MATHEMATICS FOR MACHINE LEARNING

[1]Gaurav Kumar, [1]Rishav Banerjee, [1]Deepak Kr Singh, [1]Nitesh Choubey, [1]Arnaw

[1] *Institute of Engineering & Management, EP Block, Sector V, Salt Lake, Kolkata, India*

Email: gaurav.gayan01@gmail.com

Abstract

Machine learning is a way to study the algorithm and statistical model that is used by computer to perform a specific task through pattern and deduction [1]. It builds a mathematical model from a sample data which may come under either supervised or unsupervised learning. It is closely related to computational statistics which is an interface between statistics and computer science. Also, linear algebra and probability theory are two tools of mathematics which form the basis of machine learning. In general, statistics is a science concerned with collecting, analysing, interpreting the data. Data are the facts and figure that can be classified as either quantitative or qualitative. From the given set of data, we can predict the expected observation, difference between the outcome of two observations and how data look like which can help in better decision making process [2]. Descriptive and inferential statistics are the two methods of data analysis. Descriptive statistics summarize the raw data into information through which common expectation and variation of data can be taken. It also provides graphical methods that can be used to visualize the sample of data and qualitative understanding of observation whereas inferential statistics refers to drawing conclusions from data. Inferences are made under the framework of probability theory. So, understanding of data and interpretation of result are two important aspects of machine learning. In this paper, we have reviewed the different methods of ML, mathematics behind ML, its application in day to day life and future aspects.

Keywords: *Algorithm, Statistical model, Computational statistics, Descriptive statistics, Inferential statistics*

INTRODUCTION

Learning is a transformative process that leads to change as a result of experience and making the sense of future problems. The term machine learning was introduced in 1959 by Arthur Samuel. Arthur Samuel used the game of checkers to create the first self-learning program. Checkers playing program become a better player after many games against variety of human players in a supervised learning. The program observed winning strategies and develop its own algorithm. According to Gerald Dejong, prior knowledge of world is provided by training examples which analyses the training data and discards irrelevant information to form a general rule to follow [3].

In traditional computing, algorithms are set of programmed instructions used by computers to solve the problem. Machine learning algorithms instead allow computer to train on data inputs and use statistical analysis in order to get output values. Because of this, machine learning enables the computer in building model from sample data so that it can change when exposed to new data. It's primarily goal is to understand and recognise patterns and follow the methods by using algorithms to do task automatically without any human assistance. Machine learning is a multidisciplinary field. It draws on results from artificial intelligence, probability and statistics, computational complexity theory, control theory, information theory. In this paper, we will focus on machine learning methods, mathematics for ML, its application and future aspects.

MACHINE LEARNING METHODS

On the basis of approach, how learning is received and the type of problem they are intended to solve:

1. Supervised learning - It builds a mathematical model from a set of data that contain both input and the output [1]. The data is known as training data and consists of a set of training example. Each example is labelled with their desired output. For example – an algorithm may contain a data with image of sharks which is labelled as fish. Using the supervised learning algorithm, we should able to identify unlabelled shark image as fish [4]. It trains mathematical model to generate reasonable predictions for the response to new data. It uses historical data to predict statistically likely future events. The purpose of this method is to learn by comparing its output with the correct, intended output to find errors, and modify the model accordingly. It uses classification and regression algorithm to develop predictive models.
   - 1.1 Classification is a process of categorizing a given set of data into class. It can be performed on structured or unstructured data. It is used when the outputs are restricted to a limited set of values or when data can be tagged, categorized, or separated into specific groups or classes. It predicts discrete responses for example, whether an email is genuine or spam, or whether a tumour is cancerous or benign. K-nearest neighbour and neural networks are two common classification algorithms.
   - 1.2 Regression is a form of predicative modelling technique that attempts to show the relationship between two variable (independent and dependent). It involves graphing a line over a set of data points that most closely fits the overall shape of the data. It is used when the outputs may have any numerical value within a range or the nature of response is a real number such as temperature or the time until failure for a piece of equipment. It predicts continuous responses—for example, changes in temperature or fluctuations in power demand. Linear and logistic regression are two common regression algorithms.
2. Unsupervised learning – It builds a mathematical model from set of data that contains only inputs and no desired labelled output [1]. While data is unlabelled, algorithm is left to find commonalities among its input data. As unlabelled data are more abundant, machine learning methods that enables unsupervised learning are valuable. It is used to find structure and pattern in the dataset. It has a feature which allows the computational machine to automatically discover the representations that are needed to classify raw data. It is generally used for transactional data and can look at complex data that is more seemingly unrelated in order to organize it in meaning ways. Clustering and association

