Consider a learning problem where there are two features x1 and x2 that are ''duplicate'' of one another. What will be the effect of such features on the decision tree learning? What will be the effect of such features on previously learned linear classifiers such as logistic regression or perceptron?

A. DT may use x1 or x2, or neither, but will not use both
B. DT may use x1 or x2 or both, or neither
C. Linear classifiers such as perceptron or logistic regression will have identical weights for both x1 and x2
D. Linear classifiers such as perceptron or logistic regression may have different weights for x1 and x2

Answer: assuming the feature is binary – between A and B, A is correct. But if multinomial or continuous, then B is correct because DT could potentially test on one value for X1 and test another for X2 at different nodes of the tree.
Ignoring the initialization (or assuming initializing with zero vector, C is correct. All the updates will be identical for X1 and X2, so the weights will be the same.

Consider the information gain of a feature x for label y, which of the following statements are true:

A. Information gain can be negative
B. Information gain is bounded ($\leq$) by $H(x)$
C. Information gain is bounded ($\leq$) by $H(y)$
D. The information gain on y from x is the same as the information gain on x from y.

Answer: B, C, D

How will bagging influence different parts in the expected loss?

A. Bagging can reduce the bias
B. Bagging can reduce the variance
C. Bagging can reduce the noise

Answer: B