Department of Computer Science

QUESTION BANK
(Academic Year 2018 - 19)

Course: **B.Tech** Branch: **CSE**
Year/Sem: **III-II**
Subject: **Data Warehousing and Data Mining (DWDM)** Regulation: **R-16**

UNIT – I

1) (a) Explain the steps in data mining process with a neat diagram. ---5M
(b) Explain the architecture of data mining with the help of a neat diagram. ---5M
2) On what kind of data the mining can be performed. Explain in detail ---10M
3) (a)Enumerate the applications of data mining. ---3M
(b) What are the challenges in data mining that motivate the mining tasks. ---7M
4) Define Data mining. Explain in detail about the major issues that are present in the data mining? – 10M
5) (a) What are the origins of data mining? Explain it in detail. ---5M
(b) How can we perform integration of a data mining system with a database or data warehouse system? ---5M
6) Discuss in detail about different kinds of patterns that can be mined. ---10 M

UNIT-II

1) (a) Explain different forms of preprocessing–5M
(b) Explain Data Discretization with example - 5M
2) (a) What are the methods for data normalization . Explain in detail. --- 6M
(b) What is noise? What are different techniques to smoothing the data? –4M
3) Where we can apply the data reduction technique. Explain the strategies for data reduction. –10M
4) (a) what is the use of attribute subset selection. Describe the greedy methods for attribute subset selection. –5M
(b) Discuss about the methods involved in filling the missing values. –5M
5) What is the use of Correlation analysis? What is the value for a correlation between two attribute. Explain with an example. –10M
6) (a) What are the different steps that are involved in data transformation?--7M
(b) Explain various methods for the generation of concept hierarchies for categorical data. –3M

UNIT-III

1) (a) Define the term entropy, information gain and gini index. How they are useful for attribute selection?—7M
(b) Write the algorithm for decision tree induction. ---3M
2) (a) Mention different characteristics to construct decision tree .—5M
(b) Explain how cross validation is useful in classifiers of data mining..—5M
3) What is a classifier? Why we need classifier? What are the measures available for evaluating the performance of classifier? For an example data (set), briefly describe the usage of cross validation.—10M
4) (a) What is meant by model over fitting? How can over fitting done due to presence of noise?—6M
(b) How splitting is done in continuous attributes? –4M
5) Explain in detail about the decision tree induction. –10M
6) What are split points? How to find out correct split points. Explain it with an example—10M

UNIT-IV

1) (a) Why naive Bayesian classification is called naïve? Briefly outline the major ideas of naive Bayesian classification.-5M
(b) Explain Bayes Theorem.
2) Explain Naive Bayesian Classification with Example. - 10M.
3) (a) Why is Bayesians Belief Networks and Its importance- 5M
(b) Explain Bayesians Belief Networks with example - 5M.

UNIT-V

1) (a) What are different types of association rules? Give examples. –4M
(b) Explain FP growth algorithm for generation of frequent itemsets—6M
2) (a) What is Apriori property? Explain Apriori algorithm with an example. –5M
(b) What is frequent itemset generation? What are candidate itemsets?—5M
3) What is pruning? Why pruning is require ?With an example briefly describe FP- growth algorithm. –10M
4) (a) Define support confidence. What is support threshold? —5M
(b) Writ the procedure of closed frequent itemsets. –5M
5) What is support counting? Why one need support counting? With an example , briefly describe compact representation of frequent itemsets. –10M
6) (a) Explain any association mining algorithm without generating candidate itemsets. –6M
(b) Generate rules after finding frequent candidate itemsets. –4M

UNIT-VI

1) (a) Explain a basic K-means algorithm with an example. –6M
(b)What are applications of clustering? –4M
2) What is DBSCAN? What is the time and space complexity of DBSCAN algorithm? Briefly describe the process for selecting DBSCAN parameters.—10M
3) (a) What are additional issues of K-means algorithm? –6M
(b) Describe about the strength and weaknesses of traditional density approach. –4M
4) What is a cluster? For a sample data and using appropriate dissimilarity measures, briefly describe the creation of clusters using K-means cluster algorithm. –10M
5) (a) Explain about the basic agglomerative clustering algorithm. –6M
(b) What are strengths weakness of DBSCAN? –4M
6) What is bisecting K-means? In what way it is different from the basic K-means algorithm? With a neat diagram, briefly describe the usage of bisecting K-means algorithm and its initialization. –10M