CS381/780 Data Analytic Review Quiz 7

Instruction: For multiple choice questions, clearly circle one of the choice; for all other questions, write your answer right below the questions. All questions carry the same weights.

Name: James Mancuso

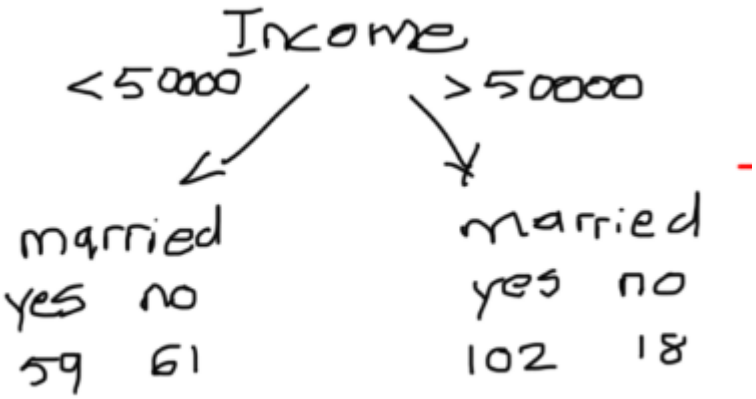
Question 1: Which of the following are true regarding decision trees algorithm?

False (opposite)

1. When we go from low to high entropy, we will have positive information gain.
2. We want to pick an attribute that has highest gain in entropy. False (want lowest)
3. Random forest may underperform decision tree because of its randomness nature

A. Only 1
B. Only 2
C. Only 3
D. 2 and 3 are true
E. None of the above

Question 2: Based on the following information, calculate the information gain



Answers:

left impurity = 1 - (59 / (59 + 61))^2 - (61 / (59 + 61))^2 = 0.499861

right impurity = 1 - (102 / (102 + 18))^2 - (18 / (102 + 18))^2 = 0.255

impurity = 1 - (120 / 240)^2 - (120/240)^2 = 0.5

information gain = 0.5 - (120 / 240) (0.499861) - (120/240) (0.255) = 0.12569