



EDU
ENGINEERING
PIONEER OF ENGINEERING NOTES

**TAMIL NADU'S BEST
EDTECH PLATFORM FOR
ENGINEERING**

CONNECT WITH US



WEBSITE: www.eduengineering.net



TELEGRAM: [@eduengineering](https://t.me/eduengineering)



INSTAGRAM: [@eduengineering](https://www.instagram.com/eduengineering)

- Regular Updates for all Semesters
- All Department Notes AVAILABLE
- Handwritten Notes AVAILABLE
- Past Year Question Papers AVAILABLE
- Subject wise Question Banks AVAILABLE
- Important Questions for Semesters AVAILABLE
- Various Author Books AVAILABLE

UNIT - II

①

Ensemble Techniques and Unsupervised Learning.

Combining Multiple Learners

* When designing a learning machine, we generally make some choices like parameters of machine, training data and representation. This implies some sort of variance in performance.

* For example in a classification setting we can use a parametric classifier or in a multilayer perceptron, we should also decide on the number of hidden units.

* Each learning algorithm dictates a certain model that comes with a set of assumptions.

* This inductive bias leads to errors if the assumptions do not hold for the data.

* Different learning algorithms have different accuracies. The no free lunch theorem asserts that no single learning algorithm always achieves the best performance in any domain.

* They can be combined to attain high accuracy. ②

* Data fusion is the process of fusing multiple records representing the same real world object into a single, consistent and clean representation.

* Fusion of data for improving prediction accuracy and reliability is an important problem in machine learning.

* Combining different models is done to improve the performance of deep learning models.

* Building a new model by combination requires less time, data and computational resources.

* The most common method to combine models is by averaging multiple models, where taking a weighted average improves the accuracy.

Generating Diverse Learners:

Different Algorithms: We can use different learning algorithms to train different base-learners. It makes different assumptions about the data and lead to different classifiers.

Different Hyper-parameters :

We can use the same learning algorithm but use it with different hyper-parameters.

Different Input Representations :

Different representations make different characteristics explicit allowing better identification.

Different Training sets :

Another possibility is to train different base-learners by different subsets of the training set.

Model Combination Schemes

Different methods are used for generating final output for multiple base learners are multiexpert and multistage combination.

Multiexpert Combination :

- * It is a method have base-learners that work in parallel.

- * Global approach : given an input all base learners generate an output and these outputs

are used such as voting and stacking. ④

local approach: in mixture of experts there is a gating model, which looks at the input and choose one (or very few) of the learners as responsible for generating the output.

Multistage Combination:

* It is a method use a serial approach where the next multistage combination base-learners are not accurate enough.

* Let's assume that we want to construct a function that maps inputs to outputs from a set of known N_{train} input-output pairs

$$D_{train} = \{(x_i, y_i)\}_{i=1}^{N_{train}}$$

Where $x_i \in x$ is a D dimensional feature input vector, $y_i \in y$ is the output.

Classification: When the output takes values in the discrete set class labels $y = \{c_1, c_2, \dots, c_K\}$

Where K is the number of different classes.

Regression consists in predicting continuous ordered outputs $y \in R$

Voting:

* The simplest way to combine multiple classifiers is by voting which corresponds to taking a linear combination of the learners.

* Voting is an ensemble machine learning algorithm.

* For regression, a voting ensemble involves making a prediction that is the average of multiple other regression models.

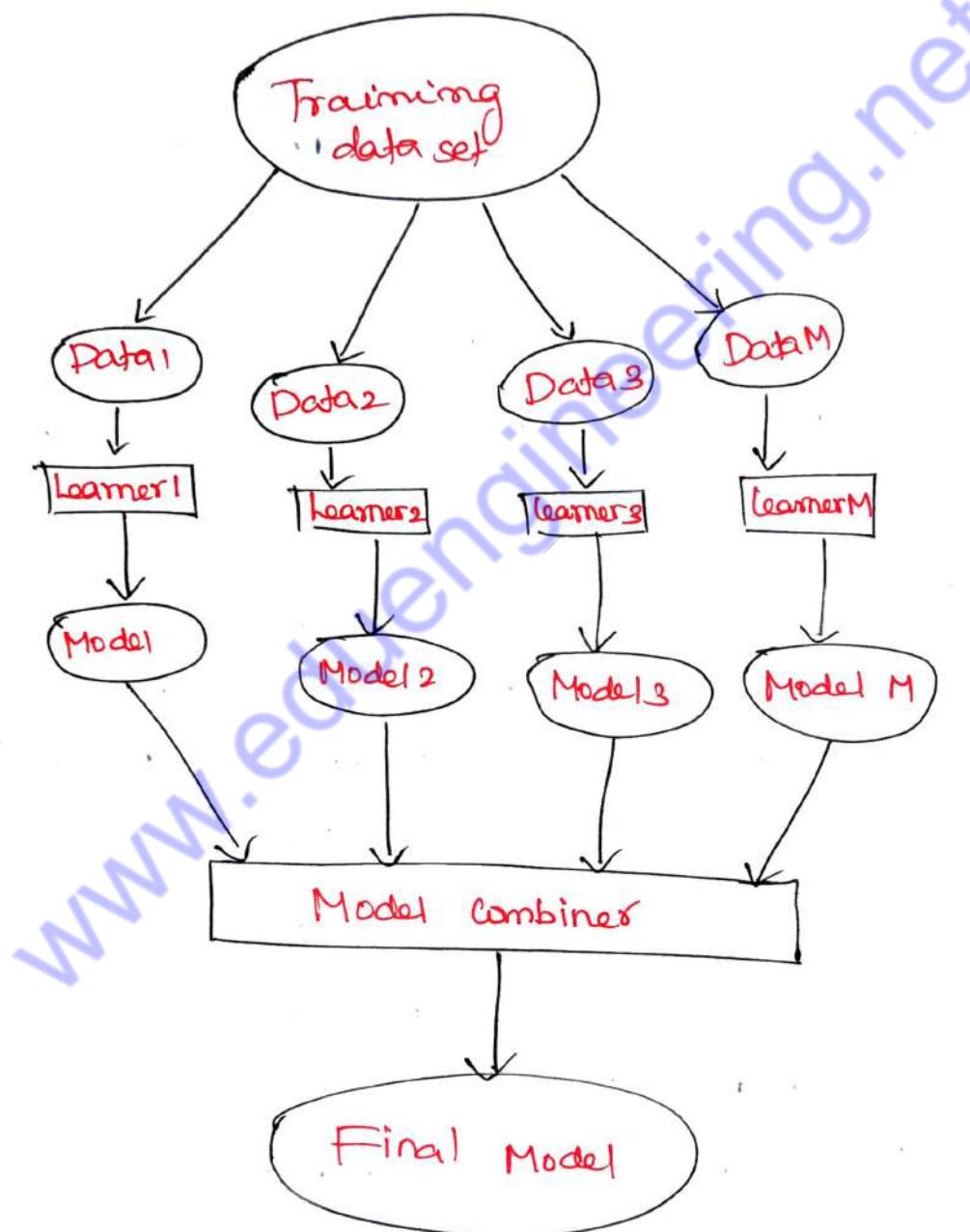
* In classification a hard voting ensemble involves summing the votes for crisp class labels from other models and predicting the class with the most votes.

