机器学习 作业2

2

2.1

有 $\mathcal{E}_D(\hat{r}) = P[C_r \setminus C_{\hat{r}}] > \varepsilon$

$\Rightarrow P[C_r \setminus C_{\hat{r}}] > \varepsilon = P[C_r \setminus C_{r'}] \Rightarrow \hat{r} < r'$

2.2

$P_{D_n}[\mathcal{E}_D(\hat{r}_{D_n}) > \varepsilon] = P_{D_n}[\hat{r} < r']$

$= P_{D_n}[\max_{(x,y)\in D_n, y=1} \|x\|_2 < r']$

不妨设 D_n 中有 k 个正例点，分别为

$(x_1^+, y_1^+), \cdots, (x_k^+, y_k^+)$, $y_i^+ = 1$. $i = 1, 2 \cdots k$.

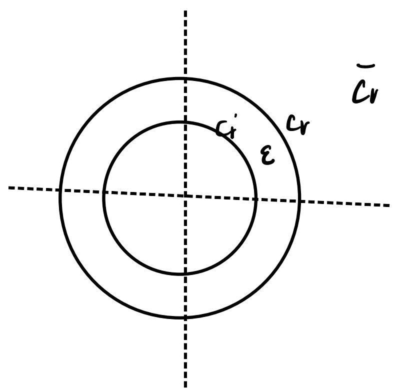
有 $n-k$ 个负例点 为

$(x_1^-, y_1^-), \cdots (x_{n-k}^-, y_{n-k}^-)$ $y_j^- = 0$. $j = 1, 2, \cdots n-k$

$P_{D_n}[\max_{(x,y)\in D_n, y=1} \|x\|_2 < r']$

$= P[\bigcap_{i=1}^{k} (\|x_i^+\|_2 < r') \bigcap_{j=1}^{n-k} (\|x_j^-\|_2 \geq r))$

$= P^k(C_{r'}) \, P^{n-k}(\bar{C}_r)$



而 $\begin{cases} P(C_{r'}) \leq 1-\varepsilon \\ P(\bar{C}_r) \leq 1-\varepsilon \end{cases}$

$\Rightarrow P_{D_n}[\max_{(x,y)\in D_n, y=1} \|x\|_2 < r'] \leq (1-\varepsilon)^n \leq (e^{-\varepsilon})^n = e^{-\varepsilon n}$

2.3.

依据定义，通过算法A，对任意 $\varepsilon > 0$、$\delta > 0$、对任意分布 D 与目标假设 $h \in \mathcal{H}$，当

$$n \geq \frac{1}{\varepsilon} \ln \frac{1}{\delta} \quad \text{时}$$

有

$$P_{D_n}[\,\mathcal{E}(\hat{r}_{D_n}) \geq \varepsilon\,] \leq e^{-n\varepsilon} \leq \delta$$，由此可得出 PAC 可学习

2.4.

定义示性变量 $\alpha(x) = \begin{cases} 1 & \|x\|_2 \leq r \text{ 且 } y' = 1 \\ 0 & y' = 0 \end{cases}$

$$\mathcal{E}_{D_n}(r) = E_{x \sim D_x}[\,\alpha(x) \neq 1[\|x\|_2 \leq r]\,]$$

$$= P(\|x\|_2 \leq r)\,\eta = P(C_r)\,\eta$$

$$\mathcal{E}_{D_n}(r') = E_{x \sim D_x}[\,\alpha(x) \neq 1[\|x\|_2 \leq r']\,]$$

$$= P(C_r / C_{r'})(1-\eta) + P(C_{r'})\,\eta$$

$$= \varepsilon(1-\eta) + (P(C_r) - \varepsilon)\,\eta$$

$$= \mathcal{E}_{D_n}(r) + \varepsilon(1-2\eta)$$

2.5. 由 2.4. $\varepsilon(\hat{r}) = \varepsilon(r) + P(C_r / C_{\hat{r}})(1-2\eta)$

则 $P_{D_n \sim D_\eta^n}[\varepsilon(\hat{r}) - \varepsilon(r) > \varepsilon]$

$= P[\,p[C_r / C_{\hat{r}}] \cdot (1-2\eta) > \varepsilon\,]$

$= P[\,p[C_r / C_{\hat{r}}] > \frac{\varepsilon}{1-2\eta}\,] \leq P[\,p[C_r / C_{\hat{r}}] > \varepsilon\,]$

$= P[\hat{r} < r']$

不妨设 D_n 中有 k 个正例点，分别为：

$(x_1^+, y_1^+), \cdots, (x_k^+, y_k^+)$，$y_i^+ = 1$，$i = 1, 2, \cdots k$。

有 $n-k$ 个负例点，为

$(x_1^-, y_1^-), \cdots (x_{n-k}^-, y_{n-k}^-)$ $y_j^- = 0$. $j = 1, 2, \cdots n-k$

