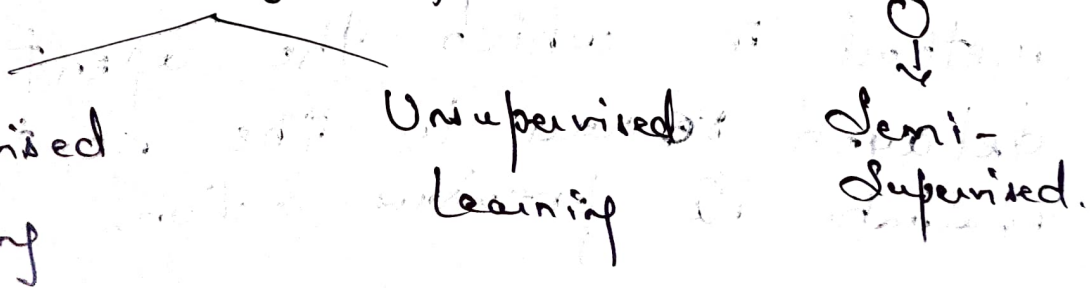
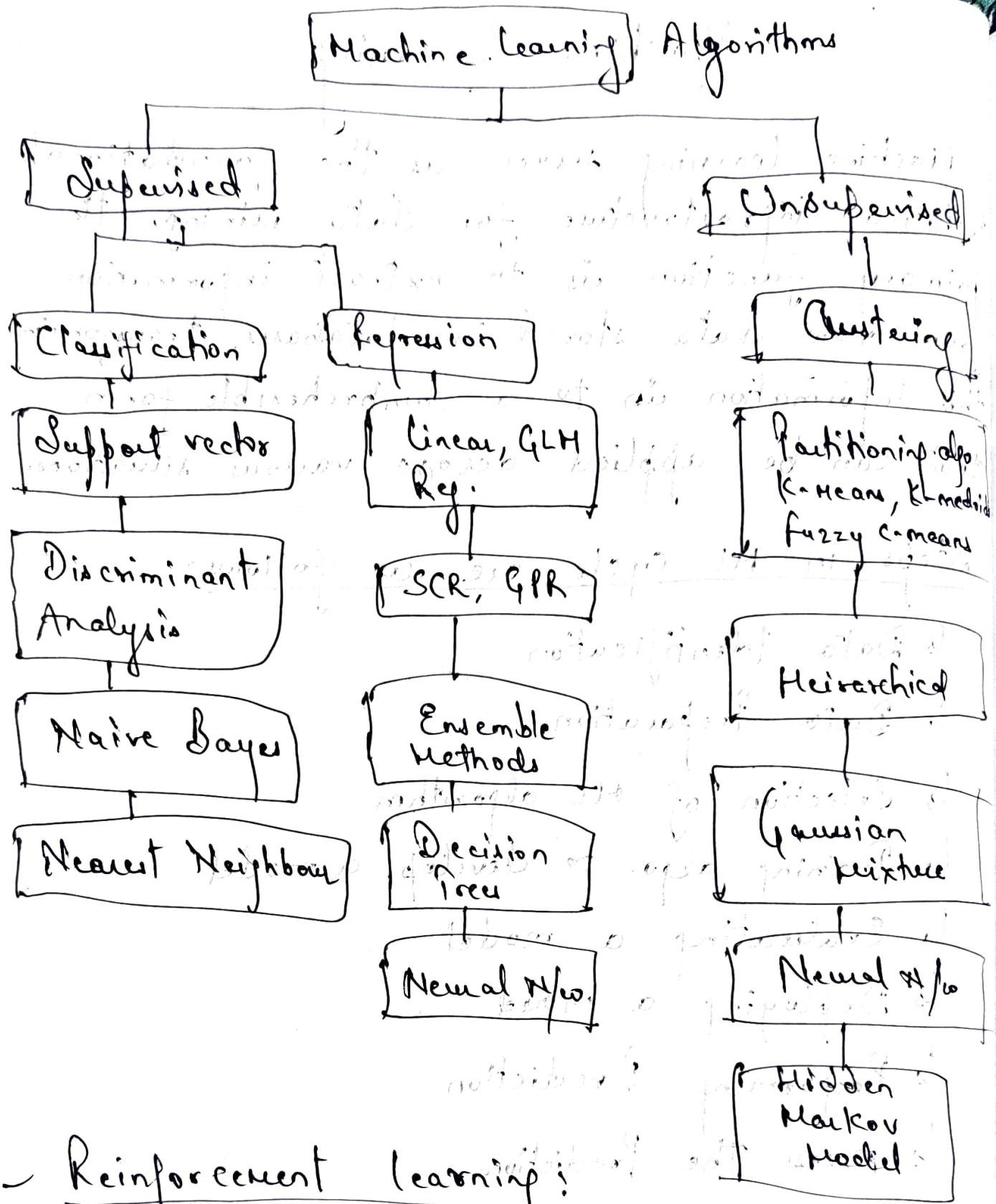
Machine Learning:

- Machine learning serves as the foundational technical infrastructure for data mining. Its primary function is to extract information from raw data stored in databases, transforming this information in to a comprehensible form that can be applied across various situations.

- Steps in ML cycle are as follows:
  - Data Identification
  - Data Preparation
  - Selection of ML algorithm
  - Training algo. to develop a model
  - Evaluating a model
  - Deploying a model
  - Performing Prediction
  - Assess the Predictions

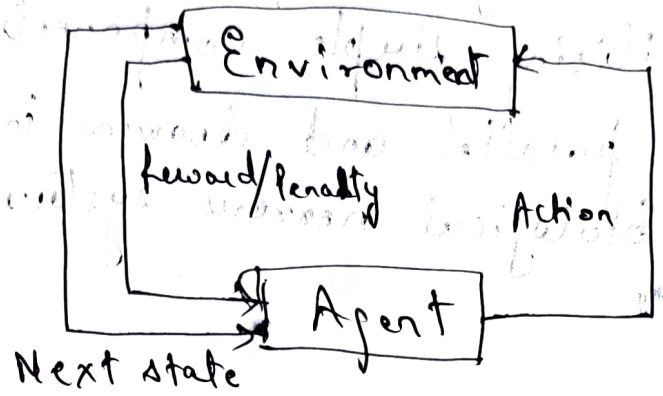
- Classes of learning algo.





- Reinforcement learning:
refers to a kind of machine learning method in which the agent recieves delayed reward in the next step to evaluate its previous action.

RL setup is composed of two components an agent and an environment.



It may be broken down into two categories: Model-free and Model-based.

Deep Learning: serves as a method for machines to acquire knowledge by utilising deep neural networks. It is used to solve practical problems in fields like computer vision (image), natural language processing (text), and automated speech (audio).

Neural Networks: are comprised of neurons. Each neuron recieves signals as input, multiplies them by weights, adds them together, and then applies a non-linear function. These neurons are organised in layers and closely stacked to each other.

- A deep neural network is characterised by the integration of multiple non linear processing layers, utilising simple components that operate in parallel and drawing inspiration from the biological nervous systems of living organisms.

- Convolutional Neural N/w's (CNN): The idea that underlies them is that rather than linking each neuron with all of the ones that come after it, we just connect it with a select few of those that come after it.

- Recurrent Neural N/w's: find application in time series forecasting due to their suitability for temporal data.

- Recursive Neural N/w: setup in a tree-like manner. As a result, they can simulate the heirarchical structures training datasets.

- Auto-Encoders: type of unsupervised technique used to that is reduce dimensionality & compress data.

Restricted Boltzmann Machines (RBM): are stochastic neural n/ws that can learn a probability distribution over their inputs & so have generative capabilities. They take the input & create a representation of it in the forward phase of the training. They rebuild the original output from the representation in the backward pass.

Generative Adversarial N/ws (GANs): construct two models: make up fake data (generator) & tell difference b/w actual and fake data (discriminator) and turned them against one another.

Graph Neural N/ws: Deep learning doesn't operate well with unstructured data in general. GANs are designed for modelling graph data, allowing them to identify and represent the relationships b/w nodes in a n/w as embeddings typically in the form of integers.
This enables their apps in various ML models for tasks such as clustering, classification & more.

— Ensemble learning : that enhances predictive performance by amalgamating predictions from diverse models. It is further categorised as:

↳ Bagging Ensemble learning : is the process of fitting multiple decision trees to various samples of same dataset and averaging the results. Popular EL algo. are based on this approach including:

↳ Bagged Decision Trees (Canonical bagging)
↳ Random forest
↳ Extra Trees.

↳ Stacking EL : seeks diversity among its members by employing various types of models fitted to the training data and utilising a model to aggregate predictions. Popular EL algo:

↳ Stacked Models (Canonical stacking)
↳ Blending
↳ Super Ensemble

Boosting EL: designed to modify the training data, emphasising instances that were incorrectly predicted by previous models.