* A soft voting ensemble involves summing the predicted probabilities for class states and predicting the class label with the largest sum probability.

In this method, the first step is to create multiple classification/regression models using some training data set.

(6)

Each base model can be created using different splits of the same training dataset and same algorithm or using the same dataset with different algorithms or any other method.



When combining multiple independent and diverse decisions each of which is at least more accurate than random guessing, random errors cancel each other out and correct decisions are reinforced. 7

Human ensembles are demonstrably better

Use a single, arbitrary learning algorithm but manipulate training data to make it learn multiple models.

Error-Correcting Output Codes

In error Correcting Output Codes main classification task is defined in terms of a number of subtasks that are implemented by all base learners

The idea is that the original task of Separating one class from all other classes may be a difficult problem.

So we want to define a set of simpler classification problems, each specializing in one aspect of the task and combining these simpler classification

We get final classifier

Base learners are binary classifiers having output -1 +1 and there is a code matrix N of $K \times L$ whose

K rows are the binary codes of classes in terms of the L base-learners dj.

Ensemble Learning

- * The idea of ensemble learning is to employ multiple learners and combine their predictions.

- * If we have a committee of M models with uncorrelated errors, simply by averaging them the average error of a model can be reduced by a factor of M

- * Unfortunately, the key assumption that the errors due to the individual models are uncorrelated is unrealistic, in practice, the errors are typically high correlated, so the reduction in overall error is generally small

- * Ensemble modelling is the process of running two or more related but different analytical models and then synthesizing the results into a single score or spread in order to improve the accuracy of predictive analytics

* Ensembles of classifiers is a set of classifiers whose individual decisions combined in some way to classify new examples.

* Ensemble methods combine several decision tree classifiers to produce better predictive performance than a single decision tree classifier.

* The main principle behind the ensemble model is that a group of weak learners come together to form a strong learner thus increasing the accuracy of the model.

ENSEMBLE METHODS WORKING

VARIANCE REDUCTION:

If the training sets are completely independent, it will always help to average an ensemble because this will reduce variance without affecting bias (eg, bagging) and reduce sensitivity to individual data points.

BIAS REDUCTION:

For simple methods, average of models has much greater capacity than single model. Averaging models can reduce bias substantially by increasing capacity and control variance by cutting one component at a time.

BAGGING:

* Bagging is also called Bootstrap aggregating.

Bagging and boosting are meta-algorithms that pool decision from multiple classifiers.

* It creates ensembles by repeatedly randomly resampling the training data

* Bagging was the first effective method of ensemble learning and is one of the simplest method of arching.

* The meta-algorithm which is a special case of the model averaging, was originally designed for classification and usually applied to decision tree models, but it can be used with any type of model for classification or regression.

* Ensemble classifiers such as bagging, boosting and model averaging are known to have improved accuracy and robustness over a single model

* Although unsupervised model, such as clustering do not directly generate label prediction for each individual they provide useful constraints for the joint prediction of a set of related objects

* For a given training set of size n create m samples of size n by drawing n examples from the original data with replacement

* Each bootstrap sample will on average contain 63.2% of the unique training examples, the rest are replicates.

* It combines the m resulting models using simple majority vote.

* In particular on each round, the base learner is trained on what is often called a bootstrap replicate of the original data set.

* Suppose training set consists of n examples.

* Then a bootstrap replicate is a new training set that also consists of n examples and which is formed by repeatedly selecting uniformly at random and with replacement n examples from the original training set.

* This means that the same example may appear multiple times in the bootstrap replicate or it may appear not at all.

* It also decreases error by decreasing the variances in the result due to unstable learner algorithms (like decision tree) whose output can change dramatically when the training data is slightly changed.

Bagging Steps:

* Suppose there are N observations and M features in training dataset

* A sample from training data set is taken randomly with replacement.

