As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 12, Special Issue 01, pages 169-179.
ISSN: 2455-3778 online
Available online at: http://ijmtst.com/ICRCSET2026.html
DOI: https://doi.org/10.5281/zenodo.19536521

Quick Overview of Machine Learning Algorithms and Techniques for Fitting an Algorithm

K. Arun Babu, K. Manideep, Sk. Mohammad Ali, R. Tarun, S. Bharath

Department of CSE, Bapatla Engineering College, Bapatla, Andhra Pradesh, India

To Cite this Article
K. Arun Babu, K. Manideep, Sk. Mohammad Ali, R. Tarun & S. Bharath (2026). Quick Overview of Machine Learning Algorithms and Techniques for Fitting an Algorithm. International Journal for Modern Trends in Science and Technology, 12(SI01), 169-179. https://doi.org/10.5281/zenodo.19536521

Article Info
Received: 02 March 2026; Revised: 01 April 2026; Accepted: 04 April 2026.

KEYWORDS

Gradient Descent, Logistic Regression, Support Vector Machine, K Nearest Neighbor, Artificial Neural Network, Decision Tree, Back Propagation Algorithm, Bayesian Learning, Naïve Bayes.

ABSTRACT

Machine learning is predominantly a neighborhood of Artificial Intelligence which has been a key component of digitalization solutions that have caught major attention within the digital arena. During this paper author intends to try to a quick review of various machine learning algorithms which are most often used and thus, are the foremost popular ones. The author intends to highlight the merits, and demerits of the machine learning algorithms from their application perspective to assist in an informed deciding towards selecting the acceptable learning algorithm to satisfy the precise requirement of the application.

In this paper, we offer a summary of the applications of ML. We classify and survey relevant literature handling the subject, and that we also provide an introductory tutorial on ML for researchers and practitioners curious about this field. Although an honest number of research papers have recently appeared, the applications of ML remains in its infancy: to stimulate further work in this area, we conclude the paper proposing new possible research directions.

I. INTRODUCTION

A good start point for this paper are going to be to start with the elemental concept of Machine Learning. In Machine Learning a computer virus is assigned to perform some tasks and it's said that the machine has learned from its experience if its measurable performance in these tasks improves because it gains more and more experience in executing these tasks. therefore, the machine takes decisions and does predictions / forecasting supported data. Take the instance of computer virus that learns to detect / predict cancer from the medical investigation reports of a patient.

It'll improve in performance because it gathers more experience by analyzing medical investigation reports of wider population of patients. Its performance are going to be measured by the count of correct predictions and defections of cancer cases as validated by an experienced

Oncologist. Machine Learning is applied in big variety of fields namely : robotics, virtual personal assistants (like Google), computer games, pattern recognition, tongue processing, data processing, traffic prediction, online transportation network (e.g. estimating surge price in peak hour by Uber app), product recommendation, share market prediction, diagnosis, online fraud prediction, agriculture advisory, program result refining (e.g. Google search engine), BoTs (chatbots for online customer support), E-mail spam filtering, crime prediction through video closed-circuit television, social media services(face recognition on Facebook). Machine Learning generally deals of those updates also can end in noisy gradients, which can cause the error rate to leap around, rather than decreasing slowly. An example application of SGD will be to gauge with three sorts of problems namely: classification, regression and clustering. Counting on the supply of types and categories of coaching data one may have to pick from the available techniques of "supervised learning", "unsupervised learning", "semi supervised learning" and "reinforcement learning" to use the acceptable machine learning algorithm. Within the next few sections, a number of the foremost widely used machine learning algorithms are going to be reviewed.

II. OVERVIEW OF MACHINE LEARNING METHODS AND TECHNIQUES

This section provides a summary of a number of the foremost popular algorithms that are commonly classified as machine learning. The literature on ML is so extensive that even a superficial overview of all the most ML approaches goes far beyond the chances of this section, and therefore, the readers can ask variety of fundamental books on the themes [16]– [20]. However, during this section we offer a high level view of the most ML techniques that are utilized in the work we reference within the remainder of this paper

III. SUPERVISED LEARNING

Supervised learning is employed during a sort of applications, like speech recognition, spam detection and visual perception. The goal is to predict the worth of 1 or more output variables given the worth of a vector of input variables x. The output variable is often endless variable (regression problem) or a discrete variable (classification problem). A training data set comprises N samples of the input variables and the corresponding output values. Different learning methods construct a function y(x) that allows to predict the value of the output variables in correspondence to a replacement value of the inputs. Supervised learning are often weakened into two main classes, described below: parametric models, where the amount of parameters to use within the model is fixed, and non- parametric models, where their number depends on the training set. Supervised learning shows in Fig.1

