Homework #2

1. $\lambda\mu+(1-\lambda)(1-\mu)$

由题可知，f 是确定的目标函数，没有噪声。如果给 f 加上噪声，$P(y|x)=\begin{cases}\lambda & y=f(x)\\ 1-\lambda & y\neq f(x)\end{cases}$，若不加入噪声，$P(y|x)=\begin{cases}1 & y=f(x)\\ 0 & y\neq f(x)\end{cases}$。

不加噪声时，hypothesis h 的错误概率（泛化误差）为μ，即：

$$\begin{aligned}\mu &= \int_{\mathcal{X}\times\mathcal{Y}} P(x,y)\,[\![y\neq h(x)]\!]dxdy\\ &= \int_{\mathcal{X}\times\mathcal{Y}} P(x)P(y|x)[\![y\neq h(x)]\!]dxdy\\ &= \int_{\mathcal{X}\times\mathcal{Y}} P(x)[P(y=f(x)|x)+P(y\neq f(x)|x)][\![y\neq h(x)]\!]dxdy\\ &= \int_{\mathcal{X}\times\mathcal{Y}} P(x)P(y=f(x)|x)[\![y\neq h(x)]\!]dxdy\\ &= \int_{\mathcal{X}\times\mathcal{Y}} P(x)[\![f(x)\neq h(x)]\!]dxdy\end{aligned}$$

$$1-\mu=\int_{\mathcal{X}\times\mathcal{Y}} P(x)[\![f(x)=h(x)]\!]dxdy$$

设有噪声的情况下错误概率为μ'，

$$\begin{aligned}\mu' &= \int_{\mathcal{X}\times\mathcal{Y}} P(x,y)\,[\![y\neq h(x)]\!]dxdy\\ &= \int_{\mathcal{X}\times\mathcal{Y}} P(x)P(y|x)[\![y\neq h(x)]\!]dxdy\\ &= \int_{\mathcal{X}\times\mathcal{Y}} P(x)[P(y=f(x)|x)+P(y\neq f(x)|x)][\![y\neq h(x)]\!]dxdy\\ &= \int_{\mathcal{X}\times\mathcal{Y}} P(x)P(y=f(x)|x)[\![y\neq h(x)]\!]dxdy+\int_{\mathcal{X}\times\mathcal{Y}} P(x)P(y\neq f(x)|x)[\![y\neq h(x)]\!]dxdy\end{aligned}$$

y、f(x)和 h(x)的取值只能是 1 或-1，$y\neq f(x)$时，$[\![y\neq h(x)]\!]=[\![f(x)=h(x)]\!]$，

$$\mu'=\int_{\mathcal{X}\times\mathcal{Y}} P(x)P(y=f(x)|x)[\![y\neq h(x)]\!]dxdy$$

$$+\int_{\mathcal{X}\times\mathcal{Y}} P(x)P(y\neq f(x)|x)[\![f(x)=h(x)]\!]dxdy$$

将有噪声情况下$P(y|x)$的值代入公式，

$$
\begin{aligned}
\mu' &= \int_{\mathcal{X}\times\mathcal{Y}} \mathrm{P(x)}\cdot\lambda\cdot[\![\mathrm{f(x)} \neq \mathrm{h(x)}]\!]dxdy + \int_{\mathcal{X}\times\mathcal{Y}} \mathrm{P(x)}\cdot(1-\lambda)\cdot[\![\mathrm{f(x)} = \mathrm{h(x)}]\!]dxdy \\
&= \lambda\int_{\mathcal{X}\times\mathcal{Y}} \mathrm{P(x)}[\![\mathrm{f(x)} \neq \mathrm{h(x)}]\!]dxdy + (1-\lambda)\int_{\mathcal{X}\times\mathcal{Y}} \mathrm{P(x)}[\![\mathrm{f(x)} = \mathrm{h(x)}]\!]dxdy \\
&= \lambda\mu + (1-\lambda)(1-\mu)
\end{aligned}
$$

2. $\lambda = 0.5$

$\mu' = \lambda\mu + (1-\lambda)(1-\mu) = 1-\lambda+\mu(2\lambda-1)$，$\mu'$与$\mu$无关，则$\lambda = 0.5$

