Feedback — 作業二

Help Center

You submitted this quiz on **Mon 26 Oct 2015 5:09 AM PDT**. You got a score of **210.00** out of **400.00**. However, you will not get credit for it, since it was submitted past the deadline.

Question 1

Questions 1-2 are about noisy targets.
Consider the bin model for a hypothesis h that makes an error with probability μ in approximating a deterministic target function f (both h and f outputs $\{-1, +1\}$). If we use the same h to approximate a noisy version of f given by

$$P(\mathbf{x}, y) = P(\mathbf{x})P(y|\mathbf{x})$$

$$P(y|\mathbf{x}) = \begin{cases} \lambda & y = f(\mathbf{x}) \\ 1 - \lambda & \text{otherwise} \end{cases}$$

What is the probability of error that h makes in approximating the noisy target y ?

Your Answer		Score	Explanation
○ $1 - \lambda$			
○ $\lambda(1 - \mu) + (1 - \lambda)\mu$			
○ μ			
◉ $\lambda\mu + (1 - \lambda)(1 - \mu)$	✔	20.00	
○ none of the other choices			
Total		20.00 / 20.00	

Question 2

Following Question 1, with what value of λ will the performance of h be independent of μ ?

Your Answer	Score	Explanation
○ 0 or 1		

1		
none of the other choices		
0.5	✔	20.00
0		
Total		20.00 / 20.00

Question 3

Questions 3-5 are about generalization error, and getting the feel of the bounds numerically. Please use the simple upper bound $N^{d_{\text{vc}}}$ on the growth function $m_{\mathcal{H}}(N)$, assuming that $N \geq 2$ and $d_{vc} \geq 2$.

For an $\mathcal{H}$ with $d_{\text{vc}} = 10$, if you want 95% confidence that your generalization error is at most 0.05, what is the closest numerical approximation of the sample size that the VC generalization bound predicts?

Your Answer		Score	Explanation
$480,000$			
$420,000$			
$500,000$			
$440,000$			
$460,000$	✔	20.00	
Total		20.00 / 20.00	

Question 4

There are a number of bounds on the generalization error ϵ, all holding with probability at least $1-\delta$. Fix $d_{\text{vc}} = 50$ and $\delta = 0.05$ and plot these bounds as a function of N. Which bound is the tightest (smallest) for very large N, say $N = 10,000$? Note that Devroye and Parrondo & Van den Broek are implicit bounds in ϵ.

Your Answer	Score	Explanation

- Original VC bound: $\epsilon \le \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}$

- Rademacher Penalty Bound:
$\epsilon \le \sqrt{\frac{2\ln(2Nm_{\mathcal{H}}(N))}{N}} + \sqrt{\frac{2}{N}\ln\frac{1}{\delta}} + \frac{1}{N}$

- Parrondo and Van den Broek: $\epsilon \le \sqrt{\frac{1}{N}(2\epsilon + \ln\frac{6m_{\mathcal{H}}(2N)}{\delta})}$

- Devroye: $\epsilon \le \sqrt{\frac{1}{2N}(4\epsilon(1+\epsilon) + \ln\frac{4m_{\mathcal{H}}(N^2)}{\delta})}$

- Variant VC bound: $\epsilon \le \sqrt{\frac{16}{N}\ln\frac{2m_{\mathcal{H}}(N)}{\sqrt{\delta}}}$

Total	0.00 / 20.00

Question 5

Continuing from Question 4, for small N, say $N = 5$, which bound is the tightest (smallest)?

Your Answer	Score	Explanation
Devroye		
Original VC bound		
Parrondo and Van den Broek		
Variant VC bound		
Rademacher Penalty Bound		
Total	0.00 / 20.00	

Question 6

In Questions 6-11, you are asked to play with the *growth function* or VC-dimension of some hypothesis sets.

What is the growth function $m_{\mathcal{H}}(N)$ of ``positive-and-negative intervals on $\mathbb{R}$''? The hypothesis set $\mathcal{H}$ of ``positive-and-negative intervals'' contains the functions which are $+1$ within an interval $[\ell, r]$ and -1 elsewhere, as well as the functions which are -1 within an interval $[\ell, r]$ and $+1$ elsewhere. For instance, the hypothesis $h_1(x) = \text{sign}(x(x-4))$ is a negative interval with -1

within $[0, 4]$ and $+1$ elsewhere, and hence belongs to $\mathcal{H}$. The hypothesis $h_2(x) = \text{sign}((x+1)(x)(x-1))$ contains two positive intervals in $[-1, 0]$ and $[1, \infty)$ and hence does not belong to $\mathcal{H}$.

Your Answer		Score	Explanation
N^2			
$N^2 + 1$			
none of the other choices.			
$N^2 + N + 2$			
$N^2 - N + 2$	✔	20.00	
Total		20.00 / 20.00	

Question 7

Continuing from the previous problem, what is the VC-dimension of the ``positive-and-negative intervals on $\mathbb{R}$''

Your Answer		Score	Explanation
4			
∞			
2			
3	✔	20.00	
5			
Total		20.00 / 20.00	

Question 8

What is the growth function $m_{\mathcal{H}}(N)$ of ``positive donuts in $\mathbb{R}^2$''? The hypothesis set $\mathcal{H}$ of ``positive donuts'' contains hypotheses formed by two concentric circles centered at the origin. In particular, each hypothesis is $+1$ within a ``donut'' region of $a^2 \leq x_1^2 + x_2^2 \leq b^2$ and -1

elsewhere. Without loss of generality, we assume $0 < a < b < \infty$.

Your Answer		Score	Explanation
none of the other choices.			
$\binom{N}{2} + 1$			
$\binom{N+1}{3} + 1$			
$N + 1$			
(selected) $\binom{N+1}{2} + 1$	✔	20.00	
Total		20.00 / 20.00	

Question 9

Consider the ``polynomial discriminant'' hypothesis set of degree D on $\mathbb{R}$, which is given by

$$\mathcal{H} = \left\{ h_{\mathbf{c}} \;\middle|\; h_{\mathbf{c}}(x) = \mathrm{sign}\left(\sum_{i=0}^{D} c_i x^i \right) \right\}$$

What is the VC-Dimension of such an $\mathcal{H}$?

Your Answer		Score	Explanation
(selected) $D + 1$	✔	20.00	
$D + 2$			
none of the other choices.			
∞			
D			
Total		20.00 / 20.00	

Question 10

Consider the ``simplified decision trees'' hypothesis set on $\mathbb{R}^d$, which is given by

$$\mathcal{H} = \{h_{\mathbf{t},\mathbf{S}} \mid h_{\mathbf{t},\mathbf{S}}(\mathbf{x}) = 2[\![\mathbf{v} \in S]\!] - 1, \text{where } v_i = [\![x_i > t_i]\!], \\ \mathbf{S} \text{ a collection of vectors in } \{0,1\}^d, \mathbf{t} \in \mathbb{R}^d \quad \}$$

That is, each hypothesis makes a prediction by first using the d thresholds t_i to locate $\mathbf{x}$ to be within one of the 2^d hyper-rectangular regions, and looking up $\mathbf{S}$ to decide whether the region should be $+1$ or -1. What is the VC-dimension of the ``simplified decision trees'' hypothesis set?

Your Answer		Score	Explanation
○ ∞			
○ 2^d			
◉ $2^{d+1} - 3$	✖	-5.00	
○ none of the other choices.			
○ 2^{d+1}			
Total		-5.00 / 20.00	

Question 11

Consider the ``triangle waves'' hypothesis set on $\mathbb{R}$, which is given by

$$\mathcal{H} = \{h_\alpha \mid h_\alpha(x) = \text{sign}(|(\alpha x) \bmod 4 - 2| - 1), \alpha \in \mathbb{R}\}$$

Here $(z \bmod 4)$ is a number $z - 4k$ for some integer k such that $z - 4k \in [0, 4)$. For instance, $(11.26 \bmod 4)$ is 3.26, and $(-11.26 \bmod 4)$ is 0.74. What is the VC-Dimension of such an $\mathcal{H}$?

Your Answer		Score	Explanation
○ none of the other choices.			
○ 2			
○ 1			
○ 3			
◉ ∞	✔	20.00	
Total		20.00 / 20.00	

Question 12

In Questions 12-15, you are asked to verify some properties or bounds on the growth function and VC-dimension.

Which of the following is an upper bound of the growth function $m_{\mathcal{H}}(N)$ for $N \geq d_{vc} \geq 2$?

Your Answer		Score	Explanation
$m_{\mathcal{H}}\left(\lfloor \frac{N}{2} \rfloor\right)$			
none of the other choices			
$\sqrt{N^{d_{vc}}}$			
$\min\limits_{1 \leq i \leq N-1} 2^i m_{\mathcal{H}}(N-i)$	✔	20.00	
$2^{d_{vc}}$			
Total		20.00 / 20.00	

Question 13

Which of the following is not a possible growth function $m_{\mathcal{H}}(N)$ for some hypothesis set?

Your Answer		Score	Explanation
none of the other choices			
2^N			
$2^{\lfloor \sqrt{N} \rfloor}$	✔	20.00	
N			
$N^2 - N + 2$			
Total		20.00 / 20.00	

Question 14

For hypothesis sets $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_K$ with finite, positive VC dimensions $d_{vc}(\mathcal{H}_k)$, some of the following bounds are correct and some are not. Which among the correct ones is the tightest bound on the VC dimension of the **intersection** of the sets: $d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k)$? (The VC dimension of an empty set or a singleton set is taken as zero)

Your Answer		Score	Explanation
$0 \leq d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq \min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$	✔	20.00	
$0 \leq d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq \max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$			
$\min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \leq d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$			
$\min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \leq d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq \max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$			
$0 \leq d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$			
Total		20.00 / 20.00	

Question 15

For hypothesis sets $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_K$ with finite, positive VC dimensions $d_{vc}(\mathcal{H}_k)$, some of the following bounds are correct and some are not. Which among the correct ones is the tightest bound on the VC dimension of the **union** of the sets: $d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k)$?

Your Answer		Score	Explanation
$\max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \leq d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k) \leq K-1+\sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$	✔	20.00	
$0 \leq d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k) \leq K-1+\sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$			
$\min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \leq d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k) \leq \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$			
$0 \leq d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k) \leq \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$			
$\max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \leq d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k) \leq \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$			
Total		20.00 / 20.00	

Question 16

For Questions 16-20, you will play with the decision stump algorithm.

In class, we taught about the learning model of ``positive and negative rays'' (which is simply one-dimensional perceptron) for one-dimensional data. The model contains hypotheses of the form:

$$h_{s,\theta}(x) = s \cdot \text{sign}(x - \theta).$$

The model is frequently named the ``decision stump'' model and is one of the simplest learning models. As shown in class, for one-dimensional data, the VC dimension of the decision stump model is 2.

In fact, the decision stump model is one of the few models that we could easily minimize E_{in} efficiently by enumerating all possible thresholds. In particular, for N examples, there are at most $2N$ dichotomies (see page 22 of class05 slides), and thus at most $2N$ different E_{in} values. We can then easily choose the dichotomy that leads to the lowest E_{in}, where ties can be broken by randomly choosing among the lowest-E_{in} ones. The chosen dichotomy stands for a combination of some `spot' (range of θ) and s, and commonly the median of the range is chosen as the θ that realizes the dichotomy.

In this problem, you are asked to implement such and algorithm and run your program on an artificial data set. First of all, start by generating a one-dimensional data by the procedure below:

(a) Generate x by a uniform distribution in $[-1, 1]$.
[b) Generate y by $f(x) = \tilde{s}(x) +$ noise where $\tilde{s}(x) = \text{sign}(x)$ and the noise flips the result with 20% probability.

For any decision stump $h_{s,\theta}$ with $\theta \in [-1, 1]$, express $E_{out}(h_{s,\theta})$ as a function of θ and s.

Your Answer		Score	Explanation
◉ $0.3 + 0.5s(\lvert\theta\rvert - 1)$	✖	-5.00	
○ $0.5 + 0.3s(1 - \lvert\theta\rvert)$			
○ none of the other choices			
○ $0.5 + 0.3s(\lvert\theta\rvert - 1)$			
○ $0.3 + 0.5s(1 - \lvert\theta\rvert)$			
Total		-5.00 / 20.00	

Question 17

Generate a data set of size 20 by the procedure above and run the one-dimensional decision stump algorithm on the data set. Record E_{in} and compute E_{out} with the formula above. Repeat the experiment (including data generation, running the decision stump algorithm, and computing E_{in} and E_{out}) $5,000$ times. What is the average E_{in}? Choose the closest option.

Your Answer		Score	Explanation
0.15			
0.35			
0.25	✖	-5.00	
0.45			
0.05			
Total		-5.00 / 20.00	

Question 18

Continuing from the previous question, what is the average E_{out}? Choose the closest option.

Your Answer		Score	Explanation
0.25			
0.45			
0.15	✖	-5.00	
0.35			
0.05			
Total		-5.00 / 20.00	

Question 19

Decision stumps can also work for multi-dimensional data. In particular, each decision stump now deals with a specific dimension i, as shown below.

$$h_{s,i,\theta}(\mathbf{x}) = s \cdot \text{sign}(x_i - \theta).$$

Implement the following decision stump algorithm for multi-dimensional data:

a) for each dimension $i = 1, 2, \cdots, d$, find the best decision stump $h_{s,i,\theta}$ using the one-dimensional decision stump algorithm that you have just implemented.
b) return the ``best of best'' decision stump in terms of E_{in}. If there is a tie, please randomly choose among the lowest-E_{in} ones.

The training data $\mathcal{D}_{train}$ is available at:

https://d396qusza40orc.cloudfront.net/ntumlone%2Fhw2%2Fhw2_train.dat

The testing data $\mathcal{D}_{test}$ is available at:

https://d396qusza40orc.cloudfront.net/ntumlone%2Fhw2%2Fhw2_test.dat

Run the algorithm on the $\mathcal{D}_{train}$. Report the E_{in} of the optimal decision stump returned by your program. Choose the closest option.

Your Answer		Score	Explanation
◉ 0.05	✖	-5.00	
○ 0.25			
○ 0.45			
○ 0.35			
○ 0.15			
Total		-5.00 / 20.00	

Question 20

Use the returned decision stump to predict the label of each example within the $\mathcal{D}_{test}$. Report an estimate of E_{out} by E_{test}. Choose the closest option.

Your Answer		Score	Explanation
○ 0.15			
○ 0.45			
◉ 0.25	✖	-5.00	
○ 0.35			

◯ 0.05

Total	-5.00 / 20.00