Questions

1. **Explain the following concepts:**

1) supervised learning,

2) unsupervised learning,

3) online learning,

4) batch learning,

5) model-based learning,

6) instance-based learning.

Answers:

1) **Supervised Learning:** Supervised Learning in Machine Leaning refers to the algorithm in which example set of labelled data and corresponding output is provided to train the system in recognizing output from example dataset. Learning is complete when the algorithm, after observing examples, is able to reach a sufficient accuracy in recognizing acceptable output. Supervised Learning can be classified into: **a) Classification** and **b) Regression**
   a. **Classification:** Classification problem deals with segregation of data into different labelled buckets. For example: Seperation of sample colors into : 'Blue' and 'Red'.
   b. **Regression:** A regression problem predicts a continuous outcome variable based on dependent parameters. So, regression helps predict y by defining it as a function of x.

2) **Unsupervised Learning:** Unsupervised Learning refers to the algorithm in which only input data is provided and the algorithm is aimed at discovering inherent structure within the input data in order to learn more about the data. Unlike Supervised Learning, there is no 'correct' solution. Unsupervised learning can be classified into: **a) Clustering** and **b) Association**
   a. **Clustering:** A clustering problem is where the algorithm is aimed at finding groups with similar features within the data. For example: Grouping customers by their age groups or buying behaviour.
   b. **Association:** An association problem is where the algorithm is aimed at trying to find relations that associate large portions of the data. For example: People who buy gloves, also buy socks.

3) **Online Learning:** Online learning is a type of machine learning algorithm which is trained incrementally over data that is available dynamically, in sequential order. This method is used for continuously updating predictions for the next set of data that is going to be available. This kind of algorithm is therefore used where data is dynamically generated and volatile like stock market, or in areas where data available is so massive that it is computationally impractical to train over the entire dataset

4) **Batch Learning:** In Batch Learning, the system must be trained using all the available data. Thus, it is not capable of learning incrementally, and for making the algorithm learn a new type of data it, needs to be retrained over the entire dataset. Since this requires a huge amount of computational power, it is mostly done offline and then launched into production, after which it doesn't learn anymore. Thus batch learning is also known as *offline learning*.

5) **Model Based Learning:** Model Based Learning is an algorithm that takes an environment and forms models. All assumptions about the domain are made in the form of a model explicitly. This model is then used for creating a model-specific to learn about the various aspects of the entire domain.

6) **Instance Based Learning:** Instance Based Learning is a technique that is used to produce a class label or prediction based on similarities of new instances with previously observed training instances. Instance based Learning therefore includes a family of algorithms that do not generalize the data to predict outcome, but rather compares instances of new data with those present in memory.