

# **e-PG Pathshala**

**Subject : Computer Science**

**Paper: Machine Learning**

**Module: Internals of Machine Learning**

**Module No: CS/ML/2**

**Quadrant 1- e-text**

Welcome to the e-PG Pathshala Lecture Series on Machine Learning.

## **Learning Objectives**

The learning objectives of this module are given below:

- To understand how Learning is carried out through an example
- To appreciate the different types of Learning output
- To know about various types of data representations that are used for machine learning
- To understand the types of learning, types of learning feedback and types of learning experience

This module explains the internals of machine learning without going into the details. This module should help you understand the types of input and output that are commonly used in the context of machine learning. Before we proceed, let us discuss what it means to learn.

### **2.1 What does it mean to Learn?**

When we talk about learning, the basic ingredients include an input data set, a model that is learned from the data and a method to know whether learning has indeed taken place. Generally, the input data set is divided into two sets, training and tests data sets, where the training data set is used to learn a hypothesis  $h$ , whereas the test set is used to evaluate the learned hypothesis. The actual model that is learned depends on the type of learning algorithm which we will discuss later. However, once learning is completed, there is a need to measure the performance of the learning. The actual evaluation parameters to be used depends on the expected output or goal. Another issue which is important is the conflicting goals of machine learning - generalization that enables learning given unknown new data and over fitting which allows the model to fit the input data but which does not necessarily work for unknown data.

#### **Example A – Document Classification**

Let us consider the example of document classification from the above perspective. As you know document classification is essentially grouping a collection of articles by topic. Given below (Figure 2.1(a)) are two articles along with their topics.



Figure 2.1(a) Document Classification

The characteristics of this very popular and important problem are the classification perspective that is assigning a class label (topic) to each document and the supervised learning approach where the system is provided with sufficient labelled articles.

### Example B – Similar/Duplicate Image Classification

Another example we will consider is the classification of images as similar or duplicate (Figure 2.1(b)). Here the features considered can include width, height, contrast, position, etc.

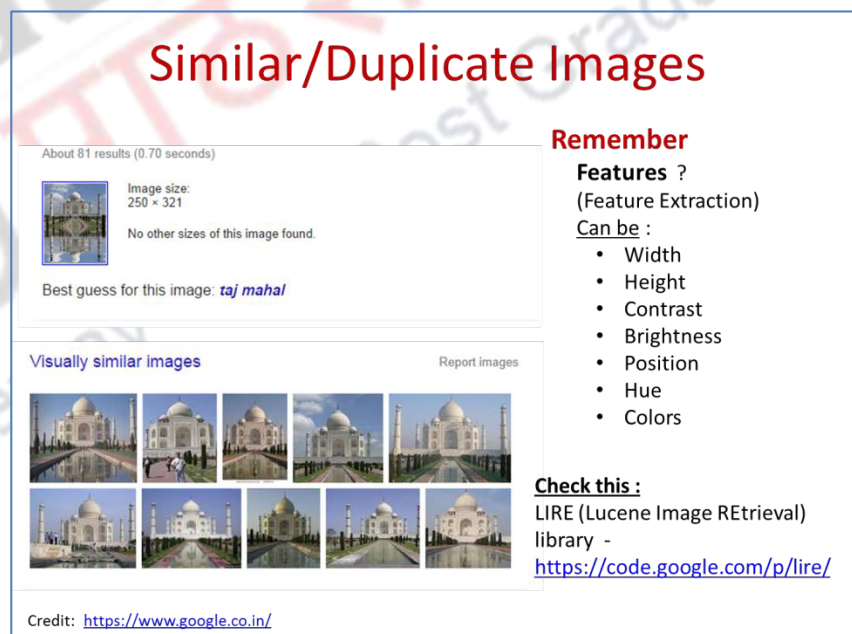


Figure 2.1(b) Image Classification

## 2.2 Internals of Machine Learning

In this module we will discuss the various components of machine learning – types of output, types of features or input, different representation of data, types of feedback used when

learning and finally some of the issues associated with machine learning. Now let us consider the components one by one.

