補足資料

joeyoji

2021 年 6 月 22 日

目次

1 発表資料の補足

1.1 (3.10)=(3.11) の補足

$P(\boldsymbol{\sigma}, S, S')$ を次のように定義する．

$$P(\boldsymbol{\sigma}, S, S') = \frac{1}{m}\sup_{g\in\mathcal{G}}\sum_{i=1}^{m}\sigma_i\left(g(z_i') - g(z_i)\right)$$

すると (3.10) $= \underset{S,S'}{\mathbb{E}}[P(\mathbf{1}, S, S')]$ は明らかに成り立つ．$\boldsymbol{\sigma} \in \{-1,+1\}^m$ を任意に固定する．ここで $S = (z_i), S' = (z_i')$ に対して $T = (w_i), T' = (w_i')$ を次のように定義する．

$$(w_i, w_i') = \begin{cases} (z_i, z_i') & (\sigma_i = +1) \\ (z_i', z_i) & (\sigma_i = -1) \end{cases}$$

と $P(\boldsymbol{\sigma}, S, S') = P(\mathbf{1}, T, T')$ と $P(\boldsymbol{\sigma}, T, T') = P(\mathbf{1}, S, S')$ が成り立つ．すなわち任意の $\boldsymbol{\sigma}$ は $S \times S' \to T \times T'$ 間で P について全単射を構成する．よって S, S' について平均をとっていれば独立性などから

$$\underset{S,S'}{\mathbb{E}}[P(\boldsymbol{\sigma}, S, S')] = \underset{T,T'}{\mathbb{E}}[P(\mathbf{1}, T, T')] = \underset{S,S'}{\mathbb{E}}[P(\mathbf{1}, S, S')]$$

が成り立つ. よって

$$(3.10) = \mathop{\mathbb{E}}_{S,S'}[P(\mathbf{1},S,S')] = \frac{1}{2^m}\sum_{\boldsymbol{\sigma}}\mathop{\mathbb{E}}_{S,S'}[P(\mathbf{1},S,S')] = \frac{1}{2^m}\sum_{\boldsymbol{\sigma}}\mathop{\mathbb{E}}_{S,S'}[P(\boldsymbol{\sigma},S,S')] = \mathop{\mathbb{E}}_{\boldsymbol{\sigma},S,S'}[P(\boldsymbol{\sigma},S,S')] = (3.11)$$

が導かれる.

1.2 同心球クラスの RC

m 歩ランダムウォークの最大位置が r である確率が $P(m,r) = \frac{1}{2^m}\binom{m}{\lfloor\frac{m-r}{2}\rfloor}$ であることの証明は [1] に記述がある. この確率の期待値を m で割ることで同心球クラスの RC を得る.

$$\mathfrak{R}_{2k+1}(\mathcal{G}) = \frac{(2k-1)!!}{(2k)!!} - \frac{1}{4k+2}\ (k=0,1,2,\ldots)$$
$$\mathfrak{R}_{2k}(\mathcal{G}) = \frac{(2k-1)!!}{(2k)!!} - \frac{1}{4k} + \frac{(2k-1)!!}{(2k)!!4k}\ (k=1,2,3,\ldots)$$

Proof. $P(m,r)$ が二項分布の確率質量関数の並び替えであることを利用して期待値を求める.$m=2k+1$ のときは

$$\begin{aligned}\sum_{r=0}^{2k+1} r\frac{1}{2^{2k+1}}\binom{2k+1}{\lfloor\frac{2k+1-r}{2}\rfloor} &= \sum_{r=0}^{k}(4k-4r+1)\frac{1}{2^{2k+1}}\binom{2k+1}{r}\\ &= \frac{4k+1}{2} - 4\sum_{r=0}^{k} r\frac{1}{2^{2k+1}}\binom{2k+1}{r}\\ &= \frac{4k+1}{2} - 2(2k+1)\sum_{r=0}^{k-1}\frac{1}{2^{2k}}\binom{2k}{r}\\ &= \frac{4k+1}{2} - (2k+1)\left(1-\frac{1}{2^{2k}}\binom{2k}{k}\right)\end{aligned}$$

よって

$$\begin{aligned}\mathfrak{R}_{2k+1}(\mathcal{G}) &= \frac{1}{2k+1}\left(\frac{4k+1}{2} - (2k+1)\left(1-\frac{1}{2^{2k}}\binom{2k}{k}\right)\right)\\ &= \frac{4k+1}{4k+2} - 1 + \frac{1}{2^{2k}}\binom{2k}{k}\\ &= -\frac{1}{4k+2} + \frac{(2k)!}{(2k)!!(2k)!!}\\ &= \frac{(2k-1)!!}{(2k)!!} - \frac{1}{4k+2}\end{aligned}$$

$m=2k$ のときは

$$\begin{aligned}\sum_{r=0}^{2k} r\frac{1}{2^{2k}}\binom{2k}{\lfloor\frac{2k-r}{2}\rfloor} &= \sum_{r=0}^{k-1}(4k-4r-1)\frac{1}{2^{2k}}\binom{2k}{r}\\ &= \frac{4k-1}{2}\left(1-\frac{1}{2^{2k}}\binom{2k}{k}\right)-4\sum_{r=0}^{k-1} r\frac{1}{2^{2k}}\binom{2k}{r}\\ &= \frac{4k-1}{2}\left(1-\frac{(2k-1)!!}{(2k)!!}\right)-4k\sum_{r=0}^{k-2}\frac{1}{2^{2k-1}}\binom{2k-1}{r}\\ &= \frac{4k-1}{2}\left(1-\frac{(2k-1)!!}{(2k)!!}\right)-2k\left(1-\frac{2}{2^{2k-1}}\binom{2k-1}{k-1}\right)\end{aligned}$$