则 $P[\hat{r} < r'] = P[\max_{(x,y) \in D_n, y=1} \|x\|_2 < \hat{r}]$

$= [(1-\eta) P(C_{r'})]^k [P(\bar{C}_r) + P(C_r)\eta]^{n-k}$

$\leq (1-\varepsilon)^k \cdot [P(\bar{C}_r) + P(C_r)\eta']^{n-k}$

$= (1-\varepsilon)^k [P(\bar{C}_r) + (P(C_{r'}) + P(C_r / C_{r'}))\eta']^{n-k}$

$\leq (1-\varepsilon)^k [1-\varepsilon + \eta'\varepsilon]^{n-k}$

$\leq (1-\varepsilon+\eta'\varepsilon)^n \leq e^{-n\varepsilon(1-\eta')}$

故有上界 $e^{-n\varepsilon(1-\eta')}$

依据定义，通过算法A，对任意 $\varepsilon>0$、$\delta>0$、对任意分布D与目标假设 $h\in\mathcal{H}$，当

$$n \geq \frac{1}{\varepsilon(1-\eta')}\ln\frac{1}{\delta}$$

有

$$P_{D_n}[\,\varepsilon(\hat{h})-\varepsilon(h)\geq\varepsilon\,] \leq e^{-n\varepsilon(1-\eta')} \leq \delta$$，由此可知 PAC 可学习

3

3.1

$$L_{D^{(i)}}(f_i)=\sum_{(x,y)\in C\times\{0,1\}} P_{D^{(i)}}[(x,y)]\cdot \mathbb{1}[y\neq f_i(x)]=0$$

3.2.

$D_n^{(i)}$ 是等概率的从 $\{S_1^{(i)}, S_2^{(i)}, \ldots, S_k^{(i)}\}$ 中抽取出的一个序列.

$$\max_{1\leq i\leq T} E_{D_n^{(i)}}[L_{D^{(i)}}(A(D_n^{(i)}))]$$

$$\geq \frac{1}{T}\sum_{i=1}^{T} E_{D_n^{(i)}}[L_{D^{(i)}}(A(D_n^{(i)}))]$$

$$=\frac{1}{Tk}\sum_{i=1}^{T}\sum_{j=1}^{k}[L_{D^{(i)}}(A(S_j^{(i)}))]$$

$$= \frac{1}{Tk} \sum_{j=1}^{k} \sum_{j=1}^{T} [L_{D^{(i)}}(A(S_j^{(i)})]$$

$$= \frac{1}{k} \sum_{j=1}^{k} \frac{1}{T} \sum_{j=1}^{T} [L_{D^{(i)}}(A\ S_j^{(i)})]$$

$$\geq \min_{1 \leq j \leq k} \frac{1}{T} \sum_{i=1}^{T} L_{D^{(i)}}(A(S_j^{(i)}))$$

3.3 首先有:

$$L_{D^i}(h) = \sum_{(x,y) \in C \times \{0,1\}} P_{D^{(i)}}[(x,y)] \cdot 1[h(x) \neq y]$$

$$= \sum_{(x,y) \in C \times \{0,1\}} P_{D^{(i)}}[(x,y)] \cdot 1[h(x) \neq f_i(x)]$$

$$\geq \sum_{r=1}^{p} P_{D^{(i)}}[(v_r, f_i(v_r))] \cdot 1[h(v_r) \neq f_i(v_r)]$$

$$= \frac{1}{2n} \sum_{r=1}^{p} 1[h(v_r) \neq f_i(v_r)] \geq \frac{1}{2p} \sum_{r=1}^{p} 1[h(v_r) \neq f_i(v_r)]$$

故 $\frac{1}{T} \sum_{i=1}^{T} L_{D^{(i)}}(A(S_j^{(i)})) \geq \frac{1}{2pT} \sum_{i=1}^{T} \sum_{r=1}^{p} 1[A(S_j^{(i)})(v_r) \neq f_i(v_r)]$

$$= \frac{1}{2pT} \sum_{r=1}^{p} \sum_{i=1}^{T} 1[A(S_j^{(i)})(v_r) \neq f_i(v_r)]$$

3.4. 由 3.2 知.

$$\max_{1\le i\le T} \mathbb{E}_{D_n^{(i)}}\left[L_{D^{(i)}}(A(D_n^{(i)}))\right] \ge \min_{1\le j\le K} \frac{1}{T}\sum_{i=1}^{T} L_{D^{(i)}}(A(S_j^{(i)}))$$

