**Machine Learning Assignment 2**

1. **What is the concept of human learning? Please give two examples.**

The process of human learning builds upon pre-existing knowledge, where the knowledge is either modified or reinforced to make it more accurate and subsequently used to improve decision making and problem solving.

1. **What different forms of human learning are there? Are there any machine learning equivalents?**

Supervised learning is similar to [**concept learning**](https://en.wikipedia.org/wiki/Concept_learning) (Bruner & Austin, 1986), where a person is required to classify new objects into existing categories, by matching the features of the new objects to examples in the categories.

Artificial neural networks follow the neuronal principle of [**Hebbian learning**](https://en.wikipedia.org/wiki/Hebbian_theory), where the algorithm centres on inputs with similar properties, just like how neurons that activate simultaneously strengthen the synaptic link between each other.

Reinforcement learning is inspired by the reward system of [operant conditioning](https://en.wikipedia.org/wiki/Operant_conditioning) (Sutton & Barto, 1998). Operant conditioning was established by [B. F. Skinner](https://en.wikipedia.org/wiki/B._F._Skinner) (1938), after [Edward Thorndike](https://en.wikipedia.org/wiki/Edward_Thorndike) (1927) observed how cats learn to escape a puzzle box more quickly after a repeated number of trials [9]. Thorndike noted that behaviours that helped the cats to escape were repeated more frequently over time compared to behaviours that did not, and he termed this the [law of effect](https://en.wikipedia.org/wiki/Law_of_effect).

1. **What is machine learning, and how does it work? What are the key responsibilities** of machine learning?

Machine learning (ML) is a type of artificial intelligence ([AI](https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence)) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning [algorithms](https://www.techtarget.com/whatis/definition/algorithm) use historical data as input to predict new output values.

The Machine Learning process starts with inputting training data into the selected algorithm. Training data being known or unknown data to develop the final Machine Learning algorithm. The type of training data input does impact the algorithm, and that concept will be covered further momentarily.

New input data is fed into the machine learning algorithm to test whether the algorithm works correctly. The prediction and results are then checked against each other.

If the prediction and results don’t match, the algorithm is re-trained multiple times until the data scientist gets the desired outcome. This enables the machine learning algorithm to continually learn on its own and produce the optimal answer, gradually increasing in accuracy over time.

Machine learning is important because it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.

1. **Define the terms “penalty” and “reward” in the context of reinforcement learning.**

A reinforcement learning algorithm, which may also be referred to as an agent, learns by interacting with its environment. The agent receives rewards by performing correctly and penalties for performing incorrectly. The agent learns without intervention from a human by maximizing its reward and minimizing its penalty.

1. **Explain the term “learning as a search”?**

Learning can be viewed as a search through the space of all sentences in a  concept description language for a sentence that best describes the data. Alternatively, it can be viewed as a search through all hypotheses in a [hypothesis space](https://doi.org/10.1007/978-0-387-30164-8_373). In either case, a generality relation usually determines the structure of the search space.

1. **What are the various goals of machine learning? What is the relationship between these and human learning?**

The Goals of Machine Learning.:

The goal of ML, in simples words, is to understand the nature of (human and other forms of) learning, and to build learning capability in computers. To be more specific, there are three aspects of the goals of ML. (1) To make the computers smarter, more intelligent. The more direct objective in this aspect is to develop systems (programs) for specific practical learning tasks in application domains. (2) To develop computational models of human learning process and perform computer simulations. The study in this aspect is also called cognitive modeling. (3) To explore new learning methods and develop general learning algorithms independent of applications.

