Q1-1: Consider Instance space $X$, Hypothesis space $H$, training examples $D$. Suppose that every $h$ in our version space $\text{VS}_{H, D}$ is consistent with $m$ training examples. Which of the following statement(s) is/are TRUE?

A. The version space $\text{VS}_{H, D}$ is $\varepsilon$-exhausted with respect to $c$ and $D$ if every hypothesis $h$ in $\text{VS}_{H, D}$ has training error < $\varepsilon$.

B. “The probability that $\text{VS}_{H, D}$ is not $\varepsilon$-exhausted $\leq |H|e^{-\varepsilon m}$” provides a bound for the probability that ANY learner will output a hypothesis $h$ with error > $\varepsilon$.

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.
Q1-1: Consider Instance space $X$, Hypothesis space $H$, training examples $D$. Suppose that every $h$ in our version space $V_{SH, D}$ is consistent with $m$ training examples. Which of the following statement(s) is/are TRUE?

A. The version space $V_{SH, D}$ is $\varepsilon$-exhausted with respect to $c$ and $D$ if every hypothesis $h$ in $V_{SH, D}$ has training error $< \varepsilon$.

B. “The probability that $V_{SH, D}$ is not $\varepsilon$-exhausted $\leq |H|e^{-m}$” provides a bound for the probability that ANY learner will output a hypothesis $h$ with error $> \varepsilon$.

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.
Q1-2: We know that Probability that the version space is not $\epsilon$-exhausted after $m$ training examples is at most $|H|e^{-\epsilon m}$ where symbols have their usual meanings. From this we can derive that if $\text{error}_{\text{train}}(h) = 0$, then with probability at least $(1 - \delta)$, $\text{error}_{\text{true}}(h) \leq 1/m \times A$ where $A$ is

1. $\ln(|H|) + \ln(\delta)$
2. $\ln(|H|) + \ln(1/\delta)$
3. $\ln(1/|H|) + \ln(\delta)$
4. $\ln(1/|H|) + \ln(1/\delta)$
Q1-2: We know that Probability that the version space is not $\epsilon$-exhausted after $m$ training examples is at most $|H|e^{-\epsilon m}$ where symbols have their usual meanings. From this we can derive that if $\text{error}_{\text{train}}(h) = 0$, then with probability at least $(1 - \delta)$, $\text{error}_{\text{true}}(h) \leq 1/m \times A$ where $A$ is

1. $\ln(|H|) + \ln(\delta)$
2. $\ln(|H|) + \ln(1/\delta)$
3. $\ln(1/|H|) + \ln(\delta)$
4. $\ln(1/|H|) + \ln(1/\delta)$

We can write the bound as follows:

\[
\Pr[\exists h \in H \text{ s.t. } (\text{error}_{\text{train}}(h) = 0) \text{ AND } (\text{error}_{\text{true}}(h) > \epsilon)] \leq |H|e^{-\epsilon m} = \delta
\]

Hence, with at least probability $(1 - \delta)$, $\text{error}_{\text{true}}(h) \leq \epsilon$, where $\epsilon$ can be obtained by solving $|H|e^{-\epsilon m} = \delta$. 


Q2-1: Which of the following statement(s) is/are TRUE?

A. PAC analysis formalizes the learning task and allows for non-perfect learning.
B. Finding a consistent hypothesis is easier for smaller concept classes.
C. In PAC analysis, we are trying to bound training error of a hypothesis.

1. True, True, True
2. True, False, False
3. True, False, True
4. False, False, False
Q2-1: Which of the following statement(s) is/are TRUE?

A. PAC analysis formalizes the learning task and allows for non-perfect learning.
B. Finding a consistent hypothesis is easier for smaller concept classes.
C. In PAC analysis, we are trying to bound training error of a hypothesis.

1. True, True, True
2. True, False, False
3. True, False, True
4. False, False, False
Q2-2: Similar to “PAC analysis example: learning decision trees”, consider the case when each instance has \( n \) features and each feature has 3 possible values. Let learned hypotheses are DTs of depth 2 using only 2 variables. What is \(|H|\)? Note: \( \text{nC2} = \binom{n}{2} \)

1. \( \text{nC3} \times 2^6 \)
2. \( \text{nC3} \times 2^9 \)
3. \( \text{nC2} \times 2^9 \)
4. \( \text{nC2} \times 2^6 \)
Q2-2: Similar to “PAC analysis example: learning decision trees”, consider the case when each instance has $n$ features and each feature has 3 possible values. Let learned hypotheses are DTs of depth 2 using only 2 variables. What is $|H|$? Note: $n\binom{\frac{n}{2}}{2}$

1. $n\binom{3}{2} \times 2^6$
2. $n\binom{3}{2} \times 2^9$
3. $n\binom{2}{2} \times 2^9$
4. $n\binom{2}{2} \times 2^6$

#possible split choices (choose 2 features out of n) = nC2
#leaves = 9. Each leaf has 2 possible labellings. So, #possible leaf labellings = $2^9$
Q3-1: Which of the following statement(s) is/are TRUE?

A. Agnostic PAC learnability requires learning a hypothesis with true error at most $\epsilon$

B. In the case of agnostic PAC learning, if $\text{errortrain}(h_{\text{best}}) = 0$, then with probability at least $(1 - \delta)$, $\text{errortrue}(h_{\text{best}}) \leq (1/2m) \sqrt{\ln |H| + \ln(1/\delta)}$

1. True, True
2. True, False
3. False, True
4. False, False
Q3-1: Which of the following statement(s) is/are TRUE?

A. Agnostic PAC learnability requires learning a hypothesis with true error at most $\varepsilon$

B. In the case of agnostic PAC learning, if $\text{error}_{\text{train}}(h_{\text{best}}) = 0$, then with probability at least $(1 - \delta)$, $\text{error}_{\text{true}}(h_{\text{best}}) \leq (1/2m) \sqrt{\ln |H| + \ln(1/\delta)}$

1. True, True
2. True, False
3. False, True
4. False, False

Being agnostic PAC learnable is a stronger condition, since for all distributions, we can get close to the optimal error.

We want to get the expression for $\varepsilon$ in agnostic PAC learning case. Solve for $\varepsilon$ using:

$$m \geq \frac{1}{2\varepsilon^2} \left( \ln |H| + \ln \left( \frac{1}{\delta} \right) \right)$$
Q3-2: What’s the VC dimension of a decision stump in $\mathbb{R}^2$? A decision stump is of the form $f(x) = \text{sign}(x_i > b)$ or $f(x) = \text{sign}(x_i < b)$ for one feature $x_i$ and some bias parameter $b$.

1. VC = 1
2. VC = 2
3. VC = 3
4. VC = 4
Q3-2: What’s the VC dimension of a decision stump in $\mathbb{R}^2$? A decision stump is of the form $f(x) = \text{sign}(x_i > b)$ or $f(x) = \text{sign}(x_i < b)$ for one feature $x_i$ and some bias parameter $b$.

1. VC = 1
2. VC = 2
3. VC = 3
4. VC = 4

VC ≥ 3: All possible labelings of 3 points can be shattered.

VC = 3: For all placements of 4 points, there exist a labeling that can’t be shattered.