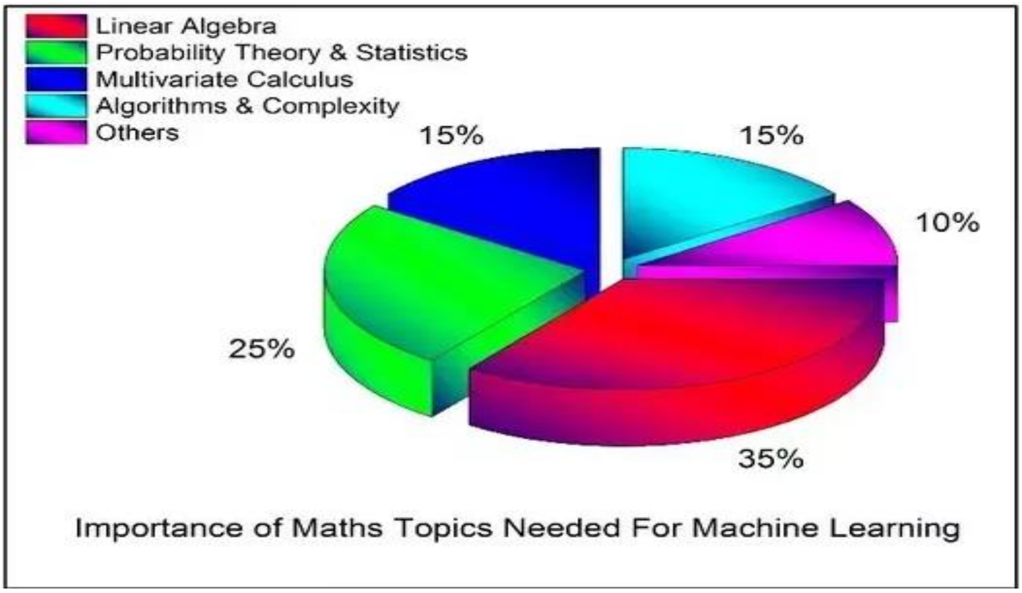
are two most common unsupervised learning technique.

1.1. Clustering is a task dividing the data points into a number of groups such that data points in the same groups are more similar than to data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them. It is used for exploratory data analysis to find hidden patterns or groupings in data. Applications for cluster analysis include gene sequence analysis, market research, and object recognition. K-means, hidden Markov models and self-organizing maps are some of the common algorithms for performing clustering.

1.2. Association is a way of finding the dependency of data objects in a large databases. It attempts to find relationships between different entities. Market based analysis is a typical example. Market Based Analysis is one of the key techniques used to show associations between items. It allows retailers to identify relationships between the items that people buy together frequently.

MATHEMATICS FOR MACHINE LEARNING

Machine learning is part of both statistics and computer science. It mainly requires the knowledge of linear algebra, probability theory and statistics, multivariate calculus, Algorithm and complexity as well as some other knowledge such as limit, topology, metric spaces, optimisation technique etc. [5].



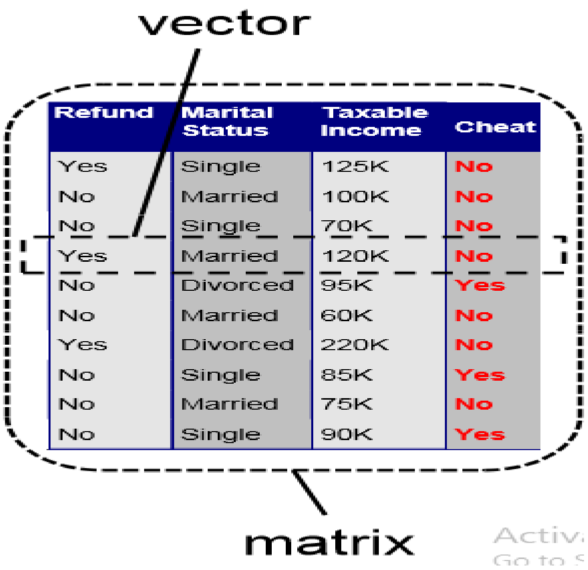
LINEAR ALGEBRA

Linear algebra is sub-field of mathematics concerned with vectors, matrices and linear transform. It is key foundation to the field of machine learning, for notations used to describe the operation of algorithms to the implementation of algorithm in code. Although Linear algebra is integral to the field of machine learning the tight relationship often remains unexplained or explained using abstract concepts such as vector spaces or specific matrix operations. Some of the obvious and concrete examples of linear algebra in machine learning are:

Dataset and data files, images and photograph, one hot encoding, regularisation, principal components analysis, singular value decomposition, latent semantic analysis, recommender systems and deep learning. It is also hugely useful for compact representation of linear transformations on data and it is also used in dimensionality reduction, coordinate transformation, operation on or matrices and vectors, linear regression, solution of linear systems of equation and many others. The two most important algebra used are vectors and matrices.

WHY VECTORS AND MATRICES?

The most common form of data organisation for machine learning is a 2-d array where row represent samples (record, items, data points) and columns represent attributes (features, variable). It is natural to think of each sample as a vector of attributes, and whole array as a matrix. Vectors are a n-tuple of values where n refer to the dimension of vector which can be any real number.



Refund	Marital Status	Taxable Income	Cheat
Yes	Single	125K	No
No	Married	100K	No
No	Single	70K	No
Yes	Married	120K	No
No	Divorced	95K	Yes
No	Married	60K	No
Yes	Divorced	220K	No
No	Single	85K	Yes
No	Married	75K	No
No	Single	90K	Yes

Vectors can be written in both row and column form but column form is conventional. Vector can thought to be a point in space or a directed line segments with magnitude and direction. There are many operations on vectors such as addition, dot product, cross product etc. [5].

Matrices are the m*n two-dimensional array of values (usually real numbers) where m is row and n is column. Matrix referenced by two-element subscript, where first element in subscript is row and the second element in subscript is column. A vector can be regarded as special case of a matrix, where one of matrix dimension is 1. There are many matrices operation such as matrix transpose means swapping of row and column, addition of two matrices, multiplication.

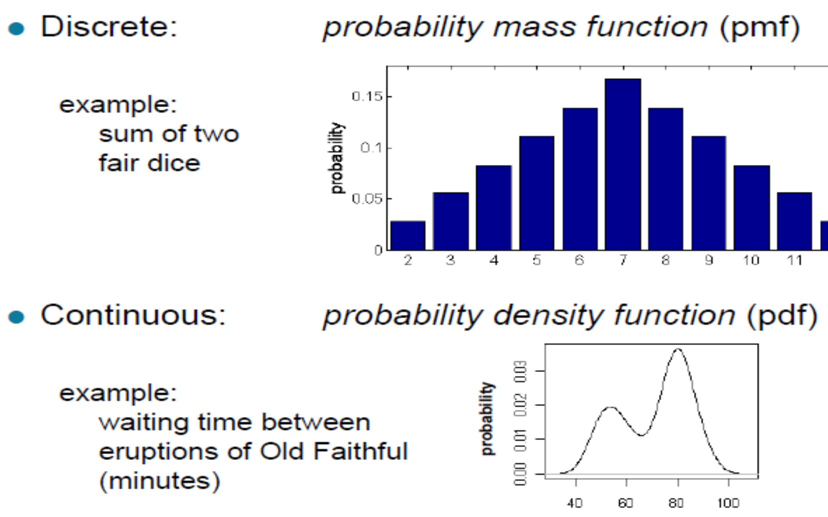
Projection in vectors are also used such as orthogonal projection of y on x can take place in any space of dimensionality is greater than or equal to 2.

PROBABILITY THEORY AND STATISTICS

Probability is the bedrock of machine learning for understanding and application of machine learning without it. Probability is about handling uncertainty which involves making decisions with incomplete information, and this is the way the world is generally operated. Probability is a field of mathematics that gives us the language and tool to quantify the uncertainty of events and reason in a principal manner hence it would be fair to say that probability is required to work through a machine learning predictive modelling project also uncertainty is fundamentals to the field of machine learning, yet it is one of the important aspects that causes the most difficulty for beginners. There are main sources of uncertainty in machine learning they are noisy data, incomplete coverage of the problem domain and imperfect models. Classification models must predict a probability of class membership, algorithms are designed using probability, learning algorithm will make decision using probability, sub-fields of study are built on probability algorithms are trained under probability frameworks shows that it is really a bedrock of machine learning.

