Homework 2

B99202064 Kuan Hou Chan

1 Ans: (D) $\lambda\mu + (1-\lambda)(1-\mu)$

Probability of error that when $y = f(x)$ will be (μ) and otherwise will be $(1-\mu)$
Hence the probability $p(err) = \lambda\mu + (1-\lambda)(1-\mu)$

2 Ans: (D) 0.5

Let the $p(err) = \lambda\mu + (1-\lambda)(1-\mu) = 0$ then we can get $\mu = \frac{-\lambda}{-2\lambda+1}$
When $\lambda = 0.5, \mu$ not exist. Hence $\lambda = 0.5$

3 Ans: (c) 460,000

By using $N \geq \frac{8}{\varepsilon^2} \ln \frac{4*(2N)^{d_{vc}}+4}{\delta}$equation, we can start from N=1000 to do iteration and the 1st iteration we will get 257251.36, 2nd iteration will get 434853 and finally it will converge to 460000.

4 Ans: (e) Devroye: $\varepsilon \leq \sqrt{\frac{1}{2N}(4\varepsilon(1+\varepsilon) + \ln\frac{4m_H(N^2)}{\delta})}$

Following results is obtain by wolframalpha:
Original VC Bound $\varepsilon = 0.63$,
Variant VC Bound $\varepsilon = 0.86$,
Rademacher Penalty Bound $\varepsilon = 0.33$
Parrondo and Van den Broek $\varepsilon = 0.22$
Devroye $\varepsilon = 0.21$

5 Ans: (D) Parrondo and Van den Broek

Following results is obtain by wolframalpha:
Original VC Bound $\varepsilon = 13.83$,
Variant VC Bound $\varepsilon = 16.26$,
Rademacher Penalty Bound $\varepsilon = 7.05$
Parrondo and Van den Broek $\varepsilon = 5.10$
Devroye $\varepsilon = 5.59$

6 Ans: (a) $N^2 - N + 2$

For growth function $m_H(N)$ of positive intervals, that is $m_H(N) = C_2^{N+1} = \frac{N^2}{2} + \frac{N}{2} + 1$, same as negative intervals.
So for positive-negative intervals, $m_H(N) = 2 * (\frac{N^2}{2} + \frac{N}{2} + 1) = N^2 + N + 2$
By considering the case of overlap between these two intervals, the overlap intervals numbers will have $2N$ intervals overlap in both positive or negative intervals since the mirror symmetry will make both intervals have same pattern except the case of $[OXX...XXO]$ in negative intervals that it won't appear in positive intervals that just 1 case won't repeat in positive-negative intervals.
Hence, $m_H(N) = N^2 + N + 2 - 2N = N^2 - N + 2$

7 Ans: (B) 3

The VC-dimension is less than 4 since the labeling $+ - + -$ is not possible to obtain using any hypothesis. Futhermore, the VC-dimension is equal to 3 since there is a set of 3 points that can be fully shattered.

8 Ans: (C) $C_2^{N+1} + 1$

We can treat the donut as a positive interval case and separate it to $N + 1$ intervals by several circles. Then it will have C_2^{N+1} can be choose. Otherwise, consider the case of that 2 interval on the same interval, then the final growth function will be $C_2^{N+1} + 1$

9 Ans: (D) $D + 1$

If we expand the term $\sum_{i=0}^{D} c_i x^i = c_0 + c_1 x^1 + ... + c_D x^D$ then it have D+1 terms appear.
For upper bound of VC-dimension, $m_H(N) \leq poly(N)$ the polymial highest term can be the upper bound of VC-dimension That we can find there are (D+1) points which are shattered by H because of linear independent and (D+2) points which are not shattered by H because it is linear dependence.
Hence the VC-dimension will be $D + 1$

10 Ans: (A) 2^d

The 'simplified decision trees' can be treated as a positive rays in each dimension that the threshold will separate x_i where i is the hypothesis of $+1 or -1$
Hence, the VC-dimension will be $2 * 2 * 2 * ... * 2$ for d-dimension and total 2^d VC-dimension.

11 Ans: (D) ∞

For a sine wave, any finite set of points can be shattered by the hypothesis set and similar for triangle waves, the VC–dimension is ∞.
We can adjust the α that every dichotomy can be implemented by α value then $m_H(N) = 2^N \Rightarrow d_{vc} = \infty$

