蔣藝萱 r07944046

2. $\because$ 当 $(N+\ell)$ 是 even 时，$g(x_{N+\ell}) \neq f(x_{N+\ell})$

$[\![g(x_{N+\ell}) \neq f(x_{N+\ell})]\!] = 1$

又 $\because$ $\{x_1, x_2 \cdots x_N\}$ 中 x_k 的 k 是 even 的个数 $= \lfloor \frac{N}{2} \rfloor$

$\{x_1, x_2 \cdots x_{N+L}\}$ 中 k 是 even 的个数 $= \lfloor \frac{N+L}{2} \rfloor$

$\therefore E_{OTS}(g,f) =$

$\frac{1}{L} \cdot (\{x_{N+1} \cdots x_{N+L}\}$ 中 k 是 even 的个数$)$

$= \frac{1}{L} \cdot [(\{x_1 \cdots x_{N+L}\}$ 中 k 是 even 的个数$) - (\{x_1 \cdots x_N\}$ k 是 even 的个数$)]$

$= \frac{1}{L} \cdot (\lfloor \frac{N+L}{2} \rfloor - \lfloor \frac{N}{2} \rfloor)$

3. $\because$ f 在 $D = \{(x_n, y_n)\}_{n=1}^{N}$ 上，$f(x_n) = y_n$.

但在 $\{x_{N+1} \cdots x_{N+L}\}$ 上，$f(x_n) = 1$ 或 -1

$\therefore$ 根据排列组合，f 有 2^L 个组合.

$\because$ 对于任意一个 A，$A(D) = g$

設 $Q = \sum_{\ell=1}^{L} [g(x_{N+\ell}) \neq f(x_{N+\ell})]$

存在一个 f 使得 Q 等于 $[0, L]$ 中任意一个整数.

$\therefore$ 当 $Q = n$ 时，$n \in [0, L]$

使得 $Q = n$ 的 f 的个数为 C_n^L $\left(C_n^L = \binom{L}{n}, \text{combination} \right)$

$\because$ f are equally likely in probability

$\therefore$ 每个 f 的机率为 $\frac{1}{2^L}$

$\therefore$

$E_f\{E_{OTS}(A(D), f)\} = \frac{1}{2^L} \cdot C_0^L \cdot \frac{1}{L} \cdot 0 + \frac{1}{2^L} \cdot C_1^L \cdot \frac{1}{L} \cdot 1 + \cdots \frac{1}{2^L} C_L^L \cdot \frac{1}{L} \cdot L$

$= \frac{1}{L} \cdot \sum_{k=0}^{L} k \cdot C_k^L (\frac{1}{2})^k (1-\frac{1}{2})^{L-k} = \frac{1}{L} \cdot L \cdot \frac{1}{2} \sum_{k=1}^{L} \frac{(L-1)!}{(k-1)![(L-1)-(k-1)]!} (\frac{1}{2})^{k-1} (\frac{1}{2})^{L-k} = \frac{1}{2} (\frac{1}{2} + \frac{1}{2})^{(L-1)}$

$= \frac{1}{2}$ #

$\because$ 对于任意一个 A，上式都恒等于 $\frac{1}{2}$，与 A 无关

$\therefore E_f\{E_{OTS}(A_1(D), f)\} = E_f\{E_{OTS}(A_2(D), f)\} = \frac{1}{2}$

4. $P(V \le 0.1) = P(V=0) + P(V=0.1)$

$= (1-0.8)^{10} + C_1^{10} \cdot (0.8) \cdot (1-0.8)^9$

$= 0.0000041984$

$= 4.1984 \times 10^{-6}$

$P(V \ge 0.9) = P(V=0.9) + P(V=1)$

$= C_9^{10} (0.8)^9 \cdot (1-0.8) + (0.8)^{10}$

$= 0.3758$

5. ∵ A和D类 dice 有 green 1

∴ 应从 A和D类中挑5个 dice

設 P(A)为挑中A类dice的概率.

P(D)为挑中D类的概率

P(A∪D)为挑中A或D类 dice 的概率

∵ A.B.C.D 四类数目一样且独立

∴ $P(A) = P(B) = P(C) = P(D) = \frac{1}{4}$

∴ $P(A \cup D) = P(A) + P(D) = \frac{1}{2}$

∴ $P(5 \text{ green 1's}) = (P(A \cup D))^5 = (\frac{1}{2})^5 = \frac{1}{32}$

$= \frac{8}{256}$

6. 由右表可得, 'some numble'全为green

可分为以下四种情况:

① 从A或D类中抽取可得 1和3全为green

② 从B或D类中抽取可得 2全为green

③ 从B或C类中抽取可得 4和6全为green

④ 从A或C类中抽取可得 5全为green

green	even	odd
A	—	1, 3, 5
B	2, 4, 6	—
C	4, 6	5
D	2	1, 3

∴ P("some numble purely green")