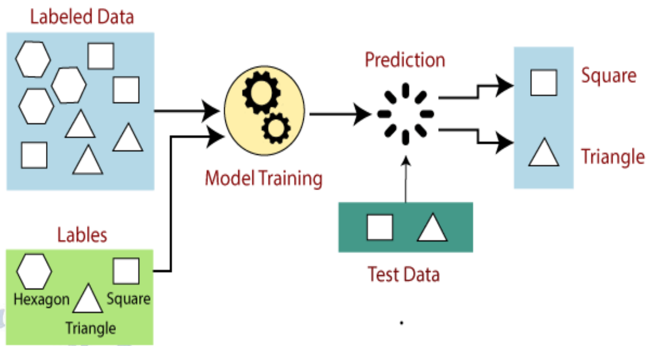


Fig.1: Supervised learning

IV. UNSUPERVISED LEARNING

Social network analysis, genes clustering and market re- search are among the foremost successful applications of un-supervised learning methods. In the case of unsupervised learning the training dataset consists only of a group of input vectors x. While unsupervised learning can address different tasks, clustering, or cluster analysis is the commonest. Clustering is the process of grouping data in order that the intra cluster similarity is high, while the inter-cluster similarity is low. The similarity is usually expressed as a distance function, which depends on the type of data. There exists a spread of clustering approaches. Here, we specialize in two algorithms, k-means and Gaussian mixture model as examples of partitioning approaches and model-based approaches, respectively, given their wide area of applicability. The reader is mentioned [27] for a comprehensive overview of cluster analysis. K-means is probably the foremost well-known clustering algorithm (see chapter X in [27]). It's an iterative algorithm starting with an initial partition of the info into k clusters. Then the center of every cluster is computed and data points are assigned to the cluster with the closest center. The procedure — center computation and data assignment — is repeated until the assignment doesn't change or a predefined maximum number of iterations is exceeded. Doing so, the algorithm may terminate at an area optimum partition. Moreover, k-means is documented to be sensitive to outliers. It's worth noting there exists ways to compute k automatically [26], and a web version of the algorithm exists. A probabilistic Gaussian Mixture Model (GMM). GMM, a linear superposition of Gaussian distributions, is one among the foremost widely used probabilistic approaches for clustering. The parameters of the model are the blending coefficient of every Gaussian component, the mean, and therefore, the covariance of every normal distribution. To maximize the log likelihood function with reference to the parameters given a dataset, the expectation maximization algorithm is employed, since no closed form solution exists during this case. The initialization of the parameters are often gone using k-means. Especially, the mean and covariance of every Gaussian component is often initialized to sample means and covariance of the cluster obtained by k-means, and therefore, the mixing coefficients are often set to the fraction of knowledge points assigned by k-means to every cluster. After initializing the parameters and evaluating the initial value of the log likelihood, the algorithm alternates between two steps. Within the expectation step, the present values of the parameters are wont to determine the “responsibility” of every component for the observed data (i.e., the contingent probability of latent variables given the dataset). The maximization step uses these responsibilities to compute a maximum likelihood estimate of the model’s parameters. Convergence is checked with reference to the log likelihood function or the parameters.

V. SEMI-SUPERVISED LEARNING

Semi-supervised learning methods are a hybrid of the previous two introduced above, and address problems in which most of the training samples are unlabeled, while only a few labeled data points are available. The obvious advantage is that in many domains a wealth of unlabeled data points is readily available. Semi-supervised learning is used for the same type of applications as supervised learning. It is particularly useful when labeled data points are not so common or too expensive to obtain and the use of available unlabeled data can improve performance.

Self-training is the oldest form of semi-supervised learning [28]. It is an iterative process; during the first stage only labeled data points are used by a supervised learning algorithm. Then, at each step, some of the unlabeled points are labeled according to the prediction resulting for the trained decision function and these points are used along with the original labeled data to retrain using the same supervised learning algorithm. This procedure is shown in Fig. 4.

Since the introduction of self-training, the idea of using labeled and unlabeled data has resulted in many semi-supervised learning algorithms. According to the classification proposed in [28], semi-supervised learning techniques can be organized in four classes: i) methods based on generative models[4]; ii) methods based on the assumption that the decision boundary should lie in a low-density region; iii) graph-based methods; iv) two-step methods (first an unsupervised learning step to change the data representation or construct a new kernel; then a supervised learning step based on the new representation or kernel).