PROBABILITY DISTRIBUTION AND BAYES THEOREM

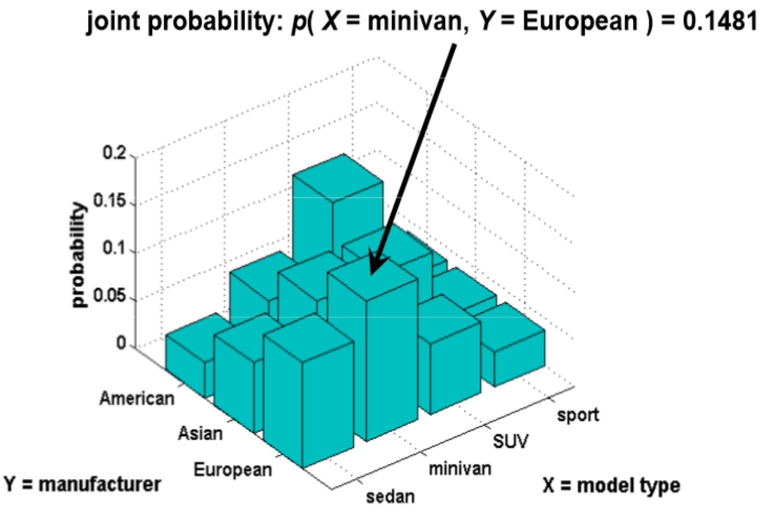
There are basically two types of probability spaces used in machine learning: discrete spaces (three consecutive flips of a coin) and continuous spaces (height of randomly chosen American male)