よって

$$\begin{aligned}\mathfrak{R}_{2k}(\mathcal{G}) &= \frac{1}{2k}\left\{\frac{4k-1}{2}\left(1-\frac{(2k-1)!!}{(2k)!!}\right)-2k\left(1-\frac{2}{2^{2k-1}}\binom{2k-1}{k-1}\right)\right\}\\ &= -\frac{1}{4k}-\frac{(2k-1)!!}{(2k)!!}\left(1-\frac{1}{4k}\right)+\frac{2}{2^{2k-1}}\binom{2k-1}{k-1}\\ &= -\frac{1}{4k}-\frac{(2k-1)!!}{(2k)!!}\left(1-\frac{1}{4k}\right)+2\frac{(2k-1)!!}{(2k)!!}\\ &= \frac{(2k-1)!!}{(2k)!!}-\frac{1}{4k}+\frac{(2k-1)!!}{(2k)!!4k}\end{aligned}$$

□

1.3 GF による汎化誤差の上界

FML 本編では式 (3.23) が証明なしで紹介されているが, この証明が [2] でなされている. しかしそこでは次の不等式の証明が省かれているので, その証明を行う.

$$\Gamma = \sum_{k\ s.t.\ |2k/m-p/m|>\epsilon/2}\frac{\binom{p}{k}\binom{2m-p}{m-k}}{\binom{2m}{m}} \le 2\exp\left\{-\frac{m\epsilon^2}{8}\right\}$$

Proof. はじめに Γ を変形する.

$$\begin{aligned}\Gamma &= \sum_{k=0}^{p}\mathbb{I}\{|k-p/2|>m\epsilon/4\}\frac{\binom{p}{k}\binom{2m-p}{m-k}}{\binom{2m}{m}}\\ &= \sum_{k=0}^{p}\mathbb{I}\{k-p/2>m\epsilon/4\}\frac{\binom{p}{k}\binom{2m-p}{m-k}}{\binom{2m}{m}}+\sum_{k=0}^{p}\mathbb{I}\{k-p/2<-m\epsilon/4\}\frac{\binom{p}{k}\binom{2m-p}{m-k}}{\binom{2m}{m}}\\ &= \sum_{k=0}^{p}\mathbb{I}\{k-p/2>m\epsilon/4\}\frac{\binom{p}{k}\binom{2m-p}{m-k}}{\binom{2m}{m}}+\sum_{k=0}^{p}\mathbb{I}\{(p-k)-p/2>m\epsilon/4\}\frac{\binom{m}{k}\binom{m}{p-k}}{\binom{2m}{p}}\\ &= \sum_{k=0}^{p}\mathbb{I}\{k-p/2>m\epsilon/4\}\frac{\binom{p}{k}\binom{2m-p}{m-k}}{\binom{2m}{m}}+\sum_{k=0}^{p}\mathbb{I}\{k-p/2>m\epsilon/4\}\frac{\binom{m}{p-k}\binom{m}{k}}{\binom{2m}{p}}\\ &= 2\sum_{k=0}^{p}\mathbb{I}\{k-p/2>m\epsilon/4\}\frac{\binom{p}{k}\binom{2m-p}{m-k}}{\binom{2m}{m}}\end{aligned}$$

ここで $\binom{p}{k}\binom{2m-p}{m-k}/\binom{2m}{m}$ について考える．これは超幾何分布の確率質量関数 $HG(k \mid 2m, m, p)$ に一致しており, 特に $2m$ 個のうち p 個が当たりであるような球の中から無作為に m 個の球を同時に抽出したとき k 個が当たりである確率を表している．抽出する方法は非復元的に 1 つずつ取り出し合計 m 個取り出すでも同じである. この過程で i 番目に取り出した球が当たりであれば $X_i = 1$, そうでなければ $X_i = 0$ とすると,

$$\mathbb{P}(\sum_{i=1}^{m} X_i = k) = HG(k \mid 2m, m, p)$$

従って $\Gamma = 2\mathbb{E}[\mathbb{I}\{\sum_{i=1}^{m} X_i - p/2 > m\epsilon/4\}]$ となる. よって $\mathbb{E}[\mathbb{I}\{\sum_{i=1}^{m} X_i - p/2 > m\epsilon/4\}] \leq e^{-m\epsilon^2/8}$ が示せれば良い. これはヘフディングの不等式の証明の手順で可能である. 但し $(X_i)_{i=1}^{m}$ は独立でないことに注意する.$t > 0$ とする.

$$\mathbb{E}[\mathbb{I}\{\sum_{i=1}^{m} X_i - p/2 > m\epsilon/4\}] \leq \mathbb{E}[\exp\{t\sum_{i=1}^{m} X_i\}]\exp\{-tp/2 - tm\epsilon/4\}$$

ここで $E_j = \mathbb{E}[\exp\{t_j \sum_{i=j}^{m} X_i\}]$ とおき条件付き期待値を考えることで漸化式を立てる.$(t_j > 0)$

$$\begin{aligned}
E_j &= \mathbb{E}[\exp\{t_j \sum_{i=j+1}^{m} X_i\}\mathbb{E}[\exp\{t_j X_j\} \mid X_{j+1}, \ldots, X_n]] \\
&= \mathbb{E}[\exp\{t_j(\sum_{i=j+1}^{m} X_i + \mathbb{E}[X_j \mid X_{j+1}, \ldots, X_m])\mathbb{E}[\exp\{t_j(X_j - \mathbb{E}[X_j \mid X_{j+1}, \ldots, X_m])\} \mid X_{j+1}, \ldots, X_m]] \\
&\leq \mathbb{E}[\exp\{t_j(\sum_{i=j+1}^{m} X_i + \frac{p - \sum_{i=j+1}^{m} X_i}{2m-(m-j)})\}]\exp\{t_j^2/8\} \\
&= \mathbb{E}[\exp\{t_j \frac{m+j-1}{m+j}\sum_{i=j+1}^{m} X_i\}\exp\{t_j^2/8 + \frac{p}{m+j}t_j\}
\end{aligned}$$

