K. K. Wagh Institute of Engineering Education and Research, Nashik.

Department of Computer Engineering

Academic Year: 2020 – 2021 Semester: I
Class & Div: BE – A and B
Course Name & Code: Laboratory Practice - I (410246)
Teaching Scheme: Practical (04 Hrs / week)
Name of Faculty: Prof. J. R. Mankar, Prof. A. V. Taware and Prof. N. G. Sharma

ASSIGNMENT 02 (DATA ANALYTICS) SAMPLE ORAL QUESTIONS

1. What is the type of dataset that is used as input to the program?
   Pima Indians Diabetes Dataset

2. How many attributes / features are there in the dataset?
   9 attributes:
   - TimesPregnant
   - GlucoseConcentration
   - BloodPrs
   - SkinThickness
   - Serum
   - BMI
   - DiabetesFunct
   - HasDiabetes

3. What is the type of each attribute?

```
0   NumTimesPrg     768 non-null    int64
1   PlGlcConc       768 non-null    int64
2   BloodP          768 non-null    int64
3   SkinThick       768 non-null    int64
4   TwoHourSerIns   768 non-null    int64
5   BMI             768 non-null    float64
6   DiPedFunc       768 non-null    float64
7   Age             768 non-null    int64
8   HasDiabetes     768 non-null    int64
```

4. How many features is numeric?
   9

5. How many features are nominal / categorical?
   0

6. Which function is used to display summary statistics in python?
   - df.describe()

7. How do you define bayes theorem?
   a theorem about conditional probabilities: the probability that an event A occurs given that another event B has already occurred is equal to the probability that the event B occurs

given that A has already occurred multiplied by the probability of occurrence of event A and divided by the probability of occurrence of event B

8. What is naive bayes classifier?
   Naive Bayes classifiers are **a collection of classification algorithms based on Bayes' Theorem**. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

9. Why do we use naive bayes classifier?

   - It is easy and fast to predict class of test data set. It also perform well in multi class prediction
   - When assumption of independence holds, a Naive Bayes classifier performs better compare to other models like logistic regression and you need less training data.
   - It perform well in case of categorical input variables compared to numerical variable(s). For numerical variable, normal distribution is assumed (bell curve, which is a strong assumption).

10. What are the applications of naive bayes classifier?
    Real time Prediction
    Multi class Prediction
    Text classification/ Spam Filtering/ Sentiment Analysis
    Recommendation System

11. What do you mean by conditional probability?
    Conditional probability refers to **the chances that some outcome occurs given that another event has also occurred**. It is often stated as the probability of B given A and is written as P(B|A), where the probability of B depends on that of A happening.

12. What do you mean by unconditional probability?
    An unconditional probability is **the chance that a single outcome results among several possible outcomes**. The term refers to the likelihood that an event will take place irrespective of whether any other events have taken place or any other conditions are present.

13. What is confusion matrix?
    A confusion matrix is a technique for summarizing the performance of a classification algorithm.Classification accuracy alone can be misleading if you have an unequal number of observations in each class or if you have more than two classes in your dataset.Calculating a confusion matrix can give you a better idea of what your classification model is getting right and what types of errors it is making.

14. How to calculate accuracy?
    TP+TN/(TP+TN+FP+FN)

15. How do we evaluate precision?
    TP/(TP+FP)
16. How do we evaluate recall?
    TP/TP+FN

17. How do we evaluate score?
    2*Precision*Recall/Precision+Recall

18. How do we define split ratio?
    rom sklearn.model_selection import train_test_split
    train_set,test_set=train_test_split(df,test_size=0.20,random_state=0)

19. What is the purpose of scaling?
    To scale data into specific range

20. What is difference between fit(), transform() and fit_transform()?
    Fit-calculate mean &variance
    Transform-x-mean/variance

n

Prof. J. R. Mankar, Prof. A. V. Taware and Prof. N. G. Sharma
Course Teacher