2.1、$\varepsilon(\hat{r}) > \varepsilon \iff P[C_r \setminus C_{\hat{r}}] > \varepsilon$（误差只出现在 C_r 以内）

$\iff C_r \subset C_{r'}$（几何视角）$\iff \hat{r} < r'$ □

2.2. 由 2.1、$\underset{D_n}{P}[\varepsilon(\hat{r}_{D_n}) > \varepsilon] = \underset{D_n}{P}[\hat{r} < r'] = \underset{D_n}{P}[D_n$ 中每一个正例都在 $C_{r'}$ 内$]$.

而 $\underset{D_n}{P}[$每一个正例都在 $C_{r'}$ 内$] \leq (1-\varepsilon)^n$

故 $\underset{D_n}{P}[\varepsilon(\hat{r}_{D_n}) > \varepsilon] \leq (1-\varepsilon)^n \leq \exp(-n\varepsilon)$

□

2.3. 易知 $\varepsilon(r) = 0$. 由 2.2、只要 $n > \frac{1}{\varepsilon}\log\frac{1}{\delta}$，就有 $\underset{D_n}{P}[\varepsilon(\hat{r}_{D_n}) - \varepsilon(r) < \varepsilon] < \delta$

而 $\frac{1}{\varepsilon}\log\frac{1}{\delta}$ 是 $O(poly(\frac{1}{\varepsilon}, \frac{1}{\delta}))$ 故该问题是 PAC-Learnable 的. □

2.4. $\varepsilon(r) = P[C_r] \cdot \eta$

$\varepsilon(r') = (P[C_r] - \varepsilon)\eta + \varepsilon(1-\eta) = P[C_r] \cdot \eta + \varepsilon(1-2\eta)$

2.5. 由 2.4 知，因为 $1-2\eta > 0$. 故 $\varepsilon(r')$ 关于 ε 单调增. 也就是说，学到的半径越小的 $C_{\hat{r}}$，期望误差越大.

现在固定 $\varepsilon > 0$，令 r'' 是使得 $P[C_r \setminus C_{r''}] = \frac{\varepsilon}{1-2\eta}$（不妨假设 $P[C_r] > \frac{\varepsilon}{1-2\eta}$，不然，对于这个 ε，一定有 $\varepsilon(\hat{r}) - \varepsilon(r) \leq P[C_r] \cdot (1-2\eta) \leq \varepsilon$. 自然满足 PAC-可学的条件）.

由上分析可知 $\varepsilon(\hat{r}) - \varepsilon(r) < \varepsilon \iff \hat{r} < r''$

故 $P[\varepsilon(\hat{r}) - \varepsilon(r) < \varepsilon] = P[\hat{r} < r''] \leq (1 - \frac{\varepsilon}{1-2\eta})^n \leq (1 - \frac{\varepsilon}{1-2\eta'})^n$

$\leq \exp(\frac{-n\varepsilon}{1-2\eta'})$. 故只要 $n > \frac{1}{\varepsilon}\log\frac{1}{\delta} \cdot (1-2\eta')$，就有 $P[\varepsilon(\hat{r}) - \varepsilon(r) < \varepsilon] < \delta$.

这也满足 PAC-Learnable 的条件.

□

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

3.2. $$\max_{1\le i\le T} \mathbb{E}_{D_n^{(i)}}[L_{D^{(i)}}(A(D_n^{(i)}))]$$

$$\geq \frac{1}{T}\sum_{i=1}^{T} \mathbb{E}_{D_n^{(i)}}[L_{D^{(i)}}(A(D_n^{(i)}))] \qquad (\max \geq \text{avg})$$

$$\geq \frac{1}{T}\sum_{i=1}^{T} \min_{D_n^{(i)}} L_{D^{(i)}}(A(D_n^{(i)})) \qquad (\text{avg} \geq \min)$$