よって $t_{j+1} = \frac{m+j-1}{m+j}t_j$ とおくと $E_j \leq E_{j+1}\exp\{t_j^2/8 + pt_j/(m+j)\}$ が成り立つ. これを再帰的に適応し $t_1 = t$ とすると

$$\mathbb{E}[\exp\{t\sum_{i=1}^{m} X_i\}] = E_1 \leq \exp\left\{\sum_{j=1}^{m} \frac{t_j^2}{8} + \frac{pt_j}{m+j}\right\}$$

となる. 今 $t_j(m-1+j)$ は j に関して一定なので $t_j = t\frac{m}{m+j-1}$ となる. 従って

$$\begin{aligned}
\sum_{j=1}^{m} \frac{pt_j}{m+j} &= \sum_{j=1}^{m} ptm\frac{1}{(m+j-1)(m+j)} \\
&= ptm\sum_{j=1}^{m}\left(\frac{1}{m+j-1} - \frac{1}{m+j}\right) = pt/2
\end{aligned}$$

更には t_j は単調現象列なので $\sum_{j=1}^{m} t_j^2/8 \leq mt^2/8$ となるから

$$E_1 \leq \exp\left\{\frac{mt^2}{8} + \frac{pt}{2}\right\}$$

従って

$$\mathbb{E}[\mathbb{I}\{\sum_{i=1}^{m} X_i - p/2 > m\epsilon/4\}] \leq \exp\left\{\frac{mt^2}{8} - \frac{tm\epsilon}{4}\right\} = \exp\left\{\frac{m}{8}\{(t-\epsilon)^2 - \epsilon^2\}\right\}$$

$t > 0$ は任意だったので $t = \epsilon > 0$ とすれば命題が示せる. □

なおここの証明では実質的にヘフディングの不等式が非復元抽出の場合でも成立することを示している．より一般の結果については [3] で紹介されている．

2　演習問題の解答

今回の発表範囲内で解ける演習問題の解答を用意した. 範囲内の問題は 2,3,5,8,9,10,11 の七つである.

Ex 3.2

問題：$\mathbb{R}$ の閾値の RC と GF

$\mathcal{H}$ を実数上の閾値関数の族とする. すなわち

$$\mathcal{H} = \{x \mapsto \mathbb{I}\{x \leq \theta\} \mid \theta \in \mathbb{R}\} \cup \{x \mapsto \mathbb{I}\{x \geq \theta\} \mid \theta \in \mathbb{R}\}$$

成長関数 $\Pi_m(\mathcal{H})$ の上界を与えよ. またこれを使って $\mathfrak{R}_m(\mathcal{H})$ の上界を与えよ.

解答：$x_1 < \cdots < x_m$ として良い. 符号のパターンは x_i を $\circ$ とおき θ を $|$ とおいたときの順列のそれぞれに対して $\lim_{x\to\infty} h(x) = 1 \ \vee \ 0$ の 2 通りがある．従って $\Pi_m(\mathcal{H}) \leq 2(m+1)$ である. なお符号が全て 1, 符号が全て 0 のパターンは $\mathbb{I}\{x \leq \theta\}$ の方でも $\mathbb{I}\{x \geq \theta\}$ の方でも得られるので重複がある．従って厳密には $\Pi_m(\mathcal{H}) = 2m$ である．
$\mathfrak{R}_m(\mathcal{H})$ の上界は補題 3.4 と系 3.8 を適応して $\mathfrak{R}_m(\mathcal{H}) \leq \frac{1}{2}\sqrt{\frac{2\log 2m}{m}}$ となる.

Ex 3.3

問題：線型結合の GF

$\mathbb{R}^d$ のベクトルの集合 X の線形分離符号割り当てとは, あるベクトル $w \in \mathbb{R}^d$ に対して X を $X^+ = \{x \in X \mid w \cdot x > 0\}$ と $X^- = \{x \in X \mid w \cdot x < 0\}$ の二つの集合に分類することである.

(a) $\{X^+, X^-\}$ を X のある二分とし $x_{m+1} \in \mathbb{R}^d$ とする. $\{X^+ \cup \{x_{m+1}\}, X^-\}$ と $\{X^+, X^- \cup \{x_{m+1}\}\}$ が原点を通る超平面で線形分離可能であることは, $\{X^+, X^-\}$ が原点と x_{m+1} を通る超平面で線形分離可能であることが必要かつ十分である.

(b) $X = \{x_1, \ldots, x_m\} \subset \mathbb{R}^d$ は, 任意の $k(\leq d)$ 個の要素を持つ部分集合が線型独立であるような集合とする. このとき X の線形分離符号割り当ての場合の数が $C(m,d) = 2\sum_{k=0}^{d-1}\binom{m-1}{k}$ であることを示せ.(ヒント：$C(m+1,d) = C(m,d) + C(m,d-1)$ を帰納法で示せ.)

