CS381 Data Analytic Regression Practice Quizz

Question 1: Which of the following are true?

```
    1. Entropy is highest when the dataset is dominated by one type of data.
    2. Gini index is almost zero when probability for one class is 99%.
    3. Information gain is calculated by substracting the weighted sum of the
entropy after the split from the entropy before the split.
```

  A. Only 1
  B. Only 2
  C. Only 3
  D. 1 and 2 are true
  E. 2 and 3 are true

Answer: E (1 is wrong, entropy is lowest when there is only one class)
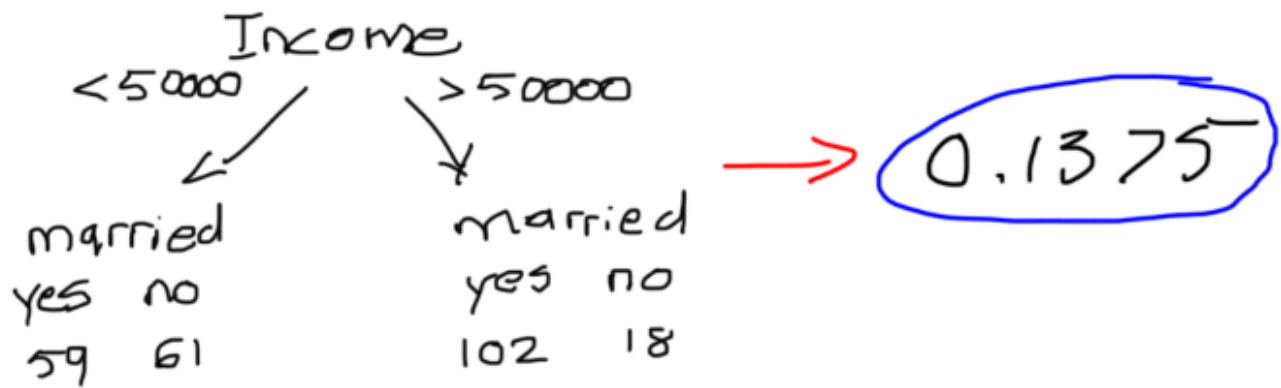
Question 2: Which of the following are true?

```
    1. Random Forest is an example of ensemble methods
    2. Decision Tree can use both entropy and gini index as the criterion
    3. Decision Tree is an example of divide-and-conquer method of solving
problems.
```

  A. Only 1
  B. Only 2
  C. 1 and 2 are true
  D. 2 and 3 are true
  E. All of them are true

Answer: E

Question 3: Based on the following information, calculate the information gain and verify that it is 0.1375

Answers:

left impurity = 1 - (59 / (59 + 61))^2 - (61 / (59 + 61))^2 = 0.49986111111

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) (0.49986111111) - (120 / 240) (0.255) = 0.12256944444