3. 453000

$$\mathbb{P}_{\mathcal{D}}\left[\left|E_{\text{in}}(g) - E_{\text{out}}(g)\right| > \epsilon\right] \quad \leq \quad 4(2N)^{d_{\text{vc}}}\exp\left(-\tfrac{1}{8}\epsilon^2 N\right)$$

带入公式计算就行了。

4. Devroye

计算结果：

Original VC bound : 0.632

Variant VC bound : 0.860

Rademacher Penalty Bound : 0.331

Parrondo and Van den Broek : 0.224

Devroye : 0.215

5. Parrondo and Van den Broek

计算结果：

Original VC bound : 13.828

Variant VC bound : 16.264

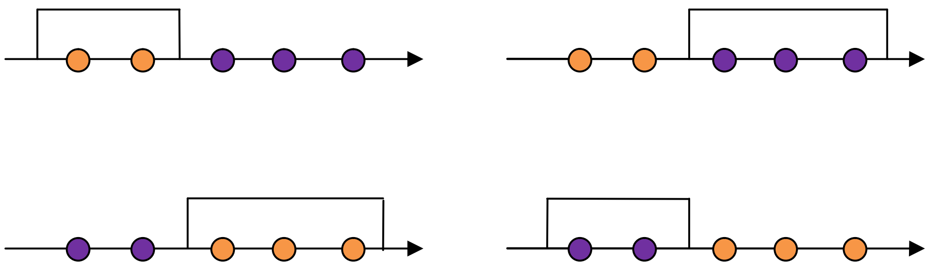
Rademacher Penalty Bound : 7.048

Parrondo and Van den Broek : 5.101

Devroye : 5.593

6. $N^2 - N + 2$

N 个点将实数轴划分为 N+1 个区间，每个区间任取一点，再从 N+1 个点中任取 2 个点构成区间，区间内部点记作+1，共有C_{N+1}^2 种分类；区间内部记作-1，也有C_{N+1}^2 种分类。这两种方式的分类结果中有重复的情形，共 2(N-1)种。



7. VC-dimension is 3

N	1	2	3	4
$N^2 - N + 2$	2	4	8	14
2^N	2	4	8	16

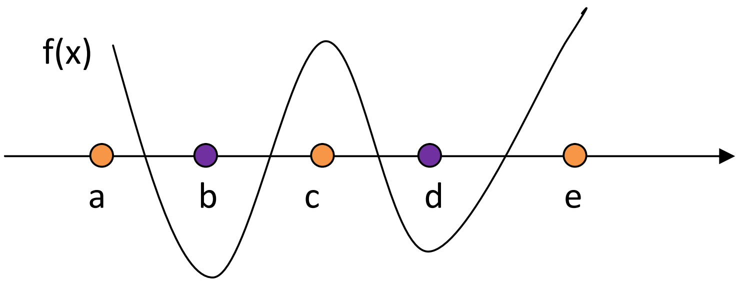
8. $m_{\mathcal{H}}(N) = C_{N+1}^2 + 1$

和课上讲的 positive internal 问题类似。

9. D+1

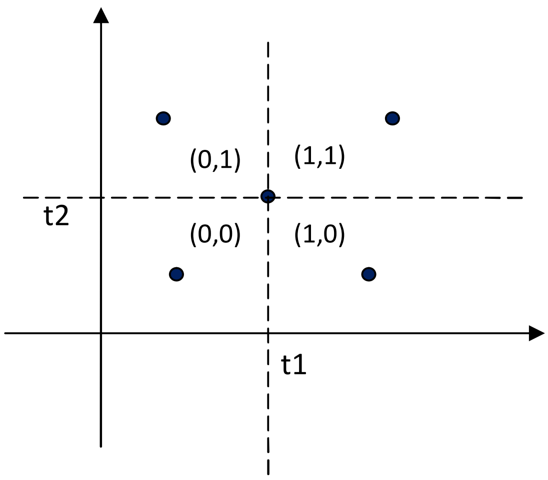
n 次多项式至多有 n 个实根。若 N=n+1，可以构造 n 次多项式将数据打散(shattering)。若 N=n+2，存在至少 2 种 n 次多项式无法实现的分类结果。