(c) $f_1, \ldots, f_p$ を $\mathbb{R}^d$ から $\mathbb{R}$ の関数とする. $\mathcal{F}$ をこれらの関数の線型結合に基づいた分類機の族とする.

$$\mathcal{F} = \left\{ x \mapsto \operatorname{sgn}\left(\sum_{k=1}^{p} a_k f_k(x)\right) \middle| a_1, \ldots, a_p \in \mathbb{R} \right\}$$

$\Psi : x \mapsto (f_1(x), \ldots, f_p(x))^\top$ とする. ある $x_1, \ldots, x_m \in \mathbb{R}^d$ があって $\{\Psi(x_1), \ldots, \Psi(x_m)\}$ の任意の要素数 p 以下の部分集合が線型独立であるとする. このとき次を示せ.

$$\Pi_{\mathcal{F}}(m) = 2\sum_{i=0}^{p-1}\binom{m-1}{i}$$

解答：(a)

Proof. X^+ または X^- が $0 \in \mathbb{R}^d$ を含むときは明らかに成り立つ. $x_{m+1} = 0$ のときは成立しないがこれは例外とする. 以上のことから X^+, X^-, x_{m+1} のノルムは 1 としても良い. また分離する w もノルム 1 として問題ない.

$\{X^+ \cup \{x_{m+1}\}, X^-\}$ と $\{X^+, X^- \cup \{x_{m+1}\}\}$ が原点を通る超平面で線形分離可能とする. このときある $w^+, w^- \in \mathbb{R}^d$ があって次が成り立つ.

$$w^+ \cdot X^+, w^+ \cdot x_{m+1}, w^- \cdot X^+ > 0 > w^+ \cdot X^-, w^- \cdot x_{m+1}, w^- \cdot X^-$$

このとき $w = -(w^- \cdot x_{m+1})w^+ + (w^+ \cdot x_{m+1})w^-$ とすると, これは x_{m+1} に直交し更に X^+ を $+$ に分類し X^- を $-$ に分類することが分かる.

逆に $\{X^+, X^-\}$ が原点と x_{m+1} を通る超平面で線形分離可能とする.w は

$$w \cdot X^+ > 0 = w \cdot x_{m+1} > w \cdot X^-$$

を満たすとする.$\delta = \min_{x \in X^+ \cup X^-} |w \cdot x|$ とおく. これに対して $w^+ = (1 - \frac{\delta}{3})w + \frac{\delta}{3}x_{m+1}$, $w^- = (1 - \frac{\delta}{3})w - \frac{\delta}{3}x_{m+1}$ とおく. $w, x \in X^+ \cup X^-, x_{m+1}$ はノルムを 1 として良いことを用いると, w^+ は $\{X^+ \cup \{x_{m+1}\}, X^-\}$ と分離して,w^- は $\{X^+, X^- \cup \{x_{m+1}\}\}$ と分離することが分かる. □

(b)

Proof. m に関する帰納法で示す.

$m=1$ のとき任意の次元 d に対して明らかに符号の割り当て方は 2 通りである. よって $C(1,d)=2=2\sum_{k=0}^{d-1}\binom{0}{k}$ で命題は真である.

$m=k(k\geq 1)$ のとき命題が真であると仮定する. このとき $m=k+1$ の符号割り当ての場合の数を考える. そのために合計要素数 k の二分に対して新しい要素 x_{k+1} を追加することを考える. $\{X^+,X^-\}$ を合計要素数 k の二分とする. これを二分する超平面の法線ベクトルを w としたとき $w\cdot x_{k+1}$ は 0 より大きいか小さいか等しいかいずれかである.(a) により等しい場合は x_{k+1} は X^+,X^- にも入れることができる. 従って

$$C(k+1,d)=C(k,d)+(x_{k+1}\text{を通る超平面で分離可能な二分の場合の数})$$

となる. 次に上式右辺第二項が $C(k,d-1)$ であることを示す.$X=\{x_1,\dots,x_k,x_{k+1}\}$ を問題文の条件を満たすベクトルの集合とする. これらはノルムが 1 に正規化されているとして良い. $x_i(1\leq i\leq k)$ に対して

$$z_i=x_i-(x_i\cdot x_{k+1})x_{k+1}\neq 0$$

とする.$Z=\{z_1,\dots,z_k\}$ とおく. このとき明らかに $\{x_1,\dots,x_k\}$ の二分 $\{X^+,X^-\}$ が x_{k+1} を通る超平面で分離可能であることは, 対応する Z の二分 $\{Z^+,Z^-\}$ が x_{k+1} に直交する部分空間 S のベクトルで分離可能であることが必要かつ十分である. また $Z\subset S,\dim(S)=d-1$ であり,Z は任意の要素数 $k(\leq d-1)$ の部分集合が線型独立である.S の正規直交基底を $y_1,\dots,y_{d-1}$ する. ここで $\Phi:S\to\mathbb{R}^{d-1}$ を $\Phi:\sum_{i=1}^{d-1}p_iy_i\mapsto(p_1,\dots,p_{d-1})^\top$ で定義すると, Φ は内積を保存する同型写像, すなわち計量同型写像である. 従って $\Phi(Z)(\subset\mathbb{R}^{d-1})$ は任意の要素数 $k(\leq d-1)$ の部分集合が線型独立であるので, その線形分離可能な場合の数は $C(k,d-1)$ である. 従って (x_{k+1} を通る超平面で分離可能な二分の場合の数)$=C(k,d-1)$ である. よって

$$\begin{aligned}C(k+1,d)=C(k,d)+C(k,d-1)&=2\sum_{i=0}^{d-1}\binom{k-1}{i}+2\sum_{i=0}^{d-2}\binom{k-1}{i}\\&=2\binom{k-1}{0}+2\sum_{i=0}^{d-2}\left\{\binom{k-1}{i+1}+\binom{k-1}{i}\right\}=2\sum_{i=0}^{d-1}\binom{k}{i}\end{aligned}$$

が成り立つので $m=k+1$ のときも命題は真である. □

(c)

Proof. (b) により定義により $\Pi_{\mathcal{F}}(m)\geq C(m,p)$ は明らかである. 従って $\Pi_{\mathcal{F}}(m)\leq C(m,p)$ を証明すれば良い. そのために任意の $X=\{x_1,\dots,x_m\}\subset\mathbb{R}^d$ について, その線形分離符号割り当ての場合の数 C_X が $C_X\leq C(m,d)$ であることを証明すれば良い.(b) と同様に帰納法によって示す.

