Machine Learning Homework #2

Qing-Cheng Li

R01922024

1. D

Error = True Negative + False Positive = $\lambda\mu + (1-\lambda)(1-\mu)$ (D)

---

2. D

Error = $\lambda\mu + (1-\lambda)(1-\mu) = (2\lambda - 1)\mu + 1 - \lambda$, let $\lambda = \frac{1}{2}$, Error = $1 - \lambda$, is independent of μ. (D)

---

3. D

The bound is $4m_H(2N)e^{-\frac{1}{8}\varepsilon^2 N} = 4(2N)^2 e^{-\frac{1}{8}\varepsilon^2 N}$, $\varepsilon = 0.05$. This bound $\leq$ 1-95% = 0.05.

Let N be 400,000, bound = 221895, let N be 420,000, bound = 697.75, let N be 440,000, bound = 2.14484, let N be 460,000, bound = 0.006458, let N be 480,000, bound = 0.000191. $N = 460000$ is the closet numerical approximation of the sample size that the VC generalization bound predicts.

---

4. D

(a) $\sqrt{\frac{8}{10000}\ln\frac{4(20000)^{50}}{0.05}} \simeq 0.63217$

(b) $\sqrt{\frac{2\ln 20000(10000)^{50}}{N}} + \sqrt{\frac{2}{10000}\ln\frac{1}{0.05}} + \frac{1}{10000} \simeq 0.33131$

(c) $\epsilon = \sqrt{\frac{1}{10000}(2\varepsilon + \ln\frac{6(20000)^{50}}{0.05}}), \epsilon \simeq 0.22370$

(d) $\epsilon = \sqrt{\frac{1}{20000}(4\epsilon(1+\epsilon) + \ln\frac{4(10000^{50})^{50}}{0.05}}, \epsilon \simeq 0.21523$

(a) > (b) > (c) > (d)

---

5. C

(a) $\sqrt{\frac{8}{5}\ln\frac{4(10)^{50}}{\delta}} \simeq 13.82816$

(b) $\sqrt{\frac{2\ln 10(5)^{50}}{N}} + \sqrt{\frac{2}{5}\ln\frac{1}{0.05}} + \frac{1}{5} \simeq 7.04878$

(c) $\epsilon = \sqrt{\frac{1}{5}(2\varepsilon + \ln\frac{6(10)^{50}}{0.05}}), \epsilon \simeq 5.10136$

(d) $\epsilon = \sqrt{\frac{1}{10}(4\epsilon(1+\epsilon) + \ln\frac{4(5^{50})^{50}}{0.05}}, \epsilon \simeq 5.59313$

(a) > (b) > (d) > (c)

---

6. A

Consider that we can select 0 (all the same), 1 (split data into 2 part), 2 (l and r) from N-1 intervals, and +1 or -1 in the interval that this question said, we have
$2\sum_{i=0}^{2}\binom{N-1}{i} = 2(\binom{N-1}{0} + \binom{N-1}{1} + \binom{N-1}{2}) = 2(1 + N - 1 + \frac{(N-1)(N-2)}{2}) =$
$2(\frac{2+2N-2+N^2-3N+2}{2}) = 2\,\frac{N^2-N+2}{2} = N^2 - N + 2$, so select (a)

7. A

Using $N^2 - N + 2$, N=3,$N^2 - N + 2 = 9 - 3 + 2 = 8 = 2^3$, N=4, $N^2 - N + 2 = 16 - 4 + 2 = 14 \leq 2^4$, N=4 is break point. $d_{vc} = k - 1 = 4 - 1 = 3$, select (a).

8. B

Consider N point on $\mathbb{R}^2$, there are N distances from $(0, 0)$ to those points. So this question can be seen as selecting 2 intervals from N+1 intervals, adding the situation that all distances are -1, not in a and b. So the $m_H(N)$ is $\binom{N+1}{2} + 1$, select (b).

9. B

一個D次多項式最多可以有D個解，其函數圖形最多可以將數線分成D+1個+,-相間區間，最少可以無解（與y=0不相交）。因此這個假設集合可以造出從無解（全+1或全-1）一直到D+1個+,-相間的結果，全部（最多也是）2^{D+1}種組合都可以被製造出來。因此VC-Dimension就是D+1。

10. A

We have 2^d hyper-rectangular, if each can be +1 or -1,there are 2^{2^d} combinations, it at most can shatter 2^d, so the VC-dimension is 2^d, select (a).

---

11. C

If there are N point on $\mathbb{R}$, from 1st point to Nth point, i^{th} point on 4^i, For $k = 1$ to 2^N, using $1 + \frac{1}{2}\left(\frac{2k-2}{2^{N+1}}\right) < \alpha_k < 1 + \frac{1}{2}\left(\frac{2k-1}{2^{N+1}}\right)$, those α_k can build all $\{+1, -1\}^N$ combinations, so this triangle wave hypothesis can shatter any N, so the VC dimension is ∞, select (c).

---

12. A,C,D

Let $N = A + B$, A!=B.