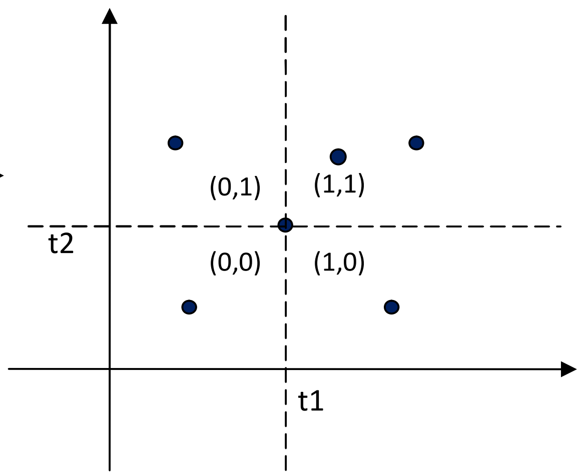
例如，当 N=5，最复杂的分类是相邻两点符号不同，如下图所示。不妨假设黄色点代表点的符号为+1，紫色为-1：



若一种函数可以实现这种分类结果，那么此函数至少有 4 个零点，也就是至少四个实根。对于多项式而言，至少是 4 次多项式。设 a<m<b<n<c<s<d<t<e，令$f(x) = (x-m)(x-n)(x-s)(x-t)$，sign(f(x))可以实现图中分类结果。

10. 2^d





假设 sign(0)=-1,hypothesis 函数可以理解为

$$h_{t,S}(x) = 2[\![(\mathrm{sign}(x-t)) \in S]\!] - 1 = \begin{cases} 1 & sign(x-t) \in S \\ -1 & sign(x-t) \notin S \end{cases}$$

以 d=2 为例，设 t=(t1,t2)，$S \subseteq \{0,1\}^2 = \{(0,0),(0,1),(1,0),(1,1)\}$，

sign(x-t)将 2 维平面划分为$2^d = 4$个部分，每个部分有相同运算结果，分别是

(0,0)、(0,1)、(1,0)、(1,1)，集合 S 的作用是决定这 4 个部分的标记是 1 还是-1。

因此，当 N=4 时，选择合适的 t 使得 4 个点落在不同部分，不同的 S 产生不同

的分类，可以打散(shattering)；当 N=5 时，至少有两个点落在同一个部分，

所以$m_{\mathcal{H}}(5) < 2^5$。综上，d=2，VC dimension 为$2^d = 4$。

同理可知，d 维数据的 VC dimension 是2^d。

11. VC dimension 是无穷大

hypothesis 函数图像是方波，α可以控制波的周期。

暂时不知道怎么证。

12. $\min_{1 \le i \le N-1} 2^i m_{\mathcal{H}}(N-i)$ 是$m_{\mathcal{H}}(N)$ 的上界

根据$m_{\mathcal{H}}(N)$的含义可知，$m_{\mathcal{H}}(N) \le 2m_{\mathcal{H}}(N-1)$，以此类推：

$$2m_{\mathcal{H}}(N-1) \le 2^2 m_{\mathcal{H}}(N-2) \le \cdots \le 2^{N-1} m_{\mathcal{H}}(1)$$

那么：$2m_{\mathcal{H}}(N-1) = \min_{1 \le i \le N-1} 2^i m_{\mathcal{H}}(N-i)$。得证。

13. $2^{\lfloor\sqrt{N}\rfloor}$不是增长函数

	X1	X2	X3	X4
a	\	\	O	O

b	\	\	O	X
c	\	\	X	O
d	\	\	X	X

注：$\mathcal{H}$对 X1 X2 X3 的标记，

a 和 b 相同，c 和 d 相同。

设假设空间$\mathcal{H}$中的 hypothesis 能够将数据分为 O、X 两类，$m_{\mathcal{H}}(N) = 2^{\lfloor\sqrt{N}\rfloor}$。

当 N=4 时，$2^{\lfloor\sqrt{N}\rfloor} = 4$。设$\mathcal{H}$可以对 X1、X2、X3 和 X4 赋予 a b c d 4 种不同标记。而当 N=3 时，$2^{\lfloor\sqrt{N}\rfloor} = 2$。可以推断出：