## 2.2.1 Types of Output

It is very important to note that an application can be specified using many types of tasks or outputs. The variation comes in the type of data available, type of features extracted from the data, type of output needed and the type of algorithm used to obtain the learning model. The machine learning systems can learn a mapping function from input to output in which case the input are features while the output is a value given by the function. The next most common problem tackled by machine learning is classification where input is again a set of features and the output is a single decision – a label. Another category of learning is sequence labeling where a sequence of features given as input is used to learn a sequence of decisions or labels. In addition there are a number of machine learning tasks including clustering, ranking, problem solving, matching, tagging, different types of prediction and finally evolution.

### 2.2.1.1 Value Function – Regression

A typical example of learning a mapping function from input to output is **Regression** where what is expected is real valued output.

*Example*

- *Prediction of the prize of tomorrow's stock*

The performance of learning is measured in terms of loss – usually as squared loss. This loss indicates how far the predicted value using the learnt function is from the actual value.

### 2.2.1.2 Classification

Classification can be of many types such as Binary, Multi-class, Multi-label, Hierarchical classification. Let us see examples of each type:

- Binary Classification – classify as either belonging to a class or not belonging to the class
  - *Example: Classify email as Spam or not Spam.*
- Multi-Class Classification – classify as belonging to a class among many classes
  - *Example: Classifying articles as sports, business, politics, movies, etc.*
- Multi-Label Classification – Classify as belonging to more than one class
  - *Example: Classifying articles as belonging to three classes – sports, movies and business.*
- Hierarchical Classification- Classifying as belonging to a subclass of a class
  - *Example: Classifying IPL matches as belonging to subclass cricket under sports*

In all the above types of classification, performance of learning is measured in terms of accuracy, how many samples of the test set have been correctly labelled.

### 2.2.1.3 Clustering

Clustering is the grouping of samples by finding similarities between actual data based on some characteristics. It is a technique that places data elements into related groups without advance knowledge of the group definitions. Clustering can also be defined as the assignment of a set of observations into subsets (called *clusters*) so that observations in the same cluster are similar in some sense. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields. Here are a few examples of clustering.

### Examples

- *Marketing: Help marketers discover distinct groups in their customer bases, and then use this knowledge to develop targeted marketing programs*
- *Insurance: Identifying groups of motor insurance policy holders with a high average claim cost*
- *City-planning: Identifying groups of houses according to their house type, value, and geographical location*

In clustering, one way of measuring the performance is in terms of purity, that is the percentage of labelled samples of the majority class in each cluster.

#### 2.2.1.4 Ranking

Another very interesting problem which has gained importance currently is the ranking problem. This is a learning algorithm that receives labelled sample of pairwise preferences and learns a scoring functions that ranks the samples. While a training set in classification is a set of samples and their class labels, in ranking a training set is an ordering of data. The ranking problem has gained importance today.

### Examples

- *Search engines rank the results by their expected relevance to a user's query using a combination of query-dependent and query-independent factors.*
- *Academic journals are ranked according to impact factor; the number of articles that cite articles in a given journal.*
- *Another important example is recommendation systems that is used by commercial sites for ranking products, movies etc.*

In ranking, the performance of the learning system is measured in terms of minimum loss which in this case is the minimum number of swapped pairs.

#### 2.2.1.5 Problem solving

Problem solving is the use of generic or *ad hoc* methods, in an orderly manner, for finding solutions to problems. Many of the early systems were designed as problem solving systems such as the General Problem Solver to solve puzzles, the Geometry Theorem Prover and the Samuel's checkers player. Basically such systems were designed to solve mathematical problems.

In problem solving, performance is measured in terms of the correctness of results.

#### 2.2.1.6 Matching

Matching is a very important component of machine learning and is often a sub-process of other learning tasks. Matching essentially tries to determine the similarity between samples based on one or more specified qualities. It is a process in which patterns in data are found, recognized or discovered. There are basically two types of pattern matching:

- **Statistical Pattern Recognition** – in this type, the data is reduced to vectors of numbers and statistical techniques are used for pattern matching.
- **Structural Pattern Recognition** – in this type, the data is converted to a discrete structure such as a graph and graph matching techniques are used for pattern matching

Examples

- *Query document matching used in Web Search*
- *Object recognition in computer vision*

In pattern matching, performance is usually measured in terms of the accuracy of the results.

