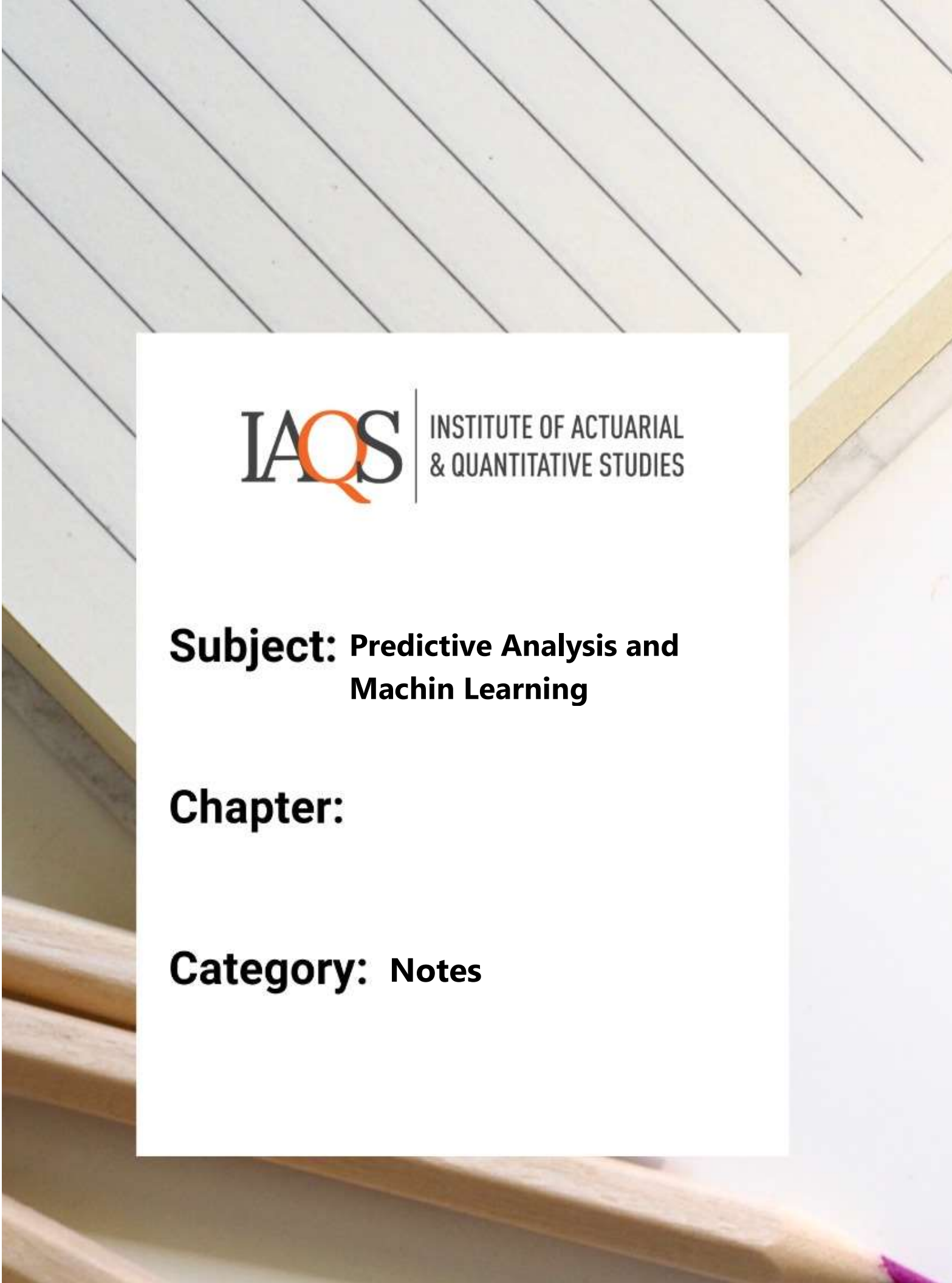


IAQS | INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Subject: Predictive Analysis and Machin Learning

Chapter:

Category: Notes

Machine Learning – Types of Learning

Supervised Learning

Supervised algorithms work by defining a set of input data , a hypothesized function and the expected results. By iteratively executing the function on the training data and involving the user to introducing control parameters, the model is improved. The algorithm is considered a success when its mappings and predictions are found to be correct.