而因为 $\{D_n^{(i)}\} \subset \{S_j^{(i)}\}_{j=1}^{k}$，（这里 $T = 2^{2n}$，$k = (2n)^n$）（也就是说从 $\{S_j^{(i)}\}_{j=1}^{k}$ 里选最小的，至少能选到与 $\{D_n^{(i)}\}$ 中最小的一样小的）

故 $$\frac{1}{T}\sum_{i=1}^{T} \min_{D_n^{(i)}} L_{D^{(i)}}(A(D_n^{(i)})) \geq \frac{1}{T}\sum_{i=1}^{T} \min_{S_j^{(i)}} L_{D^{(i)}}(A(S_j^{(i)}))$$

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

□

3.3. 先证明 $L_{D_i}(h) \geq \frac{1}{2p}\sum_{r=1}^{p} \mathbb{1}[h(v_r) \neq f_i(v_r)]$. where $h = A(S_j^{(i)})$

proof: $$L_{D_i}(h) = \frac{1}{2n}\sum_{j=1}^{2n} \mathbb{1}[h(x_j) \neq f_i(x_j)]$$

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

然后证明原命题．两边对 i 从 1 到 T 求和有（然后取平均）：

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

□

3.4. $$\max_{1\le i\le T} \mathbb{E}_{D_n^{(i)}}[L_{D^{(i)}}(A(D_n^{(i)}))] \geq \frac{1}{T}\sum_{i=1}^{T} \min_{1\le j\le k} L_{D^{(i)}}(A(S_j^{(i)})) \geq \frac{1}{2pT}\sum_{i=1}^{T}\sum_{r=1}^{p} \mathbb{1}[h(v_r) \neq f_i(v_r)] \quad (*)$$

因为 f_i 和 $f_{i'}$ 在 S_j 上的标签完全相同，故 $S_j^{(i)} = S_j^{(i')}$，又算法返回的 h 只由 $S_j^{(i)}$ 决定，故

$$\mathbb{1}_{[A(S_j^{(i)})(v_r) \neq f_i(v_r)]} + \mathbb{1}_{[A(S_j^{(i')})(v_r) \neq f_{i'}(v_r)]} = 1. \quad \text{for any } r \in [1, p]$$

（这是说，对于 $v_1, \cdots, v_p$ 中任意一个点，$A(S_j^{(i)})$ 和 $A(S_j^{(i')})$ 总有一个将其分类正确，另一个分类错误）

故 $$\frac{1}{2pT}\sum_{r=1}^{p}\sum_{i=1}^{T} \mathbb{1}[h(v_r) \neq f_i(v_r)] = \frac{1}{2pT}\sum_{r=1}^{p} \frac{T}{2} = \frac{1}{4}.$$ （一共分成了 $\frac{T}{2}$ 个 (i, i') 组）

代入 (*) 式即证

□

3.5 由3.4，我们有 $\max_{1\le i\le T} \mathbb{E}_{D_n^i}[\mathcal{L}(A(D_n^{(i)}))] \ge \frac{1}{4}$. 不妨设取到max的下标为 i，即

$\mathbb{E}_{D_n^i}[\mathcal{L}(A(D_n^{(i)}))] \ge \frac{1}{4}$. 因为 $\mathcal{L}(A(D_n^{(i)})) \in [0,1]$，故 $P_{D_n^i}[\mathcal{L}(A(D_n^{(i)})) > \frac{1}{8}]$

$$= \frac{\mathbb{E}_{D_n^i}[\mathcal{L}(A(D_n^{(i)}))] - \frac{1}{8}}{\frac{7}{8}} \ge \frac{1}{7}$$ (在Lemma中令 $a=\frac{1}{8}$)

也就是说，总存在一个分布 D^i，对于 D_n^i 的选择，有至少 $\frac{1}{7}$ 的概率使得 $\mathcal{L}(A(D_n^{(i)})) \ge \frac{1}{8}$

□

3.6. 假设该概念类是可学习的，那么对于 $\mathcal{X}$ 上的任意分布 D，任意的标签函数 f，对于任意的 $\varepsilon>0$，$\delta>0$，只要 $n > poly(\frac{1}{\varepsilon}, \frac{1}{\delta})$，就有 $P(\mathcal{L}(A(D_n)) > \varepsilon) < \delta$

然而由No-Free-Lunch Thm.，因为 $\frac{|\mathcal{X}|}{2} > n$，故总能找到一个分布 D 和一个标签函数 f_i，使得 $P(\mathcal{L}(A(D_n^i)) > \frac{1}{8}) > \frac{1}{7}$. 这与 ε, δ 的任意性矛盾.