#### **2.2.1.7 Tagging/annotation**

Tagging or annotation refers to assigning a label or meta-data to a component for identification. Tagging is normally used in the areas such as natural language processing, image processing and speech processing.

Examples

- In the area of natural language processing, Part-of-Speech Tagging is the process of marking a word in a text with a particular part of speech based on both its definition, as well as its context of usage that is the context of neighboring words and their part-of-speech.
- Image annotation is the process by metadata is automatically assigned to the components of a digital image.

In tagging/annotation, performance is usually measured in terms of the accuracy of the results.

#### **2.2.1.8 Prediction**

Prediction can be defined as generalization from data. In case if the output we want to predict is numeric then regression analysis is the statistical technique most often used for numeric prediction. Predictive models exploit make use of patterns found in historical and transactional data to identify risks and opportunities. Prediction models capture relationships that exist among a variety of factors to assess of the risk or potential associated with a particular set of conditions. Prediction methods however tend to be highly dependent on the particular problem to be solved.

Examples

- Credit card fraud detection
- Predicting the length of stay in the hospital
- Predicting which locations in the brain are likely to be affected by a certain disease

Again in prediction, performance is usually measured in terms of the accuracy of the results.

#### **2.2.1.9 Structured Prediction**

Structured prediction involves prediction of structured objects such as trees, sequences, graphs, etc. rather than discrete or numerical values. Structured prediction uses observed data for training and adjusts model parameters during the learning process.

Examples

- For example, the problem of converting a natural language sentence into a syntactic tree can be viewed as a structured prediction problem of choosing the correct parse tree from the set of all possible parse trees.
- Predicting structure from protein sequence is another example

In prediction, performance is usually measured in terms of the accuracy of the results.

#### **2.2.1.10 Time Series – Prediction**



Time series Prediction is similar to prediction except that the predictions are based on data gathered over a period of time that is historical data and in some cases future output is also predicted.

Examples

- *Predicting rainfall based on historical data*
- *Predicting real estate prices based on historical data*

Again, performance is usually measured in terms of the accuracy of the results.

### 2.2.1.11 Sequence analysis

A sequence is an ordered list of objects (or events). Here sequences are compared to find similarity and dissimilarity between them. One of the key aspects of sequence comparison is sequence alignment. Sequence alignment maximizes the number of positions that are in agreement in two sequences. By analysing such sequences, for an unseen sequence we can predict to which class the sequences belong to.

Examples

- *Given a sequence of packets, we can label the session as an intrusion or as normal.*
- *Sequence analysis in bioinformatics is an examination of characteristic fragments, e.g. of a DNA strand.*

Again, performance is usually measured in terms of the accuracy of the results.

### 2.2.1.12 Evolution

One of the important traits of learning is adaptation. Adaptation is defined as change in the structure or functioning of an organism that makes it better suited to its environment. Machine learning that allows this adaptation is called evolution.

Examples

- *Speech recognition*
- *Language learning*

Again, performance is usually measured in terms of the accuracy of the results.