$m=1$ のとき任意の d に関して $C_X=2=C(1,d)$ であるから命題が成り立つ.

$m=k(k\geq 1)$ のとき任意の d に関して $C_X\leq C(k,d)$ が成立すると仮定する. $m=k+1$ について考えるが, これは X に新しい要素 x_{k+1} を付け加えることを考えればよい. $\{X^+,X^-\}$ を X のある二分とし, 二分する超平面の法線ベクトルを w とすると $w\cdot x_{k+1}$ は 0 より大きいか小さいか等しいので,

$$C_{X\cup\{x_{k+1}\}}=C_X+(x_{k+1}\text{を通る超平面で分離可能な } X \text{ の二分の場合の数})$$

上式右辺第二項は (b) と同様の議論により X に対応する Z の計量同型写像 Φ で送った像 $\Phi(Z)\subset\mathbb{R}^{d-1}$ の線形分離割当の場合の数 $C_{\Phi(Z)}$ である. 仮定により $C_X\leq C(k,d)$,$C_{\Phi(Z)}\leq C(k,d-1)$ が成り立つので $C_{X\cup\{x_{m+1}\}}\leq C(k+1,d)$ が導ける. 従って $m=k+1$ の時命題が成り立つ.

よって $\Pi_{\mathcal{F}}(m)\leq C(m,p)$ が得られる. □

Ex 3.5

問題：より良い RC の上界

$\Pi(\mathcal{G}, S)$ を標本集合 S に対する符号の割り当て方の場合の数とする．$\mathcal{G}$ のラデマッハ複雑度に対して $\mathbb{E}_S[\Pi(\mathcal{G}, S)]$ に関するより良い上界を与えよ．

解答：本発表資料に解答あり．

Ex 3.8

問題：RC の性質

$m \geq 1$ を固定する. 任意の $\alpha \in \mathbb{R}$ と関数 $\mathcal{X} \to \mathbb{R}$ の二つの仮説集合 $\mathcal{H}, \mathcal{H}'$ について次の性質を証明せよ．

(a) $\mathfrak{R}_m(\alpha\mathcal{H}) = |\alpha|\mathfrak{R}_m(\mathcal{H})$.

(b) $\mathfrak{R}_m(\mathcal{H} + \mathcal{H}') = \mathfrak{R}_m(\mathcal{H}) + \mathfrak{R}_m(\mathcal{H}')$

(c) $\mathfrak{R}_m(\{\max(h, h') \mid h \in \mathcal{H}, h' \in \mathcal{H}'\}) \leq \mathfrak{R}_m(\mathcal{H}) + \mathfrak{R}_m(\mathcal{H}')$,
ここで $\max(h, h')$ とは $x \mapsto \max_{x \in \mathcal{X}}(h(x), h'(x))$ を表す．（ヒント：任意の $a, b \in \mathbb{R}$ に対して $\max(a, b) = \frac{1}{2}(a + b + |a - b|)$ が成り立つという性質とタラグランドの縮約の補題 (補題 5.7) を使え.)

解答：(a)

Proof. 本発表資料に証明あり． □

(b)

Proof.

$$\begin{aligned}
\mathfrak{R}_m(\mathcal{H} + \mathcal{H}') &= \mathbb{E}_S\left[\mathbb{E}_\sigma\left[\frac{1}{m}\sup_{h \in \mathcal{H}, h' \in \mathcal{H}'}\sum_{i=1}^m \sigma_i(h(z_i) + h'(z_i))\right]\right] \\
&= \mathbb{E}_S\left[\mathbb{E}_\sigma\left[\frac{1}{m}\sup_{h \in \mathcal{H}}\sum_{i=1}^m \sigma_i h(z_i)\right]\right] + \mathbb{E}_S\left[\mathbb{E}_\sigma\left[\frac{1}{m}\sup_{h' \in \mathcal{H}'}\sum_{i=1}^m \sigma_i h'(z_i)\right]\right] \\
&= \mathfrak{R}_m(\mathcal{H}) + \mathfrak{R}_m(\mathcal{H}')
\end{aligned}$$

□

(c)

Proof.

$$
\begin{aligned}
\mathfrak{R}_m(\{\max(h,h') \mid h \in \mathcal{H}, h' \in \mathcal{H}'\}) &= \mathop{\mathbb{E}}_{S}\left[\mathop{\mathbb{E}}_{\sigma}\left[\frac{1}{m}\sup_{h\in\mathcal{H},h'\in\mathcal{H}'}\sum_{i=1}^{m}\sigma_i \max(h,h')(z_i)\right]\right] \\
&= \mathop{\mathbb{E}}_{S}\left[\mathop{\mathbb{E}}_{\sigma}\left[\frac{1}{m}\sup_{h\in\mathcal{H},h'\in\mathcal{H}'}\sum_{i=1}^{m}\sigma_i \frac{1}{2}\left\{h(z_i)+h'(z_i)+|h(z_i)-h'(z_i)|\right\}\right]\right] \\
&\leq \frac{1}{2}\mathfrak{R}_m(\mathcal{H})+\frac{1}{2}\mathfrak{R}_m(\mathcal{H}')+\frac{1}{2}\mathfrak{R}_m(|\mathcal{H}|)+\frac{1}{2}\mathfrak{R}_m(|\mathcal{H}'|) \\
&\leq \mathfrak{R}_m(\mathcal{H})+\mathfrak{R}_m(\mathcal{H}')
\end{aligned}
$$

