SNS COLLEGE OF TECHNOLOGY
(An Autonomous Institution)
COIMBATORE-35
DEPARTMENT OF COMPUTER APPLICATIONS

19CAT703 Machine Learning
UNIT I

1) What is Machine learning?

Machine learning is a branch of computer science which deals with system programming in order to automatically learn and improve with experience. For example: Robots are programmed so that they can perform the task based on data they gather from sensors. It automatically learns programs from data.

2) Mention the difference between Data Mining and Machine learning?

Machine learning relates with the study, design and development of the algorithms that give computers the capability to learn without being explicitly programmed. While, data mining can be defined as the process in which the unstructured data tries to extract knowledge or unknown interesting patterns. During this process machine, learning algorithms are used.

3) What is 'Overfitting' in Machine learning?

In machine learning, when a statistical model describes random error or noise instead of underlying relationship 'overfitting' occurs. When a model is excessively complex, overfitting is normally observed, because of having too many parameters with respect to the number of training data types. The model exhibits poor performance which has been overfit.

4) Why overfitting happens?

The possibility of overfitting exists as the criteria used for training the model is not the same as the criteria used to judge the efficacy of a model.

5) How can you avoid overfitting ?

By using a lot of data overfitting can be avoided, overfitting happens relatively as you have a small dataset, and you try to learn from it. But if you have a small database and you are forced to come with a model based on that. In such situation, you can use a technique known as **cross validation**. In this method the dataset splits into two section, testing and training datasets, the testing dataset will only test the model while, in training dataset, the datapoints will come up with the model.

6) What is inductive machine learning?

The inductive machine learning involves the process of learning by examples, where a system, from a set of observed instances tries to induce a general rule.

7) What are the five popular algorithms of Machine Learning?

- Decision Trees
- Neural Networks (back propagation)
- Probabilistic networks
- Nearest Neighbor
- Support vector machines

8) What are the different Algorithm techniques in Machine Learning?

The different types of techniques in Machine Learning are

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning
- Transduction
- Learning to Learn

9) What are the three stages to build the hypotheses or model in machine learning?

- Model building
- Model testing
- Applying the model

10) What is the standard approach to supervised learning?

The standard approach to supervised learning is to split the set of example into the training set and the test.

11) What is 'Training set' and 'Test set'?

In various areas of information science like machine learning, a set of data is used to discover the potentially predictive relationship known as 'Training Set'. Training set is an examples given to the learner, while Test set is used to test the accuracy of the hypotheses generated by the learner, and it is the set of example held back from the learner. Training set are distinct from Test set.

12) List down various approaches for machine learning?

The different approaches in Machine Learning are

- Concept Vs Classification Learning
- Symbolic Vs Statistical Learning
- Inductive Vs Analytical Learning

13) What is not Machine Learning?

- Artificial Intelligence
- Rule based inference

14) Explain what is the function of 'Unsupervised Learning'?

- Find clusters of the data
- Find low-dimensional representations of the data
- Find interesting directions in data
- Interesting coordinates and correlations
- Find novel observations/ database cleaning

15) Explain what is the function of 'Supervised Learning'?

- Classifications
- Speech recognition
- Regression
- Predict time series
- Annotate strings

16) What is algorithm independent machine learning?

Machine learning in where mathematical foundations is independent of any particular classifier or learning algorithm is referred as algorithm independent machine learning?

17) What is the difference between artificial learning and machine learning?

Designing and developing algorithms according to the behaviours based on empirical data are known as Machine Learning. While artificial intelligence in addition to machine learning, it also covers other aspects like knowledge representation, natural language processing, planning, robotics etc.

18) What is classifier in machine learning?

A classifier in a Machine Learning is a system that inputs a vector of discrete or continuous feature values and outputs a single discrete value, the class.

19) What are the advantages of Naive Bayes?

In Naïve Bayes classifier will converge quicker than discriminative models like logistic regression, so you need less training data. The main advantage is that it can't learn interactions between features.

20) In what areas Pattern Recognition is used?

Pattern Recognition can be used in

- Computer Vision
- Speech Recognition
- Data Mining
- Statistics
- Informal Retrieval
- Bio-Informatics

21) What is bias?

While making predictions, a difference occurs between prediction values made by the model and actual values/expected values, and this difference is known as bias errors or Errors due to bias.

22)Write the types of bias.

- **Low Bias:** A low bias model will make fewer assumptions about the form of the target function.
- **High Bias:** A model with a high bias makes more assumptions, and the model becomes unable to capture the important features of our dataset. **A high bias model also cannot perform well on new data.**

23) What do you meant by variance? Define the types.

Variance tells that how much a random variable is different from its expected value.

Low variance means there is a small variation in the prediction of the target function with changes in the training data set. At the same time, **High variance** shows a large variation in the prediction of the target function with changes in the training dataset.