Homework-5 Solutions

Question 1

As discussed in class the concept of a medium built man can be expressed as an axis parallel rectangle. Therefore, it can be defined in terms of 4 real numbers: a, b, c, d such that:

$$a \leq \text{height} \leq b, \qquad c \leq \text{weight} \leq d$$

We assume that a consistent algorithm is available.

Part I

In this part we take as hypotheses 4 floating point numbers, each represented in terms of 4 bytes (32 bits).

a How many hypotheses are in the hypotheses class?

Answer: The number of hypotheses in the hypotheses class is: $r = 2^{32 \cdot 4} = 2^{128}$

b How many randomly obtained examples are needed for PAC learning with accuracy parameter $\epsilon = 0.1$ and confidence parameter $\delta = 0.05$?

Answer: At least 918 randomly obtained examples are needed.

$$m \geq \frac{1}{\epsilon} \ln(r/\delta) = 10 \ \ln(2^{128}/0.05) = 917.2$$

c What is the confidence level (what is δ) if it is known that 1000 randomly chosen examples were used, and the desired accuracy is level is $\epsilon = 0.1$?

Answer:

$$\delta \geq r(1-\epsilon)^m = 2^{128} \times 0.9^{1000} = 5.95E - 8$$

Answer can also be calculated based on formula $m \geq 1/\epsilon \ln(r/\delta)$, which gives us $\delta \geq 1.266E - 5$. This value is not as sharp as the one above.

d What is the accuracy level (what is ϵ) if it is known that 1000 randomly chosen examples were used, and the desired confidence level is $\delta = 0.05$?

Answer:

$$\begin{aligned} \delta \geq r(1-\epsilon)^m \Rightarrow (1-\epsilon)^m \leq \delta/r \Rightarrow m\ln(1-\epsilon) \leq (\ln\delta - \ln r) \\ \Rightarrow 1-\epsilon \leq e^{(\ln\delta - \ln r)/m} \Rightarrow \epsilon \geq 1 - e^{(\ln\delta - \ln r)/m} \\ \Rightarrow \epsilon \geq 1 - e^{(\ln 0.05 - \ln 2^{128})/1000} = 0.0876 \end{aligned}$$

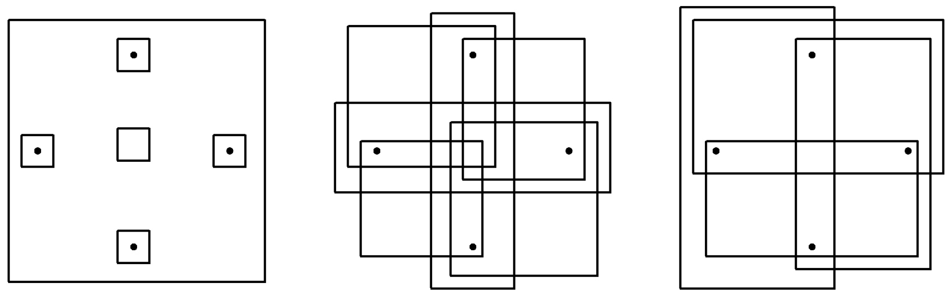
Answer can also be calculated based on formula $m \geq 1/\epsilon \ln(r/\delta)$, which gives us $\epsilon \geq 0.0917$. This value is not as sharp as the one above.

Part II

In this part we assume that we are able to represent arbitrarily accurate axis parallel rectangles.

a Prove that the VC dimension of the concept class of all axis parallel rectangles is $d = 4$.

Answer: The Following drawings show that the VC dimension of the concept class of all axis parallel rectangles is at least 4. There is no instances set of size 5 that can be shattered by the concept class, thus VC dimension of the concept class is $d = 4$



b How many rendomly obtained examples are needed for PAC learning with accuracy parameter $\epsilon = 0.1$ and confidence parameter $\delta = 0.05$?

Answer:

$$m \geq \frac{1}{\epsilon}(4\log_2(2/\delta) + 8d\log_2(13/\epsilon)) = 10(4\log_2(2/0.05) + 32\log_2(13/0.1)) = 2460.03$$

Thus, at least 2461 rendomly obtained examples are needed.

