PGP in AI/ML

Feature Engineering -Assignment 2

Submission Date: 23 59hrs on 29-08-2019

Total Marks: 17

The following questions are to enhance your understanding of basic concepts and definitions. You are expected to answer these questions on your own without referring to internet.

Q1. Pokémon video gaming company wishes to analyze their data and has captured the data of each game played by the customer. One of the attribute is number of times the customer has played the game. We have 16 examples given as follows: { 22, 12, 61, 57, 30, 1, 32, 37, 37, 68, 42, 11, 25, 7, 8, 16 }.

Apply equal frequency and equi-width binning using number of bins=4 and explain the difference between the two methods. **[2+2+1=5M]**

Q2. Given the data set of 4 objects/instances/examples having described using 3 attributes/features (F1, F2 and F3). Answer question a and b.

Note: Assume that F1 is symmetric binary, F2 is Categorical and F3 is Ratio type of feature.

	Gender (F1)	**Programming Language Preference (F2)**	**Time Taken for completing coding assignment in Mins (F3)**
Student-1	Female	C++	15
Student-2	Male	C++	10
Student-3	Male	R	23
Student-4	Female	Python	16

a) Compute a 4X4 similarity matrix. **[4+4=8M]**

b) Given that feature F2 has a weight 0.6 while F1 and F3 have 0.2 each weight, compute a 4X4 similarity matrix.

Q3. For each of the following scenarios given below identify the appropriate filters for feature selection:

a) Feature and the class labels are Continuous values. **[4 M]**

b) Feature is a Continuous value while the class label is a discrete value.

c) Feature is a discrete value while the class label is a Continuous value.

a) Feature and the class labels are discrete values.