CHAPTER – 2

MACHINE LEARNING

This chapter briefs about Machine learning and its classification, particularly Classification on the basis of underlying learning strategies used, Classification on the basis of the representation of knowledge or skill acquired by the learner and Classification in terms of the application domain in the performance system for which knowledge is acquire. Furthermore this chapter deals with supervised learning that is Classification and Regression, Unsupervised learning that is clustering.

2.1 Learning

Learning, like intelligence, covers such a broad range of processes that it is difficult to define precisely. A dictionary definition includes phrases such as “to gain knowledge, or understanding of, or skill in, by study, instruction, or experience," and modification of a behavioral tendency by experience". Learning denotes changes in the system that is adaptive in the sense that they enable the system to do the same task drawn from the same population more efficiently and more effectively the next time (Nils J Nilsson, 1998).

Learning is a many-faceted phenomenon. Learning processes include the acquisition of new declarative knowledge, the development of motor and cognitive skills through instruction or practice the organization of new knowledge into general, effective representations and the discovery of new facts and theories through observation and experimentation (Tom M Mitchell et al 1997). Since the inception of the computer era, researchers have been striving to implant such capabilities in computers. Solving this problem has been and remains a most challenging and fascinating long range goal in artificial intelligence (AI) (Jaime G Carbonell et al 1998). The study and computer modeling of learning processes in their multiple manifestations constitutes the subject matter of machine learning (Ryszard S Michalski et al 1994). Machine learning is programming computers to optimize a performance criterion using example data. It uses the theory of statistics in building mathematical model because the core task is making inference from a sample.
As regards machines, we might say, very broadly, that a machine learns whenever it changes its structure, program, or data (based on its inputs or in response to external information) in such a manner that its expected future performance improves. Some of these changes, such as the addition of a record to a database, fall comfortably within the province of other disciplines and are not necessarily better understood for being called learning. But, for example, when the performance of a speech-recognition machine improves after hearing several samples of a person's speech, we feel quite justified in that case saying that the machine has learned.

We are entering the era of big data. For example, there are about 1 trillion web pages, one hour of video is uploaded to YouTube every second, amounting to 10 years of content every day, the genomes of 1000s of people, each of which has a length of $3.8 \times 10^9$ base pairs, have been sequenced by various labs; Walmart handles more than 1M transactions per hour and has databases containing more than 2.5 petabytes ($2.5 \times 10^{15}$) of information and so on (Kevin P Murphy, 2012).

This deluge of data calls for automated methods of data analysis, which is what machine learning provides. In particular, machine learning are a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (such as planning how to collect more data!).

Machine learning usually refers to the changes in systems that perform tasks associated with artificial intelligence (AI). Such tasks involve recognition, diagnosis, planning, robot control, prediction, etc. The “changes” might be either enhancements to already performing systems or ab-initio synthesis of new systems (Kevin P Murphy, 2012).

Machine learning is organized around three primary research foci (Ryszard S Michalski et al 1997) that are

- Task-oriented studies – The development and analysis of learning systems to improve performance in a pre-determined set of tasks (also known as engineering approach)
- Cognitive simulation – The investigation and computer simulation of human learning processes
Theoretical analysis – The theoretical exploration of the space of possible learning methods and algorithms independent of application domain

2.2 Classification of Machine learning

Machine learning can be classified (Tom M Mitchell et al 1997) based on the following three dimensions

A. Classification on the basis of the underlying learning strategies used – the process themselves are ordered by the amount of inference the learning system performs on the available information. It is more difficult yet to program a computer to perform a complex task than to instruct a person to perform the task; as programming requires explicit specification of all requisite detail, whereas a person receiving instruction can use prior knowledge and common sense to fill in most mundane details. The taxonomy below captures this notion of trade-offs in the amount of effort required of the learner and the teacher.

- Rote learning and direct implanting of new knowledge – No inference or other transformation of the knowledge is required on the part of the learner. Variants of this knowledge acquisition method include – learning by being programmed and learning by memorization
- Learning from instruction – Acquiring knowledge from a teacher or other organized source, such as textbook, requiring that the learner transform the knowledge from the input language to an internally usable representation and that the new information be integrated with the prior knowledge for effective use.
- Learning by analogy – Acquiring new facts or skills by transforming and augmenting existing knowledge that bears strong similarity to the desired new concept or skill into a form effectively useful in a new situation.
- Learning from examples – Given a set of examples and counterexamples of a concept, the learner induces a general concept description that describes all of the positive examples and none of the counterexamples.
- Learning from observation and discovery (also called unsupervised learning) – This is a very general form of inductive learning that includes discovery systems, theory
formation tasks, the creation of classification criteria to form taxonomic hierarchies and similar tasks without the benefit of an external teacher.