Popular EL algo.:

- AdaBoost (canonical boosting)
- Gradient Boosting Machines
- Stochastic Gradient Boosting (XGBoost and similar)

Supervised learning: entails training a model using data where both inputs and outputs are already identified, enabling the model to make prediction about future outputs. Eg: Medical imaging, speech recognition etc.

Unsupervised learning: analysis data to uncover previously unknown patterns or structures. It is used to infer conclusions from sets of data that contain input but no tagged answers. Eg: Market Research etc.

Semi-Supervised learning: algo. are trained on small sets of labelled data before being applied to unlabeled data, like in UL.

- Naive Bayes : represents a case of statistical classification, wherein it assesses the probability that a specific sample belongs to a particular group based on given sample. When used to big Databases, the Bayesian classification demonstrates both improved accuracy and increased speed.

Steps to perform naive Bayes Algorithm:

↳ Handling data: Data is loaded from CSV file and spread into training & tested assets

↳ Summarising the data: A particular prediction is made using a summarise of the data set to make a single prediction.

↳ Making a Prediction: ——↑

↳ Summarise the properties in the training dataset to calculate the probabilities & make predictions

↳ Making all the Predic: Generate prediction given a test dataset and a summarise data set.

↳ Evaluate Accuracy: Accuracy of the predic. model for the test data set as a percentage correct out of them all the predictions made

↳ Tying all together: finally, we tie all the steps together and form our own model of Naive Bayes Classifier.

- K-Nearest Neighbours (K-NN): This method assigns items to the class to which they are closest in terms of proximity to their Neighbours. It involves the calculations of the distance b/w an item and a class to make such determinations. Classes are represented by centroid (central value) and invidual points.

- Decision Trees: Basic steps are as follows:

↳ Building the tree by using the training set dataset/database.

↳ Applying the tree to the new dataset/database.

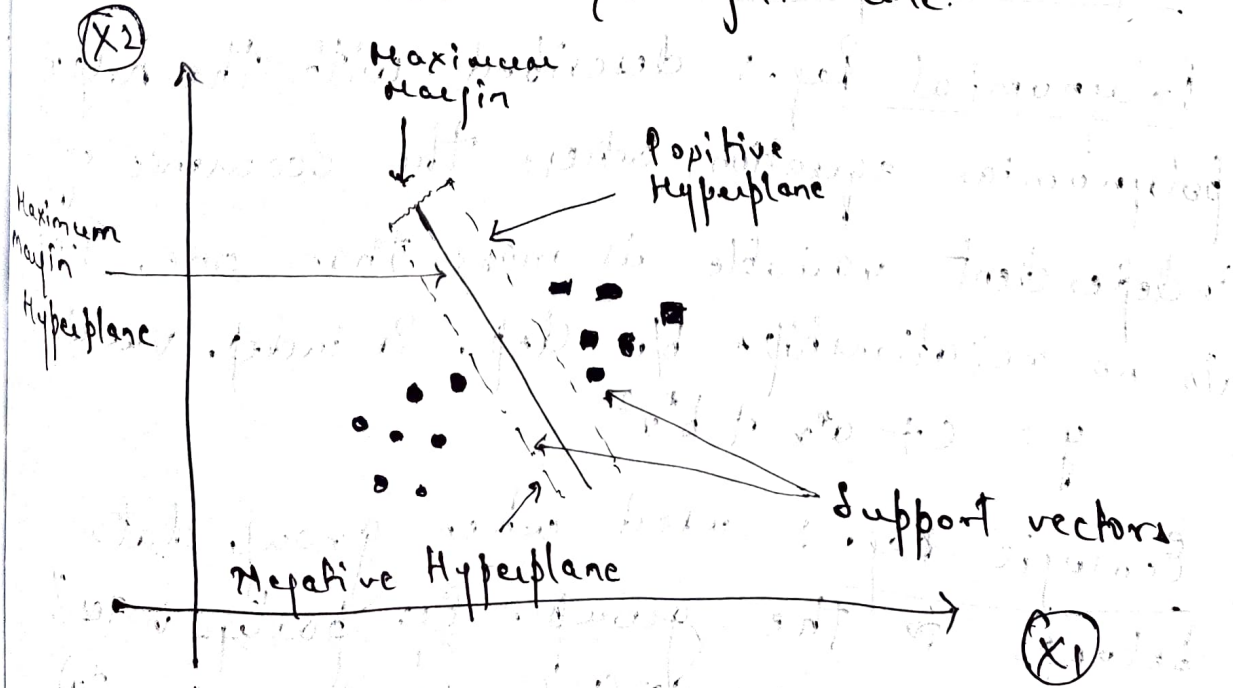
- Logistic Regression: A key component of the supervised learning approach. Its purpose is to forecast the categorical dependent variables.
- LR predicts the outcome of a dependent variable that has a 'yes' or 'no' answer.
- LR shares similarities with Linear Regression, yet their application differs.
- In LR, we fit a S-shaped logistic function which predicts two maximum values, instead of a regression line (0 or 1).
- The curve from the logistic function shows how likely something is, like whether the cells are cancerous or not, whether a mouse is overweight or not, etc.
- Types of LR:
  - Binomial: dependency variables can be either 0 or 1, pass or fail, etc.
  - Multinomial: dependent variables can be one of three types that are not in order, such as cats, dogs, or sheep.
  - Ordinal: dependent variables can be ranked such as low, medium, high.

