**1) Name three main approaches for outlier detection methods of high dimensional data ?**

1- Extending Converntional Outlier Detection

2- Finding Outliers in Subspaces

3- Modeling High-Dimensional Outliers

**2) Name three categories of constraints for clustering?**

1- Constraints on instances  
 2- Constraints on clusters  
 3- Constraints on similarity measurements

**3) Compare the differences between SCAN algorithm and DBSCAN. What are their similarities and differences?**

DBSCAN :a density-based clustering algorithm that produces a partitional clustering in which the number of clusters is automatically determined by the algorithm.

SCAN:A scan is a collection of n block-subsamples of the sequence X1,... ,Xn with the following two properties

**4) Defrenciate contexual and global outlier** **?**

global outlier :data object is a global outlier if it deviates significantly from the rest of the data set. Global outliers are sometimes called point anomalies, and are the simplest

type of outliers.

Contextual outlier : a data object is a contextual outlier if it deviates significantly with respect to a specific context of the object. Contextual outliers are also known as conditional outliers

**5)What is the proximity outlier detection method? give example.**

The proximity of an outlier deviates significantly from that of most of the others in the data set

Example**:** Distance-based outlier detection

**6) What do we mean by regression tree**?

predict the value of a response (dependent) variable from one or more predictor (independent) variables where the variables are numeric

forms of regression: linear, multiple, weighted, polynomial, nonparametric, and robust

MCQs:

1. Applications for outlier are:

a. Credit card fraud detection.

b. Telecom fraud detection.

c. Medical analysis.

**d. All**

2. Which of these statistics is unaffected by outlayers:

a. Mean.

b. Standard deviation

c. Range

d. **inter quartile range**

3. What data mining can help:

a. New data mining algorithm for intrusion detection.

b. Analysis of stream data

c. Disrupted data mining

**d. All**

T/F:

1. Detection of collecting a outlier consider not only the behavior of individual objects but also that of groups of objects.

**TRU**E

2. In global outlier object is Og if it signification only not deviated from the rest set.

**FALSE**

Fill blanks:

1. **contextual** attributes defines the context for example time and location.

2. **Multi variate** data is a data set involving two or more attributes of variables.