ただしここでの $|\mathcal{H}|$ は集合 $\{|h| \mid h \in \mathcal{H}\}$ を意味する．最後の不等式は絶対値が $1-$ リプシッツ関数であることからタラグランドの縮約の補題を用いた． □

Ex 3.9

問題：コンセプトの共通部分の RC

$\mathcal{H}_1, \mathcal{H}_2$ は $X \to \{0,1\}$ の関数族とする．$\mathcal{H} = \{h_1 h_2 \mid h_1 \in \mathcal{H}_1, h_2 \in \mathcal{H}_2\}$ と定義する．任意の要素数 m の標本集合 S の $\mathcal{H}$ 経験ラデマッハ複雑度が次のように上から抑えられることを示せ．

$$\widehat{\mathfrak{R}}_S(\mathcal{H}) \leq \widehat{\mathfrak{R}}_S(\mathcal{H}_1) + \widehat{\mathfrak{R}}_S(\mathcal{H}_2)$$

(ヒント：$x \mapsto \max(0, x-1)$ がリプシッツ関数ということとタラグランドの縮約の補題を使え.)
またこれを使って,$c_1 \in \mathcal{C}_1, c_2 \in \mathcal{C}_2$ の共通部分からなる族 $\mathcal{U}$ のラデマッハ複雑度 $\mathfrak{R}_m(\mathcal{U})$ を $\mathcal{C}_1$ 及び $\mathcal{C}_2$ のラデマッハ複雑度に関して上から抑えよ.

解答：

Proof.

$$
\begin{aligned}
\widehat{\mathfrak{R}}_S(\mathcal{H}) &= \mathop{\mathbb{E}}_{\sigma}\left[\frac{1}{m}\sup_{h_1\in\mathcal{H}_1,h_2\in\mathcal{H}_2}\sum_{i=1}^{m}\sigma_i h_1(z_i)h_2(z_i)\right] \\
&= \mathop{\mathbb{E}}_{\sigma}\left[\frac{1}{m}\sup_{h_1\in\mathcal{H}_1,h_2\in\mathcal{H}_2}\sum_{i=1}^{m}\sigma_i \max(0, h_1(z_i)+h_2(z_i)-1)\right] \\
&\leq \mathop{\mathbb{E}}_{\sigma}\left[\frac{1}{m}\sup_{h_1\in\mathcal{H}_1,h_2\in\mathcal{H}_2}\sum_{i=1}^{m}\sigma_i (h_1(z_i)+h_2(z_i))\right] \\
&= \widehat{\mathfrak{R}}_S(\mathcal{H}_1)+\widehat{\mathfrak{R}}_S(\mathcal{H}_2)
\end{aligned}
$$

□

またこの不等式は $\mathcal{H} = \mathcal{U}, \mathcal{H}_1 = \mathcal{C}_1, \mathcal{H}_2 = \mathcal{C}_2$ としてそのまま適応できる.S に対する期待値を取ることにより

$$\mathfrak{R}_m(\mathcal{U}) \leq \mathfrak{R}_m(\mathcal{C}_1) + \mathfrak{R}_m(\mathcal{C}_2)$$

を得る.

Ex 3.10

問題：予測ベクトルの RC

$S=(x_1,\ldots,x_m)$ を要素数 m の標本として $h:X\to\mathbb{R}$ を固定する．

(a) h を使った S の予測ベクトルを $\boldsymbol{u}$ で表す．すなわち $\boldsymbol{u}=\begin{bmatrix}h(x_1)\\ \vdots\\ h(x_m)\end{bmatrix}$．$\|\boldsymbol{u}\|_2$ に関する $\mathcal{H}=\{h,-h\}$ の経験ラデマッハ複雑度 $\widehat{\mathfrak{R}}_S(\mathcal{H})$ の上界を与えよ．(ヒント：$\widehat{\mathfrak{R}}_S(\mathcal{H})$ を絶対値の期待値として表しイェンセンの不等式を適応せよ.) $h(x_i)\in\{0,-1,+1\}(^\forall i\in[m])$ と仮定する．そのラデマッハ複雑度の上界をスパース測度 $n=|\{i\mid h(x_i)\neq 0\}|$ に関して表せ. その上界はスパース測度が極端な値を取るときどうなるか.

(b) $\mathcal{F}$ を $X\to\mathbb{R}$ の関数の族とする.$\widehat{\mathfrak{R}}_S(\mathcal{F})$ と $\|\boldsymbol{u}\|_2$ に関する $\mathcal{F}+h=\{f+h\mid f\in\mathcal{F}\}$ と $\mathcal{F}\pm h=(\mathcal{F}+h)\cup(\mathcal{F}-h)$ の経験ラデマッハ複雑度の上界を与えよ.

解答：(a) 問題文にはイェンセンの不等式と書いてあるが, 絶対値を交換する凸関数と見做すと $\mathbb{E}[|\cdot|]\geq|\mathbb{E}[\cdot]|$ となるので上界は得られない．$\|\boldsymbol{u}\|_2$ に関する上界なので恐らくコーシー・シュワルツの不等式の間違いである.

$$\widehat{\mathfrak{R}}_S(\mathcal{H})=\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\sup_{\pi=\pm1}\frac{\pi}{m}\boldsymbol{\sigma}\cdot\boldsymbol{u}\right]=\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\left|\frac{\boldsymbol{\sigma}\cdot\boldsymbol{u}}{m}\right|\right]\leq\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{\|\boldsymbol{\sigma}\|_2\|\boldsymbol{u}\|_2}{m}\right]=\frac{\|\boldsymbol{u}\|_2}{\sqrt{m}}$$

$\|\boldsymbol{u}\|_2$ はスパース測度を用いると $\sqrt{n}$ となるから上界は $\sqrt{n/m}$ と表せる．$n=m$ のときは上界は 1 となり,$n=0$ のときは上界は 0 である.