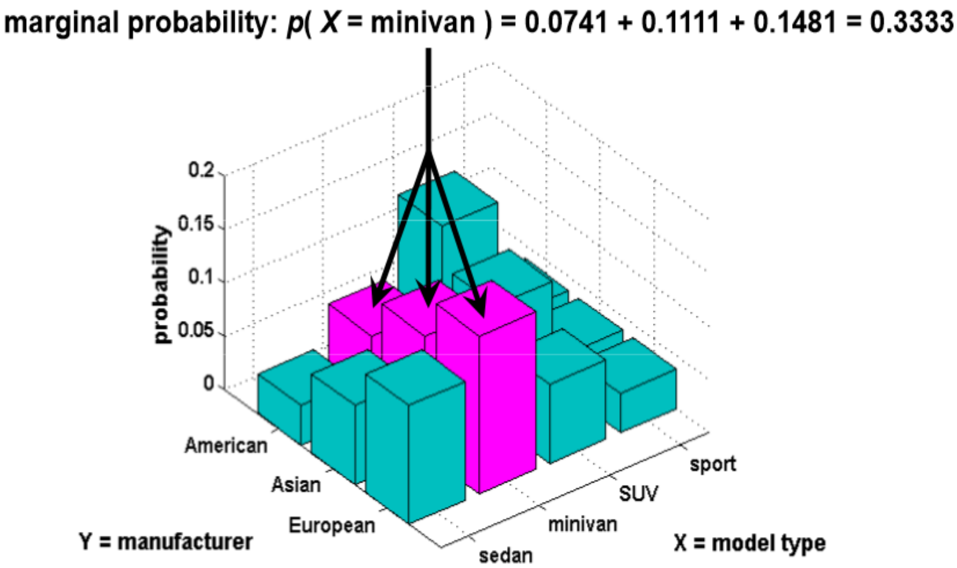
MULTIVARIATE PROBABILITY DISTRIBUTION

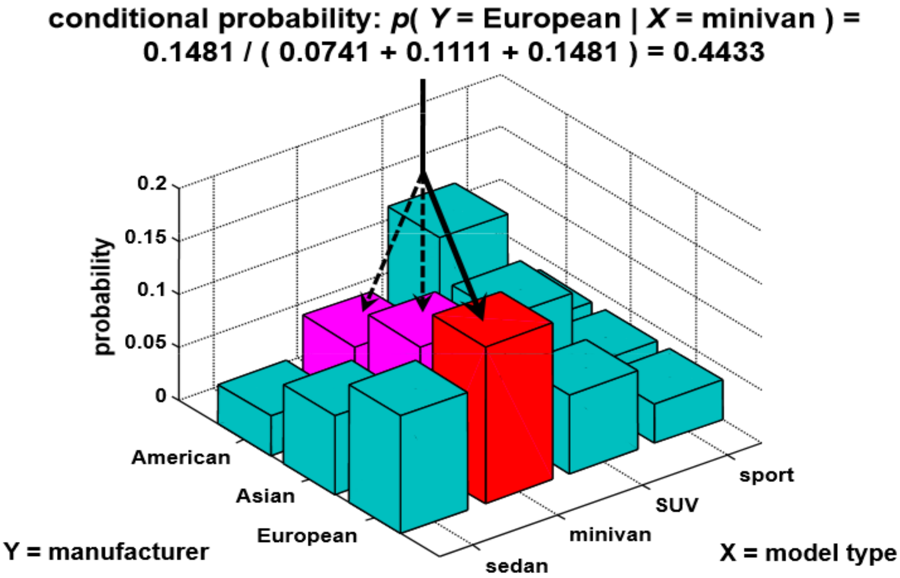
In multivariate probability distribution the main scenario is that several random processes occur (does not matter whether in parallel or in sequence) and if we want to know probability for several random variables for example two processes whose outcomes are represented by random variables x and y. probability that process x has outcome X and process Y has outcome y is denoted as:

$$P(x=X, Y=y)$$



There are two types of multivariate probability: - marginal probability (for single variable in a joint distribution) and continuous probability (one variable given that another variable takes a certain value) example of marginal and conditional probability are:-





CONTINUOUS MULTIVARIATE DISTRIBUTION

Some concepts of joints, marginal, and conditional probability apply (except use integrals) for example three-component Gaussian mixture in two dimensions.

OPTIMISATION THEORY

It mainly includes maximum likelihood estimation, Expectation maximisation, gradient descent.

Maximum likelihood estimation:- It is a method of estimation in which likelihood function is maximised so that under the assumed statistical model the observed data is more probable and the algorithm to find maximum likelihood ,where the model depends on unobserved latent variables is expectation maximisation.

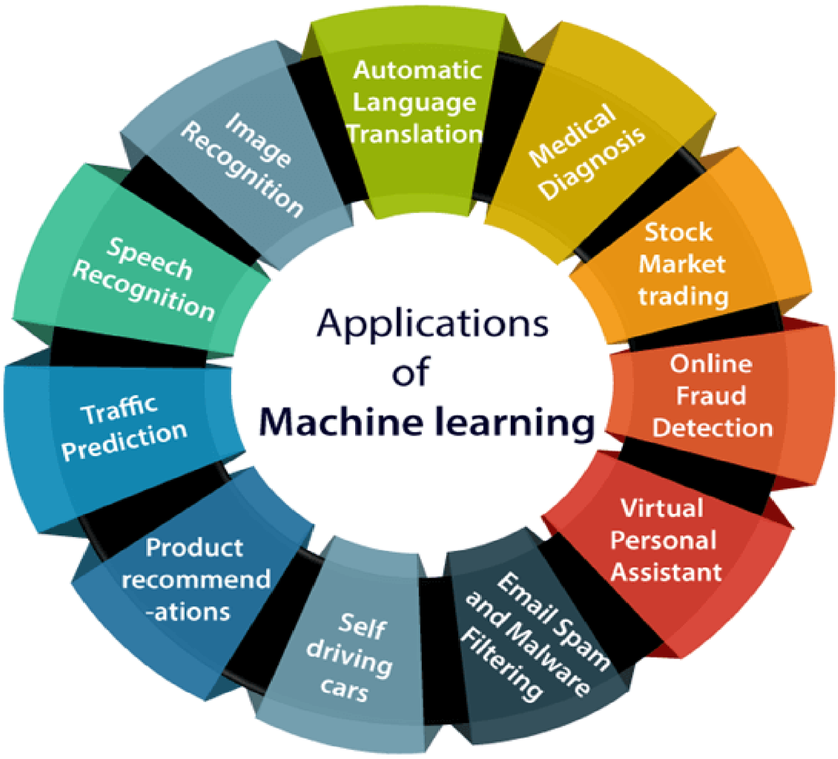
Gradient descent:- It is a first order iterative optimisation algorithm for finding local minimum of a function. To find local minimum of a function using gradient descent, one takes steps proportional to the negative of gradient of a function at that current point. In machine learning, we use gradient descent to update parameters of our model [6].

FUTURE OPEN FIELD OF RESEARCH

These are some of the fields open for further research works in the field of mathematics in machine learning.

