

# Quiz 1

September 12, 2013

Name: \_\_\_\_\_ NetID: \_\_\_\_\_

**Question 1. (1 point)** Suppose we are trying to learn the concept of "good weather". We formulate a binary classification problem where instances are *weather reports* and a positive label corresponds to a "good" weather report. Each instance has 3 attributes:

- Temperature ( $^{\circ}\text{F}$ ):  $\{< 20, [20 - 39], [40 - 59], [60 - 79], 80 <\}$
- Precipitation:  $\{\text{None, Rain, Snow}\}$
- Windy:  $\{\text{Yes, No}\}$

What is the **size of the instance space**? What is the **size of the hypothesis space**, if it consists of all possible functions  $h : (\text{Temperature, Precipitation, Windy}) \rightarrow \{+1, -1\}$ ?

- (a)  $2^{5 \cdot 3 \cdot 2}$ ,  $5 \cdot 3 \cdot 2$       (b)  $5 * 3 * 2$ ,  $2^{5 \cdot 3 \cdot 2}$       (c)  $2^{5+3+2}$ ,  $5 + 3 + 2$       (d)  $5 + 3 + 2$ ,  $2^{5+3+2}$

**Question 2. (1 point)** Consider a set of training examples  $S$ , a hypothesis space  $H$ , and a version space  $VS_{H,S}$ . Which of the following is **always** true?

- (a)  $|H| = |VS_{H,S}|$       (b)  $|H| \neq |VS_{H,S}|$       (c)  $|H| \leq |VS_{H,S}|$       (d)  $|H| \geq |VS_{H,S}|$

**Question 3. (1 point)** Suppose we have a binary classification problem where instances have two integer-valued attributes. We want to apply **unweighted kNN** (with Euclidean distance as the similarity metric) to classify new points. Given the following training data and  $k = 3$ , what would be the output labels for test points  $(0, 0)$  and  $(10, 10)$ ?

$x_1$	$x_2$	Label
1	6	-1
2	9	-1
7	4	+1

- (a) +1, +1      (b) -1, -1      (c) +1, -1      (d) -1, +1

**Question 4. (1 point)** Decision tree learning methods are generally well-suited to classification problems with which of the following properties?

- (a) Training data may contain errors (noise)  
 (b) Instances are represented as sets of attribute-value pairs  
 (c) Target function has discrete output values  
 (d) All of the above