

CS381 Data Analytic Quiz 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:

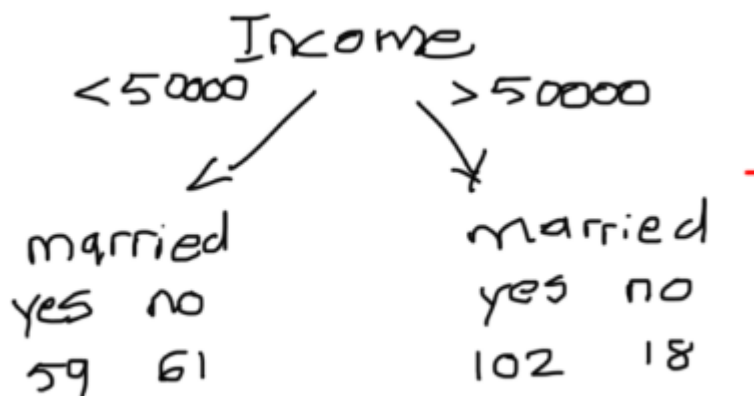
Question 1: Which of the following regarding Support Vector Machine (SVM) and Logistic Regression are true?

1. SVM finds its optimal model parameters by maximizing the Likelihood functions (i.e. the probability of the training data)
2. Logistic Regression finds its optimal model parameters by using the maximizing the margin between the predictions and the observed data.
3. The difference between Soft margin and Hard margin is that soft margin allows some mis-classification some data points.

- A. Only 1
- B. Only 2
- C. Only 3
- D. 1 and 2
- E. 2 and 3

Answers: C

Question 2: Based on the following information, calculate the information gain and verify that it is 0.1225



Answers:

$$\text{left impurity} = 1 - (59 / (59 + 61))^2 - (61 / (59 + 61))^2 = 0.49986111111$$

$$\text{right impurity} = 1 - (102 / (102 + 18))^2 - (18 / (102 + 18))^2 = 0.255$$

$$\text{impurity} = 1 - (120 / 240)^2 - (120 / 240)^2 = 0.5$$

$$\text{information gain} = 0.5 - (120 / 240) (0.49986111111) - (120 / 240) (0.255) = 0.12256944444$$