□

4.1 $P(\exists i, 1\le i\le k, \|\hat{P}_i - P_i\|_1 > \varepsilon)$

$= P(\|\hat{P}_1 - P_1\|_1 > \varepsilon \vee \cdots \vee \|\hat{P}_k - P_k\|_1 > \varepsilon)$

$\le \sum_{i=1}^{k} P(\|\hat{P}_i - P_i\|_1 > \varepsilon)$ (Union Bound)

而对于 $\forall i$, $1\le i\le k$，有

$P(\|\hat{P}_i - P_i\| > \varepsilon) \le P(|\hat{P}_{i,1} - P_{i,1}| > \frac{\varepsilon}{m} \vee |\hat{P}_{i,2} - P_{i,2}| > \frac{\varepsilon}{m} \vee \cdots \vee |\hat{P}_{i,m} - P_{i,m}| > \frac{\varepsilon}{m})$

(至少有一项要大于 $\frac{\varepsilon}{m}$)

$\le \sum_{j=1}^{m} P(|\hat{P}_{i,j} - P_{i,j}| > \frac{\varepsilon}{m})$ (Union Bound)

$= \sum_{j=1}^{m} P\left(\left|\frac{1}{n}\sum_{s=1}^{n} \mathbb{1}[X_i^{(s)}=j] - \mathbb{E}[\mathbb{1}[X_i^{(s)}=j]]\right| > \frac{\varepsilon}{m}\right)$

$\le \sum_{j=1}^{m} 2e^{-\frac{2n\varepsilon^2}{m^2}}$ (Hoeffding's ineq.)

$= 2me^{-\frac{2n\varepsilon^2}{m^2}}$. 故 $P(\exists i, 1\le i\le k, \|\hat{P}_i - P_i\|_1 > \varepsilon) \le 2mke^{-\frac{2n\varepsilon^2}{m^2}}$.

令 $2mke^{-\frac{2n\varepsilon^2}{m^2}} = \delta$. 得 $\varepsilon = m\sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}$.

即 $\exists i$, s.t. $\|\hat{p}_i - p_i\|_1 > m\sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}$ 的概率不超过 δ

即 $\forall i$, $\|\hat{p}_i - p_i\|_1 \le m\sqrt{\frac{1}{2n}\log\frac{2mk}{\delta}}$ 的概率至少为 $1-\delta$

$\square$

4.2. 我们直接 Bound $P(\|\hat{p}_i - p_i\|_1 > \varepsilon)$

$$P(\|\hat{p}_i - p_i\|_1 > \varepsilon) = P\left(\sup_{u\in\{\pm1\}^m} u^T(\hat{p}_i - p_i) > \varepsilon\right) = P\left(u_1^T(\hat{p}_i - p_i) > \varepsilon \vee u_2^T(\hat{p}_i - p_i) > \varepsilon \vee \cdots\right)$$

$$\le \sum_{r=1}^{2^m} P(u_r^T(\hat{p}_i - p_i) > \varepsilon) \quad (\text{Union Bound})$$

$$= \sum_{r=1}^{2^m} P\left(\sum_{j=1}^m u_{r,j}(\hat{p}_{i,j} - \mathbb{E}[\hat{p}_{i,j}]) > \varepsilon\right) = \sum_{r=1}^{2^m} P\left(\sum_{j=1}^m u_{r,j}\left(\sum_{s=1}^n \frac{1}{n}\mathbb{1}[X_i^{(s)}=j] - \mathbb{E}[\mathbb{1}[X_i^{(s)}=j]]\right) > \varepsilon\right)$$