Humans acquire knowledge through experience either directly or shared by others. Machines acquire knowledge through experience shared in the form of past data. We have the terms, Knowledge, Skill, and Memory being used to define intelligence. Just because you have good memory, that does not mean you are intelligent. And just because you are intelligent, it does not mean you should have a good memory. However, there are exceptions to these rules. Humans begin learning by memorizing. After few years, he realizes that mere capability to memorize is not intelligence. Then he practices on transforming the data stored in memory to knowledge and applies them to develop skills to solve problems faced in real life. A person with good memory and more knowledge without the required skills cannot be considered intelligent. Search engines replaces human memory and these days the focus is on acquiring intelligence by making use of data available on the web. In humans, learning speed depends on individuals and in machines, learning speed depends on the algorithm selected and the volume of examples exposed to it.

1. **Illustrate the various elements of machine learning using a real-life illustration.**

There are three main elements to every machine learning algorithm, and they include:

* Representation: what the model looks like; how knowledge is represented
* Evaluation: how good models are differentiated; how programs are evaluated
* Optimization: the process for finding good models; how programs are generated

1. **Provide an example of the abstraction method.**

Abstraction is defined as dealing with ideas instead of events. In the context of AI, that means worrying more about what the right algorithm is and less about how to implement it.

1. **What is the concept of generalization? What function does it play in the machine learning process**?

In machine learning, generalization is a definition to demonstrate how well is a trained model to classify or forecast unseen data. Training a generalized machine learning model means, in general, it works for all subset of unseen data.

1. **What is classification, exactly? What are the main distinctions between classification and regression?**

Classification and Regression are two major prediction problems that are usually dealt with in Data mining and machine learning.

**Classification** is the process of finding or discovering a model or function which helps in separating the data into multiple categorical classes i.e. discrete values. In classification, data is categorized under different labels according to some parameters given in input and then the labels are predicted for the data.   
The derived mapping function could be demonstrated in the form of “IF-THEN” rules. The classification process deal with the problems where the data can be divided into binary or multiple discrete labels.

**Regression** is the process of finding a model or function for distinguishing the data into continuous real values instead of using classes or discrete values. It can also identify the distribution movement depending on the historical data. Because a regression predictive model predicts a quantity, therefore, the skill of the model must be reported as an error in those predictions

1. **What is regression, and how does it work? Give an example of a real-world problem that was solved using regression.**

A regression model determines a relationship between an independent variable and a dependent variable, by providing a function. Formulating a regression analysis helps you predict the effects of the independent variable on the dependent one.

**Example**: we can say that age and height can be described using a linear regression model. Since a person’s height increases as its age increases, they have a linear relationship.

Regression models are commonly used as a statistical proof of claims regarding everyday facts.

1. **Describe the clustering mechanism in detail.**

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters.

Clustering Methods :

* Density-Based Methods: These methods consider the clusters as the dense region having some similarities and differences from the lower dense region of the space. These methods have good accuracy and the ability to merge two clusters. Example DBSCAN (Density-Based Spatial Clustering of Applications with Noise), OPTICS (Ordering Points to Identify Clustering Structure), etc.
* Hierarchical Based Methods: The clusters formed in this method form a tree-type structure based on the hierarchy. New clusters are formed using the previously formed one. It is divided into two category
  + Agglomerative (bottom-up approach)
  + Divisive (top-down approach)
  + examples CURE (Clustering Using Representatives), BIRCH (Balanced Iterative Reducing Clustering and using Hierarchies), etc.
  + Partitioning Methods: These methods partition the objects into k clusters and each partition forms one cluster. This method is used to optimize an objective criterion similarity function such as when the distance is a major parameter example K-means, CLARANS (Clustering Large Applications based upon Randomized Search), etc.
  + Grid-based Methods: In this method, the data space is formulated into a finite number of cells that form a grid-like structure. All the clustering operations done on these grids are fast and independent of the number of data objects example STING (Statistical Information Grid), wave cluster, CLIQUE (CLustering In Quest), etc.
  + Clustering Algorithms :   
    [K-means clustering algorithm](https://www.geeksforgeeks.org/k-means-clustering-introduction/) – It is the simplest unsupervised learning algorithm that solves clustering problem.K-means algorithm partitions n observations into k clusters where each observation belongs to the cluster with the nearest mean serving as a prototype of the cluster.