rightarrow$

$$\underline{n^{-\frac{2(\beta - \frac{d}{p} + \frac{d}{2})}{2(\beta - \frac{d}{p} + \frac{d}{2}) + d}} \lesssim R^{*}}$$

線形推定量の推定誤差の下限.

確 ① $\|g\|_{L^{p}}$

$$= (\Delta^{p} h^{d})^{\frac{1}{p}} = h^{\beta}$$

② $\|g\|_{B^{\beta}_{p,q}}$

$$= 2^{k(\beta - \frac{d}{p})} \cdot h^{\beta - \frac{d}{p}}$$

$$\simeq 1 \quad (2^{-k} = h \text{ なる } k \text{ を選ぶ})$$