$= [P(A \cup D)]^5 + [P(B \cup D)]^5 + [P(B \cup C)]^5 + [P(A \cup C)]^5 - [P(A)]^5 - [P(B)]^5 - [P(C)]^5 - [P(D)]^5$

减去每个类别的机率(e.g. $[p(A)]^5$)是因为:

$[p(A\cup D)]^5$ 和 $[p(A\cup C)]^5$ 都包含了5次全为A类dice的情况

∴应减去一次 $[p(A)]^5$

∴原式 $= (\frac{1}{2})^5 + (\frac{1}{2})^5 + (\frac{1}{2})^5 + (\frac{1}{2})^5 - 4 \cdot (\frac{1}{4})^5$

$= \frac{4}{2^5} - \frac{4}{4^5} = \frac{4}{2^5} - \frac{4}{2^{10}} = \frac{4 \cdot 2^5 - 4}{2^{10}} = \frac{2^5 - 1}{2^8} = \frac{31}{256}$ #

What I found:

可将1,2,3,4,5,6看作6个hypothesis $(h_1, h_2, \cdots h_6)$,将bag看作一个巨大的Data, A,B,C,D为此Data中四类Data dice上绿色的数字可看作: $h(x) \neq f(x)$,而橙色的为: $h(x) = f(x)$

由题意可知,每个hypothesis h_n 的 $E_{out} = \frac{1}{2}$,而 $N=5$ 时,抽到5个green dice使 $E_{in} = 1$, $\epsilon = 0.5$,可将抽中5次green dice看作bad sample

由Q5可知: $P_D[\text{BAD } D \text{ for } h_1] = \frac{8}{256}$

由Q6可知: $P_D[\text{BAD } D] = \frac{31}{256}$

$P_D[\text{BAD } D] = \frac{31}{256} \leq 4 \cdot P_D[\text{BAD } D \text{ for } h_n] = 4 \cdot \frac{8}{256}$

而根据Lecture 4 (Bound of BAD Data):

$P_D[\text{BAD } D] = P_D[\text{BAD } D \text{ for } h_1 \text{ or BAD } D \text{ for } h_2 \cdots \text{ or BAD } D \text{ for } h_6]$

$\leq P_D[\text{BAD } D \text{ for } h_1] + \cdots + P_D[\text{BAD } D \text{ for } h_6]$

$= 6 \cdot P_D[\text{BAD } D \text{ for } h_n]$

但是Q6中 $P_D[\text{BAD } D] \leq 4 \cdot P_D[\text{BAD } D \text{ for } h_n]$ 是因为:

h_1 和 h_3 是similar hypothesis $(h_1 \approx h_3)$, h_4 和 h_6 也是,因为他们在Data中分布情形一样.

∴ $h_1 \sim h_6$ 中 effective hypothesis 只有4种.

∴在hypothesis set中,一些h是相似的,通过找出effective hypothesis的数目,可更精准地估计 $P_D[\text{BAD } D]$

8.

① Prove that M exists:

+0 $\because$ $W_{t+1} = W_t + y_{n(t)} X_{n(t)} \cdot M$

又 $\because$ $y_{n(t)} W_{t+1}^T \cdot X_{n(t)} > 0$

$y_{n(t)} [W_t + y_{n(t)} X_{n(t)} M]^T X_{n(t)} > 0$

$\therefore$ $y_{n(t)} W_t^T X_{n(t)} + y_{n(t)}^2 X_{n(t)}^T M X_{n(t)} > 0$

$y_{n(t)} W_t^T X_{n(t)} + M \cdot X_{n(t)}^T X_{n(t)} > 0$

$M \cdot |X_{n(t)}|^2 > -y_{n(t)} W_t^T X_{n(t)}$

$$M > -\frac{y_{n(t)} W_t^T X_{n(t)}}{|X_{n(t)}|^2}$$

$\because$ $y_{n(t)} W_t^T X_{n(t)} \leq 0$

$\therefore$ $\dfrac{-y_{n(t)} W_t^T X_{n(t)}}{|X_{n(t)}|^2} \geq 0$

$\because$ 当 $M=0$ 时，$W_{t+1} = W_t$，无法 update

$\therefore$ $M \neq 0$ 且 $M=1$ 当 $y_{n(t)} W_t^T X_{n(t)} = 0$ 时

$\therefore$ $M_t = \left\lceil \dfrac{-y_{n(t)} W_t^T X_{n(t)}}{|X_{n(t)}|^2} \right\rceil > 0$

$\because$ $\min(|X_{n(t)}|^2) \geq 1$

又 $\because$ the data set is linear separable

$\therefore$ M_t is finite

$\therefore$ $M_t = \left\lceil \dfrac{-y_{n(t)} W_t^T X_{n(t)}}{|X_{n(t)}|^2} \right\rceil \geq 1$ ×

更正：$M_t = \left\lfloor \dfrac{-y_{n(t)} W_t^T X_{n(t)}}{|X_{n(t)}|^2} \right\rfloor$