CS381/780 Machine Learning Review Exercise 5

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: Jonathan Yulan - Cogollo

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

```
   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.
   3. Random forest may underperform decision tree because of its randomness
nature
```

A. Only 1

Information Gain = 1 - Entropy thus is entropy goes from 0 to 1 the Information Gain will decrease

B. Only 2

Lesser the entropy higher the information gain, which will lead to more homogeneous or pure nodes.

C. Only 3

Higher Entropy means purer nodes, you want to pick the split attribute with the most information gain not Entropy thus '2' here is not true

D. 2 and 3 are true

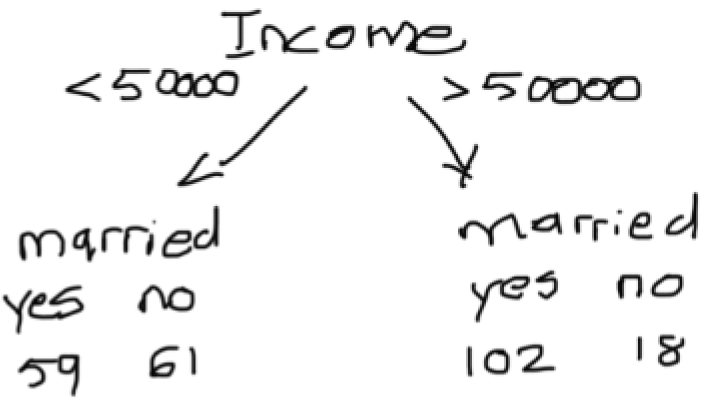
E. None of the above

Aggregated/ensemble models are not universally better than their "single" counterparts, they are better if and only if the single models suffer of instability. Particularly for highly nonlinear partitioning models (such as the decision trees), leaving training space will typically rather sooner than later lead to disaster.

Answers: E

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

INFORMATION GAIN = 0.0642013899



Answers:

1st Step: Obtain Left Impurity

----------> Left Impurity = 1 - (59/59+61)^2 - (61/59+61)^2 = 1 - 0.2417361111 - 0.258402777777=.499861111123

2nd Step: Obtain Right Impurity

----------> Right Impurity = 1 - (102/120)^2 - (18/120)^2 = 1 - 0.7225 - 0.0225 = 0.255

3rd Step: Obtain Gini Impurity

----------> Impurity = 1 - (59+102/59+61+102+18)^2 - (61+18/240)^2 = 1 - (161/240) - (79/240) = 1 - .45001736 - .10835069444 = 0.44163194556

4th Step: Obtain Information Gain

----------> Information Gain = Step 3 - (LeftImpurity * 120/240) - (RightImpurity * 120/240) = 0. - (.499861111123*0.25) - (.0225*0.25) = 0.44163194556 - (.499861111123 * .5) - (0.255 * .5) = 0.1275

Question 3: Which of the following is/are true regarding Ensemble Learning methods?

```
1. Performance of Hot Voting will often outperform Soft Voting as it is based on
the majority.
2. Random Forest Classifier is an example of Bagging Classifier
3. One of reasons why Gradient Boosting is popular is because it can be run in
parallel making it very efficient.
```

A. Only 1

B. Only 2

C. 1 and 2

D. 2 and 3

E. All of the above

Answers:

B;
A is not true because shallow trees(high bias, low variance, less depth) and deep trees both have their place and thus depending on what we are classifying; difference types of voting may be beneficial
C is not true; Boosting can not be done in parallel, but bagging can be making it an advantage of Random Forest