Quiz 2

October 1, 2013

Name: _____ **NetID:** _____

Question 1. (1 point) What is the definition of **generalization error** for a hypothesis h on a prediction task $P(X, Y)$ if Δ is the 0-1 loss function?

(a) $\sum_{x \in X, y \in Y} \Delta(h(x), y)$

(b) $\sum_{x \in X, y \in Y} \Delta(h(x), y) P(Y = y)$

(c) $\sum_{x \in X, y \in Y} \Delta(h(x), y) P(X = x, Y = y)$

(d) None of the above

Question 2. (1 point) Suppose we have a binary classification problem where instances have 3 boolean attributes. We want to construct a **one-level decision tree** to classify new points. Given the following training data, which attribute should our decision tree split on to maximize **information gain**?

x_1	x_2	x_3	Label
false	false	true	+
true	false	false	+
true	true	true	-
true	true	false	-

(a) x_1 (b) x_2 (c) x_3 (d) It doesn't matter

Question 3. (1 point) If we hard-coded the decision tree from the previous question to split on attribute x_1, what would its **training error** be?

(a) 25% (b) 50% (c) 75% (d) 100%

Question 4. (1 point) Which of the following techniques is **not** commonly used to reduce the risk of overfitting in decision trees?

(a) Early stopping

(b) Rule post-pruning

(c) Reduced error pruning

(d) Increased error pruning