Q1-1: Are these statements true or false?
(A) More data can help reduce the estimation error.
(B) Choosing hypothesis class of higher capacity can help reduce the approximation error.

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

Q1-1: Are these statements true or false?
(A) More data can help reduce the estimation error.
(B) Choosing hypothesis class of higher capacity can help reduce the approximation error.

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

(A) More data can make the training set be a better approximation for the data distribution, thus reducing the estimation error.
(B) Hypothesis class of higher capacity can be more expressive and give a more powerful modeling for the problem, thus reducing the approximation error.

Q1-2: Are these statements true or false?

(A) Since $err(\hat{h}_{opt}) - err(h_{opt}) \leq 2 \sup_{h \in \mathcal{H}} |err(h) - \widehat{err}(h)|$, a larger hypothesis class will give us a tighter bound on the estimation error.

(B) With well control on the generalization gap, our model which performs well on the training data would also perform well on unseen data.

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

Q1-2: Are these statements true or false?

(A) Since $err(\hat{h}_{opt}) - err(h_{opt}) \leq 2 \sup_{h \in \mathcal{H}} |err(h) - \widehat{err}(h)|$, a larger hypothesis class will give us a tighter bound on the estimation error.

(B) With well control on the generalization gap, our model which performs well on the training data would also perform well on unseen data.

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

(A) The value of R.H.S increases as the hypothesis class $\mathcal{H}$ grows larger, so it would give a looser bound.

(B) We can decompose the $err(\hat{h})$ to the training error and generalization gap. When both training error and generalization gap are small, the true error would be small, thus performing well on unseen data.

Q2-1: Are these statements true or false?
(A) Bias decreases with models of higher capacity.
(B) Variance decreases with models of higher capacity.

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

Q2-1: Are these statements true or false?

(A) Bias decreases with models of higher capacity.
(B) Variance decreases with models of higher capacity.

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

(A) Model of higher capacity is more expressive, thus providing better predictions in expectation and having smaller bias.
(B) Model of higher capacity is more sensitive to data and would vary more from its expected prediction, thus having larger variance.

Q2-2: Are these statements true or false?
(A) If we increase the training data, the optimal point of bias-variance tradeoff would not change.
(B) Models of higher capacity are more likely to overfit because they tend to have higher variance and are more sensitive when data changes.

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

Q2-2: Are these statements true or false?
(A) If we increase the training data, the optimal point of bias-variance tradeoff would not change.
(B) Models of higher capacity are more likely to overfit because they tend to have higher variance and are more sensitive when data changes.

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

(A) As is shown in the lecture.
(B) That’s true. So the model would have worse generalization performance on those data even with slightly difference from the training data.

Q3-1: Are these statements true or false?
(A) There is always a unique hypothesis to be consistent with the training set.
(B) It’s usually unrealistic to expect our learning method to have zero true error with finite training data.

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

Q3-1: Are these statements true or false?
(A) There is always a unique hypothesis to be consistent with the training set.
(B) It's usually unrealistic to expect our learning method to have zero true error with finite training data.

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

(A) When training data is finite, there can be multiple hypotheses consistent to the training set.
(B) As is shown in the lecture. Apart from the reason in (A), it's also possible that training data is not representative.

Q3-2: Are these statements true or false?
(A) If $\mathcal{C}$ over instance space $\mathcal{X}$ is PAC-learnable, then for some $c \in \mathcal{C}$, distribution $\mathcal{D}$ over $\mathcal{X}$ and $\varepsilon > 0$, we can always learn a hypothesis h whose true error bounded by ε.

(B) If we have a hypothesis space $\mathcal{H}$, for any $c \in \mathcal{C}$ over instance space $\mathcal{X}$, distribution $\mathcal{D}$ over $\mathcal{X}$, $0 < \varepsilon < 0.5$ and $0 < \delta < 0.5$, we can always output a $h \in \mathcal{H}$ whose true error bounded by ε with probability at least $(1 - \delta)$ in time $M_{\mathcal{X},c} \exp(\frac{1}{\varepsilon}) \frac{1}{\delta^2}$, where $M_{\mathcal{X},c}$ is dependent on the length of $\mathcal{X}$ and size of c, then $\mathcal{C}$ is PAC-learnable.

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

Q3-2: Are these statements true or false?

(A) If $\mathcal{C}$ over instance space $\mathcal{X}$ is PAC-learnable, then for some $c \in \mathcal{C}$, distribution $\mathcal{D}$ over $\mathcal{X}$ and $\varepsilon > 0$, we can always learn a hypothesis h whose true error bounded by ε.

(B) If we have a hypothesis space $\mathcal{H}$, for any $c \in \mathcal{C}$ over instance space $\mathcal{X}$, distribution $\mathcal{D}$ over $\mathcal{X}$, $0 < \varepsilon < 0.5$ and $0 < \delta < 0.5$, we can always output a $h \in \mathcal{H}$ whose true error bounded by ε with probability at least $(1 - \delta)$ in time $M_{\mathcal{X},c} \exp(\frac{1}{\varepsilon}) \frac{1}{\delta^2}$, where $M_{\mathcal{X},c}$ is dependent on the length of $\mathcal{X}$ and size of c, then $\mathcal{C}$ is PAC-learnable.

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

(A) In PAC framework, there would be $0 < \delta < 0.5$ such that we find such hypothesis with probability at least $(1 - \delta)$. It can happen that we fail to find such hypothesis, as the probability is not guaranteed to be 1. Sometimes even 1 cannot guarantee as well!

(B) The time is not polynomial in $\frac{1}{\varepsilon}$, which is exponential, so it violates our PAC definition.