PGP in AI/ML

Feature Engineering -Assignment 1

Submission Date: 2359hrs on 25-07-2019

Total Marks: 24

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.

a) For each of the following examples given below, indicate the type of data and the scale of measurement. **[6 M]**

	Continuous or Discrete	Scale of Measurement
Amount of sugar in a Coke bottle		
Birth place of BITS university students		
The size of Easter eggs represented as Small, Medium, Large, Extra Large, Jumbo		
Pan card number		
Dates		
Number of defective shirts in packs of 100 manufactured by color plus		

b) Raju measures the pressure of all tires coming into his garage and record the values. Unknown to him, his tire gauge is miscalibrated and adds 3 psi to each reading. Using the definition of noise used in the textbook, is this error introduced by the tire gauge considered noise? Answer "yes" or "no" and justify your answer in one line. **[2 M]**

c) Consider the dataset collected from 5 students of a class as given below:

Student ID	Height(cm)	Weight(kg)
S1	169	60
S2	171	70
S3	150	80
S4	180	75
S5	150	60

Suggest an appropriate data transformation method to scale both the variables in the range of [0,1] and using the suggested method what will be the weight of s4 in the new range. **[3 M]**

d)In this programming assignment, we apply basic feature engineering techniques to the regression dataset.

To-do list

1. Import the necessary libraries. **[1M]**
2. Import the dataset (final_dataset.csv) **[1M]**
3. Remove columns with missing values greater than 40% of dataset size. **[2M]**
4. Extract numeric data from columns property_price, square_area & carpet_area using regular expressions. Make sure that units are standardized (all the values in property_price should be in lakhs) **[2M]**

5. In the above columns, fill empty values with mean of remaining data points. **[1M]**
6. Convert columns facing, overlooking, ownership, transaction, furnishing into categorical variables by assigning values [0,1,2…]. **[2M]**
7. In the above columns, fill empty values with 0. **[1M]**
8. Convert all values in the floor column to decimals. For example, convert "10 out of 14 floors" as 10/14 = 0.71. **[2M]**
9. Fill empty values in floor and balcony with 0. **[1M]**

Submission Details

1. Final dataset - id_final_dataset.csv

Contact Details

You should put up queries in the discussion forum of the corresponding assignment folder only.

1. Himanshu - himanshuchawla437@gmail.com
2. Mohit - mohitrathore15@gmail.com
3. Kartheek - f20160015@hyderabad.bits-pilani.ac.in
4. Shristy - f20160688@hyderabad.bits-pilani.ac.in
5. **d)** Agam - f20160484@hyderabad.bits-pilani.ac.in

********** *That's all folks* **********