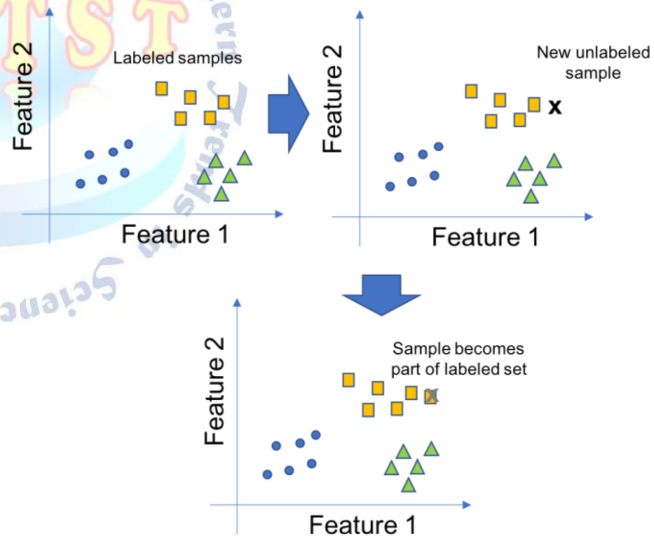
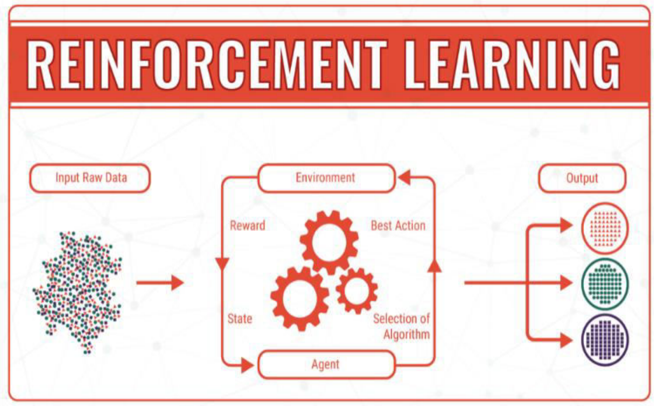


Fig. 2: Sample step of the self-training mechanism, where an unlabeled point is matched against labeled data to become part of the labeled data set.

VI. REINFORCEMENT LEARNING

Reinforcement Learning (RL) is used, in general, to address applications such as robotics, finance (investment decisions), inventory management, where the goal is to learn a policy, i.e., a mapping between states of the environment into actions to be performed, while directly interacting with the environment. The RL paradigm allows agents to learn by exploring the available actions and refining their behavior using only an evaluative feedback, referred to as the reward. The agent's goal is to maximize its long-term performance. Hence, the agent does not just take into account the immediate reward, but it evaluates the consequences of its actions on the future.



VII. OVERFITTING, UNDERFITTING AND MODEL SELECTION

In this section, we discuss a well-known problem of ML algorithms along with its solutions. Although we focus on supervised learning techniques, the discussion is also relevant for unsupervised learning methods.

Overfitting and underfitting are two sides of the same coin: model selection. Overfitting happens when the model we use is too complex for the available dataset (e.g., a high polynomial order in the case of linear regression with polynomial basis functions or a too large number of hidden neurons for a neural network). In this case, the model will fit the training data too closely[5], including noisy samples and outliers, but will result in very poor generalization, i.e., it will provide inaccurate predictions for new data points. At the other end of the spectrum, underfitting is caused by the selection of models that are not complex enough to capture important features in the data (e.g., when we use a linear model to fit quadratic data). Fig. 5 shows the difference between underfitting and overfitting, compared to an accurate model.

Since the error measured on the training samples is a poor indicator for generalization, to evaluate the model performance the available dataset is split into two, the training set and the test set. The model is trained on the training set and then evaluated using the test set. Typically, around 70% of the samples are assigned to the training set and the remaining 30% are assigned to the test set. Another option that is very useful in case of a limited dataset is to use cross-validation so that as much of the available data as possible is exploited for training. In this case, the dataset is divided into k subsets. The model is trained k times using each of the k subset for validation and the remaining (k - 1) subsets for training. The performance is averaged over the k runs. In case of overfitting, the error measured on the test set is high and the error on the training set is small. On the other hand, in the case of underfitting, both the error measured on the training set and the test set are usually high.

There are different ways to select a model that does not exhibit overfitting and underfitting. One possibility is to train a range of models, compare their performance on an independent dataset (the validation set), and then select the one with the best performance. However, the most common technique is regularization. It consists of adding an extra term - the regularization term - to the error function used in the training stage. The simplest form of the regularization term is the sum of the squares of all parameters, which is known as weight decay and drives parameters towards zero. Another common choice is the sum of the absolute values of the parameters (lasso). An additional parameter, the regularization coefficient λ, weighs the relative importance of the regularization term and the data-dependent error.

A large value of λ heavily penalizes large absolute values of the parameters. It should be noted that the data-dependent error computed over the training set increases with λ. The error computed over the validation set is high for both small and high λ values. In the first case, the regularization term has little impact potentially resulting in overfitting. In the latter case, the data-dependent error has little impact resulting in a poor model performance. A simple automatic procedure for selecting the best λ consists of training the model with a range of values for the regularization parameter and select the value that corresponds to the minimum validation error.

