CSIE 5432/5433 HW2 PDF

YiWenLai

TOTAL POINTS

5 / 0

QUESTION 1

1 Problem 3 **0 / 0**

+ **10 pts** Proved $$d_{VC} < 3$$ and $$d_{VC} >= 2$$

Proof is insufficient

✓ **+ 0 pts Proof for $$ d_{VC} < 3$$ is insufficient / incorrect.**

✓ **+ 0 pts Proof for $$d_{VC} >= 2$$ is insufficient / incorrect.**

+ **0 pts** Wrong answer / No answer

QUESTION 2

2 Problem 6 **15 / 0**

+ **20 pts** Correct. Explanation is complete and without mistake.

✓ **+ 15 pts Got correct answer but made **one** mistakes**

+ **10 pts** Got correct answer but made **two** mistakes

+ **5 pts** Got correct answer but made **three** mistakes

+ **0 pts** Got correct answer but your explanation makes no sense to me

+ **0 pts** leave blank and choose the "I don't have any idea" option during submission

- **10 pts** Incorrect (satisfy at least one of the following criteria)
(0) wrong answer
(1) little to no explanation
(2) leave blank but choose an answer outside of the "I don't have any idea" option during submission

+ **0 pts** some(but not limit to) examples of mistakes:
(0) wrong Triangle Inequality expression (write < instead of <=)
(1) obvious typos: description is correct but write the wrong math expression (ex: "g has best E_in" but writes "E_in(g)>=E_in(g*)")
(2) no "more than 1-\delta probability" statement or equivalent math expression
(3) no explanation on "E_in(g) < E_in(g*)"
(4) skip too many steps

QUESTION 3

3 Problem 10 **10 / 0**

+ **20 pts** Correct

✓ **+ 10 pts Almost Correct**

+ **5 pts** Almost Incorrect

+ **0 pts** Without choices.

- **10 pts** Incorrect

- **20 pts** Violate the class policy.

cool insight, yet lacking a more detailed description

QUESTION 4

4 Coding Part **-20 / 0**

+ **0 pts** Without judgement

✓ **- 20 pts Without complete code**

Machine Learning Foundations HW2

b05901060 賴繹文

1. c
    a. There are three points in a line, which can't be shattered if the middle point has different sign with others,
    b. There are three points in a line and another point. Same as (a), (b) can't be shattered.
    c. There are two points in a line, and two other points which can't form a plane with the line. Thus, any combination of inputs can be separated by a plane.
    d. Two lines with two points on it. These two lines can form a plane and thus four points can't be shattered by a plane.
    e. There are four points in a line and another point. Same as (a), it can't be shattered.
2. d, 4n-2.
   Consider the max dichotomies of patterns from 1 positive to N positive, then multiplied by two, which is the answer. On a 2D plane, there will be N choices of horizontal or vertical line, respectively. Thus, 2N patterns of inputs are shown. While the case that all N points are positive, the horizontal one and the vertical one should be calculated once. So there are 2N - 1 patterns. After multiplying 2, 4N - 2 is the result. Mentioned that 2N - 1 is the maximum case, which means that it is possible for other cases to exist.
3. c
   First consider the case “2 inputs” with different signs. There are infinite numbers of lines to separate two points with different slopes. And given a w0 > 0, we can always find a line matching the conditions to classify two points, by shifting or rotating the line. 2 inputs with the same sign is the much easier case, which is also found in a line to classify. Thus, “2 inputs case” can be shattered. Further, consider the case with three inputs and three aren't on the same line. Based on the discussion with case “two inputs”, the procedure shifting or rotating the line can't make sure that the other one point to be classified as the same sign. In other words, to meet the condition w0 >0, this case can't be classified correctly, same as the case “three inputs on the same line”. As a result, there is no case shattered when there are three inputs, so the VC dimension is 2.
4. b
   Same as the case “Positive Intervals”. Calculate all the distances to (0,0,0), and map the value to a 1D line. Select two from N+1 spots and plus one for a = b = 0.
5. b
   Same as “Positive Intervals”, its VC dimension is 2. If there are three points on the same line and the middle one with different signs with others. This case can't be shattered.
6. d