## 2.3 Feature Representations

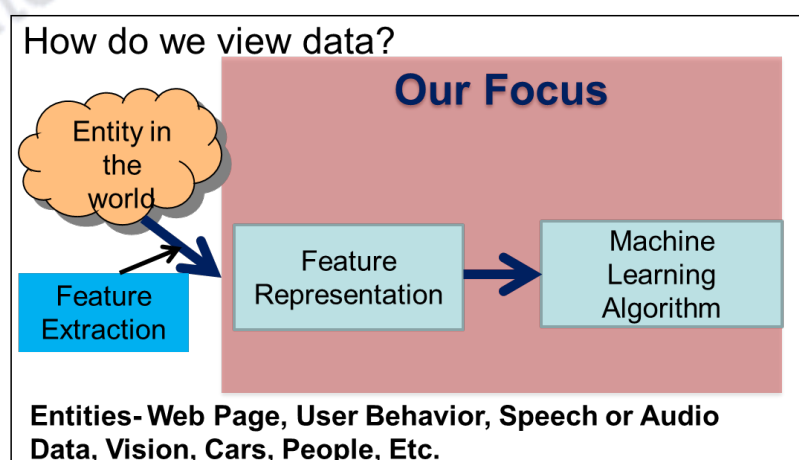


Figure 2.2 Context of Feature Representation

As we have already seen in order to carry out machine learning about an entity in the world there is a need to extract features about it. These entities could be Web Pages, User Behavior, documents, electronic health records, farm data, speech or audio Data, cars, people, etc. In general we do not consider feature extraction as our focus although machine learning methods could be used to carry out this task. Our focus is basically feature representation and the machine learning algorithms used (Figure 2.2).

Given below is an example of a feature representation in the form of a table (Table 2.1). This set of features can be used to classify people.

Height	Weight	Eye Color	Gender
66	170	Blue	Male
73	210	Brown	Male
72	165	Green	Male
70	180	Blue	Male
74	185	Brown	Male
68	155	Green	Male
65	150	Blue	Female
64	120	Brown	Female

**Table 2.1 Feature Representation**

### 2.3.1 Types of Data

The following are the three common types of data that can be used to represent features:

- **Discrete Data:** A set of data having finite number of values or data points is called Discrete Data. In other words discrete data only include integer values. 'Count' data, derived by counting the number of events or animals of interest, are a type of discrete data. Examples of discrete data include marks, number of registered cars, and number of children in a family and number of students in a class.
- **Ordinal Data:** Ordinal data are inherently categorical in nature, but have an intrinsic order to them. A set of data is said to be ordinal if the values / observations belonging to it can be ranked or have an attached rating scale. Examples include academic grades (i.e. O, A+, A, B+), clothing size and positions in an organization.
- **Continuous Data:** Continuous data can take any of a range of values and the possible number of different values which the data can take is infinite. Examples of types of continuous data are weight, height, and the infectious period of a pathogen. Age may be classified as either discrete (as it is commonly measured in whole years) or continuous (as the concept of a fraction of a year is possible).

In addition, Structural, Temporal, Spatial, Spatio-temporal and graphical models can also be used as input data to represent features. We will discuss these types later.

**2.3.2 Data Dependencies:** Another important aspect to be considered when talking about input data is the data dependencies between the different features. In many cases of machine learning a simplifying assumption is made regarding data dependency. This assumption is the so called **Independent and Identically Identified (IID)** - IID which refers to sequences of random variables. "Independent and identically distributed" implies an element in the sequence is independent of the random variables that came before it. In other words we usually assume that data points are *sampled independently and from the same distribution*. The most common example is the repeated tossing of a coin several times. The sequence of results (head/tail) you get are said to be IID. They are *independent* since every time you flip a coin, the previous result does not influence your current result. They are *identically distributed*, since every time you flip a coin, the chances of getting head (or tail) are identical (0.5), whether it is the 1st toss or the 100th toss.

**2.3.3 Linear and Non-Linear dependencies:** The dependencies between different input data can sometimes be represented by linear functions or may be complex and the dependencies can be represented only by a non-linear function.

**2.3.4 Observable vs latent variables:** Another important concept about the data is whether the variables are directly observable or are latent. Latent variables or hidden variables are not directly observable but are inferred from observable variables. One advantage of using latent variables is that it reduces the dimensionality of data. Examples include abstract concepts, like categories, behavioral aspects or mental states.

**2.3.5 Representation:** The way the input is represented by a model is also important for the machine learning process. The representations may be the instances themselves but may also include decision tree, graphical models, sets of rules or Logic programs, neural networks, etc. These representations are usually based on some mathematical concepts. While decision trees are based on Propositional logic, Inductive Logic Programming is based on First order logic. Bayesian Networks are based on Probabilistic descriptions whereas, Neural networks includes linear weighted polynomials.

**2.3.6 Availability of Prior Knowledge:** Another important aspect of the input data is the presence or absence of prior knowledge. In general, prior knowledge affects how the learner perceives new information. A majority of learning systems do not have prior knowledge while statistical learning systems use prior knowledge. The prior domain knowledge is all of the auxiliary information about the learning task that can be used to guide the learning process, and this prior information comes from either some other discovery processes or from domain experts. For example incorporating prior knowledge about the representation of objects profoundly influences the effectiveness of the image classification process.