（1）$\mathcal{H}$对 X1 X2 X3 赋予了 2 种不同标记。在这 2 种标记中，X1 X2 X3 中至少一个既标记为 O，又标记为 X，假定 X3 可以，如表中所示。

（2）无论 X1 X2 X3 是哪种标记，X4 都可以标记为 O 和 X，如表中所示。

所以，从表中可以看出，$\mathcal{H}$可以对 X3 和 X4 赋予 4 种不同标记。这与 N=2 时，$2^{\lfloor\sqrt{N}\rfloor} = 2$矛盾。

14. $0 \le d_{vc}\left(\cap_{k=1}^{K} \mathcal{H}_k\right) \le \min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$

下边界：若$\cap_{k=1}^{K} \mathcal{H}_k$ 是空集，VC dimension 最小，为 0。

上边界：$\cap_{k=1}^{K} \mathcal{H}_k$ 是任一 hypothesis set 的子集。

15.$\max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \le d_{vc}\left(\cup_{k=1}^{K} \mathcal{H}_k\right) \le K - 1 + \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$

下边界：任一 hypothesis set 是$\cup_{k=1}^{K} \mathcal{H}_k$ 的子集

上边界：

先考虑 $K=2$，设 $H = H_1 \cup H_2$. $d_{vc}(H_1) = m$

$d_{vc}(H_2) = n$.

从增长函数定义可知:

$$m_H(N) \le m_{H_1}(N) + m_{H_2}(N) \qquad (1)$$

当 $N \ge m$ 且 $N \ge n$ 时,

$$
\begin{aligned}
\text{式(1)} &\le \sum_{i=0}^{m} \binom{N}{i} + \sum_{i=0}^{n} \binom{N}{i} \\
&= \sum_{i=0}^{m} \binom{N}{i} + \sum_{i=0}^{n} \binom{N}{N-i} \\
&= \sum_{i=0}^{m} \binom{N}{i} + \sum_{i=N-n}^{N} \binom{N}{i} \qquad (2)
\end{aligned}
$$

当 $N = m+n+1$ 时,

$$\text{式(2)} = \sum_{i=0}^{m} \binom{N}{i} + \sum_{i=m+1}^{N} \binom{N}{i} = 2^N$$

即 $m_H(N) \le 2^N$

当 $N = m+n+2$ 时,

$$\text{式(2)} = \sum_{i=0}^{m} \binom{N}{i} + \sum_{i=m+2}^{N} \binom{N}{i} = 2^N - \binom{N}{m+1} < 2^N$$

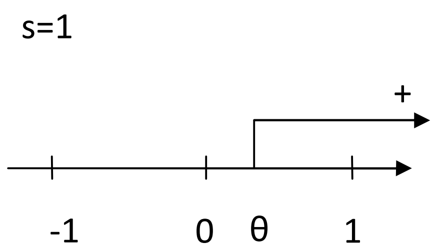
即 $m_H(N) < 2^N$.

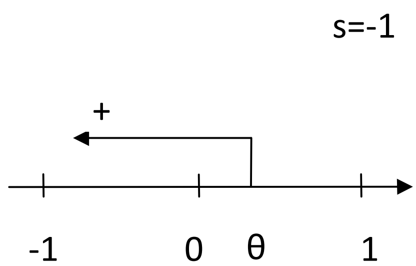
综上所述. $d_{vc}(H) \le m+n+1$.

以此类推，可证明上边界。

16. $0.5 + 0.3s(|\theta| - 1)$

由第一题的结论可知，$E_{out}(h_{s,\theta}) = \lambda\mu + (1-\lambda)(1-\mu)$，其中，$\lambda = 0.8$，$\mu$是当目标函数为$\tilde{s} = \mathrm{sign}(x)$，没有噪声时$h_{s,\theta}$的错误概率。





可以求得$\mu=\begin{cases}\frac{|\theta|}{2} & s=1\\ 1-\frac{|\theta|}{2} & s=-1\end{cases}$，合并起来，$\mu=\frac{|\theta|}{2}\left(\frac{1+s}{2}\right)+\left(1-\frac{|\theta|}{2}\right)(\frac{1-s}{2})$。

17 至 20：编程题