1 Problem 3 **0 / 0**

+ **10 pts** Proved $$d_{VC} < 3$$ and $$d_{VC} >= 2$$

Proof is insufficient

✓ + **0 pts** **Proof for $$ d_{VC} < 3$$ is insufficient / incorrect.**

✓ + **0 pts** **Proof for $$d_{VC} >= 2$$ is insufficient / incorrect.**

+ **0 pts** Wrong answer / No answer

Machine Learning Foundations HW2

b05901060 賴繹文

1. c
    a. There are three points in a line, which can't be shattered if the middle point has different sign with others,
    b. There are three points in a line and another point. Same as (a), (b) can't be shattered.
    c. There are two points in a line, and two other points which can't form a plane with the line. Thus, any combination of inputs can be separated by a plane.
    d. Two lines with two points on it. These two lines can form a plane and thus four points can't be shattered by a plane.
    e. There are four points in a line and another point. Same as (a), it can't be shattered.
2. d, 4n-2.
   Consider the max dichotomies of patterns from 1 positive to N positive, then multiplied by two, which is the answer. On a 2D plane, there will be N choices of horizontal or vertical line, respectively. Thus, 2N patterns of inputs are shown. While the case that all N points are positive, the horizontal one and the vertical one should be calculated once. So there are 2N - 1 patterns. After multiplying 2, 4N - 2 is the result. Mentioned that 2N - 1 is the maximum case, which means that it is possible for other cases to exist.
3. c
   First consider the case "2 inputs" with different signs. There are infinite numbers of lines to separate two points with different slopes. And given a w0 > 0, we can always find a line matching the conditions to classify two points, by shifting or rotating the line. 2 inputs with the same sign is the much easier case, which is also found in a line to classify. Thus, "2 inputs case" can be shattered. Further, consider the case with three inputs and three aren't on the same line. Based on the discussion with case "two inputs", the procedure shifting or rotating the line can't make sure that the other one point to be classified as the same sign. In other words, to meet the condition w0 >0, this case can't be classified correctly, same as the case "three inputs on the same line". As a result, there is no case shattered when there are three inputs, so the VC dimension is 2.
4. b
   Same as the case "Positive Intervals". Calculate all the distances to (0,0,0), and map the value to a 1D line. Select two from N+1 spots and plus one for a = b = 0.
5. b
   Same as "Positive Intervals", its VC dimension is 2. If there are three points on the same line and the middle one with different signs with others. This case can't be shattered.
6. d

$$P[E_{in}(g) - E_{out}(g) > \varepsilon] < 4\, m_H(2N)\exp\left(-\frac{1}{8}\varepsilon^2 N\right)$$

$$P[E_{out}(g^*) - E_{in}(g^*) > \varepsilon] < 4\, m_H(2N)\exp\left(-\frac{1}{8}\varepsilon^2 N\right)$$

$$E_{out}(g) \le E_{in}(g) + \sqrt{\frac{8}{N}\ln\left(\frac{4m(2N)}{\delta}\right)}$$

$$E_{in}(g^*) \le E_{out}(g^*) + \sqrt{\frac{8}{N}\ln\left(\frac{4m(2N)}{\delta}\right)}$$

$$E_{out}(g) - E_{out}(g^*) \le \underbrace{E_{in}(g) - E_{in}(g^*)}_{<0} + 2\sqrt{\frac{8}{N}\ln\left(\frac{4m(2N)}{\delta}\right)}$$

Here we assume that both data outside or inside the training area are large enough.The second line means that we change the character of testing data and the training data. At the last line, there is a term less than 0, which means that there is an upper bound, the latter term.

7. d

   $2^N = M$ is the case with the maximum number hypothesis sets. N is the number of the inputs, which is also VC dimension. $N = log_2 M$

8. d

   From zero of -1 to all numbers are -1, there are k+1 patterns. Thus VC dimension is k+1.

9. c

   “some set of d distinct inputs is shattered by H”, “some set of d + 1 distinct inputs is not shattered by H”, and “any set of d + 1 distinct inputs is not shattered by H”. The first one is by the definition of break point mentioned in lecture 5, p24. The second and third one are the same as what’s in lecture 7 p13. It provide the condition that $d_{VC} <= d+1$.

10. c

   Apply the concept of Polar coordinate system, and the sine value stands for the imaginary part. Given all the inputs with different signs, we can map those to a circle on the polar coordinate system. There are different points with different signs in the circle. Taught in class, this case is a “convex set”. Any combination of inputs can be shattered by a convex region. Thus the VC dimension of c is infinite.

2 Problem 6 **15 / 0**

+ **20 pts** Correct. Explanation is complete and without mistake.

✓ + **15 pts** **Got correct answer but made **one** mistakes**

+ **10 pts** Got correct answer but made **two** mistakes

+ **5 pts** Got correct answer but made **three** mistakes

+ **0 pts** Got correct answer but your explanation makes no sense to me

+ **0 pts** leave blank and choose the "I don't have any idea" option during submission

- **10 pts** Incorrect (satisfy at least one of the following criteria)

(0) wrong answer

(1) little to no explanation

(2) leave blank but choose an answer outside of the "I don't have any idea" option during submission

+ **0 pts** some(but not limit to) examples of mistakes:

(0) wrong Triangle Inequality expression (write < instead of <=)

(1) obvious typos: description is correct but write the wrong math expression (ex: "g has best E_in" but writes "E_in(g)>=E_in(g*)")

(2) no "more than 1-\delta probability" statement or equivalent math expression

(3) no explanation on "E_in(g) < E_in(g*)"

(4) skip too many steps

$$P[E_{in}(g) - E_{out}(g) > \varepsilon] < 4\, m_H(2N) \exp\left(-\frac{1}{8}\varepsilon^2 N\right)$$

$$P[E_{out}(g^*) - E_{in}(g^*) > \varepsilon] < 4\, m_H(2N) \exp\left(-\frac{1}{8}\varepsilon^2 N\right)$$

$$E_{out}(g) \le E_{in}(g) + \sqrt{\frac{8}{N} \ln\left(\frac{4m(2N)}{\delta}\right)}$$

$$E_{in}(g^*) \le E_{out}(g^*) + \sqrt{\frac{8}{N} \ln\left(\frac{4m(2N)}{\delta}\right)}$$

$$E_{out}(g) - E_{out}(g^*) \le \underbrace{E_{in}(g) - E_{in}(g^*)}_{<0} + 2\sqrt{\frac{8}{N} \ln\left(\frac{4m(2N)}{\delta}\right)}$$

Here we assume that both data outside or inside the training area are large enough.The second line means that we change the character of testing data and the training data. At the last line, there is a term less than 0, which means that there is an upper bound, the latter term.

7. d

   $2^N = M$ is the case with the maximum number hypothesis sets. N is the number of the inputs, which is also VC dimension. $N = log_2 M$

8. d

   From zero of -1 to all numbers are -1, there are k+1 patterns. Thus VC dimension is k+1.

9. c

   “some set of d distinct inputs is shattered by H”, “some set of d + 1 distinct inputs is not shattered by H”, and “any set of d + 1 distinct inputs is not shattered by H”. The first one is by the definition of break point mentioned in lecture 5, p24. The second and third one are the same as what’s in lecture 7 p13. It provide the condition that $d_{VC} <= d+1$.

10. c

   Apply the concept of Polar coordinate system, and the sine value stands for the imaginary part. Given all the inputs with different signs, we can map those to a circle on the polar coordinate system. There are different points with different signs in the circle. Taught in class, this case is a “convex set”. Any combination of inputs can be shattered by a convex region. Thus the VC dimension of c is infinite.

3 Problem 10 10 / 0

+ **20 pts** Correct

✓ + **10 pts** **Almost Correct**

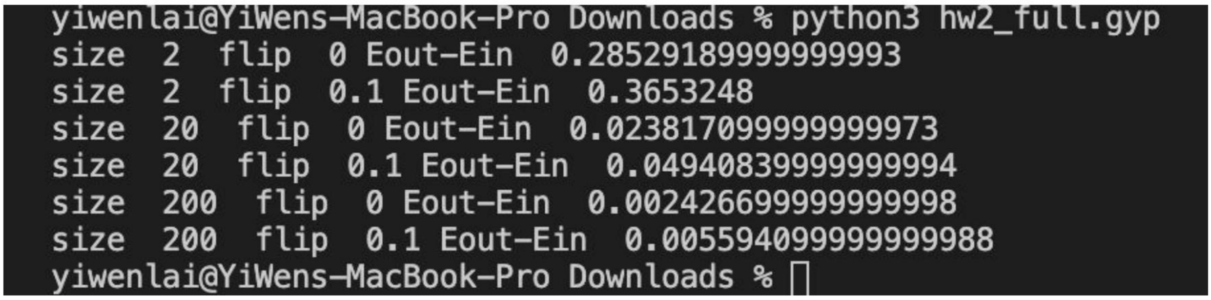
+ **5 pts** Almost Incorrect

+ **0 pts** Without choices.

- **10 pts** Incorrect

- **20 pts** Violate the class policy.

