**12.1 Ans:**

When a data object is deviating from the other objects and the deviation of that data object is significant comparing to the remaining objects is called outlier.

For instance, let’s consider the average temperature of Atlanta for each day and will do it for over a period of 30days. The average temperatures in Fahrenheit for 30 days are shown below.

50, 51, 52, 54, 56,…….., 56, 78, 80, 90

Here, the above example satisfies all types of outliers.

Global Outlier: In the above example, the average temperature is 78F on 28th day, which is deviates from all the data i.e. from day 1 to day 27.

Contextual Outlier: It is also called as conditional outlier. In the above example the temperature for 28th day is 78F, from which we can assume that the season has changed to the summer. Hence we could see that huge change in the temperature.

Contextual attributes: Here April 28th is the date and location is Atlanta are called Contextual attributes

Behavioural attributes: here, the temperature is a behavioural attribute.

Collective Outlier: Here in this case, the average temperatures from 28th day onwards changed substantially. It means that 28, 29 30 days of temperatures which is a subset of entire data is deviating suddenly to the remaining data objects.

The relationship among the data objects in Collective outlier can be seen as two different data sets generated from the 30 days of data. One set of data generated for day 1 to day 27 and the second set of data is generated for 28th day onwards till the end.

**12.2 Ans:**

Intrusion is the one example where it is not possible to separate the regular data and outlier data. It is very hard to distinguish them by using border or separation. The intrusions are usually unpredictable, unavoidable, and dangerous enough to create problems within a system. The border between the regular data and outlier data often lies on grey area. If we consider examples of clinic data and market analysis data, it is pretty even overall and occasionally changes or large fluctuations occur over the period.

Noise detection is also one of the biggest challenges. Noise is not only distinguishing the regular data with the outlier data and create problem with in the data but also falsify the regular data. So the machine should understand which one should be treated as outlier and which one should be considered as regular data. It just needs certain level of estimation criteria in order to separate the regular data with the outlier data.

**12.3 Ans:**

Spam detection is one scenario where most of the applications are trying to get rid of. Suppose in case of any mail application where the users get mails from different users or various servers. It is hard to find out the spam in them. So many applications use machine learning models to get away from them. Here in semi supervised model only few regular samples will be labelled or only few set of samples belongs to the outliers are labelled.

Case 1: Only some labelled samples of normal objects:

Here in this case, apply a machine learning model to train the labelled normal objects and try to predict the unlabelled normal objects. The samples fitted with the model are identified as regular samples and the samples not fitted are considered as outliers. Same in the case of the spam detection, the regular mails fitted with the model created by normal mails are considered as regular mails and those mails which are not classified as regular mails identified as outliers.

Case 2: Only some labelled samples of outliers:

Similar to the above situation, create a model and train with the labelled samples of outliers. Apply the model for our spam detection criteria and try to predict the outlier samples. In case of regular samples apply unsupervised methods to identify the regular mails.

As mentioned in the above the two scenarios we can separate the spam mails from regular mails.