Learning From Data
Caltech
http://work.caltech.edu/telecourse.html
2013

Online Homework # 4

Collaboration in the sense of discussions is allowed, but you should NOT discuss your selected answers with anyone. Books and notes can be consulted. All questions will have multiple choice answers ([**a**], [**b**], [**c**], **...***). You should enter your solutions online by logging into your account at the course web site.*

Note about the homeworks

- The goal of the homeworks is to facilitate a deeper understanding of the course material. The questions are not designed to be puzzles with catchy answers. They are meant to make you roll your sleeves, face uncertainties, and approach the problem from different angles.
- The problems range from easy to hard, and from theoretical to practical. Some problems require running a full experiment to arrive at the answer.
- The answer may not be obvious or numerically close to one of the choices. The intent is to prompt discussion and exchange of ideas.
- Speaking of discussion, you are encouraged to take part in the forum

  http://book.caltech.edu/bookforum

  where there are many threads about each homework. We hope that you will contribute to the discussion as well.
- Please follow the forum guidelines for posting answers (see the "BEFORE posting answers" announcement at the top there).

Generalization Error

In Problems 1-3, we look at generalization bounds numerically. For $N > d_{\text{VC}}$, use the simple approximate bound $N^{d_{\text{VC}}}$ for the growth function $m_{\mathcal{H}}(N)$.

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

[**a**] 420,000

[**b**] 430,000

[**c**] 440,000

[**d**] 450,000

[**e**] 460,000

2. 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 smallest for very large N? Note that [**c**] and [**d**] are implicit bounds in ϵ.

[**a**] Original VC bound: $\epsilon \leq \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}$

[**b**] Rademacher Penalty Bound: $\epsilon \leq \sqrt{\frac{2\ln(2Nm_{\mathcal{H}}(N))}{N}} + \sqrt{\frac{2}{N} \ln \frac{1}{\delta}} + \frac{1}{N}$

[**c**] Parrondo and Van den Broek: $\epsilon \leq \sqrt{\frac{1}{N}(2\epsilon + \ln \frac{6m_{\mathcal{H}}(2N)}{\delta})}$

[**d**] Devroye: $\epsilon \leq \sqrt{\frac{1}{2N}(4\epsilon(1+\epsilon) + \ln \frac{4m_{\mathcal{H}}(N^2)}{\delta})}$

[**e**] They are all equal.

3. For small N, say $N = 5$, which bound is the smallest?

[**a**] Original VC bound: $\epsilon \leq \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}$

[**b**] Rademacher Penalty Bound: $\epsilon \leq \sqrt{\frac{2\ln(2Nm_{\mathcal{H}}(N))}{N}} + \sqrt{\frac{2}{N} \ln \frac{1}{\delta}} + \frac{1}{N}$

[**c**] Parrondo and Van den Broek: $\epsilon \leq \sqrt{\frac{1}{N}(2\epsilon + \ln \frac{6m_{\mathcal{H}}(2N)}{\delta})}$

[**d**] Devroye: $\epsilon \leq \sqrt{\frac{1}{2N}(4\epsilon(1+\epsilon) + \ln \frac{4m_{\mathcal{H}}(N^2)}{\delta})}$

[e] They are all equal.

Bias and Variance

Consider the case where the target function $f : [-1, 1] \to \mathbb{R}$ is given by $f(x) = \sin(\pi x)$ and the input probability distribution is uniform on $[-1, 1]$. Assume that the training set has only two examples (picked independently), and that the learning algorithm picks the hypothesis that minimizes the mean squared error on the examples.

4. Assume the learning model consists of all hypotheses of the form $h(x) = ax$. What is the expected value of the hypothesis, $\bar{g}(x)$, that the learning algorithm produces (expected value with respect to the data set)? Coefficients are given to second decimal digit accuracy.