1. **Non-convex optimization:** Not all cost functions would be convex in nature, some of them are arbitrary (non-convex) and it is highly complex to find global minima in such functions. Research is still going in this area.
2. **Transfer learning:** If you have applied DL/ML you would be knowing that currently all training models start from scratch or randomly initialized weights. Researches are being done mathematically in terms of transfer learning even today, to build models which leverage memories/experience of similar models to prevent DL models to start training from scratch every time a new problem is thrown at a neutral network. This is a part of the reinforcement deep learning which is still very complex to build because of its open mathematical complexities.
3. **Global Vs Local minima:** Though we have had enough amounts of research in terms of DL and optimization but one problem is still open. Identifying whether a reached minima is local or global? There are many methods like compare f(x-a) and f(x+a) with f(x)[brute forcing] but there is no clear algorithm to handle this, hence preventing an optimization function getting stuck in a local minima.
4. **Regularisation schemes:** DL usually have L2 or dropout techniques of regularization but more and more research is being done mathematically to prevent learning from over fitting [6].

APPLICATION OF MACHINE LEARNING



Machine learning is actively being used today perhaps in many more places than one would expect. We probably use a learning algorithm dozens of time without even knowing it. Applications of Machine Learning include:

1. Web Search Engine : One of the reasons why search engines like google, bing etc. work so well is because the system has learnt how to rank pages through a complex learning algorithm.
2. Photo tagging Applications : Be it facebook or any other photo tagging application, the ability to tag friends makes it even more happening. It is all possible because of a face recognition algorithm that runs behind the application.
3. Spam Detector : Our mail agent like Gmail or Hotmail does a lot of hard work for us in classifying the mails and moving the spam mails to spam folder. This is again achieved by a spam classifier running in the back end of mail application.

FUTURE OF MACHINE LEARNING

1. Improved cognitive services

With the help of machine learning services like SDKs and APIs, developers are able to include and hone the intelligent capabilities into their applications. This will empower machines to apply the various things they come across, and accordingly carry out an array of duties like vision recognition, speech detection, and understanding of speech and dialect. Alexa is already talking to us, and our phones are already listening to our conversations— how else do you think the machine “wakes up” to run a google search on 9/11 conspiracies for you? Those improved cognitive skills are something we could not have ever imagined happening a decade ago, yet, here we are. Being able to engage humans efficiently is under constant alteration to serve and understand the human species better.

We already spend so much time in front of screens that our mobiles have become an extension of us- and through cognitive learning, it has literally become the case. Your machine learns all about you, and then accordingly alters your results. No two people’s Google search results are the same: why? Cognitive learning.

2. The Rise of Quantum Computing

“Quantum computing”— sounds like something straight out of a science fiction movie, no? But it has become a genuine phenomenon. Satya Nadella, the chief executive of Microsoft Corp., calls i7t one of the three technologies that will reshape our world. Quantum algorithms have the potential to transform and innovate the field of machine learning. It could process data at a much faster pace and accelerate the ability to draw insights and synthesize information.

Heavy-duty computation will finally be done in a jiffy, saving so much of time and resources. The increased performance of machines will open so many doorways that will elevate and take evolution to the next level. Something as basic as two numbers- 0 and 1 changed the way of the world, imagine what could be achieved if we ventured into a whole new realm of computers and physics?

3. Rise of Robots

With machine learning on the rise, it is only natural that the medium gets a face on it— robots! The sophistication of machine learning is not a 'small wonder' if you know what I mean. Multi-agent learning, robot vision, self-supervised learning all will be accomplished through robotisation. Drones have already become a normality, and have now even replaced human delivery men. With the rapid speed technology is moving forward, even the sky is not the limit. Our childhood fantasies of living in an era of the Jetsons will soon become reality. The smallest of tasks will be automated, and human beings will no longer have to be self-reliant because you will have a bot following you like a shadow at all times.

REFERENCES

1. Machine learning available at wikipedia.org/wiki/ Machine learning
2. Statistical methods for machine learning. Discover how to transform data into knowledge from Machine learning mastery.com
3. Gerald Dejong, University of Illinois. Explanation based learning
4. Lisa Tagliaferri – an introduction to machine learning , Kress Digital Humanities Fellow, The Renaissance Afterlife of Catherine of Siena.
5. Mathematics for machine learning - Garrett Thomas (Department of Electrical Engineering and Computer Sciences University of California, Berkeley)
6. Machine learning maths essentials. Available at : https://courses.washington.edu/css490/2012.Winter/lecture_slides/02_math_essentials.pdf