Support Vector Machine : a type of algo: that transforms the primary training data into a new format that has a higher dimension by making use of a non linear mapping.

Two main categories that can be applied to SVM:

↳ Linear SVM: can use linear SVM if the data can be completely separated into linear categories.

↳ Non-linear SVM: when the data is not linearly separable means we apply advanced techniques like kernel tricks to categorise the data points that can't be divided into two classes uses a straight line.



SVM

Various Types of regression algorithms:

↳ **Linear Regression**: used when there is a single dependent variable, and there can be one or more independent variables. In linear reg. the relationships b/w the dependent and independent variables are linear, i.e. of type $y_i = a + b^* x_i$. Eg.: Child height = a+b* (parent height)

↳ **Multiple Linear regression**: when there is only one dependent variable and more than one independent variables, then it results in MLR. ie; $y = a + bx_1 + cx_2 + dx_3$. Eg: a+b*(daily meal) + c* (daily exercise).

↳ **Logistic Regression**: Discussed earlier.

↳ **Polynomial Reg.**: described with the help of polynomial equation where the occurence of independent variable is more than one. There is no relationships b/w dep. & indep. variables.

$$y = c + a^*x + b^*x^2$$

↳ **Ecologic Reg.**: used when group data belongs to the group. Eg: party_votes % = .2 + .5* (below_poverty_people_votes)

↳ Ridge Regression: is a regularisation technique employed in scenarios where data variables exhibit strong correlations.

↳ Lasso Regression: coefficients are penalised, & this involve the application of the least absolute shrinkage and selection operator.

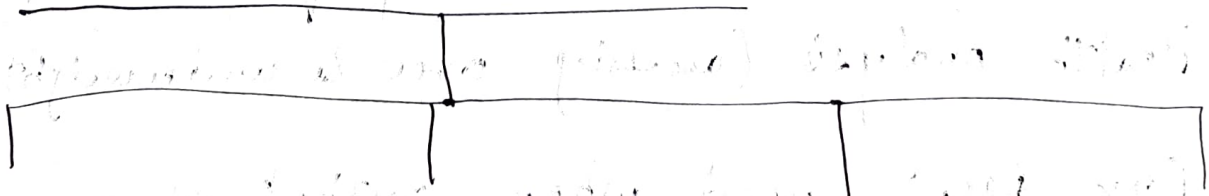
↳ Logic Reg: predictor and response variable both are binary in nature and applicable to both classification and regression problem.

↳ Bayesian Rg.: algo. relies on Bayesian statistics, utilising random variables as parameter for estimates. In this approach, when data is unavailable, the algo. incorporates prior data as input.

↳ Quantile Reg.: when the focus is on determining the boundaries of quantiles. Eg: Health Analysis (assessing over & underweight).