B. Classification on the basis of the representation of knowledge or skill acquired by the learner. A learning system may acquire rules of behavior, descriptions of physical objects, problem solving heuristics, classification taxonomies over a sample space and many other types of knowledge useful in the performance of a wide variety of tasks.

- Parameters in algebraic expressions - learning in this context consists of adjusting numerical parameters or coefficients in algebraic expressions of a fixed functional form so as to obtain desired performance.
- Decision trees – Some systems acquire decision trees to discriminate among classes of objects. The nodes in a decision tree correspond to selected object attributes, and the edges correspond to predetermined alternative value for these attributes. Leaves of the tree correspond to sets of objects with an identical classification.
- Formal grammars – in learning to recognize a particular language, formal grammars are induced from sequence of expressions in the language. These grammars are typically represented as regular expressions, finite-state automata, context-free grammar rules, or transformation rules.
- Production rules – A production rule is a condition-action pair \( \{C \Rightarrow A\} \), where \( C \) is a set of condition and \( A \) is a sequence of actions. If all the condition in a production rules are satisfied, then the sequence of action is executed. The four basic operations whereby production rules may be acquired and refined are: Creation, Generalization, Specialization and Composition.
- Formal logic-based expressions and related formalisms. These general purpose representations have been used to formulate descriptions of individual objects and to formulate resultant concepts descriptions.
- Graphs and networks – In many domains graphs and networks provide a more convenient and efficient representation than logical expressions, although the expressive power of network representations is comparable to that of formal logic expressions.
Frames and schemas - These provide larger units of representation than a single logical expressions or production rules. Frames and schemas can be viewed as collections labeled entities (“Slots”), each slots playing a certain prescribed role in the representation.

Computer programs and other procedural encodings – The objective of several learning systems is to acquire an ability to carry out a specific process efficiently rather than to reason about the internal structure of the process.

Taxonomies – learning from observation may result in global structuring of domain objects into a hierarchy or taxonomy. Clustering objects description into newly proposed categories and forming hierarchical classification require the system to formulate relevant criteria for classification.

Multiple representations – Some knowledge acquisition systems use several representation schemes for the newly acquired knowledge. Most notably, some discovery and theory – formation systems that acquire concepts, operations on those concepts, and heuristic rule for a new domain must select appropriate combination of representation schemes applicable to the different forms of knowledge acquired.

C. Classification in terms of the application domain in the performance system for which knowledge is acquired. – A useful dimension for classifying learning system is their area of application.

- Cognitive modeling (Simulating human learning process)
- Computer programming
- Expert system (High performance, domain-specific AI programs)
- Game playing
- Image recognition
- Medical diagnosis
- Natural language processing
- Physical object characterization
- Planning and problem solving
- Robotics
2.3 Types of Machine Learning

2.3.1 Supervised Learning

The concept of supervised learning comes from the supervisor, acting as a teacher in the learning process. Supervised learning is the machine learning task of inferring a function from labeled training data. The training data consists of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples (Kevin P Murphy, 2012). An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way. The parallel task in human and animal psychology is often referred to as concept learning. Supervised learning generates a function that maps inputs to desired outputs (also called labels, because they are often provided by human experts labeling the training examples).

Supervised learning is probably the most commonly used learning paradigm. In fact, given experience" under the form of examples of a target function, i.e. input-output pairs, it allows to devise practical solutions through a large spectrum of learning algorithms (Kevin P Murphy, 2012). The need for such large spectrum of learning algorithms is, in part, due to the many real-world learning problems which, falling under the supervised umbrella, are characterized by heterogeneous tasks and problem-specific learning algorithms for their solution. These include classification and regression problems (including multi-label and multi-class classification, and multivariate regression), as well as ranking-based (either label or instance ranking) and ordinal regression problems. The typical approach followed to cope with these complex problems is to map them into a series of simpler, well-known settings and then to combine the resulting predictions. Often, however, these solutions lack a principled theory and/or require too much computational resources to be practical for real-world applications.
There are two types of supervised machine learning – Regression and classification. **Classification** - Classification is considered an instance of supervised learning, i.e. learning where a training set of correctly-identified observations is available. The goal is to learn a mapping from inputs $x$ to outputs $y$, where $y \in \{1, \ldots, C\}$, with $C$ being the number of classes. If $C = 2$, this is called binary classification (in which case we often assume $y \in \{0, 1\}$); if $C > 2$, this is called multiclass classification. If the class labels are not mutually exclusive (e.g., somebody may be classified as tall and strong), we call it multi-label classification, but this is best viewed as predicting multiple related binary class labels (a so-called multiple output model). When we use the term “classification”, we will mean multiclass classification with a single output, unless we state otherwise (Kevin P Murphy, 2012).