$$= \sum_{r=1}^{2^m} P\left(\frac{1}{n}\sum_{s=1}^n \sum_{j=1}^m u_{r,j}\mathbb{1}[X_i^{(s)}=j] - \sum_{j=1}^m u_{r,j}\mathbb{E}[\mathbb{1}[X_i^{(s)}=j]] > \varepsilon\right) \quad (\text{交换}\textstyle\sum\text{顺序})$$

$$= \sum_{r=1}^{2^m} P\left(\frac{1}{n}\sum_{s=1}^n \left(\sum_{j=1}^m u_{r,j}\mathbb{1}[X_i^{(s)}=j]\right) - \mathbb{E}\left[\sum_{j=1}^m u_{r,j}\mathbb{1}[X_i^{(s)}=j]\right] > \varepsilon\right) \quad (\text{期望线性性})$$

记 $Y_s = \sum_{j=1}^m u_{r,j}\mathbb{1}[X_i^{(s)}=j]$. $Y_s \in [-1, +1]$.

$$= \sum_{r=1}^{2^m} P\left(\frac{1}{n}\sum_{s=1}^n Y_s - \mathbb{E}[Y_s] > \varepsilon\right) \le \sum_{r=1}^{2^m} e^{-\frac{2n\varepsilon^2}{4}} \overset{\text{Hoeffding's ineq.}}{=} \sum_{r=1}^{2^m} e^{-\frac{n\varepsilon^2}{2}} = 2^m e^{-\frac{n\varepsilon^2}{2}}.$$

故 $P(\exists i, 1\le i\le k, \|\hat{p}_i - p_i\|_1 > \varepsilon) \le k\cdot 2^m \cdot e^{-\frac{n\varepsilon^2}{2}}$

令 $k\cdot 2^m\cdot e^{-\frac{n\varepsilon^2}{2}} = \delta$ 即得结果!

$\square$

5、Hoeffding's Ineq. 给出的界为：

$$P(\hat{\varepsilon}_{D_n}(h)-\varepsilon(h)>\varepsilon)\le 2e^{-2n\varepsilon^2}$$ (For a certain h) . 令 $2e^{-2n\varepsilon^2}=\delta\cdot p(h)$. 得 $\varepsilon=\sqrt{\frac{\log\frac{1}{p(h)}+\log\frac{2}{\delta}}{2n}}$

也即对于任意一个(确定的) $h\in\mathcal{H}$. 至少以 $1-\delta p(h)$ 的概率成立 $\hat{\varepsilon}_{D_n}(h)-\varepsilon(h)\le\sqrt{\frac{\log\frac{1}{p(h)}+\log\frac{2}{\delta}}{2n}}$

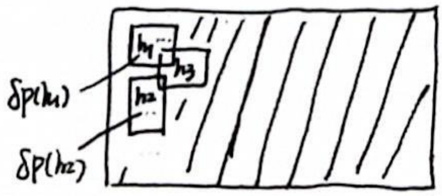
也即对于某一个 $h\in\mathcal{H}$. 至多有 $\delta p(h)$ 的概率成立 $\varepsilon(h)-\hat{\varepsilon}_{D_n}(h)>\sqrt{\frac{\log\frac{1}{p(h)}+\log\frac{2}{\delta}}{2n}}$

故至多有 $\sum_h\delta p(h)=\delta$ 的概率

成立 $\exists h\in\mathcal{H}$, $\hat{\varepsilon}_{D_n}(h)-\varepsilon(h)>\sqrt{\frac{\log\frac{1}{p(h)}+\log\frac{2}{\delta}}{2n}}$ (Union Bound)

也即至少有 $1-\delta$ 的概率，对于 $\forall h\in\mathcal{H}$, $\hat{\varepsilon}_{D_n}(h)-\varepsilon(h)\le\sqrt{\frac{\log\frac{1}{p(h)}+\log\frac{2}{\delta}}{2n}}$ $\square$

与课件上给出的误差界进行比较：



$P(\text{阴影})\ge 1-\delta[p(h_1)+\cdots+p(h_{|\mathcal{H}|})]$
$=1-\delta$

1、首先，二者都是误差界的多假设情况.