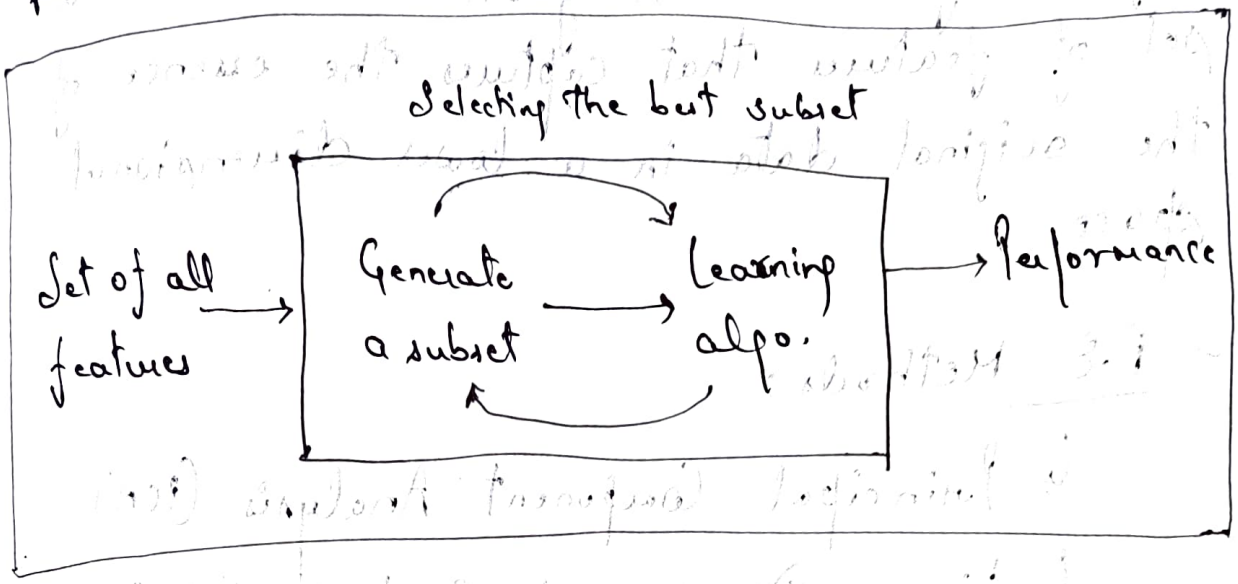
↳ Cox Reg.: used when output of a variable depends on set of independent variables.
Eg: patient_survival_after_surgery (Survived, Died) = (age, condition, BMI)

- Dimensionality Reduction: is a technique used to reduce the number of features in a dataset while retaining as much of the important information as possible.
- There are two main approaches to dimensionality reduction:
  - Feature Selection
  - Feature Extraction
- Feature Selection: involves selecting a subset of the original features that are the most relevant to the problem at hand. The goal is to reduce the dimensionality of the dataset while retaining the most important features.
- Feature Selection Methods:



- Forward feature Selection: first step in FS is to evaluate each individual feature and

choose the one that results in the most effective algo. model. This is referred to as 'forward feature selection'.



- Procedure:

↳ Train the model with each feature being treated as a separate entity, then evaluate its overall performance.

↳ Select the variable that results in the highest level of performance.

↳ Carry on with the process while gradually introducing each variable.

↳ The variable that produced the greatest amount of improvement is the one that gets kept.

↳ Perform the entire process once more until the performance of model doesn't show any meaningful signs.

- Feature Extraction: involves creating new features by combining or transforming the original features. The goal is to create a set of features that captures the essence of the original data in a lower dimensional space.

- F.E Methods:
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - t-distributed stochastic neighbour embedding (t-SNE)

- PCA: is a way to get important variables from a large set of variable in a dataset. PCA is more useful when you have data with three or more dimensions.

- LDA: is the recommended linear classification method for situation involving two or more than two classes.

- Problems associated to logistic regression:
  - Binary Class Prob.
  - Unstable, but with well-defined classes
  - It is prone to instability when there are only few occurences

Association Rules : play a crucial role and find extensive applications in scenarios such as market basket analysis.

AR do not reveal the customer's preferences for specific items. Instead, they identify relationships among items that are typically purchased together by customer.

Concept :

↳ K-itemset : a set of k items. Eg: 2-itemset can be {pencil, eraser} or {bread, butter}, etc. 3-itemset can be {bread, butter, milk}

↳ Support : $\text{support}(x) = \frac{\text{No. of transaction containing } x}{\text{Total no. of considered transac.}}$

↳ Confidence :

$$\text{Confidence}(y) = \frac{\text{No. of transac. containing } x \text{ \& } y}{\text{No. of considered containing } x}$$

↳ Frequent Itemset : An item that has a support value equal to or greater than the minimum support threshold is termed a frequent itemset.

Eg: Apriori Algorithm

FP Tree Growth: Despite a notable decrease in the number of candidate itemsets, the Apriori Algo. may still face slowness as it involves the iterative scanning of the entire transaction set. So, FP growth, employs divide and conquer strategy to generate frequent itemsets.

Algorithm:

↳ Create FP Tree, by compressing the transaction database. Along with preserving the information about the itemsets. The tree structure also retains the association among the itemsets.

↳ Divide transac. database into a set of conditional databases where each associated with one frequent item or 'pattern fragment' and examines each db separately.