(b)Ex 3.8(b) の結果は経験ラデマッハ複雑度で置き換えても成り立つ. 従ってこのことと (a) の結果から

$$\widehat{\mathfrak{R}}_S(\mathcal{F})=\widehat{\mathfrak{R}}_S(\mathcal{F}+h)\leq\widehat{\mathfrak{R}}_S(\mathcal{F}\pm h)\leq\widehat{\mathfrak{R}}_S(\mathcal{F})+\frac{\|\boldsymbol{u}\|_2}{\sqrt{m}}$$

上式の最初の等号は 3.6(a)(本発表資料に証明あり) の結果から得られる.

Ex 3.11

問題：正規化された神経回路網の RC

入力空間を $X=\mathbb{R}^{n_1}$ とする. この問題では, 以下の関数 $X\to\mathbb{R}$ の集合によって定義される正規化された神経回路網の族を考える.

$$\mathcal{H}=\left\{\boldsymbol{x}\mapsto\sum_{j=1}^{n_2}w_j\sigma(\boldsymbol{u}_j\cdot\boldsymbol{x})\middle|\|\boldsymbol{w}\|_1\le\Lambda',\|\boldsymbol{u}_j\|_2\le\Lambda,{}^\forall j\in[n_2]\right\}$$

ここで σ はある $L-$ リプシッツ関数である. 例として,σ は $1-$ リプシッツであるシグモイド関数などが当てはまる.

(a) $\widehat{\mathfrak{R}}_S(\mathcal{H})=\frac{\Lambda'}{m}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\|\boldsymbol{u}\|_2\le\Lambda}|\sum_{i=1}^m\sigma_i\sigma(\boldsymbol{u}\cdot\boldsymbol{x}_i)|\right]$ を示せ.

(b) 任意の仮説集合 $\mathcal{H}$ と $L-$ リプシッツ関数 Φ に対して有効な以下の形式のタラグランドの補題を使って,$\mathcal{H}'$ の経験ラデマッハ複雑度に関して $\widehat{\mathfrak{R}}_S(\mathcal{H})$ を上から抑えよ.

$$\frac{1}{m}\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\sup_{h\in\mathcal{H}}\left|\sum_{i=1}^m\sigma_i(\Phi\circ h)(x_i)\right|\right]\le\frac{L}{m}\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\sup_{h\in\mathcal{H}}\left|\sum_{i=1}^m\sigma_ih(x_i)\right|\right]$$

ここで $\mathcal{H}'$ は次によって定義される.

$$\mathcal{H}'=\{\boldsymbol{x}\mapsto s(\boldsymbol{u}\cdot\boldsymbol{x})\mid\|\boldsymbol{u}\|_2\le\Lambda,s\in\{-1,+1\}\}$$

(c) コーシーシュワルツの不等式を用いて以下を示せ.

$$\widehat{\mathfrak{R}}_S(\mathcal{H}')=\frac{\Lambda}{m}\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\left\|\sum_{i=1}^m\sigma_i\boldsymbol{x}_i\right\|_2\right]$$

(d) イェンセンの不等式によって成り立つ不等式 $\mathbb{E}_{\boldsymbol{v}}[\|\boldsymbol{v}\|_2]\le\sqrt{\mathbb{E}_{\boldsymbol{v}}[\|\boldsymbol{v}\|_2^2]}$ を使って $\widehat{\mathfrak{R}}_S(\mathcal{H}')$ を上から抑えよ.

(e) 任意の $\boldsymbol{x}\in S$ について, ある r があって $\|\boldsymbol{x}\|_2\le r$ であると仮定する. 前問を使って r に関する $\mathcal{H}$ のラデマッハ複雑度の上界を得よ.

解答：(a)

Proof. 両方の不等号が成り立つことを示す. まず (左辺)≤(右辺) を示す.

$$\begin{aligned}\widehat{\mathfrak{R}}_S(\mathcal{H})&=\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{1}{m}\sup_{\|\boldsymbol{w}\|_1\le\Lambda',\|\boldsymbol{u}_j\|_2\le\Lambda}\sum_{i=1}^m\sigma_i\sum_{j=1}^{n_2}w_j\sigma(\boldsymbol{u}_j\cdot\boldsymbol{x}_i)\right]\\&\le\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{1}{m}\sup_{\|\boldsymbol{w}\|_1\le\Lambda',\|\boldsymbol{u}_j\|_2\le\Lambda}\sum_{j=1}^{n_2}|w_j|\left|\sum_{i=1}^m\sigma_i\sigma(\boldsymbol{u}_j\cdot\boldsymbol{x}_i)\right|\right]\\&\le\frac{\Lambda'}{m}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\|\boldsymbol{u}\|_2\le\Lambda}\left|\sum_{i=1}^m\sigma_i\sigma(\boldsymbol{u}\cdot\boldsymbol{x}_i)\right|\right]\end{aligned}$$

次に (左辺)≥(右辺) を示す.