12 Ans: (A)(C)(D)

We know that the upper bound of growth bound is 2^N, and it will become the condition for the following option.
(A) $m_H(\lfloor \frac{N}{2} \rfloor) m_H(\lceil \frac{N}{2} \rceil) = 2^{\lfloor \frac{N}{2} \rfloor} * 2^{\lceil \frac{N}{2} \rceil} = 2^{\lfloor \frac{N}{2} \rfloor + \lceil \frac{N}{2} \rceil}$, by $\lfloor \frac{N}{2} \rfloor + \lceil \frac{N}{2} \rceil = N$, then $m_H(\lfloor \frac{N}{2} \rfloor) m_H(\lceil \frac{N}{2} \rceil) = m_H(N) = 2^N$
(B) Since $N \geq d_{vc} \geq 2, 2^{d_{vc}} \leq 2^N$, then it will become upper bound when $d_{vc} = N$. Hence for general case, it isn't an upper bound
(C) $2^i m_H(N - i) = 2^i * 2^{N-i} = 2^{i+N-i} = 2^N, then min_{1 \leq i \leq N-1} 2^i m_H(N - i) = 2^N = m_H(N)$, then it is an upper bound
(D) Since $m_H(N) = \sum_{i=0}^{d_{vc}} \begin{pmatrix} N \\ i \end{pmatrix} = N_{vc}^{d}$, then it is an upper bound.
(E) Since $m_H(N - 1) + N * d_{vc} = 2^{N-1} + N * d_{vc} < 2^N$, then it is not an upper bound.

13 Ans: (A) 2^N

(A) Obviously 2^N is growth function for all hypothesis set.
(B) $2^{\lfloor \sqrt{N} \rfloor} < 2^N$, then it is not possible growth function.
(C) $2^{\lfloor \frac{N}{2} \rfloor} < 2^N$, then it is not possible growth function.
(D) $2^{\lceil \frac{N}{2} \rceil} < 2^N$, then it is not possible growth function.
(E) $1 + N + \frac{N(N-1)(N-2)}{6} < 2^N$, then it is not possible growth function.

14 Ans: (B) $0 \leq d_{vc}(\cap_{k=1}^{K} H_k) \leq min\{d_{vc}(H_k)\}_{k=1}^{K}$

For intersection of $d_{vc}(H_k)$, intersect means min operation that if some $d_{vc}(H_k)$ is empty set or a singleton, then $d_{vc}(H_k) = 0$, if there haven't exist empty set or singleton, then $d_{vc}(H_k)$ will be bounded at $min\{d_{vc}(H_k)\}_{k=1}^{K}$

15 Ans: (E) $max\{d_{vc}(H_k)\}_{k=1}^{K} \leq d_{vc}(\cap_{k=1}^{K} H_k) \leq K - 1 + \sum_{k=1}^{K} d_{vc}(H_k)$

For union of $d_{vc}(H_k)$, union means max operation that if others $d_{vc}(H_k)$ are empty set or a singleton and just one is not empty, then $d_{vc}(H_k) = 0$ and for upper bound, the union of $d_{vc}(H_k)$ not only bounded by $\sum_{k=1}^{K} d_{vc}(H_k)$ and union of $d_{vc}(H_k)$ maybe will $> \sum_{k=1}^{K} d_{vc}(H_k)$.
Hence upper bound must be something greater than $\sum_{k=1}^{K} d_{vc}(H_k)$ and bounded by $K - 1 + \sum_{k=1}^{K} d_{vc}(H_k)$.

16 Ans: (C) $0.5 + 0.3s(|\theta| - 1)$

Since there's always been 20% flipped the results. Then we consider the following cases first:
For case 0 to case 2 we consider s is fixed and assume $s = 1$, $[-1 - - - - - - -0 - - - - - - -1]$
Case 0:
$\theta = 1 or - 1$ will always get 0.5 error without consider 0.2 possiblility of flipped results. If we consider the 0.2 in, then both the error and correct part will have 0.2 possibility of result flip and $error = 0.5 + 0.2 - 0.2 = 0.5$ which remain same.
Case 1:
For extreme case of $\theta = 0$, there always 0.5 possibility of 1 and possibility of -1 exists and will have 0 possiblility of error without consider 20% flipped result. There will always 0 error on both +1/-1 side. Once consider the 20% in, then result will be flipped by 0.2 possibility and error will become $0.5 - 0.3 = 0.2$ where 0.3 as the result didn't flipped in $-1 \leq x \leq 1$
.
Case 2:
Consider $\theta = 0, 1, -1$, 0.2 % error of result will be complement each other in correct or error region. So we need to consider the 0.3 possibility of not flip result. For $theta > 0$, Err will become $0.5 + 0.3 * (theta - 1)$ where 0.5 is $x > 0$ region and $0.3 * (theta - 1)$ is for the label correct possiblility. We also can derive for $theta < 0$ and $x < 0$ region and the error will same if theta take absolute. In general for theta and x, $Err = 0.5 + 0.3(|\theta| - 1)$
Case 3:
Consider $s = +1 or - 1$, if Err change to $0.5 + 0.3s(|\theta| - 1)$ then it is still can be used in counting its error.
Hence, $E_{out} = 0.5 + 0.3s(|\theta| - 1)$