To put it simply, for supervised learning we need input data (x_i data) , a function *f* and expected results y_i . We can define the control parameter (Maximum Likelihood or Least Square etc) based on the select algorithms and the computer will calculate the best fit parameters and their asymptotic distributions if possible.

Unsupervised Learning

While Supervised algorithms work on user labelled data for output predictions, these train machines explicitly on unlabelled data with little to no user involvement. Algorithms are left with the data to classify and group them to identify some hidden or undiscovered pattern and is often used a preliminary step for Supervised Learning.

In other words, *Unsupervised* learning is often used to identify patterns in the input data that are often difficult for humans to process without visualizing in higher dimensions or a logical deduction. Posting grouping data based on *Unsupervised Learning*, a predictive *Superivsed* model can be used in order to make predictions for any given input.

Reinforcement Learning

Reinforcement learning algorithms aim to find a perfect balance between exploration and exploitation without requiring labelled data or user intervention. These algorithms work by choosing an action and observing the consequences, based on that, it learns how optimal the result is. This process is repeated time and again until the algorithm evolves and chooses the right strategy.

Intutively speaking, this sounds a lot like how human babies evolve. At first, they have no data and thus take actions randomly, but slowly based on reactions to their actions they start adapting to the world. As a result, reinforcement learning is usually used in environments, where decisions have to be taken on a moments notice like autonomous driving, robotics and even simulations of real world activity such as getting AI to play a Boxing video game.

Deciding to Choose Between the Type of Learning

When a decision has to be made , between choosing which type of learning to be applied for the problem at hand, we have to answer a few important questions?

1. Do we already have a labelled expected output? If Yes, we can consider Supervised Learning but otherwise Supervised learning cannot be performed.
2. Is the functional relationship between the input and output clear? If you have data on income and inflation, its quite logical to say there will be a positive correlation. If such a relationship can be understood, its usually preferable to use supervised learning as opposed to unsupervised learning.
3. Is the data complete? If a lot of data is NA or incomplete, we need to take extreme caution when training models.

Generally, regression and classification is performed using supervised algorithms whereas pattern recognition is done by unsupervised algorithms.

TOP MACHINE LEARNING ALGORITHMS

1. **Linear Regression – Supervised Learning**

Linear Regression is a supervised ML algorithm that helps find a suitable approximate linear fit to a collection of points. At its core, linear regression is a linear approach to identifying the relationship between two variables with one of these values being a dependent value and the other being independent. The idea behind this is to understand how a change in one variable impacts the other, resulting in a relationship that can be positive or negative.

The regression line is represented by a linear equation :

$$Y_i = a + bX_i$$

Where

a is the intercept and b is the Slope.

This algorithms is applied for cases where the predicted output is continuous and has a constant slope like estimate sales based on price. Assessing risks for certain investments etc.

Linear Regression can be taken further by including more and more input data to predict the output but as number of input variables increase, so does the complexity of the model. So a computer program is usually used to fit a multiple linear regression model.

Linear Regression is not without its flaws. It fails when the dataset has a high multicollinearity (Correlation between predictor variables) and when the data is too scarce. As a result, **ridge regression** is used to create a parsimonious model under such conditions.

[Read More on Ridge Regression at : https://www.statisticshowto.com/ridge-regression/#:~:text=Ridge%20regression%20is%20a%20way,(correlations%20between%20predictor%20variables).]

Further modifications on linear regression allows the introduction of different weights to different observations to improve the fit of the model across all inputs rather than just where the most data of the training dataset is.