即以 $1-\delta$ 的概率一致地成立 $\hat{\varepsilon}_{D_n}(h)\le\varepsilon(h)+\sqrt{\frac{\log\frac{1}{p(h)}+\log\frac{2}{\delta}}{2n}}$
(对任意 $h\in\mathcal{H}$)

2、其次，二者的 bounding 一个与 h 本身的性质(测度)无关，一个则与 $p(h)$ 有关. 也即虽然 bound 是一致成立的，但每个 h 的 bound 不同，在这道习题给出的结果里

3、如果在 $\mathcal{H}$ 中每个 h 的测度都是一样的(即 $\frac{1}{|\mathcal{H}|}$)，那么二者是一样的.
可以将课件上结果看作本题结果中，每个 h 被选择概率一样的例子

6.1. 先证明经验 Rademacher 复杂度的下述性质

1. 若 $\mathcal{H} \subseteq \mathcal{H}'$，则 $\hat{R}_{S_n}(\mathcal{H}) \le \hat{R}_{S_n}(\mathcal{H}')$

proof: $\hat{R}_{S_n}(\mathcal{H}) = \mathbb{E}_{\sigma}\left[\sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i h(x_i)\right] \le \mathbb{E}_{\sigma}\left[\sup_{h\in\mathcal{H}'} \frac{1}{n}\sum_{i=1}^{n}\sigma_i h(x_i)\right] = \hat{R}_{S_n}(\mathcal{H}')$

2. $\hat{R}_{S_n}(\alpha\mathcal{H}) = |\alpha|\hat{R}_{S_n}(\mathcal{H})$，其中 $\alpha\mathcal{H} = \{\alpha h \mid \forall h\in\mathcal{H}\}$

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

$$= \begin{cases} \mathbb{E}_{\sigma}\left[\alpha \sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i h(x_i)\right], & \alpha \ge 0 \\ \mathbb{E}_{\sigma}\left[\alpha \inf_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i h(x_i)\right], & \alpha < 0 \end{cases}$$

而由 σ 的对称性，$\mathbb{E}_{\sigma}\left[\inf_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i h(x_i)\right] = -\mathbb{E}_{\sigma}\left[\sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i h(x_i)\right]$

(对于某个 σ，当找到了某个 $h\in\mathcal{H}$ s.t. $\sum\sigma_i h(x_i)$ 最小时，其恰好也是使 $\sigma' = -\sigma$ 时，$\sum\sigma_i h(x_i)$ 最大的那个 h，而且这时 $\sum\sigma_i h(x_i) = -\sum\sigma_i' h(x_i)$)

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

3. $\hat{R}_{S_n}(\mathcal{H}+\mathcal{H}') \le \hat{R}_{S_n}(\mathcal{H}) + \hat{R}_{S_n}(\mathcal{H}')$. 其中 $\mathcal{H}+\mathcal{H}' = \{h+h' \mid \forall h\in\mathcal{H}, \forall h'\in\mathcal{H}'\}$

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

$$= \mathbb{E}_{\sigma}\left[\sup_{g\in\mathcal{H}+\mathcal{H}'} \frac{1}{n}\sum_{i=1}^{n}\sigma_i h(x_i) + \frac{1}{n}\sum_{i=1}^{n}\sigma_i h'(x_i)\right]$$

$$\le \mathbb{E}_{\sigma}\left[\sup_{g\in\mathcal{H}+\mathcal{H}'} \frac{1}{n}\sum_{i=1}^{n}\sigma_i h(x_i)\right] + \mathbb{E}_{\sigma}\left[\sup_{g\in\mathcal{H}+\mathcal{H}'} \frac{1}{n}\sum_{i=1}^{n}\sigma_i h'(x_i)\right] \quad \text{(sup的半可加性)}$$