In case A and B all bigger than d_{vc}:

$m_{\mathcal{H}}(A) \leq A^{d_{vc}}$, $m_{\mathcal{H}}(B) \leq B^{d_{vc}}$, $m_{\mathcal{H}}(A)m_{\mathcal{H}}(B)$ is bound by $(AB)^{d_{vc}}$ which is bigger than $(A+B)^{d_{vc}}$.

In case A and B all smaller than d_{vc}:

$m_{\mathcal{H}}(A) \leq 2^A$, $m_{\mathcal{H}}(B) \leq 2^B$, $m_{\mathcal{H}}(A)m_{\mathcal{H}}(B)$ is bound by $2^{A+B} = 2^N$ which is bigger than $(A+B)^{d_{vc}}$.

In case A>B, B is smaller than d_{vc} but A is not:

$m_{\mathcal{H}}(A) \leq A^{d_{vc}}$, $m_{\mathcal{H}}(B) \leq 2^B$, $m_{\mathcal{H}}(A)m_{\mathcal{H}}(B)$ is bound by $A^{d_{vc}}2^B$ which is bigger than $(A+B)^{d_{vc}}$.

So, $m_{\mathcal{H}}(A)m_{\mathcal{H}}(B)$ can bound $m_{\mathcal{H}}(A+B)$

(a) $\lfloor \frac{N}{2} \rfloor$ and $\lceil \frac{N}{2} \rceil$ can be seen as split N into 2 parts. $m_{\mathcal{H}}(\lfloor \frac{N}{2} \rfloor)m_{\mathcal{H}}(\lceil \frac{N}{2} \rceil)$ can bound $m_{\mathcal{H}}(N)$

(b) $2^{d_{vc}}$ is smaller than the bound for $m_{\mathcal{H}}(N) = N^{d_{vc}}$, so is can not be the upper bound.

(c) It can seen as the case i is smaller than d_{vc}, $N - i$ is bigger than d_{vc}, so $m_{\mathcal{H}}(i)m_{\mathcal{H}}(N-i)$ can bound $m_{\mathcal{H}}(N)$

(d) $N^{d_{vc}}+1$ is bigger than the bound for $m_{\mathcal{H}}(N) = N^{d_{vc}}$, so it can be a upper bound.

---

13. A

If there is no break point, the growth function is 2^N, so select (a).

And we know that if there is a break point k, then the growth function will be bounded by N^{k-1}.

For (b), break point $k = 2$, $2^{\lfloor\sqrt{2}\rfloor} = 2 \leq 2^2 = 4$, but when $N = 25$, $2^{\lfloor\sqrt{25}\rfloor} = 32 > 25^{2-1} = 25$, out of the bound, so (b) can not be a growth function.

For (c), break point $k = 1$, $2^{\lfloor\frac{1}{2}\rfloor} = 2^0 = 1 \leq 2^1 = 2$, but when $N = 2$, $2^{\lfloor\frac{2}{2}\rfloor} = 2^1 = 2 > 2^{1-1} = 1$, out of the bound, so (c) can not be a growth function.

For (d), break point $k = 2$, $1 + 2 + \frac{2(2-1)(2-2)}{6} = 3 \leq 2^2 = 4$, but when $N = 3$, $1 + 3 + \frac{3(3-1)(3-2)}{6} = 5 > 3^{2-1} = 3$, out of the bound, so (d) can not be a growth function.

---

14. B

$\cap_{k=1}^{K}\mathcal{H}_k$ must be subset of all $\mathcal{H}$, the smallest subset is empty set, $d_{vc}(\emptyset) = 0$, the biggest subset is smallest set in $\mathcal{H}$, so $d_{vc}(\cap_{k=1}^{K}\mathcal{H}_k) \geq 0$, $d_{vc}(\cap_{k=1}^{K}\mathcal{H}_k) \leq min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$. Select (b).

---

15. D

Lower bound:

Any hypothesis can not help another hypothesis to shatter more data node, so the worst case of union VC dimension is $max\{d_{vc}(\mathcal{H}_{k=1}^{K})\}$.

Upper bound:

//Using the $k+$ and $k'-$ example which TA provide on CEIBA, we know that $d_{vc}(k+) = k$, $d_{vc}(k'-) = k'$, $d_{vc}(k + \cap k'-) > k + k'$, so there is only (d) can be select orz.

//But I do not know how to prove it Q___Q.

16.

Error = True Negative + False Positive

If $s = +1$, Error = $0.2\ \frac{2-|\theta|}{2} + 0.8\ \frac{|\theta|}{2} = 0.2 + 0.3|\theta|$,

If $s = -1$, Error = $0.2\ \frac{|\theta|}{2} + 0.8\ \frac{2-|\theta|}{2} = 0.8 - 0.3|\theta|$,

Combine s = +1 and -1, Error = $0.5 + 0.3s(|\theta| - 1)$

17.

```
$ python hw2_17.py
```

Decision stump algorithm in *dsa.py*

18,19,20

```
$ python hw2_18.py
```