2. Logistic Regression – Supervised Learning

Logistic Regression algorithm is often used in the **binary classification problems** where the events in these cases commonly result in either of the two values, pass or fail, true or false. It is best suited for situations where there is a need to predict probabilities that the dependent variable will fall into one of the two categories of the response. Common use cases for this algorithm would be to identify whether the given handwriting matches to the person in question, will the prices of oil go up in coming months.

Logistic Regression is the same as fitting a Generalized Linear Model with logit link function and a Binomial Family.

In general, regressions can be used in real-world applications such as:

- Credit Scoring
- Cancer Detection
- Geographic Image Processing
- Handwriting recognition
- Image Segmentation and Categorization
- Measuring the success rates of marketing campaigns
- Predicting the revenues of a certain product
- Is there going to be an earthquake on a particular day?

3. Decision Trees – Supervised Learning

Decision Tree algorithm comes under supervised ML and is used for solving regression and classification problems. The purpose is to use a decision tree to go from observations to processing outcomes at each level. Processing decision trees is a top-down approach where the best suitable attribute from the training data is selected as the root and, the process is repeated for each branch. Decision Trees are commonly used for:

- Building knowledge management platforms
- Selecting a flight to travel

- Predicting high occupancy dates for hotels
- Suggest a customer what car to buy

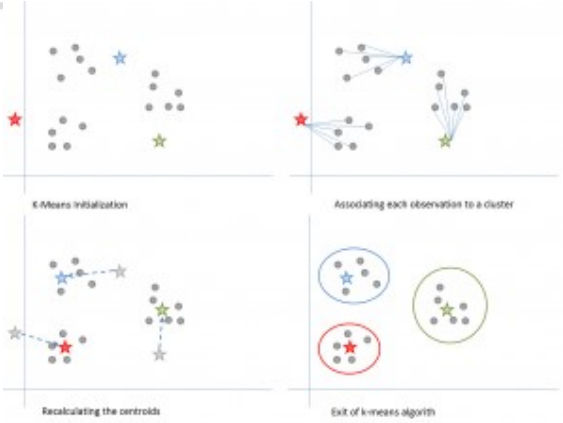
So the question is , when performing classification problems, what should be used? *Logistic Regression* or *Decision Trees*?

The decision on which algorithm is appropriate changes with the data provided. If the input numeric data has a wide spread or a few common but outlier values, decision trees can minimize the impact of such outlier input while still giving appropriate predictions. Similarly, if some data is missing or NA or Blank, then decision trees are much better at allowing for such categories then *Logistic Regression*. As a result, under such circumstances, Decision trees are preferred over Logistic Regression.

Random Forest Classifers are more advanced based on the same math as Decision Trees , where multiple decision trees are constructed and used to make a prediction.

4. **K – Means Clustering : Unsupervised Learning**

   k-means clustering is an iterative unsupervised learning algorithm that partitions n observations into k clusters where each observation belongs to the nearest cluster mean.



Steps of the K-means algorithm(source)

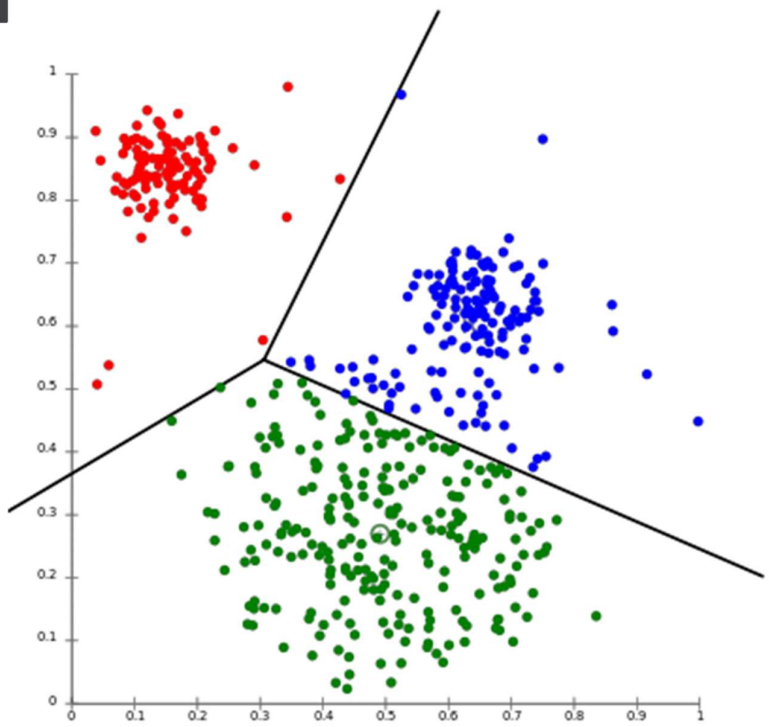
In simpler terms, this algorithm aggregates a collection of data points based on their similarity. Its applications range from clustering similar and relevant web search results, in programming languages and libraries such as Python, SciPy, **Sci-Kit Learn**, and **data mining**.

1. Identifying fake news
2. Spam detection and filtering
3. Classify books or movies by genre
4. Popular transport routes while town planning

5. K Nearest Neighbours – Unsupervised Learning

K-nearest neighbours is a supervised ML algorithm used for both regression and classification problems. Usually implemented for pattern recognition, this algorithm first stores, and identifies the distance between all inputs in the data using a distance function, selects the k specified inputs closest to query and outputs:

- The most frequent label (for classification)
- The average value of k nearest neighbours (for regression)



K-nearest Neighbours

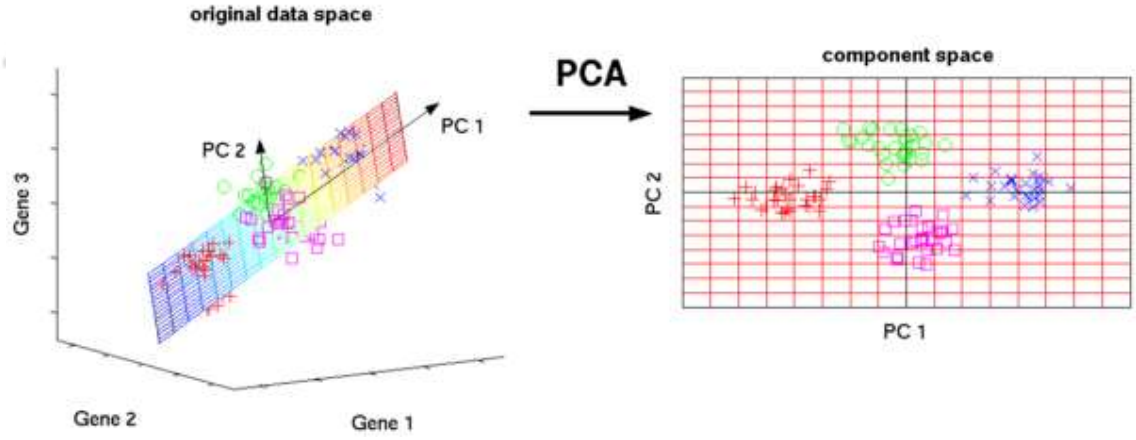
Real-life applications of this algorithm include —

- Fingerprint detection
- Credit rating
- Forecasting the stock market
- Analyzing money laundering
- Bank bankruptcies
- Currency exchange rate

6. Principal Component Analysis

Principal Component Analysis is one of the unsupervised algorithms for ML and is primarily used for reducing dimensions of your feature space by using either Feature Elimination or Feature Extraction. It is also used as a tool for exploratory data analysis and building predictive models. Requiring normalized data, PCA can help with:

- Image Processing
- Movie recommendation system
- Calculating data covariance matrix
- Perform eigenvalue decomposition on the covariance matrix
- Optimize power allocation in multiple communication channels



Principal Component Analysis

PCA aims to reduce redundancies from the datasets, making it simpler without compromising on accuracy. It is commonly deployed in image processing and risk management sectors.

[Want to learn more : https://www.youtube.com/watch?v=FgakZw6K1QQ]

7. **Naïve Bayes – Supervised Learning**

Naïve Bayes classifies are categorised as a highly effective supervised ML algorithm and are one of the simplest Bayesian Network Models.
It works by applying the Bayes' theorem on the data with a naïve assumption of conditional independence between every pair of features, given the value of the class variable. It is very useful when you already know the prior probabilities of a data point occurring in that class.

For example, you know that 1% of all mails in your inbox is spam. Naïve Bayes can be used to estimate the probability that a mail is spam based on some features of the mail.
In simpler terms, it helps find the probability of an even A happening , given that event B has occurred using the formula $P(A|B) = P(B|A) * \frac{P(A)}{P(B)}$

Naive Bayes is best for —

- Filtering spam messages
- Recommendation systems such as Netflix
- Classify a news article about technology, politics, or sports
- Sentiment analysis on social media
- Facial recognition software

8. **Gradient Boosting & AdaBoost**

Boosting is a technique for ensemble ML algorithms converting weak learners to strong learners. Boosting algorithms are required when data is abundant, and we seek to reduce the bias and variance in supervised learning. Below are two of the popular boosting algorithms.

- **Gradient Boosting**

Gradient Boosting algorithm is used for classification and regression problems by building a prediction model typically in an iterative manner such as the decision trees. It improves the weak learners by training it on the errors of the strong learners resulting in an overall accurate learner.

- **AdaBoost**

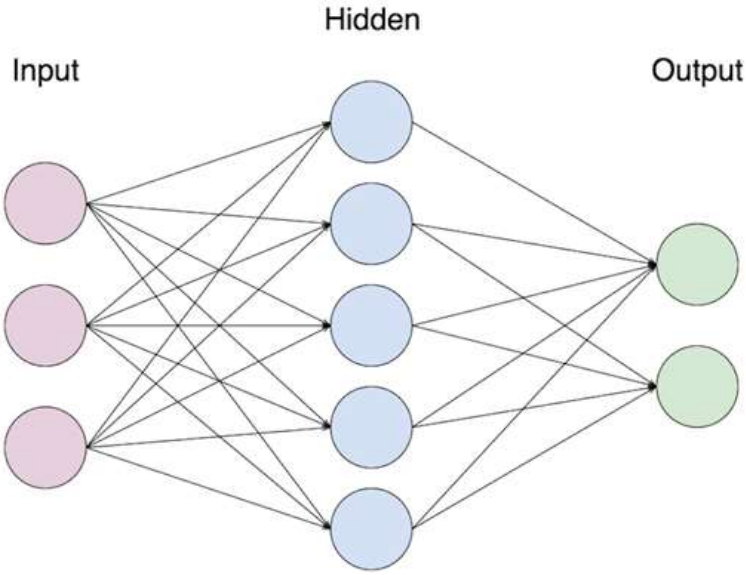
Short for Adaptive Boosting, AdaBoost improves the model when the weak learners fail. It does so by modifying the weights attached to the instances in the sample to focus more on the hard ones, later, the output from the weak learners is combined to form a weighted sum, and is considered the final boosted output.

[Learn More at : https://youtu.be/LsK-xG1cLYA]

9. Artificial Neural Networks

Modelled after the human brain, the Artificial Neural Network acts as an enormous labyrinth of neurons or simply called nodes , moving information to and from each other. The interconnected nodes pass data instantaneously to other nodes via the edges for swift processing , facilitating smoother learning. ANNs learn with examples, instead of being programmed with a specific set of rules. ANN are able to model non linear processes and can be implemented in areas such as –

- Pattern recognition
- Cybersecurity
- Data Mining
- Detecting varieties of cancer in patients



Artificial Neural Networks

[Learn More at : https://www.tutorialspoint.com/artificial_neural_network/artificial_neural_network_basic_concepts.htm#:~:text=Neural%20networks%20are%20parallel%20computing,faster%20than%20the%20traditional%20systems.]