

AIML FEATURE ENGINEERING
ASSIGNMENT 1

Submission Date: 22 December 2019 11.59 PM

Weightage: 8%

Q1. What is the purpose of aggregation when preprocessing raw data to obtain a dataset? [1%]

Q2. Why is it, sometimes, desirable to aggregate sets of attributes/ objects into a single attribute/object? [1%]

Q3. Suppose all students in the first cohort are friends on Facebook and the instructor wants to reach out to all students about a notice. To complete this task, the instructor sends the message to the most influential person (connected to many students of first cohort) so that he/she can propagate the message to all students. In this application what type of data would you use and how will you create the dataset? [2%]

Q4. Classify the following three attributes as binary, discrete or continuous. Further classify the attributes as qualitative (nominal or ordinal) or quantitative (interval or ratio)

- (a). decay in a radioactive element _____, _____
- (b). pair-wise distances between cities _____, _____ [2%]

Q5. You are given a dataset for classifying songs into classical and non-classical. The size of the data is given in Table-1. Out of 5000 samples there are 100 samples of classical songs (class label 0) and remaining are nonclassical (class label 1). You are asked to sample training data for building a classification model. What sort of sampling technique you will use and justify how your approach will help in reducing the training error? [2%]

Song Id	F1	F2	...	F10	Class Label
1					0
2					1
..					
..					
5000					1

Table -1.

Submission Details

1. Final document - IDS_Assignment1_<Student_ID>.doc