Classification is probably the most widely used form of machine learning, and has been used to solve many interesting and often difficult real-world problems. Given below are a few real world applications.

- Document classification and email spam filtering
- Image classification and handwriting recognition
- Face detection and recognition

**Regression** - Regression is just like classification except the response variable is continuous. Some examples of real-world regression problems (Kevin P Murphy, 2012).

- Predict tomorrow’s stock market price given current market conditions and other possible side information.
- Predict the age of a viewer watching a given video on YouTube.
- Predict the location in 3d space of a robot arm end sector, given control signals (torques) sent to its various motors.
- Predict the amount of prostate specific antigen (PSA) in the body as a function of a number of different clinical measurements.
- Predict the temperature at any location inside a building using weather data, time, door sensors, etc.

Furthermore supervised learning can be categorized into offline and online learning. In offline learning, there is a data set available where the learner can build the internal model without any limits in accessing the data. In contrast to offline learning, online learning has
access to a sample only once. The goal is the same, predicting targets as accurate as possible.

In general, there is very much data available in an online learning setup, the data set grows continuously. Offline learning has equal or superior accuracy compared to online learning when the same amount of data is used.

### 2.3.2 Unsupervised Learning

In machine learning, unsupervised learning refers to the problem of trying to find hidden structure in unlabeled data. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution. This distinguishes unsupervised learning from supervised learning and reinforcement learning (Kevin P Murphy, 2012).

Unsupervised learning is closely related to the problem of density estimation in statistics. However, unsupervised learning also encompasses many other techniques that seek to summarize and explain key features of the data. Many methods employed in unsupervised learning are based on data mining methods used to preprocess.

It is also more widely applicable than supervised learning, since it does not require a human expert to manually label the data. Labeled data is not only expensive to acquire, but it also contains relatively little information, certainly not enough to reliably estimate the parameters of complex models.

An approach to unsupervised learning is clustering. Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters) (Kevin P Murphy, 2012).

Here are some real world applications of clustering.

- In astronomy, the auto class system discovered a new type of star, based on clustering astrophysical measurements.
- In e-commerce, it is common to cluster users into groups, based on their purchasing or web-surfing behavior, and then to send customized targeted advertising to each group.
- In biology, it is common to cluster flow-cytometry data into groups, to discover different sub-populations of cells.
2.3.3 Reinforcement Learning

Reinforcement learning is an area of machine learning, concerned with how an agent ought to take actions in an environment so as to maximize some notion of cumulative reward. Reinforcement learning learns how to act given an observation of the world.

Every action has some impact in the environment, and the environment provides feedback in the form of rewards that guides the learning algorithm. The environment is typically formulated as a Markov decision process (MDP), and many reinforcement learning algorithms for this context are highly related to dynamic programming techniques (Kevin P Murphy, 2012).

2.3.4 Evolutionary Learning

Evolutionary learning has been developing rapidly in the last decade. It is a powerful and general learning approach which has been used successfully in both symbolic systems - for example, rule-based systems - and sub-symbolic systems - for example, artificial neural networks. However, most evolutionary learning systems have paid little attention to the fact that they are population-based learning. The popularity of evolutionary learning and population-based learning in general, has increased rapidly in recent years. They offer distinct advantages over other machine-learning approaches in dealing with complex and changing environments. In evolutionary learning, the most popular approach is to represent a complete learning system as an individual. The learning system can be either a rule-base system or an artificial neural network.

2.4 Summary

Machine learning is a rich area as it has got its application from finance to biology, medicine to physics and chemistry and so on. This chapter justifies the importance of machine learning and discussed various classification of machine learning based on various factors. The concept of classification and regression of supervised and unsupervised learning has also been discussed. The approach of machine learning is used for process discovery in the forthcoming chapters.