不妨设 $\min_{1\le j\le K} \frac{1}{T}\sum_{i=1}^{T} L_{D^{(i)}}(A(S_j^{(i)})) = \frac{1}{T}\sum_{i=1}^{T} L_{D^{(i)}}(A(S_{j_0}^{(i)}))$

则有:

$$\max_{1\le i\le T} \mathbb{E}_{D_n^{(i)}}\left[L_{D^{(i)}}(A(D_n^{(i)}))\right] \ge \frac{1}{T}\sum_{i=1}^{T} L_{D^{(i)}}(A(S_{j_0}^{(i)}))$$

由 3.3 知, $\frac{1}{T}\sum_{i=1}^{T} L_{D^{(i)}}(A(S_{j_0}^{(i)})) \ge \frac{1}{2pT}\sum_{r=1}^{p}\sum_{i=1}^{T} \mathbb{1}\left[A(S_{j_0}^{(i)})(v_r) \ne f_i(v_r)\right]$

固定 r, 下证 $\sum_{i=1}^{T} \mathbb{1}\left[A(S_{j_0}^{(i)})(v_r) = f_i(v_r)\right] = \frac{T}{2}$.

对于任何 i, 总能找到 i' 使得

f_i 与 $f_{i'}$ 在 S_{j_0} 上标签完全相同. 且 f_i 与 $f_{i'}$ 在 $v_1, \dots, v_r$ 上标签完全相反. 这样将 T 分为了 $\frac{T}{2}$ 组.

此时 $A(S_{j_0}^{(i)}) = A(S_{j_0}^{(i')})$.

则 $A(S_{j_0}^{(i)})(v_r) = A(S_{j_0}^{(i')})(v_r)$

且 $f_i(v_r) + f_{i'}(v_r) = 1$

故 $\mathbb{1}\left[A(S_{j_0}^{(i)})(v_r) = f_i(v_r)\right] + \mathbb{1}\left[A(S_{j_0}^{(i')})(v_r) = f_{i'}(v_r)\right] = 1$

$$\Rightarrow \sum_{i=1}^{T} \mathbb{1}\left[A(S_{j_0}^{(i)})(v_r) = f_i(v_r)\right] = \frac{T}{2}.$$

故有

$$\max_{1\le i\le T} E_{D_n^{(i)}}[L_{D^{(i)}}(A(D_n^{(i)}))] \ge \frac{1}{2pT}\sum_{r=1}^{p}\sum_{i=1}^{T} 1[A(S_{j_0}^{(i)})(v_r) \ne f_i(v_r)]$$

$$= \frac{1}{2pT}\sum_{r=1}^{p}\cdot\frac{T}{2} = \frac{1}{4}.$$

3.5. 由引理，

$$p\left(L_{D^{(i)}}(A(D_n^{(i)})) \ge \frac{1}{8}\right) \ge \frac{E_{D_n^{(i)}}[L_{D^{(i)}}(A(D_n^{(i)}))] - \frac{1}{8}}{1-\frac{1}{8}}$$

$$= \frac{1}{7}\left(8E_{D_n^{(i)}}[L_{D^{(i)}}(A(D_n^{(i)}))] - 1\right)$$

取分布 $D_n^{(i_0)}$ 使得 $E_{D^{(i)}}[L_{D^{(i)}}(A(D_n^{(i)}))]$ 达到 max,

由3.4 有

$$p\left(L_{D^{(i_0)}}(A(D_n^{(i_0)})) \ge \frac{1}{8}\right) \ge \frac{1}{7}\left(8\cdot\frac{1}{4} - 1\right) = \frac{1}{7}$$

即存在这样一个糟糕的分布 $D_n^{(i_0)}$

3.6. 对于有限次采样n，必有 $n < \frac{|X|}{2}$，应用 NFL Theorem，即存在一种分布 $D^{(i)}$ 使得：

$$p\left(L_{D^{(i)}}(A(D_n^{(i)})) \ge \frac{1}{8}\right) \ge \frac{1}{7}$$

即 $$p\left(L_{D^{(i)}}(A(D_{n^{(i)}})) - L_{D^{(i)}}(f_i) \ge \frac{1}{8}\right) \ge \frac{1}{7}$$

取 $\varepsilon = \frac{1}{8}$、$\delta = \frac{1}{7}$，即无法通过 $poly(|H|, \frac{1}{\varepsilon}, \frac{1}{\delta})$ 次采样使得

$$p\left(L_{D^{(i)}}(A(D_n^{(i)})) - L_{D^{(i)}}(f_i) \ge \varepsilon\right) < \delta$$

此时该问题并不是 PAC 可学习的.

4

4.1.

即证 $\exists i$ 使得 $\|\hat{p}_i - p_i\|_1 > m\sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}$ 的概率小于 δ,

即 $$p\left(\max_{1\le i\le k}\|\hat{p}_i - p_i\|_1 > m\sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}\right) < \delta$$

而 $$p\left(\max_{1\le i\le k}\|\hat{p}_i - p_i\|_1 > m\sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}\right)$$

$$\le p\left(\bigcup_{i=1}^{k}\left(\|\hat{p}_i - p_i\|_1 > m\sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}\right)\right)$$

由 Union bound,

$$p\left(\bigcup_{i=1}^{k}\left(\|\hat{p}_i - p_i\|_1 > m\sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}\right)\right) \le \sum_{i=1}^{k} p\left(\|\hat{p}_i - p_i\|_1 > m\sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}\right)$$

而 $$\sum_{i=1}^{K} p\left(\|\hat{p}_i - p_i\|_1 > m\sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}\right)$$

$$= \sum_{i=1}^{K} P\left(\frac{1}{m}\sum_{j=1}^{m}|\hat{p}_{i,j} - p_{i,j}| > \sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}\right)$$

$$= \sum_{i=1}^{K} P\left(\max_{1\le j\le m}|\hat{p}_{i,j} - p_{i,j}| > \sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}\right)$$

再应用一次 Union bound.

$$\sum_{i=1}^{K} P\left(\max_{1\le j\le m}|\hat{p}_{i,j} - p_{i,j}| > \sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}\right)$$

$$\le \sum_{i=1}^{k}\sum_{j=1}^{m} P\left(|\hat{p}_{ij} - p_{ij}| > \sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}} \right)$$

$$\le \sum_{i=1}^{k}\sum_{j=1}^{m} P\left(\left|\sum_{s=1}^{n} \left(1[X_i^{(s)}=j] - \mathbb{E}\,1[X_i^{(s)}=j]\right)\right| > \sqrt{\frac{n}{2}\log\frac{2mk}{\delta}} \right)$$

由 Hoeffding's Inequality.

$$\sum_{i=1}^{k}\sum_{j=1}^{m} P\left(\left|\sum_{s=1}^{n} \left(1[X_i^{(s)}=j] - \mathbb{E}\,1[X_i^{(s)}=j]\right)\right| > \sqrt{\frac{n}{2}\log\frac{2mk}{\delta}} \right)$$

$$\le \sum_{i=1}^{k}\sum_{j=1}^{m} 2e^{-\frac{2\cdot\frac{n}{2}\cdot\log\frac{2mk}{\delta}}{\sum_{s=1}^{n}(1-0)^2}} = 2km\cdot\frac{\delta}{2km} = \delta \quad \text{证毕.}$$

4.2.

同 4.1 的证明思路，可以得到.

$$P\left(\max_{1\le i\le k}\|\hat{p}_i - p_i\|_1 > \sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\right) \le \sum_{i=1}^{k} P\left(\|\hat{p}_i - p_i\|_1 > \sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\right)$$

注意到. $\forall v\in\mathbb{R}^m$, $\|v\|_1 = \sup_{u\in\{-1,+1\}^m} u^T v$

即

$$\sum_{i=1}^{k} P\left(\|\hat{p}_i - p_i\|_1 > \sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\right)$$

$$\le \sum_{i=1}^{k} P\left(\sup_{u\in\{-1,+1\}^m} u^T(\hat{p}_i - p_i) > \sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\right)$$

$$\le \sum_{i=1}^{k} P\left(\bigcup_{u\in\{-1,+1\}^m} u^T(\hat{p}_i - p_i) > \sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\right)$$

union bound

$$\le \sum_{i=1}^{k}\sum_{u\in\{-1,+1\}^m} P\left(u^T(\hat{p}_i - p_i) > \sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\right)$$

对于任一 $u\in\{-1,+1\}^m$，可以找到 u' 使得 $u'+u=\{0\}^m$、

即 $u^T(\hat{p}_i-p_i)>0$ 与 $u^T(\hat{p}_i-p_i)>0$ 不可能同成立.

而若 $u^T(\hat{p}_i-p_i)\le 0$. $P\left(u^T(\hat{p}_i-p_i)>\sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\right)=0$.

故可以对 $\{-1,+1\}^m$ 作筛选得到集合 $U_i^+=\{u\in\{-1,+1\}^m \mid u^T(\hat{p}_i-p_i)>0\}$.

且 $|U_i^+|\le 2^{m-1}$

$$\sum_{i=1}^{k} P\left(\|\hat{p}_i-p_i\|_1>\sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\right)$$

$$\le \sum_{i=1}^{k}\sum_{u\in U_i^+} P\left(u^T(\hat{p}_i-p_i)>\sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\right)$$

即 对于固定的 u，假设其值为1的所有维度的索引组成集合 $S_{u,1}$.

即 $S_{u,1}=\{r\in[1,m] \mid u^{(r)}=1\}$.

$u^{(r)}$ 为 u 的第 r 个维度.

则

$$\sum_{i=1}^{k}\sum_{u\in U_i^+} P\left(u^T(\hat{p}_i-p_i)>\sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\right)$$

$$=\sum_{i=1}^{k}\sum_{u\in U_i^+} P\left(\sum_{r\in S_{u,1}}(\hat{p}_{i,r}-p_{i,r})-\sum_{r\notin S_{u,1}}(\hat{p}_{i,r}-p_{i,r})>\sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\right)$$

$$=\sum_{i=1}^{k}\sum_{u\in U_i^+} P\Bigg(\sum_{r\in S_{u,1}}\left(\frac{1}{n}\sum_{s=1}^{n}1[X_i^{(s)}=r]-E\,1[X_i^{(s)}=r]\right)$$

$$-\sum_{r\notin S_{u,1}}\left(\frac{1}{n}\sum_{s=1}^{n}1[X_i^{(s)}=r]-E\,1[X_i^{(s)}=r]\right)>\sqrt{\frac{2}{n}\log\frac{k\cdot 2^m}{\delta}}\Bigg)$$

$$= \sum_{i=1}^{K} \sum_{u \in U_i^+} P\Big(\sum_{s=1}^{n} \frac{1}{n}\big(1[X_i^{(s)} \in S_{u,1}] - E\, 1[X_i^{(s)} \in S_{u,1}] \big)$$

$$- \sum_{s=1}^{n} \frac{1}{n}\big(1[X_i^{(s)} \notin S_{u,1}] - E\, 1[X_i^{(s)} \notin S_{u,1}] \big) > \sqrt{\frac{2}{n} \log \frac{k \cdot 2^m}{\delta}} \Big)$$

再引入 $\begin{cases} 1[X_i^{(s)} \in S_{u,1}] + 1[X_i^{(s)} \notin S_{u,1}] = 1 \\ E\, 1[X_i^{(s)} \in S_{u,1}] + E\, 1[X_i^{(s)} \notin S_{u,1}] = 1. \end{cases}$

$$= \sum_{i=1}^{K} \sum_{u \in U_i^+} P\Big(\frac{2}{n} \sum_{s=1}^{n} \big(1[X_i^{(s)} \in S_{u,1}] - E\,1[S_i^{(s)} \in S_{u,1}] \big) > \sqrt{\frac{2}{n} \log \frac{k \cdot 2^m}{\delta}} \Big)$$

$$= \sum_{i=1}^{K} \sum_{u \in U_i^+} P\Big(\Big| \sum_{s=1}^{n} \big(1[X_i^{(s)} \in S_{u,1}] - E\,1[S_i^{(s)} \in S_{u,1}] \big) \Big| > \sqrt{\frac{n}{2} \log \frac{k \cdot 2^m}{\delta}} \Big)$$

Hoeffding's Inequality.

$$\leq \sum_{i=1}^{K} \sum_{u \in U_i^+} 2e^{-\frac{2 \cdot \frac{n}{2} \cdot \log \frac{k \cdot 2^m}{\delta}}{\sum_{s=1}^{n} (1-0)^2}}$$

$$= \sum_{i=1}^{K} \sum_{u \in U_i^+} 2 \cdot \frac{\delta}{k \cdot 2^m} \leq k \cdot 2^{m-1} \cdot 2 \frac{\delta}{k \cdot 2^m} = \delta$$

证毕.

5

由 Hoeffding's Inequality,

$$P\left(\left|\sum_{i=1}^{n}\frac{1}{n}(X_i - EX_i)\right| \geq \sqrt{\frac{1}{2n}\log\frac{2}{\delta}}\right) \leq \delta$$

其中 $X_i = \ell(h(x_i), y_i)$, $x_i \in \mathcal{X}$. $y_i \in [0,1]$

X_i为 i.i.d 采样.

令 $\delta' = p(h)\delta$, 有:

$$P\left(\left|\sum_{i=1}^{n}\frac{1}{n}(X_i - EX_i)\right| \geq \sqrt{\frac{1}{2n}\log\frac{2}{p(h)\delta}}\right) \leq p(h)\delta$$

即 $P\left(\left|\hat{\varepsilon}_{D_n}(h) - \varepsilon(h)\right| \geq \sqrt{\frac{1}{2n}\left(\log\frac{1}{p(h)} + \frac{2}{\delta}\right)}\right) \leq p(h)\delta$

而由 union bound,

$$P\left(\exists h \in \mathcal{H},\ \left|\hat{\varepsilon}_{D_n}(h) - \varepsilon(h)\right| \geq \sqrt{\frac{1}{2n}\left(\log\frac{1}{p(h)} + \frac{2}{\delta}\right)}\right)$$

$$= P\left(\sup_{h \in \mathcal{H}} \left|\hat{\varepsilon}_{D_n}(h) - \varepsilon(h)\right| \geq \sqrt{\frac{1}{2n}\left(\log\frac{1}{p(h)} + \log\frac{2}{\delta}\right)}\right)$$

$$\leq \sum_{h \in \mathcal{H}} P\left(\left|\hat{\varepsilon}_{D_n}(h) - \varepsilon(h)\right| \geq \sqrt{\frac{1}{2n}\left(\log\frac{1}{p(h)} + \log\frac{2}{\delta}\right)}\right)$$

$$\leq \sum_{h \in \mathcal{H}} p(h)\delta = \delta$$

即证明了 $\forall \delta > 0$, 以至少 $1-\delta$ 的概率使得

$$\varepsilon(h) \leq \hat{\varepsilon}_{D_n}(h) + \sqrt{\frac{1}{2n}\left(\log\frac{1}{p(h)} + \log\frac{1}{\delta}\right)}$$

对比 $\frac{1}{p(h)}$ 与 $|H|$，我们可以得出如下结论：

① 由 $|H|$ 约束的泛化误差界并未引入假设先验概率对泛化误差的影响，是平均意义上的泛化误差界。

② 由 $\frac{1}{p(h)}$ 约束的泛化误差界说明了：假设先验概率越小，则在其上只能给出较大的泛化误差界；反之，假设先验概率越大，则可以给出更紧的泛化误差界。

6

6.1

由 $R_n(g) = \mathbb{E}_{S_n \sim D^n} \hat{R}_{S_n}(g)$ 以及期望的线性可加性知，只需验证 $\hat{R}_{S_n}(g)$ 满足以下性质即可。

6.1.1 由 $H' \supseteq H$，有 $\sup\limits_{h \in H'} \frac{1}{n}\sum\limits_{i=1}^{n} \sigma_i h(z_i) \geq \sup\limits_{h \in H} \frac{1}{n}\sum\limits_{i=1}^{n} \sigma_i h(z_i)$

即 $\hat{R}_{S_n}(H') = \mathbb{E}_\sigma\left[\sup\limits_{h \in H'} \frac{1}{n}\sum\limits_{i=1}^{n} \sigma_i h(z_i)\right]$

$$\geq \mathbb{E}_\sigma\left[\sup_{h \in H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)\right] = \hat{R}_{S_n}(H)$$

6.1.2. 若 $\alpha \geq 0$

$$\hat{R}_{S_n}(\alpha H) = \mathbb{E}_\sigma\left[\sup_{h \in \alpha H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)\right]$$

$$= \mathbb{E}_\sigma\left[\sup_{h \in H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i \alpha h(z_i)\right]$$

$$= \alpha\, \mathbb{E}_\sigma\left[\sup_{h \in H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)\right] = \alpha \hat{R}_{S_n}(H)$$

若 $\alpha < 0$,

对于一固定的 σ，可以找到 σ' 使得 $\sigma_i + \sigma_i' = 0$. 由此将所有 σ 分为若干对.

则 $\sup_{h \in \alpha\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i) = \sup_{h \in -\alpha\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i' h(z_i)$

则 $$\hat{R}_{S_n}(\alpha\mathcal{H}) = E_\sigma\left[\sup_{h \in \alpha\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)\right]$$

$$= E_\sigma\left[\sup_{h \in -\alpha\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} (-\alpha)\sigma_i' h(z_i)\right]$$

$$= (-\alpha) E_\sigma\left[\sup_{h \in \mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i' h(z_i)\right]$$

$$= (-\alpha) E_\sigma\left[\sup_{h \in \mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i' h(z_i)\right] = (-\alpha) E_\sigma\left[\sup_{h \in \mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)\right]$$

$$= -\alpha \hat{R}_{S_n}(\mathcal{H})$$

6.1.3.

$$\hat{R}_{S_n}(\mathcal{H}+\mathcal{H}') = E_\sigma\left[\sup_{h \in \mathcal{H}+\mathcal{H}'} \frac{1}{n}\sum_{i=1}^{n} \sigma_i g(z_i)\right] = E_\sigma\left[\sup_{h \in \mathcal{H}, h' \in \mathcal{H}'} \frac{1}{n}\sum_{i=1}^{n} \sigma_i (h(z_i)+h'(z_i))\right]$$

$$= E_\sigma\left[\sup_{h \in \mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i) + \sup_{h' \in \mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h'(z_i)\right]$$

$$= E_\sigma\left[\sup_{h \in \mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)\right] + E_\sigma\left[\sup_{h' \in \mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h'(z_i)\right]$$

$$= \hat{R}_{S_n}(\mathcal{H}) + \hat{R}_{S_n}(\mathcal{H}')$$

6.1.4.

首先. 显然有: convex-hull$(\mathcal{H}) \supseteq \mathcal{H}$. 由 6.1.1 知, $\hat{R}_{S_n}(\text{convex-hull}(\mathcal{H})) \geq \hat{R}_{S_n}(\mathcal{H})$

而

$$\hat{R}_{S_n}(\text{convex-hull}(\mathcal{H})) = E_\sigma\left[\sup_{h \in \text{convex-hull}(\mathcal{H})} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)\right]$$

由 sub 次可加性

$$\leq E_\sigma\left[\sum_{j=1}^{k} \sup_{h_j \in \mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i \lambda_j h_j(z_i)\right]$$

$$= \sum_{j=1}^{k} \lambda_j E_\sigma\left[\sup_{h_j \in \mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h_j(z_i)\right] = \sum_{j=1}^{k} \lambda_j \hat{R}_{S_n}(\mathcal{H}) = \hat{R}_{S_n}(\mathcal{H})$$

故 $\hat{R}_{S_n}(\text{convex-hull}(\mathcal{H})) = \hat{R}_{S_n}(\mathcal{H})$

6.2

$$\Pi_{\mathcal{H}}(n) = \max_{\{x_1,\cdots,x_n\}\subseteq \mathcal{X}} \left|\{(h(x_1),\cdots,h(x_n)): h\in\mathcal{H}\}\right|$$

$$= \max_{x_1<\cdots<x_n} \left|\{(h_{[a,b]}(x_1),\cdots, h_{[a,b]}(x_n)): h\in\mathcal{H}\}\right|$$

$$= \sum_{k=0}^{n} \left|\{(h_{[a,b]}(x_1),\cdots, h_{[a,b]}(x_n)): x_1\cdots x_n\text{中有}k\text{个在}[a,b]\text{中}\}\right|$$

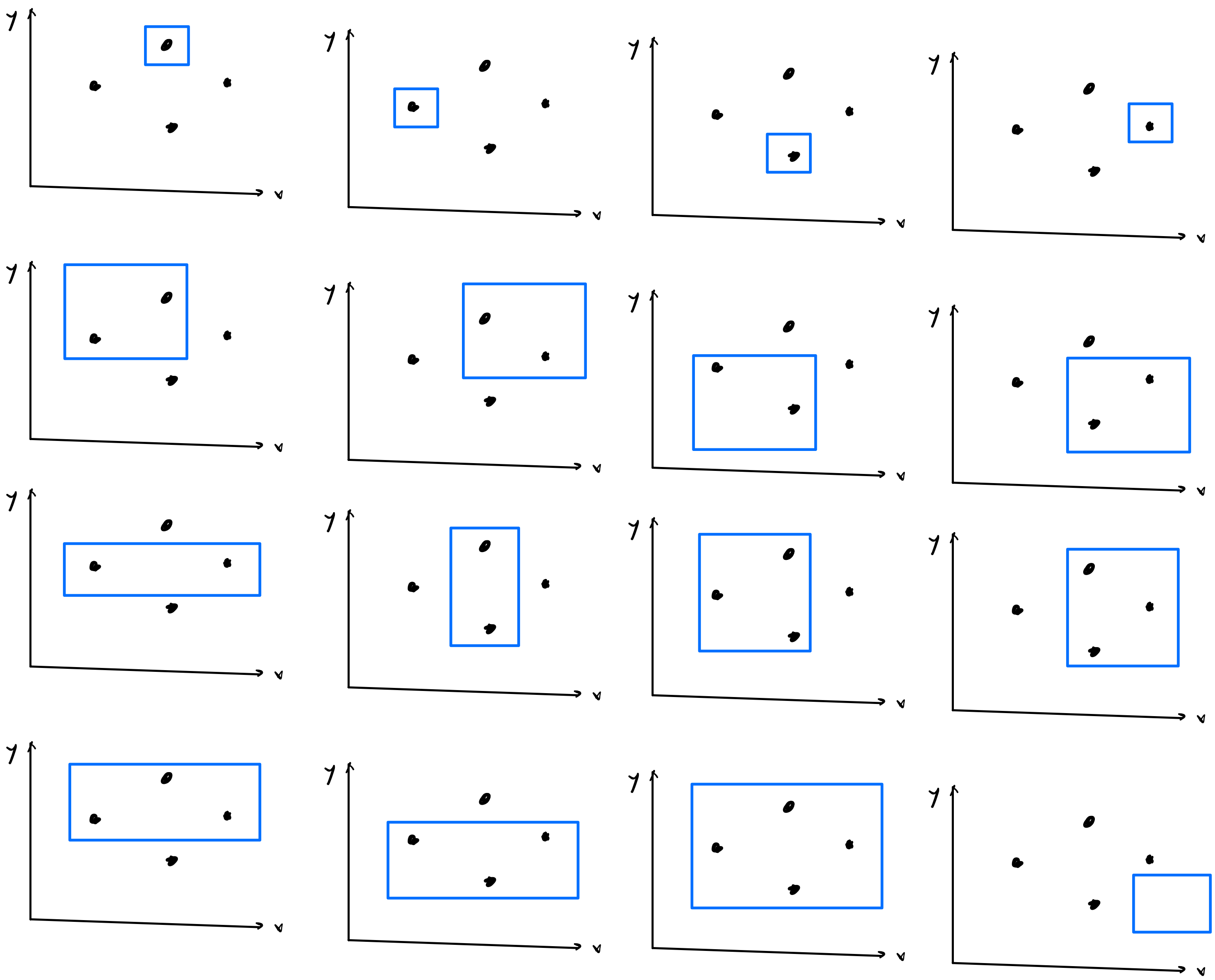
$$= 1+\sum_{k=1}^{n}(n-k+1)$$

$$= 1+(n+1)n-\frac{n(n+1)}{2} = \frac{n^2+n+2}{2}$$

故 $R_n(\mathcal{H}) \le \sqrt{\frac{2\log \Pi_{\mathcal{H}}(n)}{n}} = \sqrt{\frac{2\log\frac{n^2+n+2}{2}}{n}}$

6.3

6.3.1. ① 首先证明 $VCdim(H) \geq 4$. 有以下例子:



由此知, H 可以打散 4 个数据点 $VCdim(H) \geq 4$.

② 接着证明 $VCdim(H) \leq 4$.

若考察 5 个数据点的情形

寻找横轴坐标最小、最大的两个点, 记为 x_1, x_2.

寻找纵轴坐标最小、最大的两个点, 记为 x_3, x_4

余下一点记为 x_5.

考察以下一个分割：$h(x_1) = h(x_2) = h(x_3) = h(x_4) = 1$　$h(x_5) = 0$.

则有　$a \le x_{1.1} < x_{2.1} \le b$.

$c \le x_{3.2} < x_{4.2} \le d$.

则必有　$x_{1.1} \le x_{5,1} \le x_{2.1}$，即 $x_{5.1} \in [a.b]$

$x_{3.2} \le x_{5.2} \le x_{4.2}$. 即 $x_{5.2} \in [c.d]$

即　$h(x_5) = 1$．即以上分割不存在！

故 H 不可打散 5 个及以上的点，$VC\dim(H) \le 4$

结合 ① ②，$VC\dim(H) = 4$.

6.3.2

对任意数据点 $x \in \mathbb{R}$，定义映射 $f(x) = x + k\pi \in [0,\pi]$．$k \in \mathbb{Z}$.

即将 x 映射到 $[0,\pi]$ 上，且有

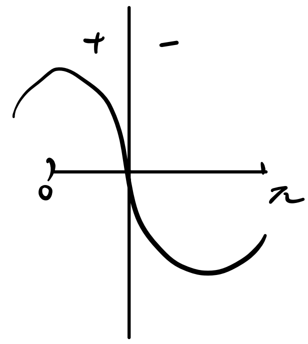
$$h_a(x) = \mathbb{1}[\sin(x+a) > 0] = \begin{cases} \mathbb{1}[\sin(f(x)+a)] = h_a(f(x)) & k = 2t \quad t \in \mathbb{Z} \\ 1 - \mathbb{1}[\sin(f(x)+a)] = 1 - h_a(f(x)) & k = 2t+1 \quad t \in \mathbb{Z} \end{cases}$$

即　$(h(x_1), \cdots, h(x_n))$ 与 $(h(f(x_1)), \cdots, h(f(x_n)))$ 一一映射，

从而有　$$\max_{\{x_1,\cdots,x_n\} \in \mathbb{R}} \left| \{ (h(x_1), \cdots, h(x_n)) : h \in H \} \right|$$

$$= \max_{\{x_1,\cdots,x_n\} \in \mathbb{R}} \left| \{ (h(f(x_1)), \cdots, h(f(x_n))) : h \in H \} \right|$$

考察 $[0,\pi]$ 区间，容易发现 h 将 $f(x_1),\cdots,f(x_n)$ 以一个超平面分割开，由课件上的引理，



$$VC\dim(H) = d+1 = 2$$

即 VC 维为 2.