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

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

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

proof: $\hat{R}_{S_n}(\text{convex-hull}(\mathcal{H})) = \mathbb{E}_{\sigma}\left[\sup_{h\in \text{convex-hull}(\mathcal{H})} \frac{1}{n}\sum_{i=1}^{n}\sigma_i h(x_i)\right] = \mathbb{E}_{\sigma}\left[\sup_{\sum\lambda_j h_j \in \text{convex-hull}(\mathcal{H})} \frac{1}{n}\sum_{i=1}^{n}\sigma_i \sum_{j=1}^{k}\lambda_j h_j(x_i)\right]$

$= \mathbb{E}_{\sigma}\left[\sum_{j=1}^{k} \sup_{\lambda_j\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i\lambda_j h_j(x_i)\right]$ ($\sum$取最大 $\Leftrightarrow$ 求和的每一项都取最大)

$= \sum_{j=1}^{k}\lambda_j \mathbb{E}_{\sigma}\left[\sup_{\lambda_j\in\mathcal{H}}\sum_{i=1}^{n}\sigma_i h_j(x_i)\right]$ (期望的线性性, $\lambda_j>0$)

$= \sum_{j=1}^{k}\lambda_j\cdot\hat{R}_{S_n}(\mathcal{H}) = \hat{R}_{S_n}(\mathcal{H})$

然后, 因为 $R_n(\mathcal{H}) = \mathbb{E}_{S_n\sim D^n}[\hat{R}_{S_n}(\mathcal{H})]$, 在1.中应用"若$X\leq Y$,那么$\mathbb{E}X\leq\mathbb{E}Y$".
在2.中应用"$\mathbb{E}[aX]=a\mathbb{E}X$", 在3.中应用期望的线性性和1, 即可得到期望 Rademacher 复杂度的如上性质.

□

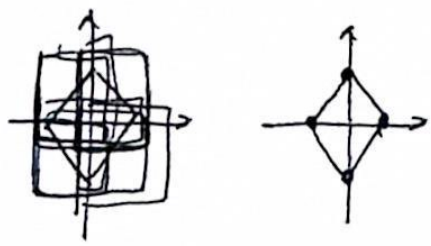
6.2

对于实轴上的n个点, 我们可以从中选择连续的k个点, 将其标为正例, 这有

$$\underbrace{1}_{(k=0)} + \sum_{k=1}^{n}(n-k+1) = 1+\frac{n(n+1)}{2}$$ 也即 $\Pi_{\mathcal{H}}(n) = 1+\frac{n(n+1)}{2}$

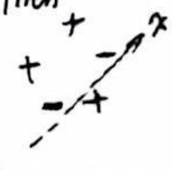
利用 Growth Function Bound, 我们可以得到 $R_n(\mathcal{H}) \leq \sqrt{\frac{2\log\left[1+\frac{n(n+1)}{2}\right]}{n}}$

6.3 1、$VCdim(\mathcal{H}) = 4$. 下界:



上界: 可以利用几何知识证明$\mathcal{H}$无法打散任一个五边形. 大致思路为:

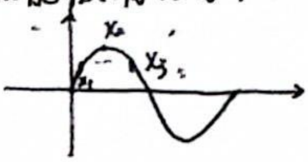
If 

Then 

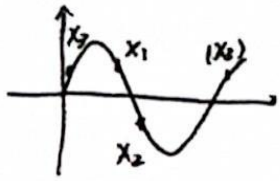
?

2、VCdim(H) = 2

只需考虑三个点的相位($\in[0,2\pi)$)，如果三个点能被标记为{+,+,+}，我们不妨按相位，将它们排序为 x_1, x_2, x_3；(实际上它们的位置顺序可能不同，这里只是表明它们的相位)。



现在，假设能将 x_1, x_2, x_3 分类为{+,-,+}，三者相位一定为：



矛盾！x_3 领先了 x_1 大于 π 的相位。

而2个点的情况是容易构造的