## 2.4 Algorithms

Once the input data is provided to the learning system, the next obvious step is the use of the machine learning algorithm. The success of a machine learning system depends on the algorithms. These algorithms control the search to find and build the knowledge structures. The main function of the learning algorithms is to extract useful information from training examples.

## 2.5 Learning Methods



As you would have observed there are many ways in which we can categorize machine learning. Here the categorization is based on from what input learning happens or what type of data is learned. Accordingly we classify as:

**2.5.1 Rote learning** (memorization) - Here facts are stored and retrieved when needed. No inference happens.

**2.5.2 Learning from instruction** - Here instructions provided to the system probably in the form of rules are used for learning. Example - Teaching a robot how to hold a cup.

**2.5.3 Learning by analogy** – Here the system learns from an example and applies that knowledge to an analogous situation. In other words it transforms existing knowledge to fit anew situation. Example - learn how to hold a cup and generalize to learn to hold objects with a handle.

**2.5.4 Learning from examples** – This is a special case of inductive learning and is a well studied method in machine learning. Example – Giving samples of good/bad credit card customers, the system learns whether a new customer is good or bad.

**2.5.5 Learning from sequential data** – Here learning is based on the sequence of data. Example - Speech recognition, DNA data analysis.

**2.5.6 Learning associations** – Here the learning system tries to learn the associations between entities Example if a man likes red, he also likes roses.

**2.5.7 Learning from observation and discovery** – This is complete unsupervised learning where with only data learning happens. This is ambitious and is in fact the goal of science such as cataloguing celestial objects.

## 2.6 Types of Feedback

As we discussed in the previous section machine learning can be classified in many ways. Another very popular classification is based on the type of feedback provided to the learning system. The general types under this category include:

### 2.6.1 Supervised Learning

In supervised learning the sufficient training data including the desired outputs is provided that is data is provided in a supervised manner. The idea is that with sufficient input-output pairs provided the learning algorithm generates a function that maps inputs to desired outputs.

Example - **classification problem**: the learner is required to learn a function which maps a vector into one of several classes by looking at several input-output examples of the function. This classification is one of the most popular supervised learning strategies.

### 2.6.2 Unsupervised Learning

In this method training data do not include the desired outputs. Example is the **clustering problem** where without the desired output, the grouping is carried out by finding common features between input samples.

### 2.6.3 Semi-supervised Learning

In this method training data includes a limited number of samples with desired outputs. This approach avoids the huge cost involved in providing large number of input-output pairs but at the same time avoids the convergence issue associated with unsupervised learning.

#### 2.6.4 Reinforcement Learning

This method is also based on the type of feedback provided. In reinforcement learning, rewards are given based on sequential actions. The system learns a policy of how to act given an observation of the world. We assume that every action has some impact in the environment, and the environment provides feedback that guides the learning algorithm.

Example – Suppose you are learning driving from your father. Each time you make a correct move you are praised, but if you make a wrong move you are reprimanded. In course of time you will try to perform actions for which you are praised and in this way learning happens.

#### 2.6.5 Transduction

This approach is similar to supervised learning, but however does not explicitly construct a function but instead, tries to predict new outputs based on training inputs, training outputs, and new inputs.

#### 2.6.6 Learning to learn

In this case the algorithm learns its own inductive bias based on previous experience. Inductive bias is the set of assumptions that the learner uses to predict outputs given inputs that it has not yet encountered

### 2.7 Choosing the Type of Training Experience

Another aspect of machine learning the way the feedback is given during the training. This aspect can again be classified as follows

- **Direct experience:** Here the given sample input and output pairs are given explicitly in order to learn a useful target function.  
Example - Checker boards labeled with the correct move, e.g. extracted from record of expert play
- **Indirect experience:** Here the given feedback which is *not* direct inputoutput pairs but this information has to be gathered from the given feedback in order to learn a useful target function. One associated issue in this context is the Credit/Blame Assignment Problem that is the problem of how to assign credit/ blame to individual moves given only indirect feedback?  
Example-Potentially arbitrary sequences of game moves and their final game results.