* A subset of M features is selected randomly and whichever feature gives the best split is used to split the node iteratively.

* The tree is grown to the largest. Above steps are repeated n times and prediction is given based on the aggregation of predictions from n number of trees.

Advantages of Bagging:

* Reduces over fitting of the model.

* Handles higher dimensionality data very well.

* Maintains accuracy of missing data.

Disadvantages of Bagging

- * Since final prediction is based on the mean predictions from subset

BOOSTING

- * Boosting is a very different method to generate multiple predictions (function and estimates) combine them linearly.

* Boosting refers to a general and provably effective method of producing a very accurate classifier by combining rough and moderately inaccurate rules of thumb

* Originally developed by computational learning theorists to guarantee performance improvements on fitting training data for a weak learner that only needs to generate a hypothesis with a training accuracy greater than 0.5

* Final result is the weighted sum of the results of weak classifiers.

* A learner is weak if it produces a classifier that is only slightly better than random guessing, while a learner

is said to be strong if produces a classifier that achieves a low error with high confidence for a given concept.

* Revised to be a practical Algorithm

Adaboost for building ensembles that empirically improves generalization performance. Examples are given weight at each iteration a new hypothesis is learned and the examples are reweighted to focus the system.

* Boosting is a bias reduction technique

It typically improves the performance of a single tree model.

* A reason for this is that we often cannot construct trees which are sufficiently large due to thinning out of observations in the terminal nodes.

* Boosting is then a device to come up with a more complex solution by taking linear combination of trees.

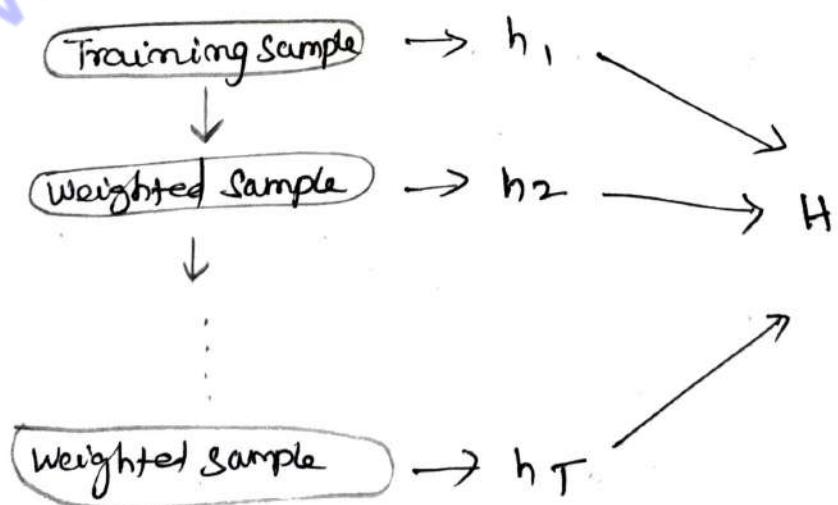
* In presence of high dimensional predictors boosting is also very useful as a regularization technique for additive or interaction modeling.

* To begin we define an algorithm for finding the rules of thumb, which we call a weak learner.

* The boosting algorithm repeatedly calls this weak learner, each time feeding it a different distribution over the training data.

* Each call generates a weak classifier and we must combine all of these into a single classifier that hopefully is much more accurate than any one of the rules.

* Training a set of weak hypothesis h_1, \dots, h_T . The combined hypothesis H is a weighted majority vote of the T weak hypotheses. During the training focus on the examples that are misclassified



Ada Boost:

* Ada Boost short for "Adaptive Boosting", is a machine learning meta-algorithm formulated by Yoav Freund and Robert Schapire who won the prestigious "Godel prize" in 2003 for their work.

* It can be used in conjunction with many other types of learning algorithms to improve their performance

* It can be used to learn weak classifiers and final classification based on weighted vote of weak classifiers.

* It is linear classifier with all its desirable properties. It has good generalization properties

* To use the weak learner to form a highly accurate prediction rule by calling the weak learner repeatedly on different distributions over the training examples

Initially all weights are set equally but each round the weights of incorrectly classified examples are increased so that those observations

that the previously classifier poorly etc predicts
~~will~~ receive greater weight on the next iteration.

Advantages of AdaBoost

- * Very Simple to implement
- * Fairly good generalization
- * The prior errors need not be known ahead of time.

Disadvantages of AdaBoost

- * Suboptimal Solution
- * Can over fit in presence of noise

Boosting Steps :

- * Draw a random subset of training samples d_1 without replacement from the training set D to train a weak learner c_1
- * Draw second random training subset d_2 without replacement from the training set and add 50 percent of the samples that were previously falsely classified / misclassified to train a weak learner c_2
- * Find the training samples d_3 in the training set D on which c_1 and c_2 disagree to train a third weak learner c_3

- * Combine all the weak learners via majority voting

Advantage of Boosting:

- * Supports different loss function
- * Works well with interactions

Disadvantages of Boosting:

- * Prone to over fitting
- * Requires careful tuning of different hyper-parameters

STACKING:

- * Stacking Sometimes called stacked generalization is an ensemble machine learning method that combines multiple heterogeneous base or component models via a meta model
 - * The base model is trained on the complete training data and then the meta-model is trained on the predictions of the base models.
 - * The advantages of stacking is the ability to explore the solution space with different models in the same problem

* The stacking based model can be visualized in levels and has at least two levels of the models.

* The first level typically trains the two or more base learners (can be heterogeneous) and the second level might be single meta learner that utilizes the base models predictions as input and gives the final result as output

A stacked model can have more than more than two such levels but increasing the levels does not always guarantee better performance.

Stacking is concerned with multiple classifiers generated by different learning algorithms L_1, \dots, L_N on a single dataset S , which is composed by a feature vector

$$S_i = (x_i, t_i)$$

The stacking process can be broken into two phases:

Generate a set of base-level classifiers C_1, \dots, C_N where $C_i = L_i(S)$

Train a meta-level classifier to combine

the output of the base-level classifiers (20)

Training set

hypotheses

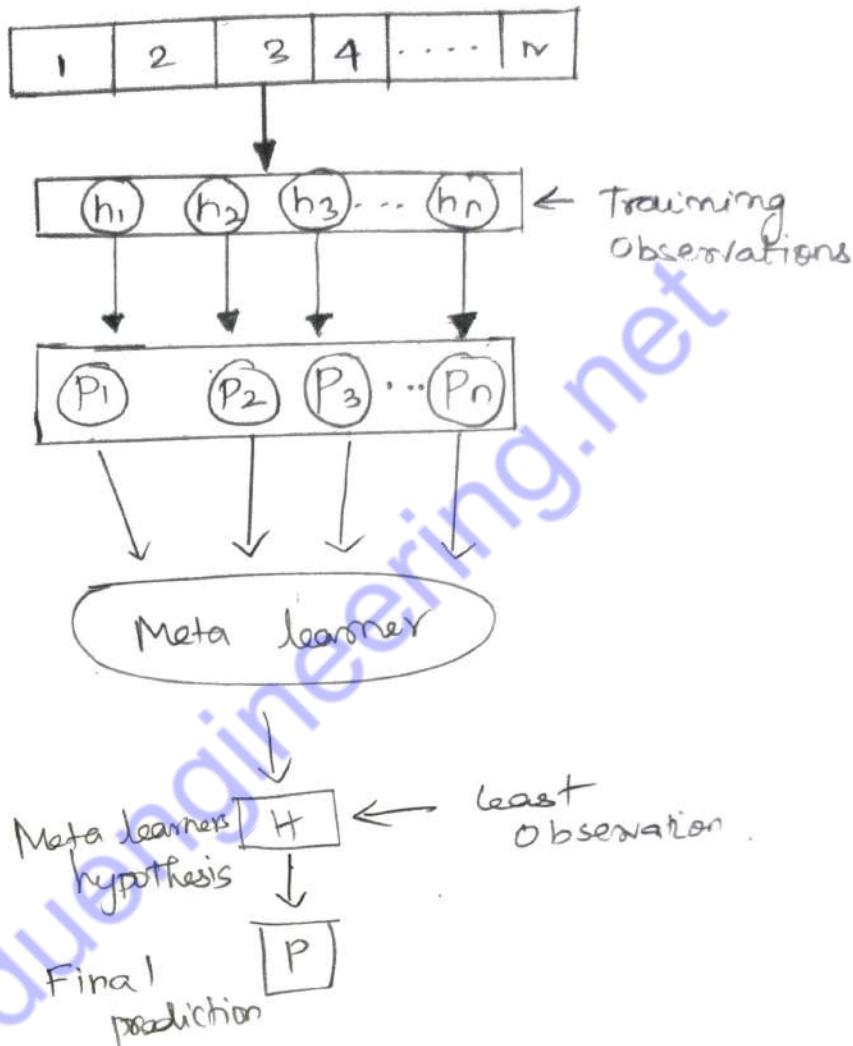


fig: Stacking frame

Based on two basic observations

Variance reduction:

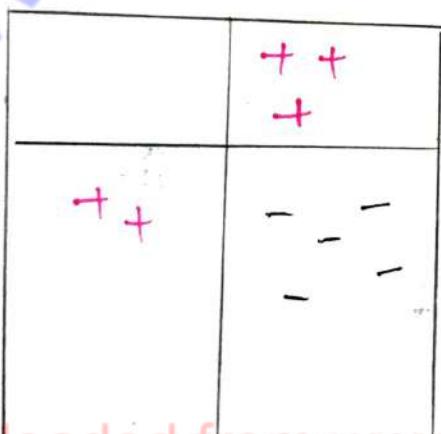
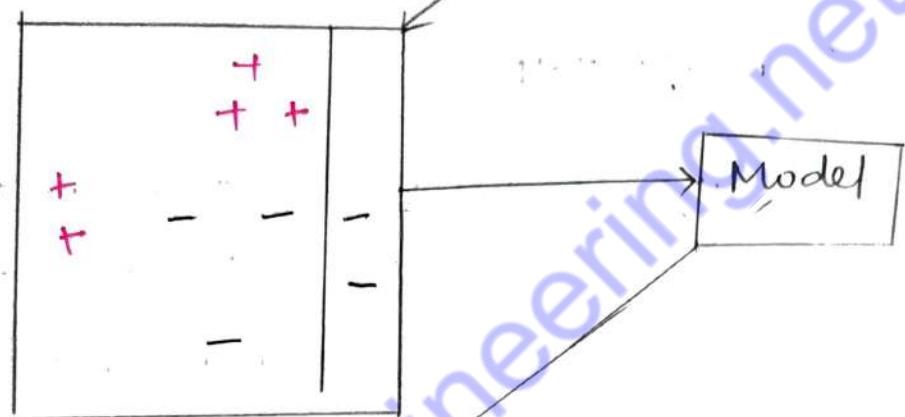
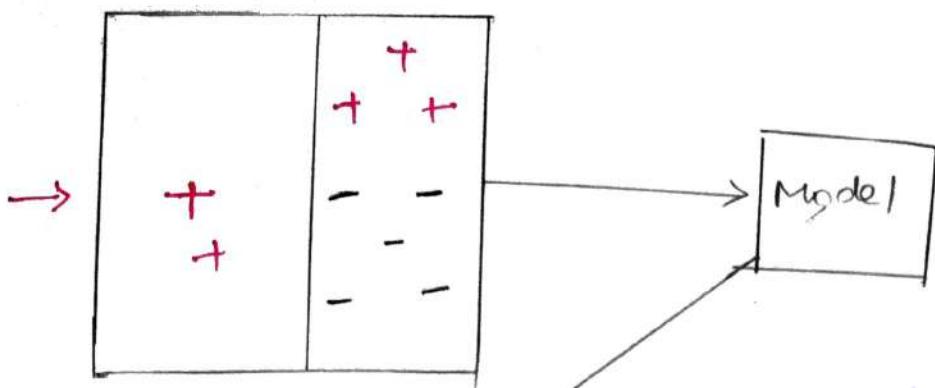
If the training sets are completely independent it will always help to average an ensemble because this will

Reduction

Downloaded from www.eduengineering.net

ADABOOST ALGORITHM:

- * Ada Boost algorithm short for adaptive algorithm. It is a Boosting technique used as an ensemble method in machine learning.
- * It is called adaptive boosting as the weights are reassigned to incorrectly classified instances.
- * Boosting is used to reduce bias as well as variance for supervised learning.
- * It works on the principle of learners growing sequentially
 Except for the first each subsequent learner is grown previously grown learners.
 In simple words, weak learners are converted into strong ones. The AdaBoost algorithm works on the same principle as boosting with a slight difference in detail.



* The maximal not usual algorithm used with AdaBoost is selection trees with one stage meaning with decision trees with most effective one split.

* Those trees are also referred to as decision stumps

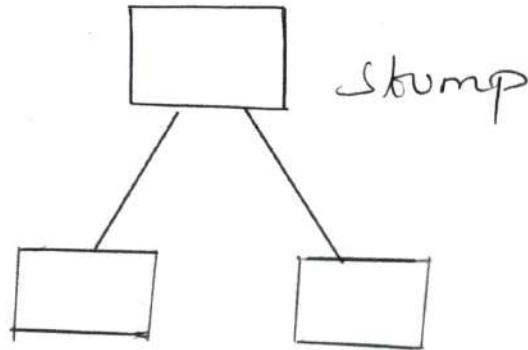
The working of Ada boost version follows the beneath referred to as decision stumps or path :

- * Creation of the base learner
- * Calculation of the total error via the beneath formulation
- * Calculation of performance of the decision stumps
- * Updating the weights in line with the misclassified factors.

Creation of new database :

AdaBoost ensemble :

In the ensemble approach we upload the Susceptible fashion sequentially and then teach them the use of weighted Schooling records.



We hold to iterate the process till we gain the advent of a pre-set range of vulnerable learners or we can not look at further improvement at the data set ..

At the end of Algorithm we are left with some vulnerable learners with a stage fee.

Difference between bagging and boosting

Bagging

It is a technique that builds multiple homogeneous models from different subsamples of the same training dataset to obtain more accurate predictions.

Boosting

It refers to a group of algorithms that utilize weighted averages to make weak learning algorithms stronger learning algorithms.

- * It helps in reducing variance
- * Every model receives an equal weight
- * It helps in reducing bias and variance
- * Models are weighted by their performance

CLUSTERING

* Give an set of objects, place them in group such that the objects in a group are similar (or related) to one another and different from (or unrelated to) the objects in other group

* Cluster analysis can be a powerful data mining tool for any organization that needs to identify discrete groups of customers, sale transactions, or other types of behaviours and things.

* For example, insurance providers use cluster analysis to detect fraudulent claims and banks used it for credit scoring

* Cluster analysis uses mathematical tool models to discover group of similar customers based on the smallest variation among customers within

each group

* Cluster is a group of objects that belong to the same class. In another words the similar object are grouped in one cluster and dissimilar grouped in other cluster.

* Clustering is a process of partitioning a set of data in a set of meaningful

Subclasses

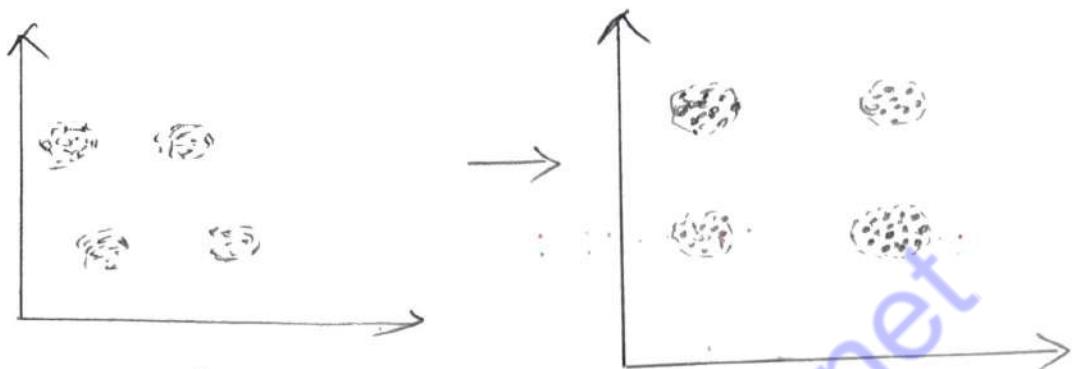
* Every data in the sub class shares a common trait. It helps a user understand the natural grouping or structure in the data set.

* Various types of clustering methods are partitioning methods, hierarchical, clustering, fuzzy clustering, density based clustering and model based clustering.

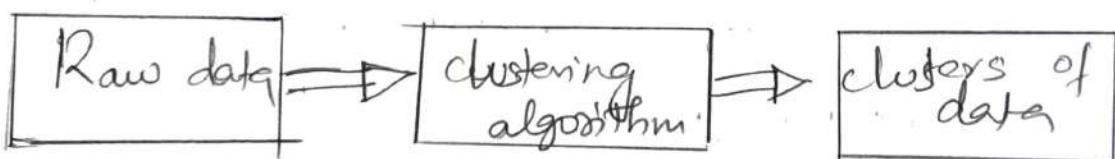
* Cluster analysis is process of grouping a set of data objects into clusters.

27

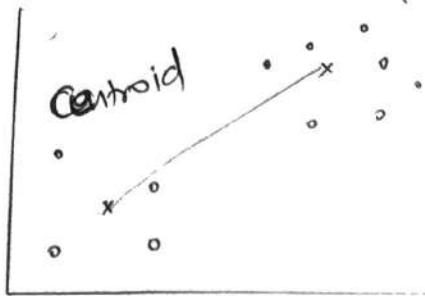
Desirable properties of a clustering algorithm
are as follows



- * Scalability (in terms of both time and space)
 - * Ability to deal with different data types
- Minimal requirements for domain knowledge to determine input parameters
- * Interpretability and usability
- Clustering of data is a method by which large sets of data are grouped into clusters of smaller sets of similar data.
- * Clustering can be considered the most important unsupervised learning problem.



- * Clustering means grouping of data or dividing a large set into smaller data sets of some similarity



CLUSTER CENTROID:

* The centroid of a cluster is a point whose parameter values are the means of the parameter values of all the points in the cluster.

* Each cluster has a well defined centroid.

DISTANCE:

The distance between two point is taken as common metric to see as the similarity among the components of population

The commonly used distance measure is the euclidean metric which defines the distance between two points

$p = (p_1, p_2, \dots)$ and $q = (q_1, q_2, \dots)$ is given by

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

* The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. But how do decide what constitutes a good clustering?

* It can be shown that there is no absolute best criterion which should be independent of the final aim of clustering. Clustering algorithm can be classified as listed below:

- ⇒ Exclusive clustering
- ⇒ Overlapping clustering
- ⇒ Hierarchical clustering
- ⇒ Probabilistic clustering

A good clustering method will produce high quality clusters intra-class similarity and low intra class similarity.

The major clustering techniques are

- * Partitioning methods
- * Hierarchical methods
- * Density Method

UNSUPERVISED K-MEANS CLUSTERING

(30)

* K-Means clustering is heuristic method. Here each cluster is represented by the center of the cluster.

* K stands for number of clusters, It is typically a user input to the algorithm some criteria can be used automatically estimate K.

* This method initially takes the number of components of the population equal to the final required number of clusters

* In this step itself the final required number of clusters is chosen such that the points are mutually farthest apart.

* Given k-means algorithm consists of four steps:

⇒ Select initial centroids at random

⇒ Assign each object to the cluster with nearest centroid

⇒ Compute each centroid as the mean of the objects assigned to it.

⇒ Repeats previous 2 steps until no change (3)

* The x_1, \dots, x_N are data points or vector of observations

Each observation (vector x_i) will be assigned to one and only one cluster. The c_i denotes cluster number for the i th observation. K-means minimizes within-cluster point scatter

$$W(C) = \frac{1}{2} \sum_{k=1}^K \sum_{c(i)=k} \sum_{c(j)=k} \|x_i - x_j\|^2 \\ = \sum_{k=1}^K N_k \sum_{c(i)=k} \|x_i - m_k\|^2$$

Where
 m_k is the mean vector of the k^{th} cluster
 N_k is the number of observations in k^{th} cluster.

K-Means Algorithm properties

- * There are always K clusters
- * There is always at least one item in each cluster
- * The clusters are non hierarchical and they do not overlap

Every member of cluster is closer to

its cluster than any other cluster (82)
because closeness does not always involve
the centers of clusters

K-Means Algorithm

1) The dataset is partitioned into K clusters and the data points are randomly assigned to the clusters that have roughly the same number of data points.

2) For each data point

* calculate the distance from the data point to each cluster.

* If the data point is closest to its own cluster, leave it where it is

* If the data point is not closest to its own cluster, move it into the closest clusters

Repeat the above step until a complete pass through all data points result in no data point moving from one cluster to another

* K Means algorithm is iterative on nature. It converges however only a local minimum is obtained. If works only for numerical data. This method is easy to implement.

Advantages of K-Means Algorithm.

- * Efficient in computation
- * Easy to Implement

Weakness

- * Applicable only when mean is defined
- * Need to specify K the number of clusters in advance.
- * Trouble with noisy data & outliers
- * Not suitable to discover clusters with non-convex shapes

KNN:

K-nearest Neighbour is one of the Machine learning algorithms based totally on Supervised learning approach.

K-NN algorithm assumes the similarity between the brand new case/facts and available instances and placed the brand new case into the category that is maximum similar to the to be had classes.

K-NN set of rules shops all of the be had facts and classifies a new statistics point based at the similarity.

This means when new data seems then it may be effortlessly categorized into a properly suite class by using K-NN algorithm

K-NN set of rules can be used for regression as well as for classification however normally its miles used for the classification troubles

KNN is a non parametric algorithm because of this it does no longer makes any assumption on underlying data

It is also referred to as a lazy learner set of rules because it does not longer research

(35)

research from the training set immediately as a substitute it drops the dataset and at the time of class it plays an movement of the dataset

The KNN set of rules at the Schooling section simply stores the dataset and when it gets new data then it classifies that statistics into a class that is an awful lot similar to the brand new data.

Example :

Suppose we have a picture of creature that looks much like cat and dog but we want both it is a cat or dog. So far this identity we are able to use the KNN algorithm, because it works on a similarity degree. Our KNN version will discover the similar features of the new facts set to the cats and dogs snap shots and primarily based on the most similar functions it will place it in both cat or canine class.

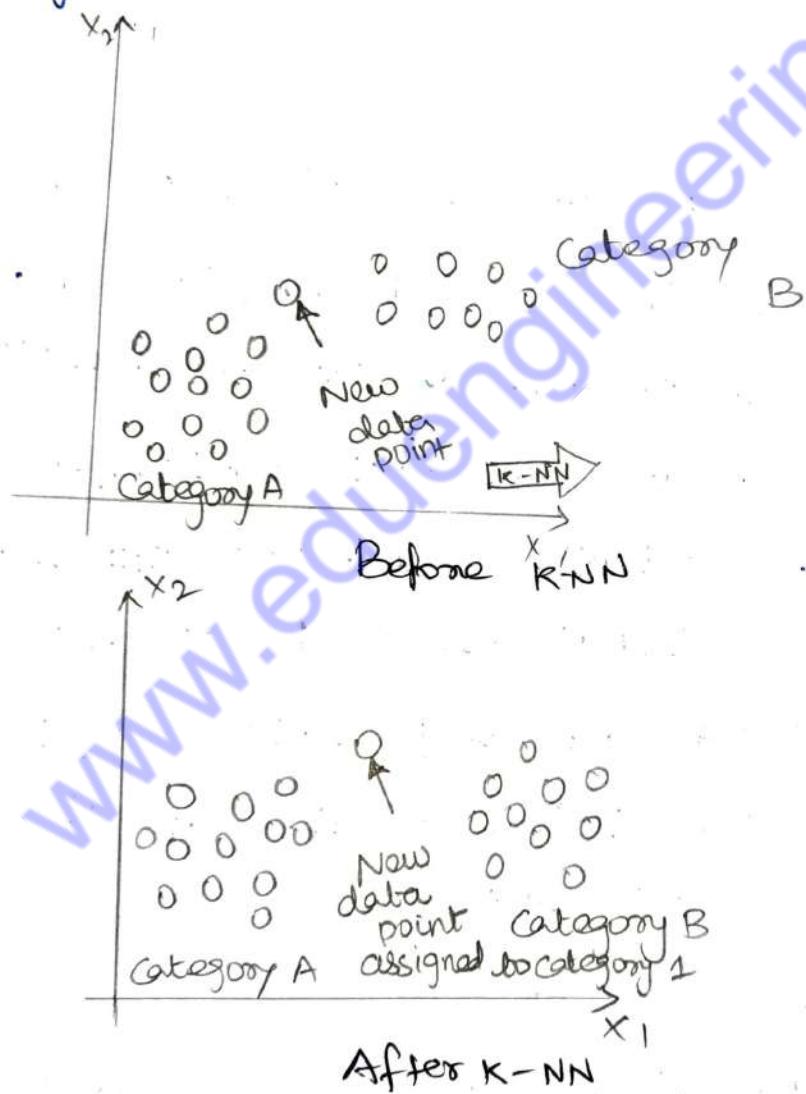
Why do we need KNN ?

Suppose there are two categories i.e. category A and category B and we have a brand new statistics point x_1 and so this

fact point will lie within of these classes.

To solve this sort of problem we need a K-NN set of rules.

With the help of K-NN we will without difficulty discover the category or class of a selected dataset. Consider the underneath diagram.

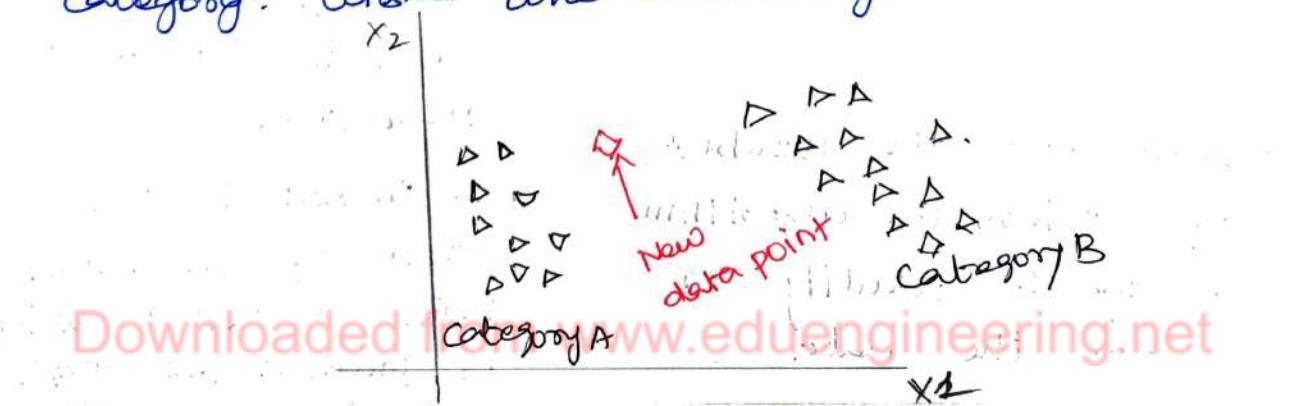


KNN Working :

(37)

The KNN Working can be explained on the basis of the below algorithm:

- 1 \Rightarrow select the wide variety K of the acquaintances
 - 2 \Rightarrow calculate the Euclidean distance of K variety of friends.
 - 3 \Rightarrow Take the K nearest neighbor's as according to the calculated Euclidean distance
 - 4 \Rightarrow Among these ok pals, count number of the data points in each class.
 - 5 \Rightarrow Assign the brand new record points to that category for which quantity of the neighbor is maximum.
 - 6 \Rightarrow Our model is ready
- Suppose we have got a brand new information point and we want to place it in the required category. Consider the under image.



Firstly we are able to pick the number of friends so we are able to select the $OK = 5$.

Next we will calculate the Euclidean distance between the fact points. The Euclidean distance is the gap between points which we have got already studied in geometry. It may be calculated as

$$\text{Euclidean distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Difference between K means and KNN

K Means

- * K Means is an unsupervised machine learning algorithm used for clustering.
- * K-means is an eager learner.
- * It is used for clustering.
- * K-means is the number of clusters the algorithm is try to identify or learn the data.

KNN

- * KNN is a supervised machine learning algorithm used for classification.
- * KNN is a lazy learner.
- * It is used for classification and sometimes even for regression.
- * K in KNN is the number of the nearest neighbours used to classify or predict a test sample.

* K Means require unlabelled data. It gathers and groups data into K number of clusters.

* KNN require labelled data and will give new data points accordingly to the K number or the closest data points.

Gaussian Mixture Models:

* Gaussian Mixture models is a soft clustering algorithm where each point probabilistically belong to all clusters. This is different than k means where each point belong to one clusters.

* The gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mix of gaussian distributions with unknown parameters.

* Gaussian mixture models consists of two parts : Mean vectors and Covariance matrices.

* A gaussian distribution is defined as a continuous probability distribution that takes a bell shaped curve. Another name of the gaussian distribution is the normal distribution.

In one dimensional space the probability density function of a gaussian distribution is given

by.

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where μ is the mean and σ^2 is the variance.

- * Gaussian mixture models can be used for a variety of use cases, including identifying customer segments, detecting fraudulent activity and clustering images.
- * GMM have variety of real world applications. They are
 - * Used for signal processing
 - * Used for customer churn analysis
 - * Used for language identification
 - * Used in video game industry
 - * Genre classification of songs

Expectation-Maximization

* In Gaussian mixture models an expectation maximization method is a powerful tool for estimating the parameters of Gaussian mixture model. This expectation is termed as E and maximization is termed M.

and maximizing (M) step which computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found in the E Step.

* In the Expectation Step, find the expected values of the latent variables (here you need to use the current parameter values)

* In the Maximization Step first plug in the expected values of the latent variables in the log-likelihood of the augmented data. Then maximize this log-likelihood to reevaluate the parameters.

* EM is a technique used in point estimation. Given a set of observable variables x and unknown (latent) variables z we want to estimate parameters θ in a model. It is a widely used maximization likelihood estimation procedure for the statistical models when the values of some of the variables in the model are not observed.

* In E step, the algorithm estimates the posterior distribution of the hidden variables q given the observed data x and the current parameter settings and in the M step algorithm calculates

* Expectation is used to find the gaussian parameters which are used to represent each component of gaussian mixture models. Maximization is termed M and it is involved in determining whether new data points can be added or not.

* This algorithm used in maximum likelihood estimation where the problem involves two sets of random variables of which one x is observable and the other z is hidden.

* The goal of the algorithm is to find the parameter vector ϕ that maximizes the likelihood of the observed values of x $L(\phi|x)$

EM Algorithm:

* It is an iterative method used to find maximum likelihood estimates of parameters in probabilistic models where the models depend on unobserved also called as latent variables

* EM alternate between performing an expectation (E) step which computes an expectation of the likelihood by including the latent variables as if they were observed

- * The ML parameter settings with q fixed
- * At the end of each iteration the lower bound on the likelihood is optimized for the given parameter setting (M-step) and the likelihood function is set to that bound E step.
- * Generally EM works best when the fraction of missing information is small and the dimensionality of the data is not too large.
- * EM require many iterations and higher dimensionality can dramatically slow down the E Step.

EM is useful for several reasons:

- * conceptual simplicity
- * ease of implementation

Sometimes the M-step is a constrained maximization which means that there are constraints on valid solutions not encoded in the function itself

- * Expectation maximization is an effective technique that is often used in data analysis

44

to manage missing data. Indeed expectation maximization overcomes some of the limitations of other techniques, such as mean Substitution or regression Substitution.

If the alternative techniques generate biased estimates and specifically underestimate the standard errors. Expectation maximization overcome this problem.



EDU
ENGINEERING
PIONEER OF ENGINEERING NOTES

**TAMIL NADU'S BEST
EDTECH PLATFORM FOR
ENGINEERING**

CONNECT WITH US



WEBSITE: www.eduengineering.net



TELEGRAM: [@eduengineering](https://t.me/eduengineering)



INSTAGRAM: [@eduengineering](https://www.instagram.com/eduengineering)

- Regular Updates for all Semesters
- All Department Notes AVAILABLE
- Handwritten Notes AVAILABLE
- Past Year Question Papers AVAILABLE
- Subject wise Question Banks AVAILABLE
- Important Questions for Semesters AVAILABLE
- Various Author Books AVAILABLE