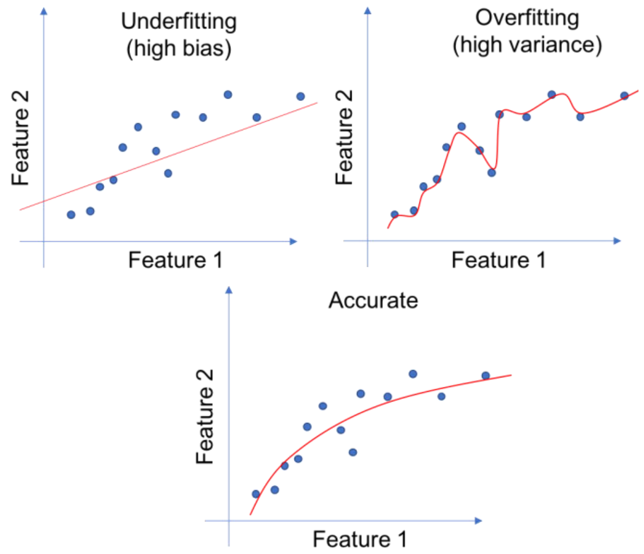


Fig.4: Difference between underfitting and overfitting

VIII. GRADIENT DESCENT ALGORITHM

Gradient Descent is an iterative method during which the target is to attenuate a price function. It should be possible to compute the partial of the function which is slope or gradient. The coefficients are computed at each iteration by taking the negative of the derivative and by reducing the coefficients at each step by a learning rate (step size) multiplied by derivative in order that the local minima are often achieved after a couple of iterations. So eventually the iterations are stopped when it converges to minimum value of the value function after which there's no further reduction in function. There are three differing types of this method: “Stochastic Gradient Descent” (SGD), “Batch Gradient Descent” (BGD), And “Mini Batch Gradient Descent” (MBGD) In BGD error is computed for each example within the training dataset, but the model is going to be updated only after the evaluation of all training examples are completed. The main advantage of BGD algorithm is computational efficiency. It produces a stable error gradient and a stable convergence. However, the algorithm has the disadvantage that the stable error gradient can sometimes end in a state of convergence that's not the simplest which the model is able to do. Also, the algorithm requires the whole training dataset to be in memory and available thereto. In SGD error is calculated for each training example within the dataset and parameters are updated for every training example. This might end in SGD to be faster than BGD, for the precise problem. SGD has the advantage that the frequent updates end in an in-depth rate of improvement. However, the frequent updates are more computationally expensive as compared to the BGD approach. The frequency performance contribution of employees to the organization which may help in creating employee incentivization scheme.

Approach of MBGD is obtained by combining the concepts of SGD and BGD. In this approach the training dataset is split into small batches and an update is performed for each of these batches. Therefore, it creates a balance between the robustness of SGD and the efficiency of BGD. This algorithm can be used to train a neural network and so this algorithm is mostly used in deep learning. The approach of Gradient Descent optimization is used in Backpropagation algorithm wherein the gradient of loss function is computed to adjust the weight of neurons.

Gradient Descent algorithm has the following disadvantage: if the learning rate for gradient descent is too fast, it is going to skip the true local minimum to optimize for time. If it is too slow, the gradient descent may never converge because it is trying really hard to find a local minimum exactly.

The learning rate can affect which minimum is reached and how quickly it is reached. A good practice is to have a changing learning rate, that slows down as the error starts to decrease.

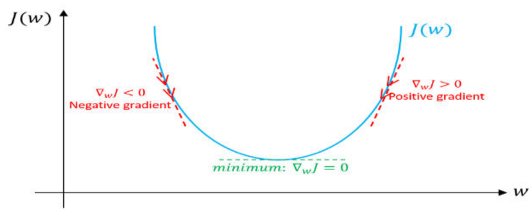


Fig:5 – Gradient descent algorithm

IX. LINEAR REGRESSION ALGORITHM

Regression is an approach of supervised learning. It can be used to model continuous variables and do the predictions. Examples of application of rectilinear regression algorithm are the following: prediction of price of real-estate, forecasting of sales, prediction of students’ exam scores, forecasting of movements within the price of stock available exchange. In Regression we've the labeled datasets and therefore the output variable value is decided by input variable values - so it's the supervised learning approach. The foremost simple sort of regression is rectilinear regression where the attempt is formed to suit a line (straight hyperplane) to the dataset and it's possible when the connection between the variables of dataset is linear.

Linear regression has the advantage that it's easy to understand and it's also easy to avoid over fitting by regularization. Also, we will use SGD to update linear models with new data. rectilinear regression may be a good fit if it's known that the connection between covariates and response variable is linear. It shifts focus from statistical modeling to data analysis and preprocessing. rectilinear regression is sweet for learning about the info analysis process. However, it's not a recommended method for many practical applications because it oversimplifies world problems.

Disadvantage of rectilinear regression is that it's not an honest fit when one must affect non-linear relationships. Handling complex patterns is difficult. Also, it's tough to feature the right polynomials appropriately within the model. Linear Regression over simplifies many world problems. The covariates and response variables usually don't have a linear relationship. Hence fitting a regression curve using OLS will give us a line with a high train RSS. In world problems there may not be relationship between mean of dependent and independent variables which linear regression expects.

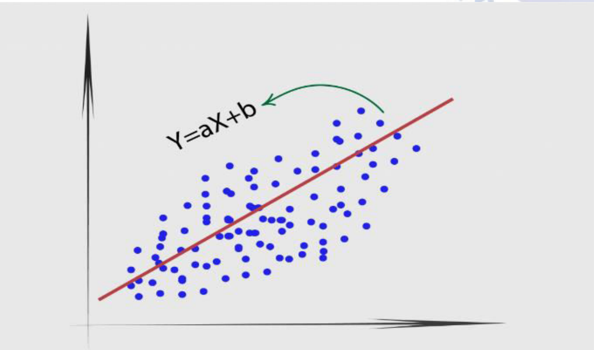


Fig:6- Linear regression algorithm

X. MULTIVARIATE REGRESSION ANALYSIS

A simple linear regression model features a variable guided by one experimental variable. However real-world problems are more complex. Generally, one dependent variable depends on multiple factors. for instance, the price of a house depends on many factors just like the neighborhood it's situated in, area of it, number of rooms, attached facilities, distance of nearest station / airport from it, distance of nearest shopping area from it, etc. In summary in simple linear regression there's a one-to-one relationship between the input variable and therefore the output variable. But in multiple linear regression, there's a many-to-one relationship, between a number of independent (input/predictor) variables and one dependent (output/response) variable. Adding more input variables doesn't mean the regression are going to be better, or will offer better predictions. Multiple and straightforward linear regression have different use cases and one isn't superior than the opposite. In some cases, adding more input variables can make things worse because it leads to over-fitting. Again, as more input variables are added it creates relationships among them. So not only are the input variables potentially associated with the output variable, they're also potentially associated with one another, this is mentioned as multicollinearity. The optimal scenario is for all of the input variables to be correlated with the output variable, but not with one another

Multivariate technique has the following merits: it gives a deep insight to the relationship between the set of independent variables and dependent variables. It also gives insight to relationship among the independent variables. This is achieved through multiple regression, tabulation techniques and partial correlation. It models the complex real-world problems in a practical and realistic way.

Multivariate technique has the following demerits: complexity of this technique is high and it requires knowledge and expertise on statistical techniques and statistical modeling. The sample size for statistical modeling needs to be high to get a higher confidence level on analysis outcome. Also, it often gets too difficult to do a meaningful analysis and interpretation of the outputs of statistical model.

This Regression Analysis technique involving multiple variables can be used in property valuation, car evaluation, forecasting electricity demand, quality control, process optimization, quality assurance, process control and medical diagnosis etc.

XI. LOGISTIC REGRESSION

Logistic regression is employed to deal a classification problem. It gives the binomial outcome because it gives the probability if an occasion will occur or not (in terms of 0 and 1) supported values of input variables. for instance, predicting if a tumor is malignant or benign or an e-mail is assessed as spam or not are the instances which may be considered as binomial outcome of Logistic Regression. There is often multinomial outcome of Logistic Regression also e.g. prediction of sort of cuisine preferred: Chinese, Italian, Mexican etc. There is often ordinal outcome also like: product rating 1 to five etc. So Logistic Regression deals with prediction of target variable which is categorical. Whereas rectilinear regression deals with prediction of values of continuous variable e.g. prediction of land price over a span of three years. Logistic Regression has the subsequent advantages: simplicity of implementation, computational efficiency, efficiency from training perspective, simple regularization. No scaling is required for input features. This algorithm is predominantly wont to solve problems of industry scale. because the output of Logistic Regression may be a probability score so to use it for solving business problem it's required to specify customized performance metrics so on obtain a cutoff which may be wont to do the classification of the target. Also, logistic regression isn't suffering from small noise within the data and multi- collinearity. Logistic Regression has the subsequent disadvantages: inability to unravel non-linear problem as its decision surface is linear, susceptible to over fitting, won't compute well unless all independent variables are identified. Some samples of application of Logistic Regression are: predicting the danger of developing a given disease, cancer diagnosis, predicting mortality of injured patients and in engineering for predicting probability of failure of a given process, system or product.

XII. DECISION TREE

Decision Tree may be a Supervised Machine Learning approach to unravel classification and regression problems by continuously splitting data supported a particular parameter. the choices are within the leaves and therefore the data is split within the nodes. In Classification Tree the choice variable is categorical (outcome within the sort of Yes/No) and in Regression tree the choice variable is continuous. Decision Tree has the subsequent advantages : it's suitable for regression also as classification problem, ease in interpretation, simple handling categorical and quantitative values, capable of filling missing values in attributes with the foremost probable value, high performance thanks to efficiency of tree traversal algorithm. Decision Tree might encounter the matter of over-fitting that Random Forest is that the solution which is predicated on ensemble modeling approach. Disadvantages of decision tree is that it are often unstable, it's going to be difficult to regulate size of tree, it's going to be susceptible to sampling error and it gives a locally optimal solution-not globally optimal solution. Decision Trees are often utilized in applications like predicting future use of library books and tumor prognosis problems.

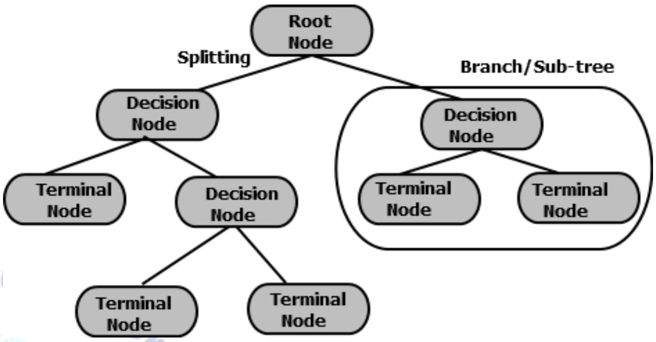


Fig. 7. Graphical representation of Decision Tree

XIII. SUPPORT VECTOR MACHINE

Support Vector Machines (SVM) can handle both classification and regression problems. In this method hyperplane needs to be defined which is the decision boundary. When there are a set of objects belonging to different classes then decision plane is needed to separate them. The objects may or may not be linearly separable in which case complex mathematical functions called kernels are needed to separate the objects which are members of different classes. SVM aims at correctly classifying the objects based on examples in the training data set. Following are the advantages of SVM : it can handle both semi structured and structured data, it can handle complex function if the appropriate kernel function can be derived. As generalization is adopted in SVM so there is less probability of over fitting. It can scale up with high dimensional data. It does not get stuck in local optima.

Following are disadvantages of SVM: its performance goes down with large data set due to the increase in the training time. It will be difficult to find appropriate kernel function. SVM does not work well when dataset is noisy. SVM does not provide probability estimates. Understanding the final SVM model is difficult. Support Vector Machine finds its practical application in cancer diagnosis, fraud detection in credit cards, handwriting recognition, face detection and text classification etc. So among the three approaches of Logistic Regression, Decision Tree and SVM the first approach to attempt will be the logistic regression approach, next the decision trees (Random Forests) can be tried to see if there is significant improvement. When the number of observations and features are high then SVM can be tried out.

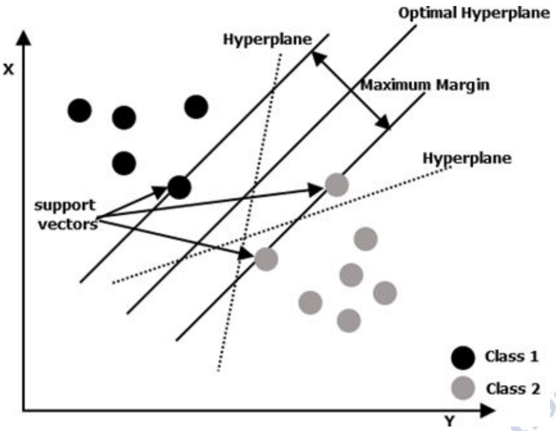


Fig. 8. Support vector machine

XIV. BAYESIAN LEARNING

In Bayesian Learning a prior probability distribution is selected and then updated to obtain a posterior distribution Later on with availability of new observations the previous posterior distribution can be used as a prior. Incomplete datasets can be handled by Bayesian network. The method can prevent over-fitting of data. There is no need to remove contradictions from data. Bayesian Learning has the following disadvantages: selection of prior is difficult. Posterior distribution can be influenced by prior to a great extent. If the prior selected is not correct it will lead to wrong predictions. It can be computationally intensive. Bayesian Learning can be used for applications like medical diagnosis and disaster victim identification etc.

NAÏVE BAYES

This algorithm is simple and is based on conditional probability. In this approach there is a probability table which is the model and through training data it is updated. The "probability table" is based on its feature values where one needs to look up the class probabilities for predicting a new observation. The basic assumption is of conditional independence and that is why it is called "naive". In real world context the assumption that all input features are independent from one another can hardly hold true.

Naïve Bayes (NB) have the following advantages: implementation is easy, gives good performance , works with less training data, scales linearly with number of predictors and data points, handles continuous and discrete data, can handle binary and multi-class classification problems, make probabilistic predictions. It handles continuous and discrete data. It is not sensitive to irrelevant features.

Naïve Bayes has the following disadvantages: Models which are trained and tuned properly often outperform NB models as they are too simple. If there is a need to have one of the feature as “continuous variable” (like time) then it is difficult to apply Naive Bayes directly, Even though one can make “buckets” for “continuous variables” it’s not 100% correct. There is no true online variant for Naive Bayes, So all data.

XV. K NEAREST NEIGHBOUR ALGORITHM

K Nearest Neighbor (KNN) Algorithm is a classification algorithm It uses a database which is having data points grouped into several classes and the algorithm tries to classify the sample data point given to it as a classification problem. KNN does not assume any underlying data distribution and so it is called non-parametric. Advantages of KNN algorithm are the following : it is simple technique that is easily implemented. Building the model is cheap. It is extremely flexible classification scheme and well suited for Multi-modal classes. Records are with multiple class labels. Error rate is at most twice that of Bayes error rate. It can sometimes be the best method. KNN outperformed SVM for protein function prediction using expression profiles.

Disadvantages of KNN are the following: classifying unknown records are relatively expensive. It requires distance computation of k-nearest neighbors. With the growth in training set size the algorithm gets computationally intensive,. Noisy / irrelevant features will result in degradation of accuracy.