$$
\begin{aligned}
\widehat{\mathfrak{R}}_S(\mathcal{H}) &= \mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{1}{m}\sup_{\|\boldsymbol{w}\|_1\le\Lambda',\|\boldsymbol{u}_j\|_2\le\Lambda}\sum_{i=1}^{m}\sigma_i\sum_{j=1}^{n_2}w_j\sigma(\boldsymbol{u}_j\cdot\boldsymbol{x}_i)\right] \\
&\ge \mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{1}{m}\sup_{\|\boldsymbol{w}\|_1\le\Lambda',\|\boldsymbol{u}\|_2\le\Lambda}\sum_{j=1}^{n_2}w_j\sum_{i=1}^{m}\sigma_i\sigma(\boldsymbol{u}\cdot\boldsymbol{x}_i)\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{1}{m}\sup_{v_j\ge 0,v_1+\cdots+v_{n_2}\le\Lambda',\pi_j=\pm1,\|\boldsymbol{u}\|_2\le\Lambda}\sum_{j=1}^{n_2}v_j\pi_j\sum_{i=1}^{m}\sigma_i\sigma(\boldsymbol{u}\cdot\boldsymbol{x}_i)\right] \\
&\ge \mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{1}{m}\sup_{v_j\ge 0,v_1+\cdots+v_{n_2}\le\Lambda'}\ \sup_{\pi_j=\pm1,\|\boldsymbol{u}\|_2\le\Lambda}\sum_{j=1}^{n_2}v_j\pi_j\sum_{i=1}^{m}\sigma_i\sigma(\boldsymbol{u}\cdot\boldsymbol{x}_i)\right] \\
&= \frac{\Lambda'}{m}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\|\boldsymbol{u}\|_2\le\Lambda}\left|\sum_{i=1}^{m}\sigma_i\sigma(\boldsymbol{u}\cdot\boldsymbol{x}_i)\right|\right]
\end{aligned}
$$

□

(b)$\mathcal{H}'$ の経験ラデマッハ複雑度は次のように表せる.

$$
\widehat{\mathfrak{R}}_S(\mathcal{H}') = \mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{1}{m}\sup_{\|\boldsymbol{u}\|_2\le\Lambda,s=\pm1}\sum_{i=1}^{m}\sigma_i s\boldsymbol{u}\cdot\boldsymbol{x}_i\right] = \mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{1}{m}\sup_{\|\boldsymbol{u}\|_2\le\Lambda}\left|\sum_{i=1}^{m}\sigma_i\boldsymbol{u}\cdot\boldsymbol{x}_i\right|\right]
$$

従ってタラグランドの補題を適応することによって

$$
\widehat{\mathfrak{R}}_S(\mathcal{H}) = \frac{\Lambda'}{m}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\|\boldsymbol{u}\|_2\le\Lambda}\left|\sum_{i=1}^{m}\sigma_i\sigma(\boldsymbol{u}\cdot\boldsymbol{x}_i)\right|\right] \le \Lambda' L\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{1}{m}\sup_{\|\boldsymbol{u}\|_2\le\Lambda}\left|\sum_{i=1}^{m}\sigma_i\boldsymbol{u}\cdot\boldsymbol{x}_i\right|\right] = \Lambda' L\widehat{\mathfrak{R}}_S(\mathcal{H}')
$$

(c)

Proof.

$$
\widehat{\mathfrak{R}}_S(\mathcal{H}') = \mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{1}{m}\sup_{\|\boldsymbol{u}\|_2\le\Lambda}\left|\left(\sum_{i=1}^{m}\sigma_i\boldsymbol{x}_i\right)\cdot\boldsymbol{u}\right|\right] \le \mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\frac{1}{m}\sup_{\|\boldsymbol{u}\|_2\le\Lambda}\left\|\sum_{i=1}^{m}\sigma_i\boldsymbol{x}_i\right\|_2\|\boldsymbol{u}\|_2\right] = \frac{\Lambda}{m}\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\left\|\sum_{i=1}^{m}\sigma_i\boldsymbol{x}_i\right\|_2\right]
$$

□

(d) $\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\|\sum_{i=1}^{m}\sigma_i\boldsymbol{x}_i\|_2\right]$ を上から評価すれば良い.

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\left\|\sum_{i=1}^{m}\sigma_i\boldsymbol{x}_i\right\|_2\right] &\le \sqrt{\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\left\|\sum_{i=1}^{m}\sigma_i\boldsymbol{x}_i\right\|_2^2\right]} \\
&= \sqrt{\mathop{\mathbb{E}}_{\boldsymbol{\sigma}}\left[\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2 + 2\sum_{i<j}\sigma_i\sigma_j\boldsymbol{x}_i\cdot\boldsymbol{x}_j\right]} = \sqrt{\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2}
\end{aligned}
$$

従って $\widehat{\mathfrak{R}}_S(\mathcal{H}') \le \frac{\Lambda}{m}\sqrt{\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2}$ を得る.

(e) 前問の結果を全て連結すればよい.

$$
\mathfrak{R}_m(\mathcal{H}) \le \frac{\Lambda' L\Lambda r}{\sqrt{m}}
$$

参考文献

[1] Kyle Siegrist, “11.6: The Simple Random Walk”, STATISTICS LibreTexts, https://stats.libretexts.org/Bookshelves/Probability_Theory/Book\%3A_Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/11\%3A_Bernoulli_Trials/11.06\%3A_The_Simple_Random_Walk , Aug 10 2020(cited Jun 4 2021)

[2] V.N.Vapnik and A. Ya.Chervonenkis, Translated by B. Seckler “On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities”, Theory of Probability and Its Applications, Vol. 16, No. 2, 1971, pp.264-280.

[3] Remi Bardenet and Odalric-Ambrym Maillard, “Concentration inequalities for sampling without replacement”, Bernoulli, Vol. 21, No. 3, 2015, pp.1361-1385.