[a] $\bar{g}(x) = 0$

[b] $\bar{g}(x) = 0.13x$

[c] $\bar{g}(x) = 0.45x$

[d] $\bar{g}(x) = 0.79x$

[e] None of the above

5. What is the closest value to the bias in this case?

[a] 0.1

[b] 0.3

[c] 0.5

[d] 0.7

[e] 1.0

6. What is the closest value to the variance in this case?

[a] 0.2

[b] 0.4

[c] 0.6

[d] 0.8

[e] 1.0

7. Now, let's change $\mathcal{H}$. Which of the following learning models has the least expected value of out-of-sample error?

[a] Hypotheses of the form $h(x) = b$.

[b] Hypotheses of the form $h(x) = ax$.

[c] Hypotheses of the form $h(x) = ax + b$.

[d] Hypotheses of the form $h(x) = ax^2$.

[e] Hypotheses of the form $h(x) = ax^2 + b$.

VC Dimension

8. Assume $q \geq 1$ is an integer and let $m_{\mathcal{H}}(1) = 2$. What is the VC dimension of a hypothesis set whose growth function satisfies: $m_{\mathcal{H}}(N+1) = 2m_{\mathcal{H}}(N) - \binom{N}{q}$ (*Hint: there is no way to choose 10 people from a group of 7 people*)

[a] $q - 2$

[b] $q - 1$

[c] q

[d] $q + 1$

[e] None of the above

9. For hypothesis sets $\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_K$ with finite VC dimensions $d_{\text{VC}}(\mathcal{H}_k)$, some of the following bounds are correct and some are not. Which among the correct ones is the tightest bound (the smallest range of values) on the VC dimension of the **intersection** of the sets: $d_{\text{VC}}(\bigcap_{k=1}^{K} \mathcal{H}_k)$?

[a] $0 \leq d_{\text{VC}}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq \sum_{k=1}^{K} d_{\text{VC}}(\mathcal{H}_k)$

[b] $0 \leq d_{\text{VC}}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq \min\{d_{\text{VC}}(\mathcal{H}_k)\}_{k=1}^{K}$

[c] $0 \leq d_{\text{VC}}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq \max\{d_{\text{VC}}(\mathcal{H}_k)\}_{k=1}^{K}$

[d] $\min\{d_{\text{VC}}(\mathcal{H}_k)\}_{k=1}^{K} \leq d_{\text{VC}}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq \max\{d_{\text{VC}}(\mathcal{H}_k)\}_{k=1}^{K}$

[e] $\min\{d_{\rm VC}(\mathcal{H}_k)\}_{k=1}^{K} \;\leq\; d_{\rm VC}(\bigcap_{k=1}^{K}\mathcal{H}_k) \;\leq\; \sum_{k=1}^{K} d_{\rm VC}(\mathcal{H}_k)$

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

[**a**] $0 \;\leq\; d_{\rm VC}(\bigcup_{k=1}^{K}\mathcal{H}_k) \;\leq\; \sum_{k=1}^{K} d_{\rm VC}(\mathcal{H}_k)$

[**b**] $0 \;\leq\; d_{\rm VC}(\bigcup_{k=1}^{K}\mathcal{H}_k) \;\leq\; K-1+\sum_{k=1}^{K} d_{\rm VC}(\mathcal{H}_k)$

[**c**] $\min\{d_{\rm VC}(\mathcal{H}_k)\}_{k=1}^{K} \;\leq\; d_{\rm VC}(\bigcup_{k=1}^{K}\mathcal{H}_k) \;\leq\; \sum_{k=1}^{K} d_{\rm VC}(\mathcal{H}_k)$

[**d**] $\max\{d_{\rm VC}(\mathcal{H}_k)\}_{k=1}^{K} \;\leq\; d_{\rm VC}(\bigcup_{k=1}^{K}\mathcal{H}_k) \;\leq\; \sum_{k=1}^{K} d_{\rm VC}(\mathcal{H}_k)$

[**e**] $\max\{d_{\rm VC}(\mathcal{H}_k)\}_{k=1}^{K} \;\leq\; d_{\rm VC}(\bigcup_{k=1}^{K}\mathcal{H}_k) \;\leq\; K-1+\sum_{k=1}^{K} d_{\rm VC}(\mathcal{H}_k)$