



A Survey on Machine Learning Algorithms

Sunpreet Kaur, Sonika Jindal,
Department of CSE,
Shaheed Bhagat Singh State Technical Campus, Ferozepur

Abstract — A large number of techniques has been developed so far to tell the diversity of machine learning. Machine learning is categorized into supervised, unsupervised and reinforcement learning. Every instance in given data-set used by Machine learning algorithms is represented same set of features. On basis of label of instances it is divided into category. In this review paper our main focus is on Supervised, unsupervised learning techniques and its performance parameters.

Index Terms — IEEEtran, journal, L^AT_EX, paper, template

I. INTRODUCTION

IMAGE Processing necessitate change of the nature of an image in order to improve its pictorial information for human explanation and contributing it to make it suitable for autonomous machine perception. The upper-hand of the Image processing machines Like systems made are CBIR, over humans is that they cover almost the entire electromagnetic spectrum where as humans cannot precept so wide range as their eye has visual limited band. An image retrieval system can operate on images of every range as computer-generated images, ultrasound, and electron microscopy. Image processing has wide range of applications in almost every area such as medicine, industry, remote sensing, astronomy, agriculture, industry and other related areas. The main key point of Image processing is Image retrieval in which from raw data reorganization of colour ,texture ,shape feature are done before any type of reasoning related to image content. An fruitful image retrieval need to operate on the collection of images to retrieve the relevant images based on the query image from the database which is near to human perception.

The image database main purpose is to store the image and image sequences that are relevant to a query. Content based image retrieval technique is a part of image processing .In this we need to extract image feature and store it in a table in row column format. Then similarity measurement techniques are applied on them to find similar kind of images to that of query image. For this Feature Extraction is done on database images and Feature of image is some visual similarity which can uniquely identify an image from others. There exist many image features like color,texture,shape etc. So Feature extraction refers to mapping the image pixels in to the feature space. With this extracted feature similarity measure are done between database and query image [1].

The main part we are focusing in this paper is machined learning in which Machine are made for correct prediction making. The main idea behind machine learning is if computer could learn from experience its usefulness increases. It studies how to automatically learn to make accurate predictions based on past observations and main part of machine learning is Classification, in which examples are classified into given set of categories. Sometime the accuracy of machine prediction are much more accurate than human-crafted rules. As human are often incapable of expressing what they know so for this evolution of machine learning system came into lime light. The primary goal of Modern Machine Learning is highly accurate predictions on test data .There are several application of Machine Learning, the main of all is Data mining. Every instance in a data-set used by these algorithms is represented using same set of features.

These features may be continuous, categorical, and binary. If instances are given with labels and correct output then it is called supervised learning otherwise if instances are not labelled it is unsupervised learning. In this paper we review different techniques of machine learning and its different algorithms. The learning algorithms encountered are categorized as supervised or unsupervised algorithms.

II. MACHINE LEARNING TECHNIQUES

Machine learning has become one of the main-stays of information technology. With the ever increasing amount of data becoming available there is good reason to believe that smart data analysis will become even more pervasive as a necessary ingredient for technological purposes. Much of the science of machine learning is to solve those problems and provide good guarantees for the solution. [2]. There are different ways an algorithm can model a problem based on its interactions with the experience or environment or input data. For this whole firstly we have to adopt a learning style that an algorithm can adopt. There are only few main learning models that an algorithm can have. The way of organizing machine learning algorithms is useful because it forces to think about the roles of input data and the model of preparation process and selecting the one which is most suitable for problem for target results. An application where machine learning helps is named entity recognition. Let's discuss the different learning styles in machine learning algorithms and its different parts:

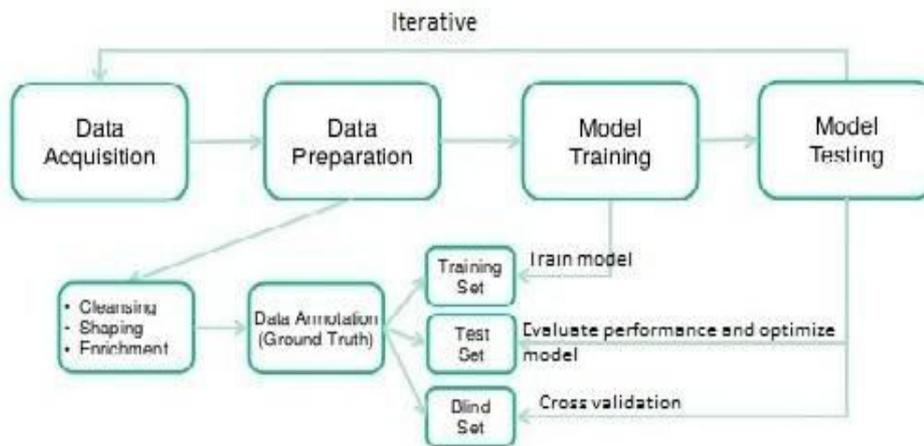


Fig. 1. Machine learning Flow Diagram

Supervised learning: In Supervised learning we learn a target function that can be used to predict the values of a discrete class attribute as approved or not-approved. Machine learning algorithm makes predictions on given set of sample whereas supervised learning algorithms searches for patterns within the value labels assigned to data points. This algorithm consists of an outcome variable which is to be predicted from, a given set of predictor's i.e. independent variables. Using these set of variables, we generate a function that map input to desired outputs. The training process continues until the model achieves level of accuracy on the training data. This whole process helps in reduction of expenditure on manual review for relevance and coding. Examples of supervised learning: Neural Networks, Regression, Decision tree, KNN, Logistic Regression, SVM, Naive Bayes etc. So predominantly it is divided into 2 parts:

Learning (training): Learn a model using the training data. Testing: Test the model using unrevealed test data to appraise the model accuracy. Itself continually using trial and error. This machine learns from its past experience and tries to capture the best possible knowledge to make accurate business decisions Such as Markov Decision Process. It learns to select an action to maximize payoff. Timely the algorithm changes its strategy to learn better and the best decision and accuracy. [3]

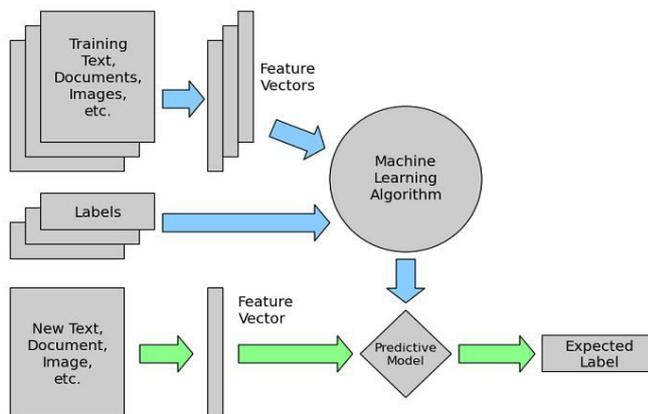


Fig. 2. Supervised learning

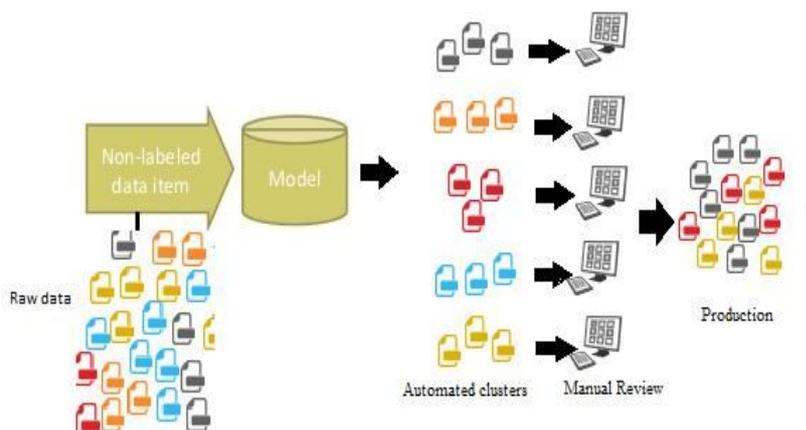


Fig. 3. Unsupervised learning

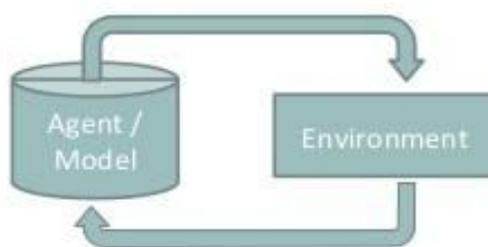


Fig. 4. Reinforcement Learning



Unsupervised learning: Learning useful structure with-out labelled classes, optimization criterion, feedback signal, or any other information beyond the raw data is referred as unsupervised learning. In this algorithm, we don't have any target variable to estimate means here we don't have any label associated with data points or we can say class label of training data are unknown. This algorithm is used for organizing the data into the group of clusters to describe its structure i.e. cluster the data to reveal meaningful partitions and hierarchies. It makes data look simple and organized for analysis. Examples: K-means, Fuzzy clustering, Hierarchical clustering. Input data is not labelled and doesn't have a known result. A model is prepared by deducing structures present in the input data. This may be to extract general rules. It may through a mathematical process to systematically reduce redundancy.

Reinforcement learning: Using this algorithm, the machine is trained to make specific decisions. These algorithms choose an action, based on each data point and later learn how good the decision was. In this the machine is exposed to an environment where it trains

A. SUPERVISED MACHINE LEARNING TECHNIQUES

Neural network, decision tree and support vector machines are some of the supervised machine learning techniques, which learn the high level concept from the low level image features

This technique performs the classification process with the help of the already categorized training data. For the training data, the input and the desired output is already known. When the supervised learning algorithms are trained with the known training data, it is able to generalize the new unseen data. Here the machine learning algorithm predicts the category of the query image which is nothing but the semantic concept of the query image. Here the matching is on the basis of query image category and the query image instead of whole database. So the retrieval results are more pure.

1) Support Vector Machine: SVM (support vector machine) are supervised learning models with associated learning algorithm that analyze data after which they are used for classification. Classification refers to which images are related to which class or data set or set of categories. In machine learning Classification is considered an instance of supervised learning which refers to task of inferring a function from labelled training data. Training data in image retrieval process can be correctly identified images that are put in an particular class. Where each class belong to different category of images. In the SVM training algorithms model is build in which the new examples are assigned to one category class or other. In this model representation of examples in categories are done with clear gaps that are as vast as possible.

Definition: The main idea of SVM is to construct the hyper plane in a high dimensional space which can be used for classification. Hyper plane refers to a subspace one dimension less than its ambient space. If there is 3-dimensional space then its hyperplane is the 2-dimensional planes. By hyper plane a good separation is achieved that has the largest distance to the nearest training-data point of any class. The separation between the hyper plane and closet data point is called the margin of separation. So more the margin of separation less be the generalization error of the classifier.

The main objective of the SVM machine is to find a particular hyper-plane for which the margin of separation is very high or which can be controlled to be maximized when this condition is met or we can under these circumstances, the decision plane which we take to differentiate between two classes, and then it is called as optimal hyper plane. The Support vectors play an important role in the operation of this class of learning machine as we can define Support vectors as the elements of training data set that would change the position of the dividing hyper-plane in SVM training algorithm if they are removed. As maximum-margin hyper-plane and margins for an SVM trained with samples from two classes and these samples on the margin are called as support vectors or we can say that these are data point that lies closest to the decision surface.

The belief which is central to the construction of the support vector machine in which learning algorithms is the inner product kernel between a support vector X and a vector X drawn from input space. The support vectors constitute a small subset of the training data extracted by the support vector learning algorithm. To keep the computational load, the mapping used by SVM scheme are designed to ensure that dot products maybe computed easily in terms of the variable in the original space, by declaring and defining them in terms of kernel function so that problem can be easily resolved.

Method: Classification of data is a common task in machine learning. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data.

Let there are some images each belong to one of two classes, and our main objective is to decide to which class a new image will be assigned to. Different images are put in (p -dimensional vector) and we need to know whether we can separate such points with hyper-plane ($p-1$). There are many hyper-planes which may classify the data. But we have to choose the best as per maximum margin of separation. The two main mathematical operations:

Nonlinear mapping of an input patten to higher dimensional feature space.

Construction of an optimal hyper plane for separating the patterns in the higher dimensional space obtained from first process.

Input: Set of training samples i.e. $x_1; x_2; x_3; \dots; x_n$ and the output result is y . In this we can take a lot of features as needed. Output: Set of weights w , one for each feature ,whose linear combination predict the value of y . Here we use the optimization of maximizing the margin to reduce the number of weights that are nonzero to just a few that correspond to the important features that matter in deciding the hyper plane .These non zero weights correspond to the support vector.

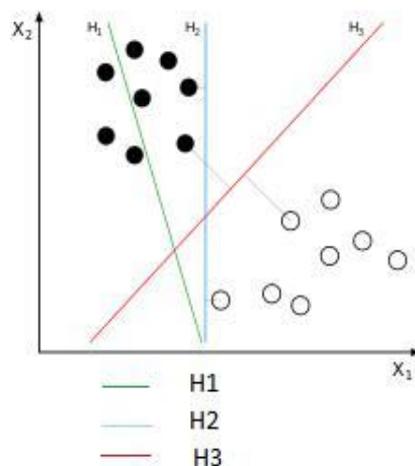


Fig. 5. H1 doesnot separate the class,H2 does but with small margin,H3 separate with maximum margin

ADVANTAGES OF USING SVM:

SVM offers best classification performance on the training data.

SVM provide more efficiency for pure classification of the future data. It doesn't make any strong assumption on data. It doesn't over fit the data.

DISADVANTAGES OF USING SVM:

More than one svm class may accept or all svm's may reject the data points .In such case data points cannot be classified.

APPLICATIONS OF SVM:

SVM is commonly used for stock market forecasting by various financial institutions. As For comparing relative performance of the stocks of different companies of same sector. So this relative comparison of stocks helps in managing investment based decisions.

Neural Network: Neural network are the models that are inspired by the structure and function of biological neural networks. They are class of pattern matching that are commonly used for regression and classification problem. They are basically based on a simple model of a neuron. A neural net is an artificial representation of the human brain that tries to simulate the learning process. This is often referred to Neural Network. Traditionally the word neural network is referred to as network of biological neurons in the nervous system that process and transmit information. Most neurons have three parts:1. A dendrite which collect input from other neurons 2. A soma which performs an important nonlinear processing step 3. An axon ,a cable-like wire along which the output signal is transmitted to the other neurons further down the processing chain.

But Artificial Neural Network is a interconnected group of artificial neurons that uses a mathematical model for information processing based on the connectionist approach to computation.

The connection site between two neurons is Synapse. In Artificial Neural Network (ANN), there is a continuous variable x_j which replaces the original spikes (short pulses of electricity). The rates x_j of all neurons which send signals to neurons i are weighted with parameters w_{ij} . Here may say that these weights as describing the efficacy of the connection from j to i . i.e. weight is therefore called as 'synaptic efficacy'. The output x_i of neuron i is a nonlinear transform of the summed input, where n is the number of input lines converging onto neuron i and θ_i is a formal threshold parameter. Basically neuron has three parameters - weights, threshold and a single activation function [4].

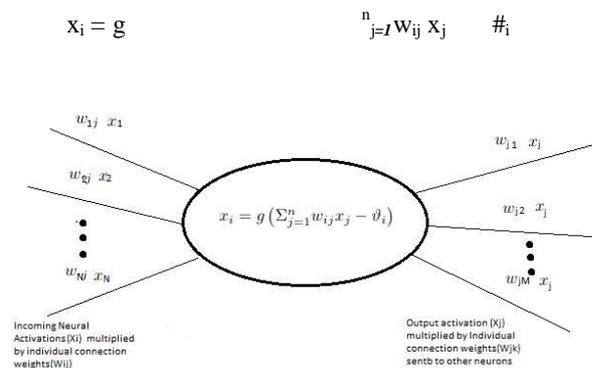


Fig. 6. Neural Network

It is a concept of learning the low level features of the segmented regions of the training set images are fed into the neural network classifiers, to establish the link between the low level image features and high level semantics. These basically follow, learn by example rule and are configured for specific applications as 1. Pattern recognition 2. Data classification. These types of machine are also called as Perception (a computer model devised to represent the ability of the brain to recognize and discriminate).

PERCEPTRON LEARNING RULE: AN ALGORITHM FOR LEARNING WEIGHTS IN SINGLE LAYERED NETWORKS.

PERCEPTRONS: LINEAR SEPARATORS, INSUFFICIENT EXPRESSIVE

ADVANTAGE OF ANN:

Adaptive system that changes its structure based on external or internal information that flows through the network

DISADVANTAGE OF ANN:

It requires a large amount of training data and is computationally intensive.

APPLICATIONS OF ANN:

Speech, driving, handwriting, fraud detection etc.

3) Decision Trees: A decision tree is a graphical representation that makes use of branching methodology to exemplify all possible a=outcomes of a decision .based on certain conditions. In this tree, the internal node represents a test on the attribute, each branch of the tree represents the outcome of the test and the leaf node represents a particular class label means the last decision after all computations. The classification rules are represented through the path from root to the leaf node.

It is supervised learning algorithm that is mainly used for classification. They works for both categorical and continuous dependent variables .Decision tree methods construct a model of decisions based on actual values of attributes in the data. Decisions split into tree structure until a prediction decision is made for a given record. Decision trees are trained on data for classification and regression problems. They are often fast, accurate and a big favourite in machine learning. This algorithm splits the data into two or more homogeneous sets .This is done based on most significant attribute to make as distinct groups as possible. For splitting the data into different heterogeneous groups, it uses various techniques like Gini, Information gain, Chi-square, entropy [5].

Input: An object or an situation described by a set of features. Output: A decision for an input value.

Representation of Decision Tree:

Each internal node tests an attribute

Each branch correspond to attribute value each leaf node assigns a Classification

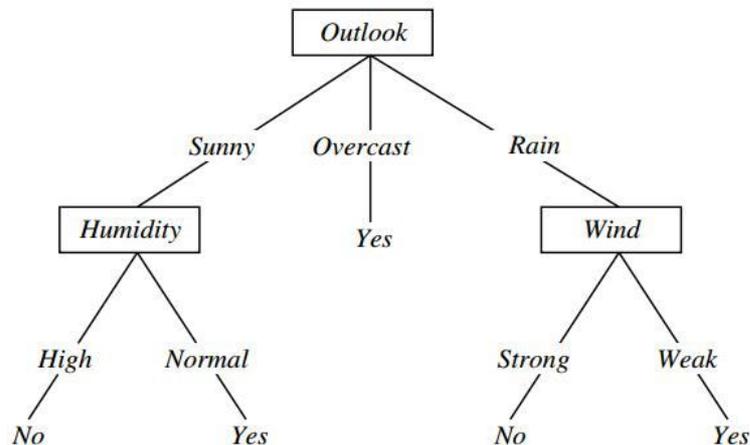


Fig. 7. Decision Tree for Play Tennis

TYPES OF DECISION TREES:

Classification Trees: These are the trees which are used to separate a data-set into different classes, based on the response variable. They are generally used when the response variable is categorical in nature.

Regression Trees: When the response or target variable is continuous or numerical, regression trees are used. These are mainly used in predictive type of problems [6], [7].

THE DIFFERENT DECISION TREE ALGORITHMS ARE:

Classification and Regression Tree (CART): It is a binary decision tree. It is constructed by splitting a node into two child nodes, repeatedly beginning with the root node that contains the whole learning sample. The CART algorithm will itself identify the most significant variables and eliminate the non-significant variables.

Iterative Dichotomiser 3 (ID3): This algorithm requires the value of the input attributes to be discrete. The ID3 finds the most useful attribute in classifying the given set. The attribute with maximum information gain is the most useful attribute.

C4.5: This algorithm can handle continuous attributes. At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits the set of samples into two subsets. The attribute with the highest normalized information gain is chosen to make decisions [8].

4) Naive Bayes: It is a classification Technique Based on Bayes' Theorem after Thomas Bayes Who proposed this Theorem. with an assumption of independence between predictors' Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. As, a fruit maybe considered to be an apple if it is red, round and about 3 inches in diameter.

Even if these features depend on each other or upon the existence of the other features, this classifier would consider all these properties to independently contribute to the probability that this fruit is an apple. This model is easy to build and mainly used for very large data sets. Along with its simpler part, it also covers highly sophisticated classification methods and performs very well. Bayesian methods are those that are explicitly apply Bayes' Theorem for problems such as classification and regression [9].

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad \text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

Fig. 8. Bayesian rule

The Bayesian Classification represents a supervised learning method and in addition to that a statistical method for classification. The probabilistic model allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. In machine learning, a probabilistic classifier is a classifier that is able to predict when given a sample input, a probability distribution over a set of classes, rather than only outputting the most likely class that the sample should belong [10].

MERITS OF NAIVE BAYES CLASSIFICATION:

It can solve diagnostic and predictive problems.

It provides practical learning algorithm and prior knowledge and observed data can be combined.

It provides a useful perspective for understanding and evaluating many learning algorithms.

It calculates explicit probabilities for hypothesis. It is robust to noise in input data.

USES OF NAIVE BAYES CLASSIFICATION:

- *This classification is used as a probabilistic learning method. This the most known and popular methods for learning to classify text documents.*
- *Spam filtering is the best known use of this classification. This is mainly used for identifying Spam e-mail. This spam filtering has become a popular mechanism to distinguish illegitimate spam email from legitimate email [11].*
- *Naive Bayes when combined with collaborative filtering makes a hybrid switching technique for filtering and prediction of resource allocation. It is more scalable, with more accuracy and performance.*

B. UNSUPERVISED MACHINE LEARNING TECHNIQUES

K-mean, Clustering approaches, genetic algorithms are some of the unsupervised machine learning techniques. In this learning technique input data is not la-belled and doesn't have a known result. A model is prepared by deducing structures present in the input data and extract general rules. It may go through a mathematical process to systematically reduce redundancy or may organize data through similarity. The main goal of the model is to determine Data patterns/grouping. The data have no target attribute and we want explore the data to find some intrinsic structure in them. Clustering is a technique for finding similarity groups in data called clusters. Means it groups data instances that are similar to each other in one cluster and data in stances that are very different from each other into different clusters. Clustering is often called as unsupervised learning as no class values denoting and a derivable grouping of the data instances are given. It refers to a problem of finding hidden structures in the labelled data. It has no measurements of outcome, to guide the learning process. Image clustering is typically unsupervised learning technique. It groups the set of images in such a way, that the similarity between different clusters must be minimized. The algorithm used is Apriori algorithm and K-mean algorithm.

1) K-MEAN ALGORITHM: K-mean is a partitional - clustering algorithm. It aims to partition the given n observations into K clusters. The mean of each cluster is found and the image is placed in an cluster, whose mean has the least Euclidean distance with the image feature vector. Due to the complex distribution of the image data, the k-mean clustering often cannot separate images with different concepts well enough. Clustering like regression describes the class of problem and the class of methods. Clustering methods are typically organized into two modelling approaches as Centroid-based and Hierarchical. The most popular among all is K-mean which basically comes under the category of clustering in unsupervised learning. K-mean is a type of unsupervised algorithm which solves the clustering problem. Its procedure follows a simple and easy way to classify a given data set through a certain number of clusters (take as K clusters). Data points inside a cluster are homogeneous and heterogeneous to peer groups.

Let the set data points be $x_1; x_2; \dots; x_n$ where $x_{1r}; x_{2r}; \dots; x_{nr}$ is a vector in a re-valued space $X \in R^r$ and here r is the number of attributes in the data. This algorithm partitions the input data into clusters. Each cluster with its centroid. Here k is specified by user.

CLUSTER FORMATION IN K-MEAN:

- *It picks k number of points for each cluster known as centroid.*
- *Each data point forms a cluster with the closest centroids means k clusters.*
- *Find the centroid of each cluster based on existing cluster members. Here we have new centroids.*
- *now we have new centroids, repeat 2 and 3 steps. Find the closest distance for each data point from new centroid and get related with new k-clusters. Continue repeating the process until we reach convergence i.e. centroids doesn't change.*

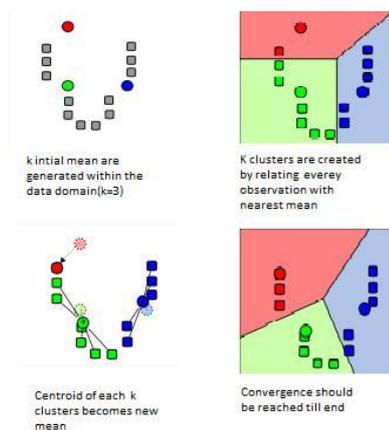


Fig. 9. K-mean clustering

DETERMINING THE VALUE OF K:

In k-means, clusters have its own centroid. Sum of square of difference between centroid and data points within a cluster constitutes within sum of square value for that cluster. When the sum of squares values for all the clusters are added, it becomes total within sum of square values for the cluster solution. As the number of cluster increases, this value keeps on decreasing but if we plot the result we can see the sum of squared distance decreases sharply up to the same value of K, after that its slows down and here we can find the optimum number of clusters.

STRENGTH OF K-MEAN:

- Easy to understand and implement.
- Efficient: Time complexity: (tkn) where n = number of data points, k = number of clusters and t = number of iterations.
- if both k and t are small it is considered as linear algorithm

WEAKNESS OF K-MEAN:

- This algorithm is only applicable if the mean is defined. The user needs to specify k .
- This algorithm is sensitive to outliers (data points that are very far away from other data points).
- Not suitable for discovering clusters that are not hyper-spheres.

REFERENCES

1. J. M. Kalyan Roy, "Image similarity measure using color histogram, color coherence vector, and sobel method," vol. Volume 2 Issue 1. International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064, January 2013.
2. A. Smola and S. Vishwanathan, INTRODUCTION TO MACHINE LEARNING. United Kingdom at the University Press, Cambridge, October 1, 2010.
3. [Online]. Available: www.analyticsvidhya.com
4. W. Gerstner, Supervised learning for neural networks: a tutorial with JAAv exercises.
5. L. Breiman, "Classification and regression trees." Belmont CA Wadsworth International group, 1984.
6. J. R. Quinlan, "Multivariate decision trees: machine learning," no. 19, 1995, pp. 45-47.
7. K. M. Dietterich, "Applying the weak learning framework to understand and improve c4.5," no. pp 96-104, san francisco:morgan
8. K. C. Chai, "Proceeding of the 13th international conference on Machine Learning, 1996.
9. H. L. C. Chai, K.; H. T. Hn, "Bayesian online classifiers for text classification and filtering." Proceedings of the 25th annual international ACM SIGIR conference on Research and Development in Information Retrieval., August 2002, pp. pp 97-104.
10. R. F. J. Hastie, Trevor; Tibshirani, "In datamining applications the interest is Often more in the class probabilities $p(x)$; $\hat{p}(x)$; $\hat{p}(x)$; themselves; rather than in performing a class assignment." T heelem
11. [Online]. Available: <http://www.statsoft.com/Textbook/Naive-Bayes-Classifier>