

CS381 Data Analytic Regression Practice Quizz

Question 1: Which of the following 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
- B. Only 2
- C. Only 3
- D. 2 and 3 are true
- E. None of the above

Answer: E (1 is wrong because we have information gain when we go from high to low entropy, 2 is wrong because we want highest gain in information, not entropy, Random forest should overperform than decision)

Question 2: Which of the following are true regarding clustering?

1. Sum of square of the distance from the data point to its centroid increases when K increases
2. Elbow graph is just a heuristic for determining the value of K, i.e. there is no absolute answer for K.
3. Clustering is considered as supervised learning because K has to be an external input (thus supervised by the user)

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

Answer: B

Question 3: Write down the pseudo code for K-means clustering algorithm

Answer: Check the class presentation