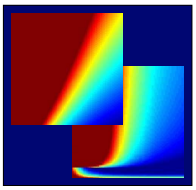
Learning From Data
Yaser Abu-Mostafa, *Caltech*
http://work.caltech.edu/telecourse
Self-paced version



Homework # 4

*All questions have multiple-choice answers (***[a], [b], [c], ...***). You can collaborate with others, but do not discuss the selected or excluded choices in the answers. You can consult books and notes, but not other people's solutions. Your solutions should be based on your own work. Definitions and notation follow the lectures.*

Note about the homework

- The goal of the homework 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 up your sleeves, face uncertainties, and approach the problem from different angles.
- The problems range from easy to difficult, and from practical to theoretical. 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, but one (and only one) choice will be correct if you follow the instructions precisely in each problem. You are encouraged to explore the problem further by experimenting with variations on these instructions, for the learning benefit.
- You are also encouraged to take part in the forum

  http://book.caltech.edu/bookforum

  where there are many threads about each homework set. 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 $d_{\text{VC}} = 10$, 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?

[a] 400,000

[b] 420,000

[c] 440,000

[d] 460,000

[e] 480,000

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, say $N = 10,000$? 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 the same values of d_{VC} and δ of Problem 2, but 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 produces 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, $\bar{g}(x)$, of the hypothesis produced by the learning algorithm (expected value with respect to the data set)? Express your $\bar{g}(x)$ as $\hat{a}x$, and round $\hat{a}$ to two decimal digits only, then match *exactly* to one of the following answers.

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

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

[c] $\bar{g}(x) = 1.07x$

[d] $\bar{g}(x) = 1.58x$

[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: $m_{\mathcal{H}}(N+1) = 2m_{\mathcal{H}}(N) - \binom{N}{q}$? Recall that $\binom{M}{m} = 0$ when $m > M$.

[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, positive 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: $d_{\text{VC}}(\bigcap_{k=1}^{K} \mathcal{H}_k)$? (The VC dimension of an empty set or a singleton set is taken as zero)

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

10. For hypothesis sets $\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_K$ with finite, positive 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 **union** of the sets: $d_{\text{VC}}(\bigcup_{k=1}^{K} \mathcal{H}_k)$?

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

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

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

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

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