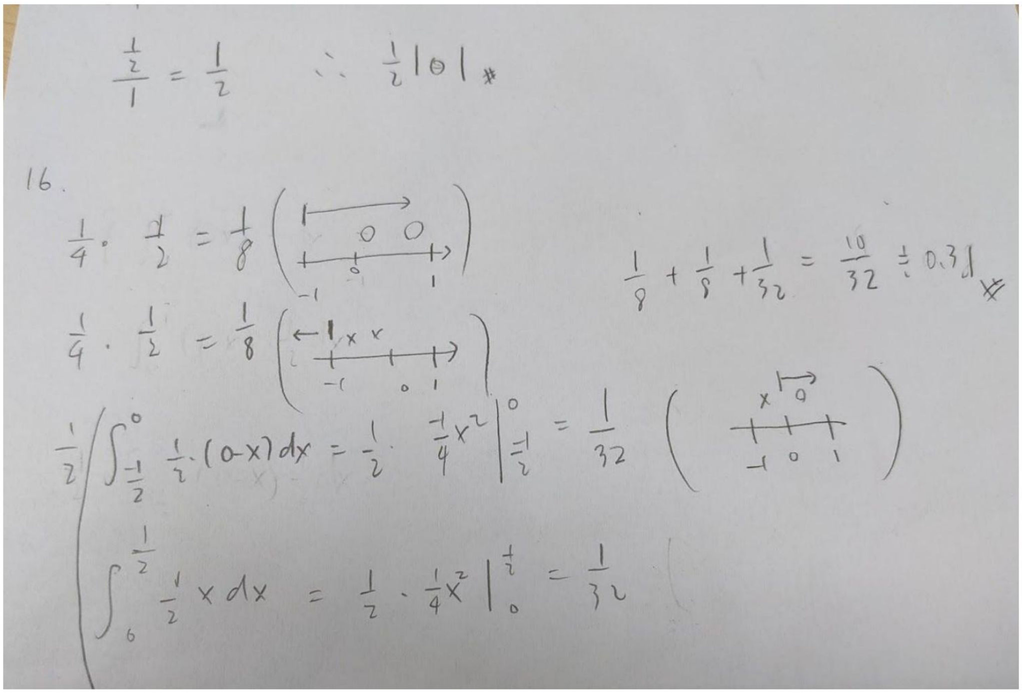
cool insight, yet lacking a more detailed description



```
yiwenlai@YiWens-MacBook-Pro Downloads % python3 hw2_full.gyp
size  2  flip  0 Eout-Ein  0.28529189999999993
size  2  flip  0.1 Eout-Ein  0.3653248
size  20  flip  0 Eout-Ein  0.023817099999999973
size  20  flip  0.1 Eout-Ein  0.04940839999999994
size  200  flip  0 Eout-Ein  0.002426699999999998
size  200  flip  0.1 Eout-Ein  0.005594099999999988
yiwenlai@YiWens-MacBook-Pro Downloads %
```

figure A

16. d, size = 2, flip = 0 in figure A
   a. Analytical proof
      Consider three cases after sorted, with probability 0.25,0.25,0.5. Multiply each probability with the expected error value. In the first case, the number less than 0 would be considered as positive; in the second case, the number greater than 0 would be considered as negative; In the third case, the value of theta will fall in (-0.5,0.5) equally. We can separate it as two cases, theta in (-0.5,0) and (0,0.5). Then calculate the expected error. Sum of those value is the answer.



$\frac{\frac{1}{2}}{1} = \frac{1}{2} \quad \therefore \frac{1}{2}|\theta|$ #

16.

$\frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$

$\frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$

$\frac{1}{2}\left(\int_{-\frac{1}{2}}^{0} \frac{1}{2}\cdot(0-x)\,dx = \frac{1}{2} \cdot \frac{-1}{4}x^2 \Big|_{\frac{-1}{2}}^{0} = \frac{1}{32} \right.$

$\int_{0}^{\frac{1}{2}} \frac{1}{2} x\,dx = \frac{1}{2} \cdot \frac{1}{4}x^2 \Big|_{0}^{\frac{1}{2}} = \frac{1}{32}$

$\frac{1}{8} + \frac{1}{8} + \frac{1}{32} = \frac{10}{32} \doteq 0.31$ #

17. b, size = 20, flip = 0 in figure A
18. e, size = 2, flip = 0.1 in figure A
19. c, size = 20, flip = 0.1 in figure A
20. a, size = 200, flip = 0.1 in figure A

4 Coding Part **-20 / 0**

+ **0 pts** Without judgement

✓ **- 20 pts** **Without complete code**