#### 2.7.1 Teacher versus Learner Controlled Experience

Associated with the type of training experience is how these training examples are provided. The following are some of the ways:

- the teacher might provide training examples;
- the learner might suggest *interesting* examples and ask the teacher for their outcome; or
- the learner can be completely on its own with no access to correct outcomes (basically unsupervised learning)

#### 2.7.2 How Representative is the Experience?

Another crucial feature that facilitates effective learning is whether the training experience is representative of the task the system will actually have to solve when learning is completed.

### 2.8 Issues in Machine Learning for achieving Generalization

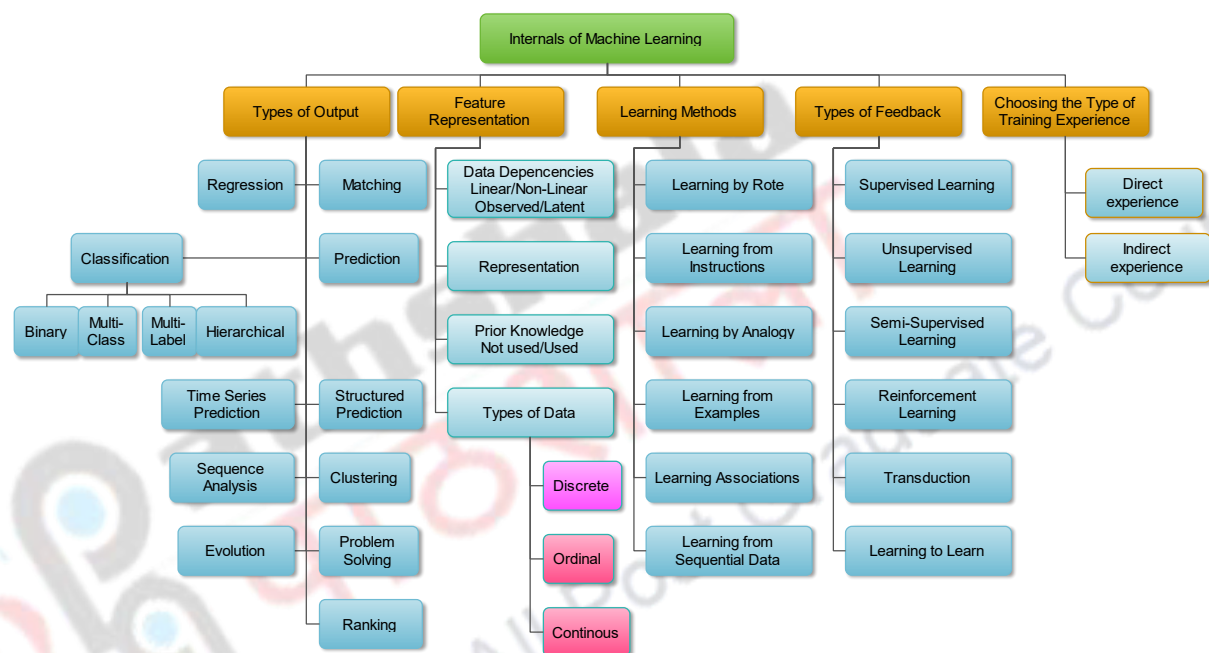
Machine learning continues to be a challenging issue with more and more interesting and new applications entering into the arena mainly due to the availability of large amounts of

data. However some of the issues that needs to be addressed while building successful machine learning applications include:

- What algorithms are available for learning a concept? How well do they perform?
- How much training data is sufficient to learn a concept with high confidence?
- When is it useful to use prior knowledge?
- Are some training examples more useful than others?
- What are best tasks for a system to learn?
- What is the best way for a system to represent its knowledge?

## Summary

Given below (Figure 2.4) is a summarization of the different aspects of machine learning along with their categorization.



**Figure 2.4 A categorization of the Internals of Machine Learning**

The aspects we covered in this module is given below:

- Outlined the different types of output expected from machine learning
- Discussed input data characteristics
- Explained the different type of feedback that aid Learning
- Discussed Types of Training Experience