For $d = 3$, we have $m \geq 1898.2$, thus at least 1899 rendomly obtained examples are needed.

c What is the confidence level (what is δ) if it is known that 1000 randomly chosen examples were used, and the desired accuracy is level is $\epsilon = 0.1$?

Answer:

$$m \geq \frac{1}{\epsilon}(4\log_2(2/\delta) + 8d\log_2(13/\epsilon)) \Rightarrow m\epsilon \geq 4\log_2(2/\delta) + 8d\log_2(13/\epsilon)$$

$$\Rightarrow \log_2\delta \geq (4 + 8d\log_2(13/\epsilon) - m\epsilon)/4$$

For $d = 4$: $\log_2\delta \geq 32.18 \Rightarrow \delta \geq 4862133971$

For $d = 3$: $\log_2\delta \geq 18.13 \Rightarrow \delta \geq 287700$ Since these values are greater than 1, it means that we have no confidence at all that such accuracy level can be obtained.

d What is the accuracy level (what is ϵ) if it is known that 1000 randomly chosen examples were used, and the desired confidence level is $\delta = 0.05$?

$$m \geq \frac{1}{\epsilon}(4\log_2(2/\delta) + 8d\log_2(13/\epsilon)) \Rightarrow m\epsilon + 8d\log_2\epsilon \geq 4\log_2(2/\delta) + 8d\log_2 13$$

The following results are computed by program,

$$d = 4, \epsilon \geq 0.2115$$

$$d = 3, \epsilon \geq 0.1713$$

Part III

Do you think that your answer to Part I can be improved based on your answer to Part II? Explain.

In this case the answer is NO, but there are cases in which the results computed using VC dimensions are more sharp than the results computed with the finite case.

Question 2

Consider the problem of learning the concept of " *possible temperature for this month*" which gives minimum and maximum temperatures for each one of the twelve months. The concept has the following form:

> The temperature in January is at least **min1** and at most **max1**.
> The temperature in February is at least **min2** and at most **max2**.
> The temperature in March is at least **min3** and at most **max3**.
> The temperature in April is at least **min4** and at most **max4**.
> The temperature in May is at least **min5** and at most **max5**.
> The temperature in June is at least **min6** and at most **max6**.
> The temperature in July is at least **min7** and at most **max7**.
> The temperature in August is at least **min8** and at most **max8**.
> The temperature in September is at least **min9** and at most **max9**.
> The temperature in October is at least **min10** and at most **max10**.
> The temperature in November is at least **min11** and at most **max11**.
> The temperature in December is at least **min12** and at most **max12**.

The values of **min1, max1, ..., min12, max12** are computed from randomly chosen training examples. A training example has the form of (date and time , temperature), e.g., (4/26/92 at 2 PM , 78).
A test example is a question such as:

> " Is 53 a possible temperature in July?"

It is assumed that the test questions and the training examples come from the same probability distribution. In the following two cases compute how many randomly chosen training examples are needed to guarantee with confidence of at least 95% that at least 90% of the test examples are answered correctly. Specify which formula you use for the computation, and what is the value of each of the variables in the formula.

a. It is known that a solution can be found in which each value of **min1, max1, ..., min12, max12** is represented by 4 bits.

Answer: The number of training examples should be at least 696. The formula used:

$$m \geq \frac{1}{\epsilon} \ln(r/\delta)$$

The variables in the formula have the values: $\epsilon = 0.1, \delta = 0.05, r = 2^{4*24}$

Each min/max value can be represented by 4 bits, so it can take one of $2^4 = 16$ values. The number of hypothesis in concept class is: $r = 2^{4*24}$.

A better answer can be obtained by considering the restriction $\min_i \leq \max_i$, in that case, each pair of $\min_i, \max_i$ can take one of $\sum_{i=1}^{16} i = 136$ pairs of values. the number of hypothesis in concept class is: $r = 136^{12}$. This gives $m \geq 620$.

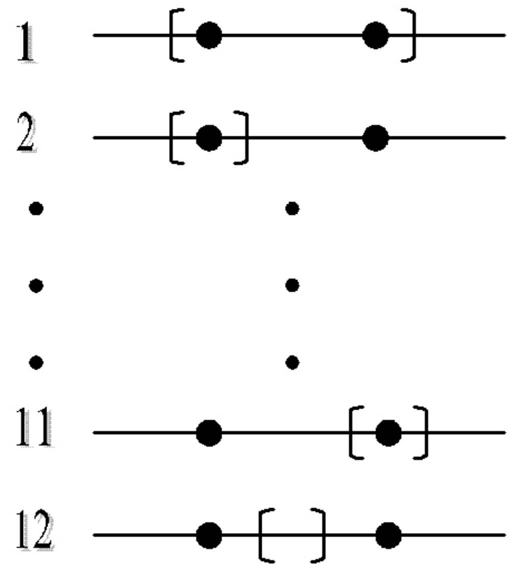
b. It is known that a solution can be found in which each value of **min1, max1, ..., min12, max12** is kept as a real number.

Answer: The number of training examples should be at least 13696. The formula used:

$$m \geq \frac{1}{\epsilon}(4 \log_2(2/\delta) + 8d \log_2(13/\epsilon))$$

The variables in the formula have the values: $\epsilon = 0.1, \delta = 0.05, d = 24$.

The following figure shows that an instance set of size 24, where each month has two examples of different temperature values, can be shattered by the hypothesis class. At the same time, it is easy to show that for any instance set if any month has more than two examples, that instance set can not be shattered by hypothesis space. Thus, the VC dimension of the concept class is $d = 24$.



Question 3

Initialization: $p_i = 1/1000$
Iteration 1:
$\epsilon_1 = 200 \times 1/1000 = 0.2$
$\alpha_1 = 0.5 \ln 0.8/0.2 = 0.693$
$q_{\text{right}} = 0.5, \quad q_{\text{wrong}} = 2$
new p_i: 0.5/1000 for 800 examples, 2/1000 for 200 examples
$Z_1 = 0.4 + 0.4 = 0.8$
normalized p_i: 0.625/1000 for 800 examples, 2.5/1000 for 200 examples
Iteration 2:
$\epsilon_2 = 200 \times 0.625/1000 = 0.125$
$\alpha_2 = 0.973$
$q_{\text{right}} = 0.38, \quad q_{\text{wrong}} = 2.65$
new p_i: 0.238/1000 for 600 examples, 1.66/1000 for 200 examples, 0.95/1000 for 200 examples
$Z_2 = 0.142 + 0.322 + 0.19 = 0.654$
normalized p_i: 0.364/1000 for 600 examples, 2.538/1000 for 200 examples, 1.45/1000 for 200 examples
Iteration 3:
$\epsilon_3 = 100 \times 0.364/1000 = 0.0364$
$\alpha_3 = 1.64$

1 What would be the weight of Classifier A?

Answer: $\alpha_1 = 0.693$

2 What would be the weight of Classifier B?

Answer: $\alpha_2 = 0.973$

3 What would be the weight of Classifier C?

Answer: $\alpha_3 = 1.64$

4 What would be the answer produced by the combined classifier if according to Classifier A the answer is POSITIVE, according to Classifier B it is POSITIVE, and according to Classifier C it is NEGATIVE?

Answer: $0.693 + 0.973 - 1.64 > 0 \quad \Rightarrow$ POSITIVE.

5 What would be the answer produced by the combined classifier if according to Classifier A the answer is POSITIVE, according to Classifier B it is NEGATIVE, and according to Classifier C it is POSITIVE?

Answer: $0.693 - 0.973 + 1.64 > 0 \quad \Rightarrow$ POSITIVE.

6 What would be the answer produced by the combined classifier if according to Classifier A the answer is NEGATIVE, according to Classifier B it is POSITIVE, and according to Classifier C it is POSITIVE?

Answer: $-0.693 + 0.973 + 1.64 > 0 \quad \Rightarrow$ POSITIVE.