It is lazy learner; it computes distance over k neighbors. It does not do any generalization on the training data and keeps all of them. It handles large data sets and hence expensive calculation. Higher dimensional data will result in decline in accuracy of regions.. KNN can be used in Recommendation system, in medical diagnosis of multiple diseases showing similar symptoms, credit rating using feature similarity, handwriting detection, analysis done by financial institutions before sanctioning loans, video recognition, forecasting votes for different political parties and image recognition.

XVI. K MEANS CLUSTERING ALGORITHM

K Means Clustering Algorithm is frequently used for solving clustering problem. It is a form of unsupervised learning. It has the following advantages: it is computationally more efficient than hierarchical clustering when variables are huge. With globular cluster and small k it produces tighter clusters than hierarchical clustering. Ease in implementation and interpretation of the clustering results are the attraction of this algorithm. Order of complexity of the algorithm is O(K*n*d) and so it is computationally efficient.

Disadvantages of K-Means Clustering Algorithm are the following: prediction of K value is hard. Performance suffers when clusters are globular. Also since different initial partitions result in different final clusters it impacts performance. Performance degrades when there is difference in the size and density in the clusters in the input data. Uniform effect often produces clusters with relatively uniform size even if the input data have different cluster size. Spherical assumption (i.e. joint distribution of features within each cluster is spherical) is hard to be satisfied as the correlation between features break it and would put extra weights on correlated features. K value is not known. It is sensitive to outliers. It is sensitive to initial points and local optimal, and there is no unique solution for a certain K value - so one needs to run K mean for a K value lots of times(20-100times) and then pick the results with lowest J.

K Means Clustering algorithm can be used for document classification, customer segmentation, rideshare data analysis, automatic clustering of IT alerts, call record details analysis and insurance fraud detection.

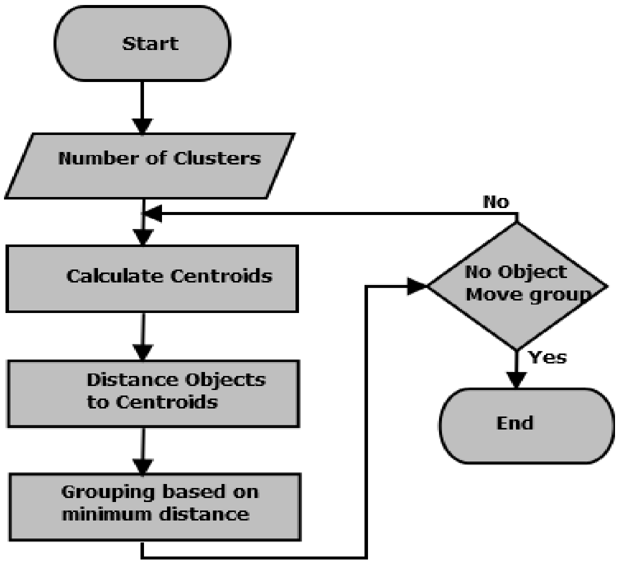


Fig. 9. Execution Flow of K-means Clustering

XVII. BACK PROPAGATION ALGORITHM

This algorithm provides a very simple and efficient way to compute the gradient in a neural network and one can use it in conjunction with stochastic gradient descent which is also quite simple. There are more complex "quasi-Newton" techniques which make a better estimate of the gradient direction and step size, but they don't perform better than backprop and SGD. Back Propagation Algorithm is used in deep learning. Neural Network (NN) has its specific applications in different industry segments and it has its merits and demerits. Scenarios where there are no well defined criteria or rules to find an answer then NN is useful. It gives the solution but it becomes difficult to explain how the solution is arrived at and so it is like a blackbox.. NN finds its application in classification of credit rating and in forecasting market dynamics in financial sector. Here are some of the applications of NN in marketing segment : in product classification, in classification of customer segments i.e. which customers will like and purchase which products, finding new market for specific product category, in associating relationship between customer and company. NN becomes instrumental in increasing revenue of a business house, in increasing the percentage of response to direct marketing. NN finds its application in Post offices for sorting the letters/parcels based on area zip code / postal code. Following are the merits of NN for which it is widely used in industry segments as mentioned above: easy adaption to new scenarios, fault tolerant, ability to handle noisy data. Shortcoming of NN are the following : training time of NN is very long and for training the NN efficiently, the sample sets need to be large. Back Propagation Algorithm encounters the Moving Target Problem which impacts its efficiency. There are number of hidden layers in the Artificial Neural Network derivative of the activation function is very small and it results in Network Paralysis. ANN with multilayer needs many repeated presentations of the input patterns, in which we need to adjust the weights so that an optimal solution is achieved with the network settling down.

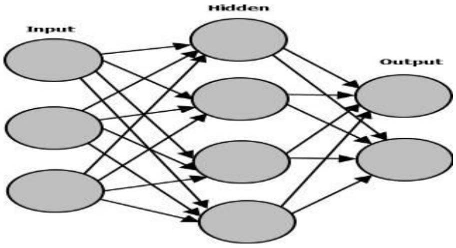


Fig. 10. Input, Hidden and Output layers in Neural Network

XVIII. Capsule Neural Networks

Capsule Neural Networks (Caps Nets) are a type of neural network architecture that aims to overcome some of the limitations of traditional neural networks, particularly in their ability to handle variations in object orientation, size, and position. The core idea of Caps Nets is to replace scalar output values of individual neurons in traditional neural networks with vectors. These vectors, known as "capsules," represent various properties of an object, such as its orientation, position, and texture. Each capsule is responsible for detecting a specific feature of the object, and the values of these capsules are combined to produce the final output of the network.

Caps Nets were first introduced in a paper by Geoffrey Hinton et al. in 2017, and they have since gained attention for their potential applications in image recognition, natural language processing, and robotics. The primary advantage of Caps Nets is their ability to handle variations in object orientation, size, and position. Traditional neural networks are limited in this respect because they treat objects as a collection of pixels, without any consideration of their spatial relationships. Caps Nets, on the other hand, can explicitly model these relationships through the use of capsules.

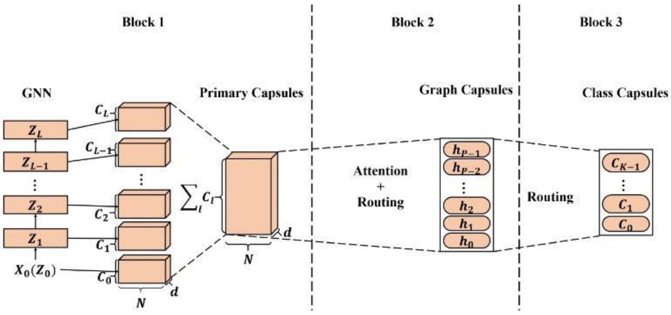


Fig. 11. Execution Flow of Capsule Neural Networks.

XIX. CONCLUSION

In this paper an endeavor was made to audit most as often as possible utilized AI calculations to explain arrangement, relapse and grouping issues. The points of interest, hindrances of these calculations have been talked about alongside examination of various calculations (at every possible opportunity) as far as execution, learning rate and so on. Alongside that, instances of down to earth uses of these calculations have been examined. Sorts of AI methods specifically managed learning, solo learning, semi regulated learning, have been talked about. It is normal that it will offer understanding to the perusers to take an educated choice in recognizing the accessible choices of AI calculations and afterward choosing the fitting AI calculation in the particular critical thinking setting.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] D. Pelleg, A. Moore (2000): "X-means: Extending K-means with Efficient Estimation of the Number of Clusters"; ICML '00 Proceedings of the Seventeenth International Conference on Machine Learning, pp. 727-734.

[2] Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya R ,"Prediction of Crop Yield using Machine Learning", International Research Journal of Engineering & Technology, Vol 5, Issue 2, Feb-2018

[3] C. Phua, V. Lee, K. Smith, R. Gayler (2010); "Comprehensive Survey of Data Mining-based Fraud Detection Research", ICICTA '10 Proceedings of the 2010 International Conference on Intelligent Computation Technology and Automation Volume 1, pp. 50-53.

[4] Sonal S. Ambalkar, S. S. Thorat2, "Bone Tumor Detection from MRI Images using Machine Learning: A Review", International Research Journal of Engineering & Technology", Vol. 5, Issue 1, Jan -2018.

[5] Dr. Rama Kishore, Taranjit Kaur, "Backpropagation Algorithm: An Artificial Neural Network Approach for Pattern

Recognition" International Journal of Scientific & Engineering Research, Volume 3, Issue 6, June-2012

[6] Kedar Potdar, Rishab Kinnerkar, "A Comparative Study of Machine Algorithms applied to Predictive Breast Cancer Data", International Journal of Science & Research, Vol. 5, Issue 9, pp. 1550-1553, September 2016.

[7] S. Selvakani and R. Rajesh, "Genetic algorithm for framing rules for intrusion detection," IJCSNS International Journal of Computer Science and Network Security, vol. 7, no. 11, pp. 285–290, 2007.

[8] P. Gupta and S. K. Shinde, "Genetic algorithm technique used to detect intrusion detection," in Advances in Computing and Information Technology. Springer, 2011, pp. 122–131.

[9] O. Depren, M. Topallar, E. Anarim, and M. K. Ciliz, "An intelligent intrusion detection system (ids) for anomaly and misuse detection in computer networks," Expert systems with Applications, vol. 29, no. 4, pp. 713–722, 2005.

[10]]A. Ahmad and L. Dey, "A k-mean clustering algorithm for mixed nu- meric and categorical data," Data & Knowledge Engineering, vol. 63, no. 2, pp. 503–527, 2007.

[11] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM computing surveys (CSUR), vol. 41, no. 3, p. 15, 2009.

[12] A. A. Aburomman and M. B. I. Reaz, "A novel svm-knn-pso ensemble method for intrusion detection system," Applied Soft Computing, vol. 38, pp. 360–372, 2016.

[13] S. Ray, "A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 35-39