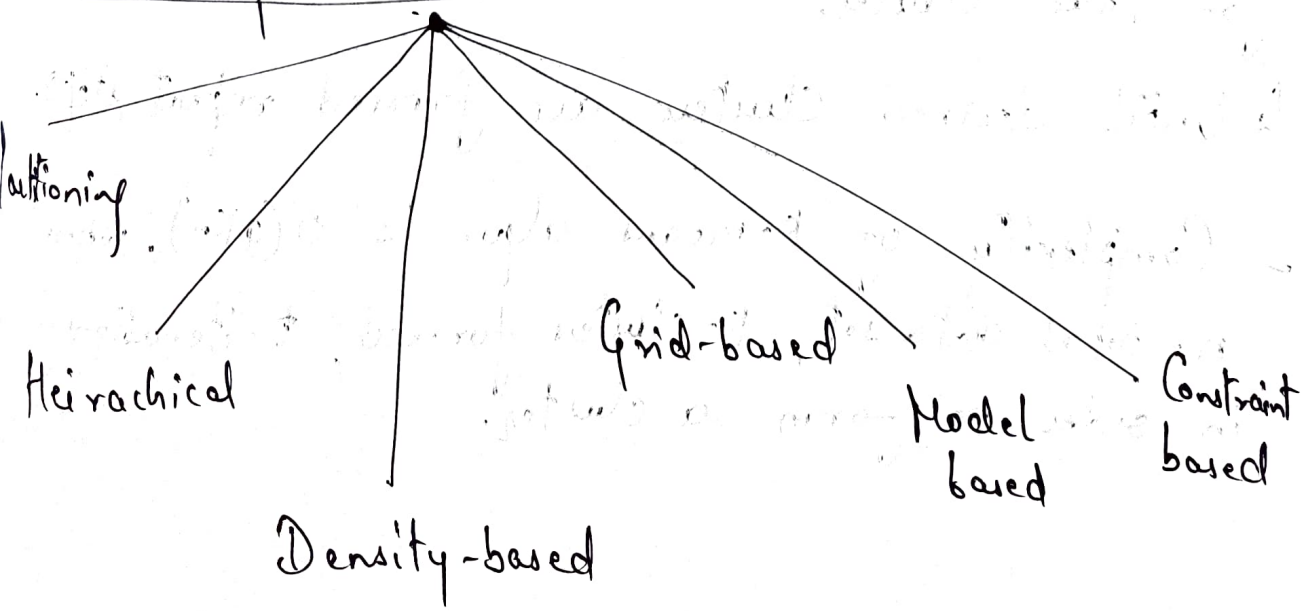
↳ for each 'pattern fragment', examine its associated itemsets only.

- Clustering: Cluster analysis, also known as clustering, is a technique used to partition a set of data objects into subsets based on individual observations.

- Example of Clustering is outlier detection where Credit Card fraud and criminal activities are monitored.

- Applications in Real-world scenarios:
  - Customer segmentation: telecom, ecommerce, sports etc.
  - Document Clustering: cluster similar docs together
  - Image Segmentation: club similar pixels in image together.

- Clustering Methods:



- k-means Algorithm: primary objective is to determine the cluster locations, minimising the distance between the cluster and dataset. Also called (Lloyd's algorithm).

Algo:

↳ Define no. of clusters (k) to be produced and identical data point centroids.

↳ The distance from every data point to all the centroids is calculated and the point is assigned to cluster with a minimum distance.

↳ follow the above setup for all the data points

↳ The average of the data points present in a cluster is calculated and can set new centroid for that cluster.

↳ Until desired clusters are formed repeat step 2.

- Complexity of k-means algo. is $O(tkn)$, where n - total data set, k - clusters formed, t - iterations in order to form a cluster.

K-medoid or PAM (Partioning Around Medoids):

In PAM, the medoid of the cluster has to be an input data point while this is not true for k-means clustering as the average of all the data points in a cluster may not belong to an input data point.

- Algo. is implemented into two steps:
  - Build: Initial Medoids are innermost objects.
  - Swap: A fn can be swapped by another fn until the fn can no longer be reduced.

- Algo:
  - Initially choose m random points as initial medoids from given data set.
  - for every data point assign a closest medoid by distance metrics.
  - Swapping cost is calculated for every chosen & unchosen object given as TC_{sn} where s is selected and n is non-selected object.
  - If $TC_{sn} < 0$, s is replaced by n
  - Until there is no change in medoids, repeat 2 and 3.

- Characteristics: