

I N D E X

NAME: B-Dheeraj Chandan STD.: _____ SEC.: _____ ROLL NO.: _____ SUB.: _____

S. No.	Date	Title	Page No.	Teacher's Sign / Remarks
		DEEP LEARNING , MACHINE LEARNING , & FOR ARTIFICIAL INTELLIGENCE		
		PART-1		
		PART-2 · PYTHON FOR DATA SCIENCE		
		AI & ML , Python Data Science		
				2024/2/22 11:
		Notes Prepared By - DHEERAJ		

3/06/2023

1. Introduction to ML

used in ML (Machine Learning) : Regression, classification, clustering, Scikit Learn & Scipy

ing (ML) :-

A field of computer science that gives computers to learn without being explicitly programmed of Artificial Intelligence (AI) and which focuses on the use of to imitate the way that humans bring its accuracy. Artificial Intelligence (AI) that to become more accurate + being explicitly programmed data as input to on as computation

(Part - I)

Deep Learning & Machine Learning

stand the fundamentals & a computational process &

computer science & statistics.

AI (Artificial Intelligence) where trained to learn from its past

techniques :-

- (i) Regression / Estimation
- (ii) Classification
- (iii) Clustering
- (iv) Associations
- (v) Anomaly detection
- (vi) Sequence mining
- (vii) Dimension Reduction
- (viii) Recommendation systems

13/06/2023

1. Introduction to ML

↳ skills used in ML (Machine Learning) : Regression, classification, clustering, Scikit Learn & Scipy

Machine Learning (ML) :-

- ↳ It is a subfield of computer science that gives computers the ability to learn without being explicitly programmed.
- ↳ It is a branch of Artificial Intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.
- ↳ It is a type of Artificial Intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so.
- ↳ Machine Learning (ML) uses historical data as input to predict new output values.
- ↳ Machine Learning theory, also known as computation learning theory.
- ↳ This theory aims to understand the fundamentals & principles of learning as a computational process & combine tools from computer science & statistics.
- ↳ It is a subset of AI (Artificial Intelligence) where the machine is trained to learn from its past experience.
- ↳ Major ML techniques :-
 - (i) Regression / Estimation
 - (ii) Classification
 - (iii) Clustering
 - (iv) Associations
 - (v) Anomaly detection
 - (vi) Sequence mining
 - (vii) Dimension Reduction
 - (viii) Recommendation systems

14/06/23

- (i) Regression / Estimation — Predicting continuous values.
- ↳ A technique for investigating the relationship between independent variables or features & a dependent variable or outcome is called ML Regression.
 - ↳ Used as a method for predicting (Predictive modelling) in ML (Machine Learning) in which an algorithm is used to predict continuous outcomes.
- (ii) Classification — Predicting the item class/category of a case
- ↳ It is a supervised machine learning method where the model tries to predict the correct label of a given input data.
 - ↳ In classification model is fully trained using the training data and then evaluated on test data before being used to perform prediction on new unseen data.
- (iii) Clustering — Finding the structure of data ; summarization
- ↳ Grouping unlabeled examples is called clustering.
 - ↳ It is an undirected technique used in data mining for identifying several hidden patterns in the data without coming up with any specific hypothesis.
 - ↳ It groups the unlabelled dataset which can be defined as a way of grouping the data points into different clusters, consisting of similar data points.
 - ↳ The objects with possible similarities remain in a group that has less or no similarities with another group.

- (iv) Associations — Associating frequent co-occurring item/events.
- ↳ Association rule learning is a type of unsupervised learning technique that checks for the dependencies of one data item on another data item and maps accordingly so that it can be more profitable.
 - ↳ It tries to find something interesting relations or associations among the variables of dataset.
 - It is based on different rules to discover the interesting relations between variables in the database.
 - ↳ It involves the use of machine learning models to analyze data patterns or co-occurrences in a database.

(v) Anomaly detection — Discovering abnormal & unusual cases

- ↳ It identifies rare events or identification of rare events, items or observations which are suspicious because they differ significantly from standard behaviours or patterns.
- ↳ It is a technique of identifying rare events or observations which can raise suspicious/suspicions by being statistically different from the rest of observations.

Anomalous behaviour translates to some kind of problem like credit card fraud detection, failing machine in a server, cyber attacks etc.

(vi) Sequence Mining — Predicting next events; click stream
(Markov Model, HMM)

- ↳ It is to discover interesting patterns in data with respect to some subjective or objective measure of how interesting it is.
- ↳ This task involves discovering frequent sequential pattern with respect to a frequency support measure.

(Vii) Dimension Reduction — Reducing the size of data (PCA)

↳ It is a Machine Learning (ML) or statistical technique of reducing the amount of random variables in a problem by obtaining a set of variables.

↳ It can be used in applied Machine Learning (ML) to simplify a classification or regression dataset in order to better fit a predictive model.

↳ Fewer features mean less complexity & requires less computation time. You will need less storage space because you have fewer data.

(Viii) Recommendation Systems — Recommending items

↳ It is a type of Machine Learning (ML) based systems that are used to predict the rating or preferences of items for a given user.

↳ It is a class of ML (Machine Learning) that uses data to help predict, narrow down & find what people are looking for among an exponentially growing number of options.

↳ There are 3 recommender systems:

a. Content based

b. Collaborative filtering

c. Hybrid

a. It relies on the similarity between items to make recommendations.

b. It is a type of ML algorithm that makes predictions about what a user might want to buy or watch based on past behaviour of other users.

c. It is combination of both

c. It is a combination of both content based & collaborative filtering approaches. The hybrid approach takes advantage of both content based and collaborative filtering by using them to supplement each other.

Difference between AI, ML & DL :-

- ↳ AI (Artificial Intelligence) is a concept of creating smart intelligent machines.
- ↳ ML (Machine Learning) is a subset of AI (Artificial Intelligence) that helps you to build AI-driven applications.
- DL (Deep Learning) is a subset of ML (Machine Learning) that uses vast volume of data and complex algorithms to train a model.
- AI (Artificial Intelligence) is basically the mechanism to incorporate human intelligence into machines through a set of rules (algorithm).
- AI is basically the study of training your machine (computers) to mimic a human brain & its thinking capabilities.
- AI focuses on 3 major aspects (skills): Learning, reasoning & self-correction to obtain maximum efficiency possible.
- ML (Machine Learning) is basically the study / process which provides the system (computer) to learn automatically on its own through experiences it had and improved accordingly without being explicitly programmed.
- ML focuses on the development of programs so that it can access data to use it for itself.
- Major aim of ML is to allow the systems to learn by themselves through experience without any kind of human intervention or assistance.

- DL (Deep Learning) is basically a subpart of the broader family of ML (Machine Learning) which makes use of Neural Networks to mimic human brain-like behaviour.
- DL algorithms focus on information processing patterns mechanism to possibly identify the patterns just like human brains does & classifies the information accordingly.
- DL works on larger set of data when compared to ML & prediction mechanism is self-administered by machines.
- Examples of AI : speech recognition, Personalized recommendations, Predictive maintenance, Medical diagnosis, Autonomous Vehicles, Virtual personal Assistance (VPA) like Siri or Alexa, Fraud detection, Image recognition, Natural language processing, Predictive analytics, Game Playing AI.
- Examples of ML : Image Recognition, Speech recognition, Natural Language Processing (NLP), Recommendation Systems, sentiment analysis, Predictive maintenance, spam filter in email, Credit risk assessment, customer segmentation, Fraud Detection, Speech recognition.
- Examples of DL : Image & Video recognition, Generative models, Autonomous vehicles, Image classification, Speech recognition, Language (Natural Language processing), Recommender systems, Fraud detection, Game playing AI, Time series forecasting

Python for ML (Machine Learning)

→ Python is one of the most popular programming language for this task & it has replaced many languages in the industry, one of the reason is its vast collection of libraries.

→ Python libraries that are used in ML (Machine Learning) are :

- * (i) Numpy (vi) Theano
- * (ii) Pandas (vii) TensorFlow
- * (iii) Matplotlib (viii) Keras
- * (iv) Scipy * (ix) Scikit-learn
- * (v) PyTorch

(i) Numpy → Numpy is a very popular python library for large multidimensional array and matrix processing, with the help of a large collection of high level mathematical functions.

→ It is very useful for fundamental scientific computation in ML (Machine Learning). It is particularly used for Linear Algebra, Fourier Transform and random number capabilities.

→ High end libraries like TensorFlow uses Numpy internally for manipulation of Tensors.

→ It is a python library used for working with arrays. It also has functions for working in domain of Linear Algebra, Fourier Transform and matrices.

(ii) Pandas → Pandas is a popular library for data analysis.

→ It was developed for data extraction & preparation.

→ It provides high level of data structures & wide variety tools for data analysis. It provides inbuilt methods for grouping, combining & filtering data.

→ Name is derived from the term "panel data" which is used for working with data sets. It has functions for analyzing, cleaning, exploring & manipulating data.

(iii) Matplotlib → Matplotlib is a popular python library used for data visualisation.

- ↳ It particularly comes when a programmer wants to visualize the patterns in the data.
- ↳ A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes etc. It provides various kind of graphs & plots for data visualization viz histograms, error charts, bar charts etc..
- ↳ It is a comprehensive library for creating static, animated and interactive visualisations in python.
- ↳ It allows users to create numerous & diverse plot types.

17/06/2023

(iv) Scipy → Scipy is a very popular library among ML (Machine Learning) as it contains different modules for optimization, Linear Algebra, integration & statistics.

- ↳ There is a difference between Scipy library & Scipy stack. Scipy is one of the core package that makes up the Scipy stack. Scipy is useful for image manipulation.
- ↳ Scipy is useful for scientific computation library that uses Numpy underneath. Scipy stands for scientific python.
- ↳ It provides more utility functions for optimization, stats and signal processing. Like Numpy, Scipy is a open source we can use it freely.
- ↳ It is a collection of mathematical algorithms and convinience functions built on the Numpy extension of Python.

2024/2/22 11:15

(V) PyTorch → It has an extensive tools choice and libraries that supports computer vision, Natural Language Processing (NLP) and many more ML programs.

- ↳ It allows developers to perform computation on Tensors with GPU acceleration and also helps in creating computational graphs.
- ↳ It is an open source Machine Learning (ML) framework based on python programming language and the Torch library.
- ↳ Torch is an open source ML library used for creating deep neural networks and is written in the Lua scripting language. Its one of the preferred platforms for deep learning research.

(vi) Theano → Theano is a python library that is used to define, evaluate and optimize mathematical expressions involving multi Dimensional arrays in an efficient manner.

- ↳ It is extensively used for unit testing & self-verification to detect & diagnose different type of errors.
- ↳ It is used in ML(Machine Learning) Model development.
- ↳ Theano execution speed is faster than TensorFlow execution speed.

(vii) TensorFlow → TensorFlow is a popular open source library for high performance numerical computation developed by Google brain team in google.

- ↳ It is a framework that involves defining and running computations involving tensors.
- ↳ It can train and run deep neural networks that can be used to develop several AI applications.
- ↳ It is widely used in the field of DL research & application.
- ↳ It is an open source platform for ML which is a rich system for managing all aspects of a ML system.
- ↳ It helps you implement best practices for data automation,

Model tracking, performance monitoring and model retraining.

(viii) Keras → Keras provides many in built methods for grouping, combining & filtering data.

↳ It makes it really for ML beginners to build and design a Neural network.

↳ It is a high level, deep learning API developed by Google for implementing Neural networks.

↳ It is written in Python & used to make implementation of neural networks easily. It also supports multiple backend neural network computation.

(ix) Scikit Learn → It is a popular ML library for classical ML algorithms. It is easy to implement.

↳ It is build on top of 2 basic python libraries i.e. NumPy and Scipy.

↳ Scikit Learn supports most of the supervised and unsupervised learning algorithms.

↳ It can also be used for data mining & data analysis, which makes it a great tool who is starting out with ML.

↳ It provides a selection of efficient tools for Machine Learning and statistical modelling including classification, regression, clustering, dimensionality reduction via a consitence in Python.

↳ Data Processing → Train/Test split → Algorithm setup

↓
Evaluation ← Prediction ← Model fitting

Model export

Supervised & Unsupervised learning :-

- Supervised learning also known as supervised ML. It is a subcategory of ML & AI.
 - It is defined by its use of labelled datasets to train algorithms that to classify data or predict outcomes accurately. It is a process of providing input data as well as correct output data to the ML (Machine Learning).
 - Supervised learning is classified into two categories :
 - (i) Classification
 - (ii) Regression
 - (i) Classification → Task for predicting a discrete class label,
(ii) Regression → Task of predicting a continuous quantity or values.
 - Unsupervised Learning is the training of a machine using information that is neither classified nor labeled & allowing the algorithm to act on that information without guidance.
 - It is also known as unsupervised ML
 - It uses ML algorithms to analyze & cluster unlabelled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention.
 - Techniques in Unsupervised learning :
 - Dimension Reduction, Density estimation,
 - Market basket analysis & Clustering
 - Unsupervised Learning is classified in two categories :
 - (i) Clustering
 - (ii) Association
 - (i) Clustering → It is a grouping of data points or objects that are somehow similar by discovering structure, summarization, Anomaly detection.
Cluster analysis is a ML technique which groups the unlabelled dataset.
 - (ii) Association → It is a rule based ML method for discovering interesting relations between variables in large databases.

Supervised Learning

- Classification : classifies labelled data.
- Regression : Predicts trends using previous labeled data.
- It has more evaluation methods than unsupervised learning.
- Controlled environment.
- Algorithms are trained using labelled data.
- Simpler method.
- High accurate.
- Uses offline analysis.
- Can test our model.

Unsupervised Learning

- Clustering : Finds patterns & grouping from unlabelled data.
- It has fewer evaluation methods than supervised learning.
- Less controlled environment.
- Algorithms are used against data that is not labelled.
- Computationally Complex.
- Less accurate.
- Uses real time analysis of data.
- Cannot test our model.

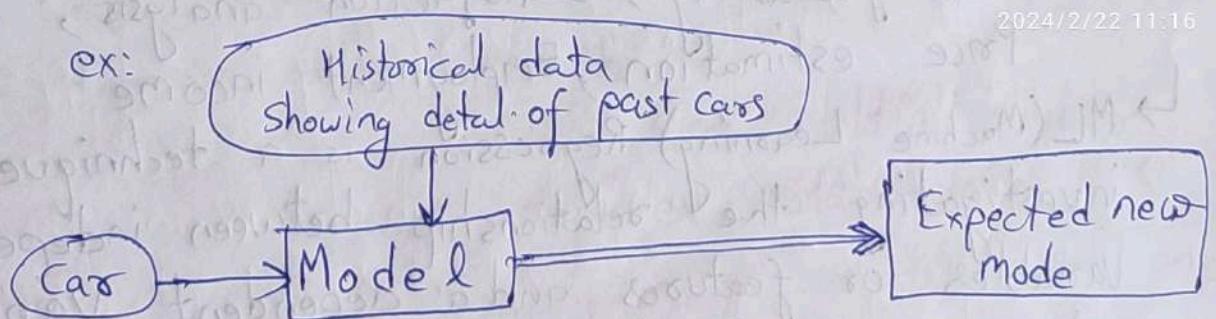
23/06/2023 2. Linear & Non-Linear Regression

- Regression is the process of predicting a continuous values. Task of predicting a continuous quantity or values.
- Regression in ML (Machine Learning) consists of mathematical methods that allows data scientist to predict a continuous outcome (y) based on the value of one or more predictor variables (x).
- The explanatory variable (or the independent variable) always belongs to on the x-axis. The responsive variable (or the dependent variable) belongs on y-axis.

Regression Model :-

- It provides a function that describes the relationship between one or more independent variables and a response, dependent or target variable.
- In statistical modeling, regression analysis is a set of statistical process for estimating the relationships between a dependent variable and one or more independent variable.

ex:



- Types of Regression Model :
 - (i) Simple Regression
 - (ii) Multiple Regression.
 - (iii) It is a regression algorithm that models the relationship between dependent & independent variable.
- If is of 2 types :
- a. Simple Linear Regression
 - b. Simple Non-Linear Regression

- a. Simple Linear Regression is a type of Regression method algorithm that models the relationship between a dependent variable & a single independent variable.
- b. It is a form of regression analysis in which data is fit to a model and then expressed as a mathematical function. It relates the two variables in a non-linear (curved) relationship.
- (ii) Multiple Linear Regression (also known as multiple Regression)
It is a statistical technique that uses several explanatory variables to predict the outcome of a response variable.
It is a regression model that estimates the relationship between a quantitative dependent variable and two or more independent variables using a straight line.

24/06/2023

Application of Regression :

- ↳ Applications of Regression analysis are in sales forecasting, satisfaction analysis, price estimation, Employment income
- ↳ ML (Machine Learning) Regression is a technique for investigating the relationship between independent variables or features and a dependent variable or outcome.
- ↳ It is used as a method for predictive modelling in ML (Machine Learning), in which an algorithm is used as to predict continuous outcome.
- ↳ The main uses of regression analysis are forecasting, time series modelling and finding the cause and effect relationship between variables.

- ↳ Regression Algorithms → Ordinal regression, Poisson regression, Fast forest quantile regression, (Linear, polynomial, Lasso, stepwise, Ridge regression), Bayesian Linear regression, Neural network regression
- Decision forest regression, Boosted decision tree regression, kNN (k-Nearest neighbours)

Simple Linear Regression :-

- ↳ It is a type of regression algorithms that models the relationship between a dependent variable & a single independent variable.
- ↳ It is a statistical method for establishing the relationship between two variables using a straight line.
- ↳ Formula for simple Linear Regression can be written as, $y_i = E(Y_i) + \epsilon_i$ $= (\beta_0 + \beta_1 x_i) + \epsilon_i$ $\therefore Y_i = \beta_0 + \beta_1 X_i$
 ϵ_i is a response variable
 X_i is a single predictor

Multiple Linear Regression :-

- ↳ Multiple Linear Regression is an extension of Simple Linear Regression which takes more than one predictor variable to predict the response variable.
- ↳ It is a statistical technique that uses multiple linear regression to model more complex relationships between two or more independent variables and one dependent variable.

- ↳ Formula for Multiple Linear regression can be written as, $y = a + b x_1 + c x_2 + d x_3 + \dots$
 $\therefore y$ is response variable & x_1, x_2, x_3, \dots are multiple predictor

- ↳ Regression Algorithms → Ordinal regression, Poisson regression, Fast forest quantile regression, (Linear, polynomial, Lasso, stepwise, Ridge regression), Bayesian Linear regression, Neural network regression
- Decision forest regression, Boosted decision tree regression, kNN (k-Nearest neighbours)

Simple Linear Regression :-

- It is a type of regression algorithms that models the relationship between a dependent variable & a single independent variable.
- It is a statistical method for establishing the relationship between two variables using a straight line.
- Formula for simple Linear Regression can be written as,

$$y_i = E(Y_i) + \epsilon_i$$

$$= (\beta_0 + \beta_1 x_i) + \epsilon_i$$

$$\therefore Y_i = \beta_0 + \beta_1 X_i$$

2024/2/22 11:17

Multiple Linear Regression :-

- Multiple Linear Regression is an extension of Simple Linear Regression which takes more than one predictor variable to predict the response variable.
- It is a statistical technique that uses multiple linear regression to model more complex relationships between two or more independent variables and one dependent variable.
- Formula for Multiple Linear regression can be written as,

$$y = a + b x_1 + c x_2 + d x_3 + \dots$$

$\therefore y$ is response variable & x_1, x_2, x_3, \dots are multiple predictor

- ## Model evaluation in Regression models :-
- ↳ Model evaluation is the process that uses some metrics which helps us to analyze the performance of the model.
 - ↳ It is a process of using different evaluation metrics to understand a ML (Machine Learning) model's performance as well as its strengths & weakness.
 - ↳ It is important to assess the efficacy of a model during initial research phases and it also plays a role in model monitoring.
 - ↳ The train test split procedure is used to estimate the performance of Machine learning algorithms when they are used to make predictions on data not used to train the model.
 - ↳ The problem of training and testing on the same dataset is that we won't realize that your model is overfitting because the performance of your model on the test set is good.

The purpose of testing on data that has not been seen during training is to allow you to properly evaluate whether overfitting is happening.

Classification Metrics :

- ↳ To predict the variable which is in the form of discrete values. To evaluate the performance of a model there are metrics : (Evaluation Metrics Classification)
 - (i) Classification accuracy
 - (ii) Logarithmic Loss
 - (iii) Area under curve (AUC)
 - (iv) F1 score $\frac{\text{Precision}}{\text{Recall}}$
 - (v) Confusion matrix

(i) Classification Accuracy — It is the accuracy we get when/while calculating the ratio of correct prediction -s to the total number of input samples.

$$\text{Accuracy} = \frac{\text{No. of correct predictions}}{\text{Total number of input Samples}}$$

(ii) Logarithmic Loss — It is also known as Log Loss. It usually works well with multi-class classification. The classifier should assign the probability for each and every class of all the samples. If there are N samples belonging to M class then, Log Loss calculated as,

$$\text{Logarithmic Loss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M [y_{ij} * \log(p_{ij})]$$

hence, y_{ij} indicates sample i belongs to class j .

p_{ij} indicates probability of sample i belongs to class j .

Note : Range of log loss is $[0, \infty]$.

When log loss is near 0 which means high accuracy & away from 0 then lower accuracy.

(iii) Area under Curve — It is basically used for binary classification.

It is defined as probability of a classifier will rank a randomly chosen positive example higher than a negative example.

↳ True positive Rate : It is also called or termed as 'sensitivity'. It is considered as a portion of positive data points that are correctly considered as positive w.r.t all data points that are true.

$$TPR = TP / (TP + FN)$$

→ True Negative : It is also termed as specificity. It is considered as a portion of negative data points that are correctly considered as negative, w.r.t all data points that are negative.

$$TNR = \frac{TN}{TN + FP}$$

→ False Positive Rate : It is considered as a portion of negative data points that are mistakenly considered as negative, w.r.t all data points that are negative.

$$FPR = \frac{FP}{FP + TN}$$

→ FPR and TPR both have values in range [0,1].

→ AUC (Area Under Curve) is curve plotted between False Positive Rate (FPR) vs True Positive Rate (TPR) at all data points with a range [0, 1].

→ Greater the values of AUC better the performance of the model.

(iv) F1 Score — It is a harmonic mean between recall and precision. Its range is [0,1].

Harmonic Mean is defined as the average of the reciprocal values of the given values.

$$\text{Harmonic Mean} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

→ Precision : Another metric named precision. It is a measure of a model's performance that tells you how many of the positive

predictions made by the model are actually correct.

It is calculated as the

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

↳ Recall : It can be expressed as

$$F1 = 2 * \left[\frac{1}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} \right]$$

(V) Confusion Matrix — It presents table layout of the different outcomes of the prediction and results of a classification problem and helps visualize its outcome.

- It is used to determine the performance of the classification models for a given set of test data.

- Showing errors in the model performance in the form of matrix known as error matrix.

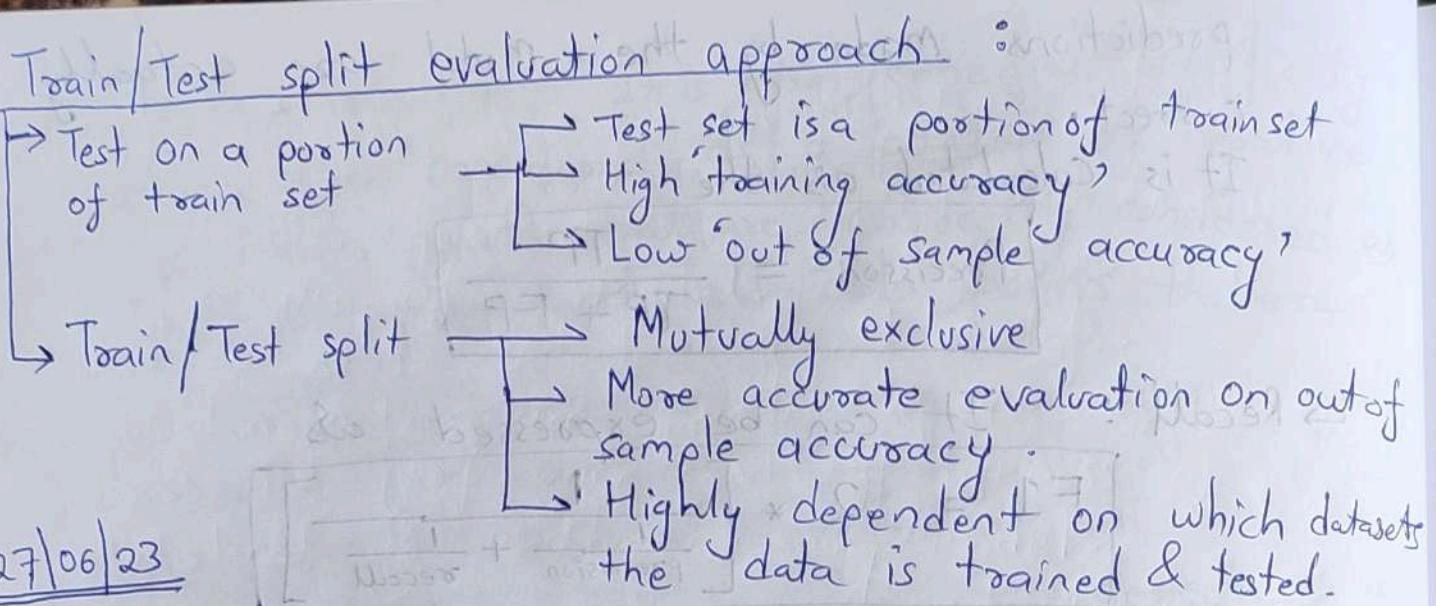
→ Matrix divided into 2 dimension : predicted value and actual value .

→ Predicted values are those values which are predicted by model . Actual values are the true values for the given observation .

$n = \text{no. of prediction}$	Predicted: NO	Predicted: Yes
Actual: No	50	10
Actual: Yes	5	100

so, $n = 165$ since $50 + 5 + 10 + 100 = n = 165$
no. of predictions are 165

↳ True Positives — If predicted is Yes & the real output is Yes.
True Negative — If predicted is No & the real output is No.
False Positive — If predicted is Yes & the real output is No.
False Negatives — If predicted is No but actual is Yes.



↳ Training Accuracy — High training accuracy isn't necessarily a good thing.

- Result of overfitting.
- Accuracy score in ML (Machine Learning) is an evaluation metric that measures the number of correct predictions made by a model in relation to the total number of predictions made.

$$\text{Accuracy} = \frac{\text{No. of correct prediction}}{\text{Total no. of Prediction}}$$

- Validation accuracy is the accuracy of a model on new data.
- Training accuracy of a model on the data it was trained on.
- Validation accuracy is lower than training accuracy as the model has never seen the validation data before.

↳ Out of sample Accuracy — It is the data that was unseen and you only produce the prediction/forecast one it.

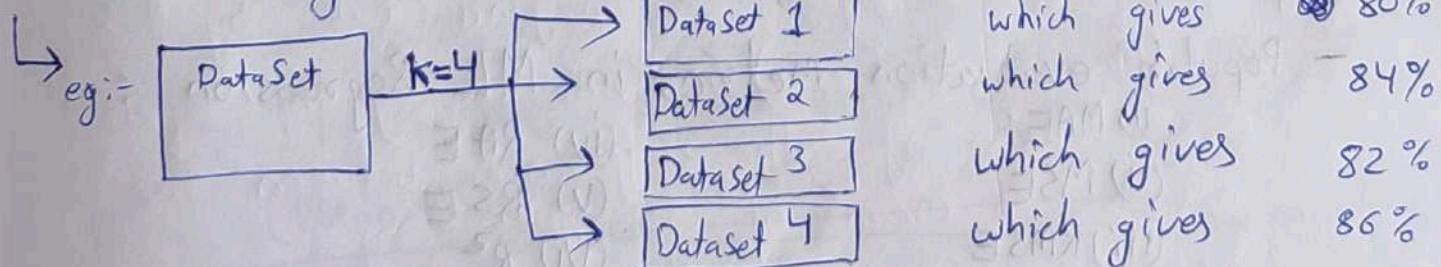
- In sample refers to the data that you have.
- Out of sample refers to the data that you don't have but want to forecast or estimate.

- 'In Sample' refers to data that is used to fit or train a model.
- 'Out of Sample' refers to data that is held back from the model training process & used to evaluate the performance of the model.

k-fold cross validation

- ↳ Cross validation sometimes called rotation estimation or out of sample testing, is a process where the dataset is split into k number of folds & is used to evaluate model's ability when given new data.
- ↳ This approach divides the input dataset into k groups of samples of equal sizes. These samples are called folds.
- ↳ Cross validation is a resampling procedure used to evaluate machine learning models on a limited data sample.

It reduces the variance of performance estimation & allowing to use more data for training -



$$\begin{aligned}
 \text{So, Accuracy} &= \text{avg}(DS_1 + DS_2 + DS_3 + DS_4) \\
 &= \text{avg}(80\% + 84\% + 82\% + 86\%) \\
 &= 83\%
 \end{aligned}$$

28/06/23

Evaluation Metrics in Regression Models:

- ↳ Evaluation metrics ~~to~~ is used to check the performance of a model. It refers to a measure that we use to evaluate different models.
- ↳ Evaluation metrics are quantitative measures used to assess the performance and effectiveness of a statistical or machine learning model.
- ↳ Regression is a method for understanding the relationship between independent variables or features and a dependent variable or outcome.
- ↳ Error of a Model — It is an action which is inaccurate or wrong.

- Error is used to see how accurately our model can predict on data it uses to learn as well as new unseen data. Based on error we choose the Machine Learning Model which performs best for a particular dataset.
- It measures of how far the data is from the fitted regression line. It is a difference between Actual value and predicted value.
- Popular evaluation metrics in ML Regression :-
 - (i) MAE
 - (ii) MSE
 - (iii) RMSE
 - (iv) RAE
 - (v) RSE
 - (vi) R^2

- (i) MAE (Mean Absolute Error) — It refers to the magnitude of difference between the prediction of an observation and the true value of that observation.

- It is the average distance between the real data and the predicted data.

$$MAE = \frac{1}{N} \left[\sum_{j=1}^N |y_j - \hat{y}_j| \right]$$

$N \rightarrow$ no. of data point
 $y_j \rightarrow$ Actual %
 $\hat{y}_j \rightarrow$ Predicted %

$\hat{y}_j - y_j \rightarrow$ residual error $|y_j - \hat{y}_j| \rightarrow$ Absolute value

(iii) Mean (MSE) Mean Square Error — It is averaged squared difference between the estimated values and the actual values.

- It measures how close a regression line is set to all data points.
- It is calculated by taking the average, specifically the mean, of squared errors from data as it refers to a function.
- It is defined as the mean or average of the squared differences between the actual and estimated values.

$$MSE = \frac{1}{n} \left[\sum_{j=1}^n (y_j - \hat{y}_j)^2 \right]$$

$n \rightarrow$ no. of data points
 $y_j \rightarrow$ Actual %
 $\hat{y}_j \rightarrow$ Predicted %
 $y_j - \hat{y}_j \rightarrow$ residual error

(iii) RMSE (Root Mean Square Error) — It is a standard way to measure the error of a model in predicting quantitative data.

- It is the square root of the MSE (Mean Square Error).

$$RMSE = \sqrt{\frac{1}{n} \left[\sum_{j=1}^n (y_j - \hat{y}_j)^2 \right]}$$

$$RMSE = \sqrt{MSE}$$

$n \rightarrow$ no. of data point
 $y_j \rightarrow$ Actual %
 $\hat{y}_j \rightarrow$ Predicted %

$\hat{y}_j - y_j \rightarrow$ residual error

(iv) RAE (Relative Absolute Error) — It is a way to measure the performance of a predictive model.

— The difference between the actual value and measured value is called absolute error.

— RAE is defined as ratio of absolute error of a measurement and the actual value of the quantity is known as relative error.

$$\text{RAE} = \frac{\sum_{j=1}^n |y_j - \hat{y}_j|}{\sum_{j=1}^n |y_j - \bar{y}|}$$

$$\bar{y} = \frac{1}{n} \left[\sum_{j=1}^n y_j \right]$$

$n \rightarrow$ no. of data points

$y_j \rightarrow$ Actual %

$\hat{y}_j \rightarrow$ Predicted %

$\bar{y} \rightarrow$ Average %

$y_j - \hat{y}_j \rightarrow$ residual error

so,

$$\text{RAE} = \frac{\sum_{j=1}^n |y_j - \hat{y}_j|}{\sum_{j=1}^n |y_j - \left(\frac{1}{n} \sum_{j=1}^n y_j \right)|}$$

(V) RSE (Relative Square Error) — We can find RSE by taking MSE (Mean Square Error) & divide it by the ~~squared~~ square of the difference between the actual and mean of the data.

— We divide MSE of our model by the MSE of a model which uses mean as the predicted value.

$$RSE = \frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n (y_j - \bar{y})^2}$$

$$\bar{y} = \frac{1}{n} \left[\sum_{j=1}^n y_j \right]$$

so,

$$RSE = \frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n \left(y_j - \left\{ \frac{1}{n} \left[\sum_{j=1}^n y_j \right] \right\} \right)^2}$$

(vi) R^2 (R square) — It is a measure that provides information about the goodness of fit of a model.

- It refers to the coefficient of determination or the coefficient of multiple determination in case of multiple regression.
- It acts as an evaluation metric to evaluate the scatter of the data points around the fitted regression line.

$$R^2 = 1 - RSE$$

$$R^2 = 1 - \left[\frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n (y_j - \bar{y})^2} \right]$$

$$R^2 = 1 - \left[\frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n \left(y_j - \left\{ \frac{1}{n} \left[\sum_{j=1}^n y_j \right] \right\} \right)^2} \right]$$

$$\bar{y} = \frac{1}{n} \left[\sum_{j=1}^n y_j \right]$$

$n \rightarrow$ no. of data points

$y_j \rightarrow$ Actual %

$\hat{y}_j \rightarrow$ Predicted %

$\bar{y} \rightarrow$ Average %

$\hat{y}_j - y_j \rightarrow$ residual error

$n \rightarrow$ no. of data point

$y_j \rightarrow$ Actual %

$\hat{y}_j \rightarrow$ Predicted %

$\bar{y} \rightarrow$ Average %

$\hat{y}_j - y_j \rightarrow$ residual error

30/06/2023

Multiple Linear Regression :

- ↳ It is a statistical technique that uses multiple linear regression to model more complex relationships between two or more independent variables and one dependent variable.
- ↳ It is simply known as multiple regression. It is a statistical technique that uses several explanatory variables to predict the outcome of a response variable.
- ↳ In simple Linear Regression one independent variable is used to predict or estimate dependent variable whereas, in Multiple Linear Regression one or more, multiple independent variable is used to predict or estimate dependent variable.
- ↳ Multiple Linear Regression is considered as a Simple Linear Regression.
- ↳ One of the important regression algorithms which models the linear relationship between a single dependent continuous variable and more than one independent variable.

Examples of Multiple Linear Regression :

- ↳ Prediction of CO_2 emission based on engine size & number of cylinders in a car.
- ↳ Independent variables effectiveness on prediction :
 - Does revision time, time anxiety, lecture attendance, & gender have any effect on the exam performance of students?

- Predicting impacts of changes :
 - How much does blood pressure go up (or down) for every unit increase (or decrease) in the BMI of a patient?

Applications of Multiple Linear Regression :

- To know: How strong the relationship is between two or more independent variables and one dependent variable.
ex: How rainfall, temperature and amount of fertilizer added affect crop growth.
 - Important for evaluating data and establishing a definite relationship between two or more variables.
 - To estimate relationship between a dependent variable and independent variables.
 - To model more complex relationships between two or more independent variables & one dependent variable.
 - It is used to predicting continuous values.
 - Independent variable also called as feature variable, predictor variable, left hand variable, target variable.
Dependent variable also called as response variable, outcome variable & right hand variable.
 - Dependent variable is the one being trained on. Independent Variable are those being used to train the model.
 - As we had taken example CO₂ emission based on engine size & no. of cylinders in a car say,
- | Engine Size | Cylinders | Fuel consumption | CO ₂ Emission |
|-------------|-----------|------------------|--------------------------|
| x_1 | x_2 | x_3 | y |

Engine Size	Cylinders	Fuel Consumption	CO2 Emission
2.0	4	8.5	196
2.4	4	9.6	221
3.5	6	5.9	136

Here, (X) Independent variables are Engine Size, Cylinders, Fuel consumption
(Y) Dependent variables is CO2 Emission.

$$\therefore \text{CO2 Emission} = \theta_0 + \theta_1 \cdot \text{EngineSize} + \theta_2 \cdot \text{Cylinders} + \theta_3 \cdot \text{FC}$$

$$\hat{y} = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \theta_3 \cdot x_3 \quad | \quad x_0 = 1$$

$$\boxed{\hat{y} = \theta^T \cdot X} \quad \begin{array}{l} \theta^T \rightarrow \text{Transpose of coefficient set} \\ X \rightarrow \text{Feature set} \end{array}$$

$$\theta = [\theta_0, \theta_1, \theta_2, \theta_3] \quad \& \quad X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

say, predicted value of CO2 emission is 140 i.e \hat{y}_j

Actual Value of CO2 emission is 196 i.e y_j

Residual error is $\Rightarrow y_j - \hat{y}_j$

$$\Rightarrow 196 - 140 \Rightarrow 56 \quad | \quad \text{where, } j=1$$

$$\text{so, } \boxed{\text{MSE} = \frac{1}{n} \sum_{i=1}^n [(y_i - \hat{y}_i)^2]} \quad \begin{array}{l} \text{For } n \rightarrow \text{multiple data points} \\ 2024/2/22 11:20 \end{array}$$

$$\boxed{\hat{y} = \theta^T \cdot X} \quad (\theta)$$

$$\boxed{\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n}$$

$$\theta = [\theta_0, \theta_1, \theta_2, \dots, \theta_n]$$

$$X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} : (x_0 = 1)$$

Estimating multiple linear Regression Parameters:

↳ The most common methods to find parameters or coefficients in Multiple Linear Regression :

(i) OLS (Ordinary Least Squares)

(ii) Optimizational algorithm

$$\begin{aligned} & - S = \sum_{t=1}^n (y_t - (\alpha x_t + \beta))^2 \\ & - \boxed{Y = mx + c} \end{aligned}$$

(i) OLS (Ordinary Least Squares) : - It tries to estimate the value of parameter or coefficients by minimizing the MSE (Mean Square Error).

- This approach uses the data as matrix and uses Linear algebra operations to estimate optimal values for data.
- It takes long time for matrix operation (large dataset). More time complexity.

(ii) Optimization algorithm : - It is the process of improving the effectiveness and accuracy of a machine learning model, usually through the tweaking of model hyperparameters.

- Process of optimizing values of coefficient by iteratively minimizing error of a model on our training data.
- Gradient descent is an optimization algorithm which is commonly used to train machine learning models and neural network.
It is the proper approach if you have very large dataset.
- It is defined as the most commonly used iterative optimization algorithms of ML (Machine Learning) to train the Machine Learning & Deep Learning models.
It helps in finding the local minimum of a function
 $b \rightarrow$ mean of independent variable

$$Y = \beta X + b$$

→ If 'y' represents the dependent variable and 'x' as the independent variable this relationship is called as regression of y on x. This relationship can be represented as by simple eq. called regression eq? i.e. $Y = mx + c$

Non-Linear Regression :-

- It is statistical technique that helps describe non-linear relationships in experimental data.
- It is a type of polynomial regression. It is a method to model a non linear relationship between the dependent and independent variables.
- It is used when the data shows the curvey trend and linear regression would not produce very accurate results when compared to non-linear regression. This is because in linear regression it is pre assumed that the data is linear.

Polynomial Regression — It is a kind of linear regression in which the relationship shared between the dependent & independent variables 'Y' & 'X' is modeled as n^{th} degree of polynomial.

- It describes the fitting of a non linear relationship between the value of 'X' and the conditional mean of 'Y'.

- Polynomial regression is expressed as :

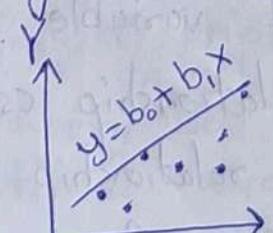
$$y = b_0 + b_1 x_1^1 + b_2 x_1^2 + b_3 x_1^3 + \dots + b_n x_1^n$$

$$= b_0 x_1^0 + b_1 x_1^1 + b_2 x_1^2 + b_3 x_1^3 + \dots + b_n x_1^n$$

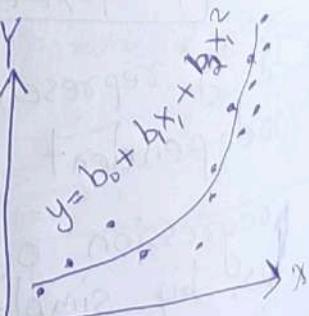
$$y = \sum_{a=0}^n b_a x_1^a$$

- It is also called special case of Multiple Linear Regression because we add some polynomial terms to multiple linear regression to convert it to Polynomial regression.

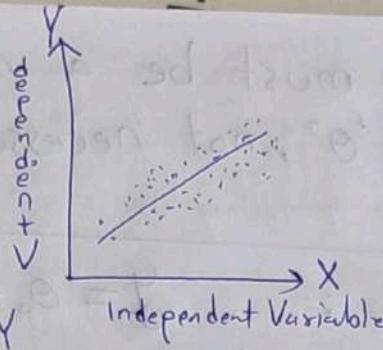
Simple Linear Model



Polynomial model :



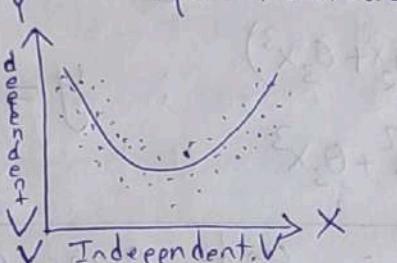
↳ Linear Regression →



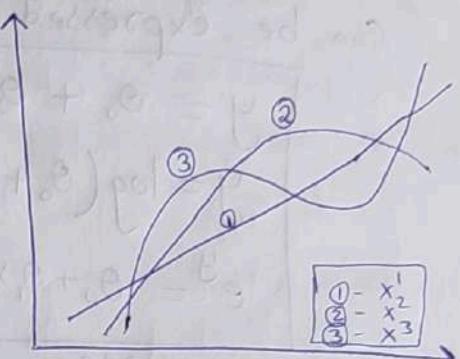
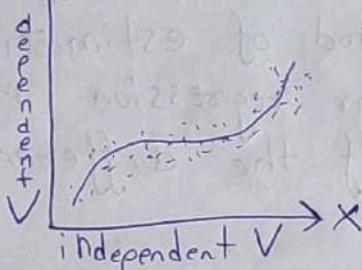
so,

Polynomial regression

Quadratic (Parabolic) Regression



Cubic Regression



↳ A polynomial regression model can be transformed into linear regression model.

as we know polynomial regression expressed as,

$$y = b_0 + b_1 x_1^1 + b_2 x_2^2 + b_3 x_3^3 + \dots + b_n x_n^n$$

So, by considering

$$x_1^1 = x, x_2^2 = x = x_3^3 = x_4^4 = \dots = x_n^n = x$$

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_n x^n$$

So, by considering $x_1^1 = x, x_2^2 = x^2, \dots, x_n^n = x^n$

polynomial regression

2024/2/22 11:21

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

Linear Regression
(Multiple)

↳ The least square method is a mathematical technique that allows the analyst to determine the best way of fitting a curve on top of chart of data points. It is widely used to make scatter plots easier to interpret & is associated with regression analysis.

↳ 'y' dependent variable must be a non linear function of the parameters ' θ ', not necessarily the features x .
Can be expressed as

$$y = \theta_0 + \theta_1 x^2 ; \quad y = \theta_0 + \theta_1 \theta_2 x$$

$$y = \log(\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3) ; \quad y = \frac{\theta_0}{1 + \theta_1^{(x-\theta_2)}}$$

$$e^y = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$

16|Q|23 and many eq? - and many more eq?

↳ Least squares is the method of estimating the unknown parameter and the Linear regression model by minimizing the sum of squares of the differences between y and \hat{y} .

↳ Least squares method is a statistical procedure to find the best fit for a set of data points. It is used to predict the behaviour of dependent points.

16/7/23

3. Decision Trees, KNN, Logistic Regression, Support Vector Machine

- ↳ Parametric methods is used to determine a probability model that is used in ML(Machine Learning) using/from a set of fixed parameters.
- ↳ Parametric method classification is completely depends on presumptions that are made about a population.
- ↳ Parametric statistics are based on assumptions about the distribution of population from which sample was taken. Non-Parametric statistics are not based on assumptions that is the data can be collected from a sample that does not follow a specific distribution.
- ↳ Non parametric methods are also known as distribution free methods.
- ↳ Non parametric methods are statistical methods that do not rely on assumptions that the data are drawn from a given probability distribution.
- ↳ Non parametric are often applied when less is known about the data.

Decision Trees :-

- ↳ It is a non-parametric supervised learning algorithm which is utilized for both classification & regression tasks.
- ↳ It is a Supervised Learning technique that can be used for both classification & Regression problems , but mostly it is preferred for solving classification problems.
- ↳ It has a hierarchical tree structure which consists of a root node , branches , internal nodes and leaf nodes.
- ↳ Decision trees usually mimic human thinking ability while making a decision , so it is easy to understand.
- ↳ It is a graphical representation for getting all the possible solutions to a problem/decision based on given condition.

→ The decision trees can be easily understood because it

→ shows a tree like structure.
→ It is one of the predictive modelling approach in statistics, determining & ML.

Root Node :-

- It is from where the decision tree starts.
- It represents the entire dataset which further gets divided into two or more homogeneous sets.
- Subtree is just like a small portion of a graph is called subgraph.
- Sub section of this decision tree is called subtree.
- The root node is the starting point of the tree, and both root & leaf nodes contain questions or criteria to be answered.

Branches :-

→ They are the arrows connecting nodes, showing the flow from question to answer

→ A tree formed by splitting on the tree.

→ It represents conjunctions of features that lead to those class labels. It represents the decision rules.

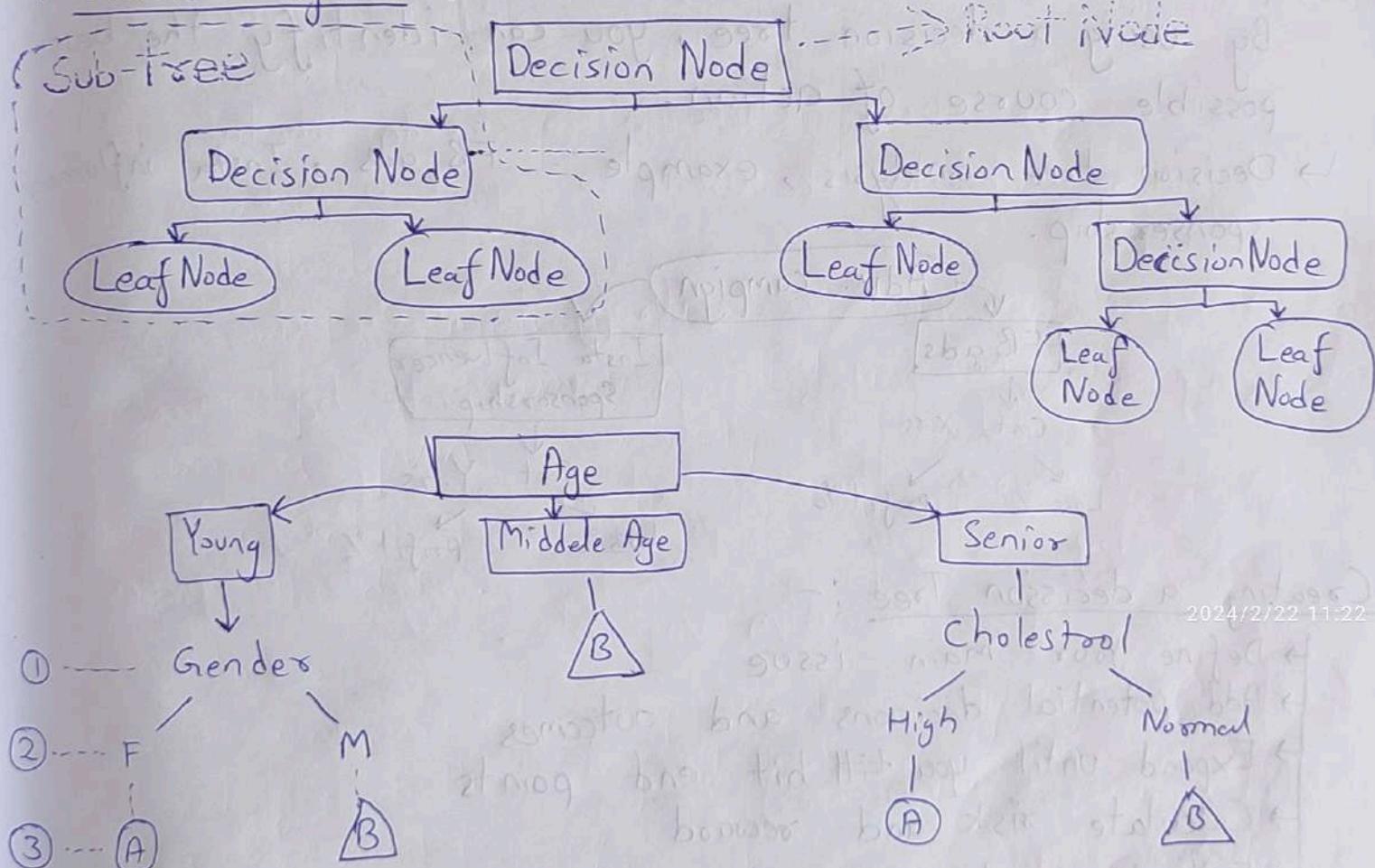
Internal Nodes :-

- A node that symbolize a choice regarding an input feature.
 - Each internal node (non-leaf node) denotes a test on an attribute, each branch represents an outcome of the test & each leaf node holds a class label.
- The top most node is the root node.

Leaf Nodes :-

- Leaf node are the final output node, and the tree cannot be segregate further after getting a leaf node.

- They are the nodes that don't split into more nodes.
- To use a classification tree, start at the root node & traverse the tree until you reach a leaf (terminal) node.
- Leaf nodes of a decision tree are labeled with a class or category.
- ↳ The root node of the tree is called parent node & other nodes are called child nodes.
- ↳ Block Diagram:



Here,
① is an Internal Node corresponds to a test

② is a branch corresponds to a result of the test

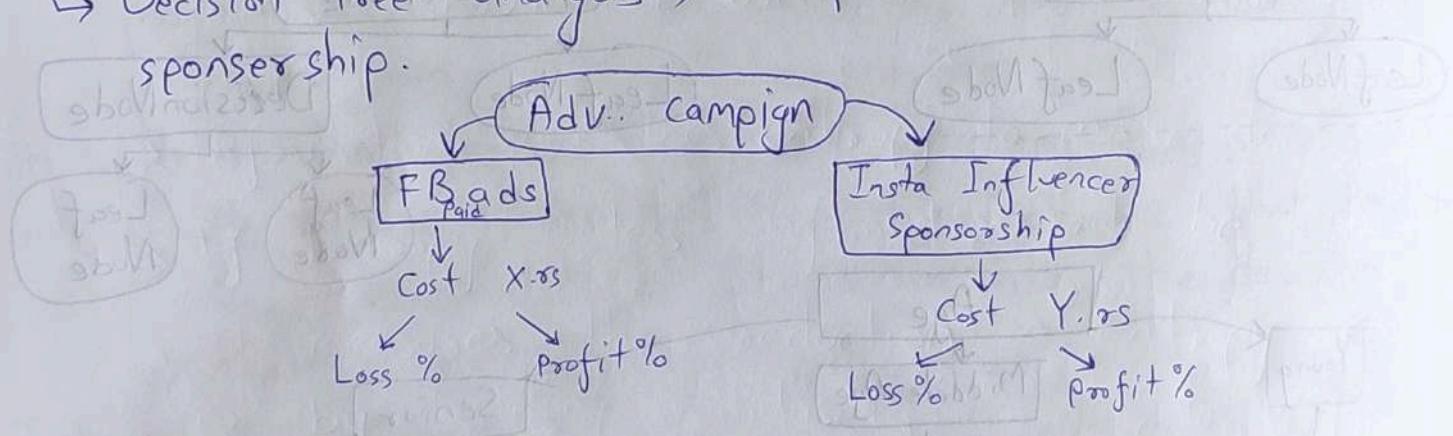
③ is a leaf node assigns a classification

- ↳ To build a decision tree:
- Step 1: choose an attribute from your dataset.
- Step 2: calculate the significance of attribute in splitting of data.
- Step 3: split the data based on the value of the best attribute.
- Step 4: Go to step 1

18/7/23

- ↳ A decision tree is a flowchart that starts with one main idea - or question - and branches out with potential outcomes of each decision. By using a decision tree, you can identify the best possible course of action.

- ↳ Decision tree analysis, example: - FB ads, Insta influencer sponsorship.



Creating a decision Tree :-

- Define your main issue
- Add potential decisions and outcomes
- Expand until you hit end points
- Calculate risk and reward
- Evaluate outcome.

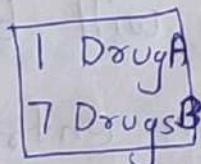
- Which attribute is the best? - When Less impurity & lower Entropy
- A choice of attribute to split the data is to get purity after splitting the data based on More predictiveness important attribute

- A node of tree is considered a pure node if 100% of cases the nodes fall into a specific category of a target field.
- The method used recursive partitioning to split the training records into segments by minimizing the impurity at each step.
- Impurity at nodes are calculated by entropy of the data in the node.

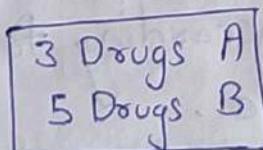
Entropy :- It is the amount of information disorder or amount of randomness or uncertainty in the data.

- ↳ It depends on the how much random data is in that node. It is calculated for each node.
- ↳ The lower the Entropy, the less uniform the distribution, the purer the node.
- ↳ It is also defined as a measure of disorder or impurity in a node.

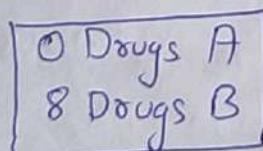
ex:-



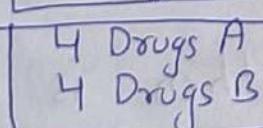
Entropy is Low (↓) ✓



Entropy is High (↑) ✗



Entropy = 0 ✓



Entropy = 1 ✗

↳ Formula :

$$E = - (P(v) * \log(P(v)) + P(x) * \log(P(x)))$$

$$P(v) = \frac{\text{Count of } 'v'}{\text{Total examples}}$$

say, $P(v) = \frac{2}{4}$ ~~0.5~~

$P(x) = \frac{2}{4}$

$$E = - (0.5 * \log(0.5) + 0.5 * \log(0.5))$$

$E = 1$

2024/2/22 11:22

↳ The tree with the higher Information Gain after splitting.

Information Gain :- It is the information that can increase the level of certainty after splitting.

↳ Formula :

$$\text{Information Gain} = \left(\text{Entropy of a tree before split} \right) - \left(\text{Weighted entropy after split} \right)$$

$$\text{Weighted Entropy} = \frac{1}{\text{Information Gain}}$$

↳ It is the basic criterion to decide whether a feature should be used to split a node or not.

↳ Information gain uses entropy to make decisions.

↳ If Entropy is less, information will be more.

↳ It measures the reduction in entropy or surprise by splitting a dataset according to a given value of a random variable.

19/07/2023

KNN (k Nearest Neighbours) :-

↳ The k-Nearest neighbours algorithm is also known as KNN or k-NN.

→ It is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

- KNN is mostly used for classification.
- k-means clustering is a method for vector quantization originally from signal processing that aims to partition n observations into k clusters
- k-means clustering also known as flat clustering algorithm.
- k-means is the fastest unsupervised ML (Machine Learning) algorithm to breakdown data points into groups even when very little information is available.
- k-means is a good choice for large dataset
- ↳ KNN is an classification or regression ML (Machine Learning) algorithm while k-means is a clustering ML (Machine Learning) algorithm.
- ↳ KNN is a lazy learner while k-means is a eager learner.

Classification

- It is supervised learning approach.
- It is a way of categorizing some unknown items into a discrete set of categories or classes.
- The target variable / target attribute in classification is a categorical variable with discrete values.
- The classification in KNN is a supervised ML (Machine Learning) method tries to predict the correct label of a given input data.

20/7/23

- ↳ kNN (k-Nearest Neighbors) algorithm is a classification algorithm that takes bunch of label points and uses them to learn how to label other points.
- ↳ KNN method for classifying cases based on their similarity to other cases.

- ↳ In kNN datapoints or cases that are near each other are said to be "neighbors".
- ↳ Based on similar cases with same class labels are near each other.

Classification in kNN algorithm :-

- ↳ We need to follow steps while doing classification in kNN algorithm -

step 1 : Pick a value of 'K'

step 2 : calculate the distance of unknown case from all cases.

step 3 : Select the K-observations in the training data that are "nearest" to the unknown data points.

step 4 : Predict the response of the unknown data points using the most popular response value from the K-nearest neighbors.

→ kNN(k-Nearest Neighbor) algorithm can also be used for regression.

→ kNN regression is a non-parametric method that, in an intuitive manner, approximates the association between independent variables and the continuous outcome by averaging the observations in the same neighbourhood.

→ We use kNN in regression because kNN models are easy to implement and handle non-linearities well.

→ kNN algorithm is a simple, supervised ML(Machine Algorithm) that can be used to solve both classification and regression problems.

21 7 23

Evaluation Metrics in classification :-

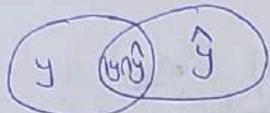
- ↳ Evaluation metrics are used to measure the quality of the statistical or ML (Machine Learning) model.
 - ↳ Evaluation metrics are quantitative measures used to assess the performance and effectiveness of a statistical or ML model.
 - ↳ These metrics provide insights into how well the model is performing and help in comparing different models or algorithms.
 - ↳ Evaluation metrics explain the performance of the model.
 - Model evaluation metrics is the process of using different evaluation metrics to understand a ML (Machine Learning) model's performance as well as its strengths and weakness.
 - Model evaluation is important to assess the efficiency of a model during initial research phase and it also plays a role in model monitoring.
 - ↳ Different Model evaluation metrics :
(accuracy measurements)
 - Jaccard Index
 - F1-Score
 - Log Loss

Jaccard Index :-

- Jaccard Index :-

 - It is also known as Jaccard similarity coefficient.
 - It is defined as the size of the intersection divided by the size of the union of two label sets.
 - It is used to compare set of predicted label for a sample to the corresponding set of labels in y_{true} .

ex:-



where,

$$y = [0, 0, 0, 0, 1, 1, 1, 1] \rightarrow \text{Actual Label}$$

$$\hat{y} = [1, 1, 0, 0, 0, 1, 1, 1] \rightarrow \text{Predicted Label}$$

so,

$$J(y, \hat{y}) = \frac{|y \cap \hat{y}|}{|y \cup \hat{y}|}$$

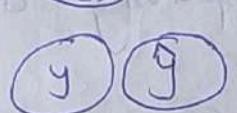
$$\text{or } J(y, \hat{y}) = \frac{|y \cap \hat{y}|}{|y \cup \hat{y}|} = \frac{|y \cap \hat{y}|}{|y| + |\hat{y}| - |y \cap \hat{y}|}$$

then, $J(y, \hat{y}) = \frac{8}{10+10-8} = 0.66$

If



then, $J(y, \hat{y}) = 1.0$



then, $J(y, \hat{y}) = 0.0$

F1-Score :-

→ F1 score is a ML (Machine Learning) evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model.

→ Accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$F1 = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

$$F1 = 2 * \left(\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \right)$$

→ Confusion matrix is a matrix that summarizes the performance of a ML (Machine Learning) model on a set of test data.

It is often used to measure the performance of a classification models, which aims to predict a categorical label for each input instance.

- ↳ Confusion matrix is a table that is used to define the performance of a classification algorithm. It visualizes and summarizes the performance of a classification algorithm.

- ↳ As we discussed earlier,

Precision	$Precision = \frac{TP}{TP + FP}$
Recall	$Recall = \frac{TP}{TP + FN}$
F1-score	$F1\text{-score} = \frac{(2 * Precision * Recall)}{(Precision + Recall)}$

TP is True Positive

FP is False Positive

FN is False Negative

If F1 is near to 0 is considered as low accuracy
F1 is nearer to 1 is considered as high accuracy

Log Loss :-

- ↳ It is also known as logarithmic Loss
- ↳ It is indicative of how close the prediction probability is to the corresponding actual/true value (0 or 1 in case of binary classification)
- ↳ Lower Log-loss indicates better model performances.

$$\text{Log Loss} = -\frac{1}{n} \sum \left([y * \log(\hat{y})] + [(1-y) * \log(1-\hat{y})] \right)$$

$y \rightarrow$ Actual Labels

$\hat{y} \rightarrow$ Predicted Labels

if Log Loss is nearer to 0 is considered as high accuracy.
Log Loss is nearer to 1 is considered as low accuracy.

22/7/22

Logistic Regression :-

- It is a classification algorithm for categorical variables.
- It is an example of supervised learning.
- It is used to calculate or predict the probability of a binary (yes:1/no:0) event occurring.
- Logistic Regression Algorithm : It is a ML (Machine Learning) algorithm that is used to predict the probability of certain classes based on some dependent variables.
- Logistic regression model computes a sum of the input features & calculates the logistic of the result.
- It is used for predicting the categorical dependent variable using a given set of independent variables.
- Linear regression is used for solving regression problems whereas, Logistic regression is used for solving the classification problems.
- It has the ability to provide probabilities and classify new data using continuous & discrete datasets.

Application of Logistic Regression :

- Predicting the probability of a person having heart attack.
- Predicting the mortality in injured patients.
- Predicting a customer's propensity to purchase a product or halt a subscription.
- Predicting the probability of failure of a given process or product.
- Predicting the likelihood of a homeowner defaulting on a mortgage.
- Predicting whether an email is spam or not.
- Forecasting future opportunities & threats.

- Logistic regression is useful when the response variable is binary but the explanatory variables are continuous.
- Non Linear problems can't be solved with logistic regression because it has a linear decision surface.
- Logistic regression is suitable if your data is binary i.e. 0/1, Yes/No, True/False

If you need probabilistic results

When you need a linear decision boundary.

If you need to understand the impact of a feature.

24/07/23

Logistic Regression v/s Linear Regression :

- Logistic regression is the type of statistical model which is also known as logit model & is often used for classification and predictive analytics.
- Logistic regression is used to calculate the probability of a binary event occurring and to deal with issues of classification.
- Linear Regression analysis is used to predict the value of a variable based on the value of another variable.
- Linear Regression is a statistical method that is used for predictive analysis.
- Linear regression shows a linear relationship between a dependent (y) & one or more independent variables.
- Linear regression is used to predict the continuous dependent variable using a given set of independent variables.
- Logistic regression is used to predict categorical dependent variable using a given set of independent variables.
- Linear Regression is used for solving regression problems.

→ Application of Linear regression : Business insights ,
Market analysis, Financial risk assessment

Applications of Logistic regression : Medicine , credit scoring
Hotel booking, Gaming, gaming, Text editing

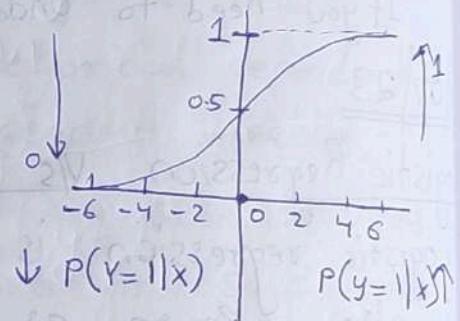
Sigmoid function : Sigmoid function performs the role of an activation function in ML(Machine Learning) which is used to add non-linearity in a ML model.

This function determines which value to pass as output & what not to pass as output.

Formula -

$$\sigma(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

$$\begin{aligned} \therefore \sigma(\theta^T x) &= 1; x \uparrow \\ \therefore \sigma(\theta^T x) &= 0; x \downarrow \end{aligned}$$



$$\downarrow P(Y=1|x) \quad P(Y=0|x)$$

$$f(x) = \frac{L}{1 + e^{-k(x - x_0)}}$$

$$S(x) = \frac{1}{1 + e^{-t}}$$

Logistic

Sigmoid

→ Sigmoid function is a special case of the logistic function when $L=1, k=1, x_0=0$.

→ Sigmoid function also known as squashing function as its domain is the set of all real numbers and its range is $(0,1)$.

If the input to function is either very large negative number or a very large positive number, the output is always between 0 and 1.

↳ ex:- $P(Y=1|X)$ & $P(Y=0|X)$

$$P(Y=1|X) = 1 - P(Y=0|X)$$

so, $\sigma(\theta^T x) \rightarrow P(Y=1|X)$
 $\sigma(\theta^T x) \rightarrow 1 - P(Y=0|X)$
 $P(Y=0|X) = 1 - \sigma(\theta^T x)$

then, Probability of
true or false
in given set X'

Training process for Sigmoid function :

→ Step 1: Initialize ' θ '.

say,

$$\theta = [-1, 2]$$

Step 2: calculate $\hat{y} = \sigma(\theta^T x)$ for a customer.

$$\hat{y} = \sigma([-1, 2] * [2, 5]) = 0.7$$

Step 3: Compare the o/p ' \hat{y} ' with actual o/p of customer, ' y ' & record it as error.

$$\text{Error} = 1 - 0.7 = 0.3$$

Step 4: Calculate the error for all customers

Step 5: Change the ' θ ' to reduce cost

Step 6: Go back to 'Step 2'.

Logistic Regression Training :

→ The main objective of training a logistic regression is to change the parameter of the model so as to be the best estimation of labels of the samples in the dataset.

→ A cost function is a measure of how well a ML (Machine Learning) model performs by quantifying the difference between predicted and actual output. Its goal is to be minimized by adjusting the model's parameters during training.

It is used to estimate how badly models are performing.

Formula -

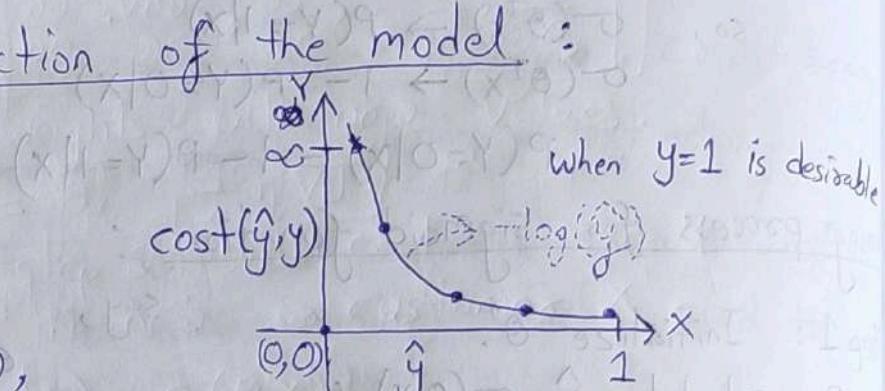
$$\text{Cost}(\hat{y}, y) = \frac{1}{2} [\sigma(\theta^T x) - y]^2$$

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \text{cost}(\hat{y}, y)$$

To set the parameters as best, need to minimize the cost function.

Plotting the cost function of the model :

↳ Model \hat{y} :



Actual value $y=1 \text{ or } 0,$

If $y=1 \& \hat{y}=1 \text{ then cost}=0.$

$y=1 \& \hat{y}=0 \text{ then cost is large.}$

↳ so, we will replace ~~the~~ cost function :

$$\cdot \text{cost}(\hat{y}, y) = \frac{1}{2} [\sigma(\theta^T x) - y]^2$$

$$\boxed{\text{cost}(\hat{y}, y) = \begin{cases} -\log(\hat{y}) & \text{if } y=1 \\ -\log(1-\hat{y}) & \text{if } y=0 \end{cases}}$$

$$\cdot J(\theta) = \frac{1}{m} \sum_{i=1}^m \text{cost}(\hat{y}_i, y_i)$$

$$\boxed{J(\theta) = -\frac{1}{m} \sum_{i=1}^m y_i \log(\hat{y}_i) + (1-y_i) \log(1-\hat{y}_i)}$$

↳ Using this function to define/find parameters of the model in such a way to minimize the cost.

→ To find the best parameters for our model we have to minimize the cost function.

→ To ~~find~~ minimize the cost function we have to find Gradient Descent.

$$\frac{1}{m} = (\theta^T \theta) + \alpha$$

$$\left(\theta^T \theta + \alpha \right) \frac{1}{m} = (\theta)^T$$

Gradient Descent → It is a technique to use the derivative of a cost function to change the parameters value in order to minimize the cost.

It is an optimization algorithm, which is commonly used to train machine learning models & neural networks.

→ To calculate the gradient of a cost function at a point

$$\boxed{\frac{\partial J}{\partial \theta_j} = -\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i) \cdot x_i^j}$$

Gradient is the slope of a surface at the point & direction of the point is the direction of the uphill.

at specific point θ , $\frac{\partial J}{\partial \theta}$ is slope at point

if $\frac{\partial J}{\partial \theta}$ is +ve which indicates $J(\theta)$ is increasing

$$\nabla J = \begin{bmatrix} \frac{\partial J}{\partial \theta_0} \\ \frac{\partial J}{\partial \theta_1} \\ \vdots \\ \frac{\partial J}{\partial \theta_k} \end{bmatrix}$$

$$\text{so, } \boxed{\text{New } \theta = \text{Old } \theta - \alpha \nabla J}$$

→ Step 1: Initialize the parameters randomly.

$$\theta^T = [\theta_0, \theta_1, \theta_2, \dots]$$

Step 2: Feed the cost function with training set & calculate the error.

$$\boxed{J(\theta) = -\frac{1}{m} \sum_{i=1}^m y_i \log(\hat{y}_i) + (1-y_i) \log(1-\hat{y}_i)}$$

Step 3: Calculate the gradient of cost function.

$$\nabla J = \left[\frac{\partial J}{\partial \theta_1}, \frac{\partial J}{\partial \theta_2}, \frac{\partial J}{\partial \theta_3}, \dots, \frac{\partial J}{\partial \theta_k} \right]$$

Step 4: Update weights with new values.

$$\theta_{\text{new}} = \theta_{\text{previous}} - \eta \nabla J$$

Step 5: Go to step 2 until cost is small enough.

Step 6: Predict the new customer 'X'.

$$P(Y=1|X) = \sigma(\theta^T X)$$

26/7/23

Support Vector Machine :

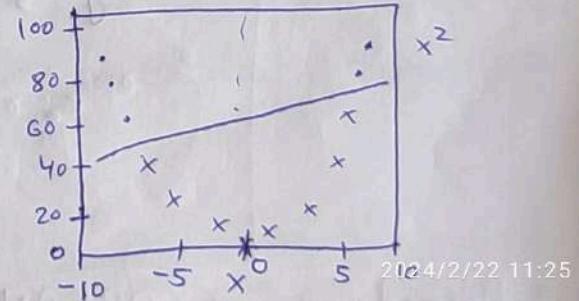
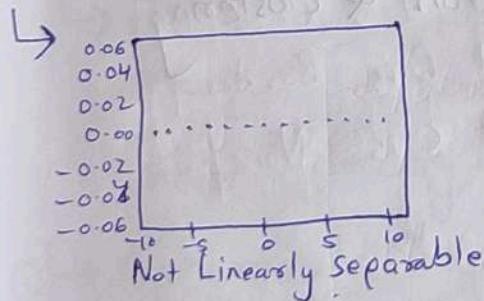
- ↳ SVM is also known as Support Vector Machine (SVM).
- ↳ It is an powerful ML (Machine Learning) algorithm used for linear or non-Linear classification, regression & even outlier detection tasks.
- ↳ It is a type of DL (Deep Learning) algorithm that performs supervised learning for classification or regression of data groups.
- ↳ In ML (Machine Learning), SVM (Support Vector Machines) are Supervised Learning models with associated learning algorithms that analyze data for classification & regression analysis.
- ↳ SVM can be used for a variety of tasks such as text classification, image classification, spam detection, hand writing identification, gene expression analysis, face detection & anomaly detection.
- ↳ SVM is a supervised algorithm that classifies case by finding a separator.
 1. Mapping data to a high dimensional feature space
 2. Finding a separator.

Data transformation :-

- It is also known as data preparation or data pre-processing.
 - It makes sure that your data is clean and ready to be used by your ML (Machine Learning) algorithm.
 - Kernelling is used to a set of mathematical functions used in SVM (Support Vector Machine) providing the window to manipulate the data.
 - Kernel function generally transforms the training set of data so that a non linear decision surface is able to transform to a linear equation in a higher number of dimension spaces.
- Mapping the data into higher dimensional spaces is called Kernelling.

$$\phi(x) = [x, x^2]$$

→ Hyperplane

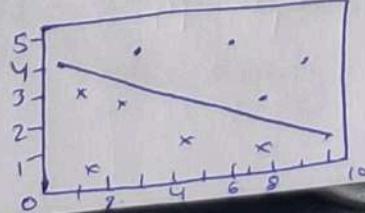


- Kernel function can be of different types:
Linear, Polynomial, RBF (Radio bases function), Sigmoid.

RBF is a popular kernel function used in various Kernelized learning algorithms.

- SVM (Support Vector Machine) are based on idea in finding the hyperplane that best divides the dataset into two classes

ex:-



↳ Hyperplane is a decision boundary that differentiates the two classes in SVM.

↳ Advantages of SVM:

↳ Accurate in high dimensional spaces

↳ Memory efficient

↳ Disadvantages of SVM:

↳ Prone to over fitting

↳ No probability estimation

↳ Small datasets

↳ Applications of SVM:

↳ Image recognition

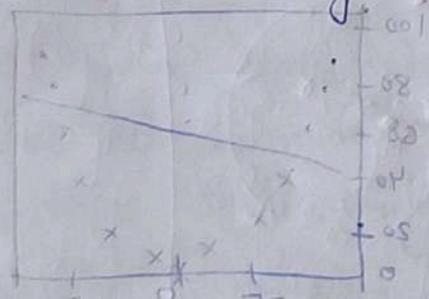
↳ Text category assignment

↳ Detecting spam

↳ Sentiment analysis

↳ Gene expression classification

↳ Regression, outlier detection & clustering



27/7/23

4. Density Based Clustering, Hierarchical clustering & k-Means clustering

Density Based Clustering :-

- ↳ Density based clustering refers to a method that is based on local cluster criterion such as density connected points.
- ↳ It refers to unsupervised ML (Machine Learning) methods that identify distinctive clusters in the data based on the idea that a cluster/group in a data space is a contiguous region of high point density, separated from other clusters by sparse regions.



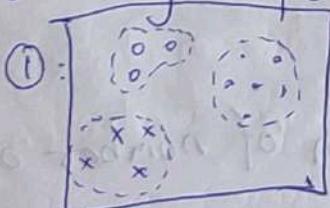
database 1



database 2

- ↳ It is an unsupervised machine learning algorithm that groups similar data points in a dataset based on their density.

- ① → Spherical Shape clustering : They are dense & consist almost exclusively of elliptical & so galaxies.
- ② → Arbitrary Shape clusters :



- ↳ k-Means clustering : k-means assigns all points to a cluster even if they do not belong to any.

Density Based Clustering : Density based clustering locates regions of high density & separated outliers.

↳ DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

- It is one of the most common clustering algorithms
- Works based on density of objects.

R (Radius of neighborhood)

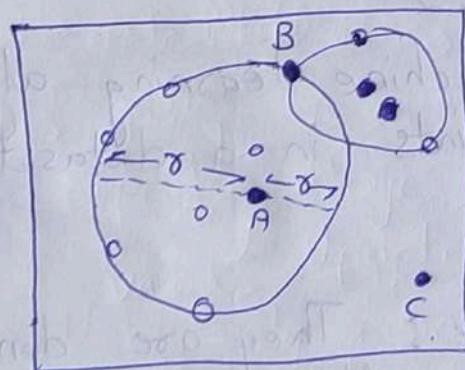
- Radius(R) that if includes enough number of points within, we call it a dense area.

M (Minimum number of neighbors)

↳ A point is a core point if it has more than M points within ϵ .

A point which has fewer than M points within ϵ but it is in the neighborhood of a core point.

Noise or outlier: A point which is not a core point or border point.



Point A is Core point
point B is Border point
point C is Noise/outlier

2024/2/22 11:25

↳ Advantages of DBSCAN:

- Arbitrarily shaped clusters
- Robust to outliers
- Does not require specification of number of clusters.

28/7/23

Hierarchical Clustering

Hierarchical clustering :-

- ↳ Hierarchical clustering algorithms build a hierarchy of clusters where each node is a cluster consists of the clusters of its daughter node.
- ↳ It is a popular method for grouping objects.
- ↳ It creates groups so that objects within a group are similar to each other & different from objects in other groups.
- ↳ Clusters are visually represented in a hierarchical tree called a dendrogram.
- ↳ Strategies for hierarchical clustering generally fall in two types :
 - Divisive
 - Agglomerative

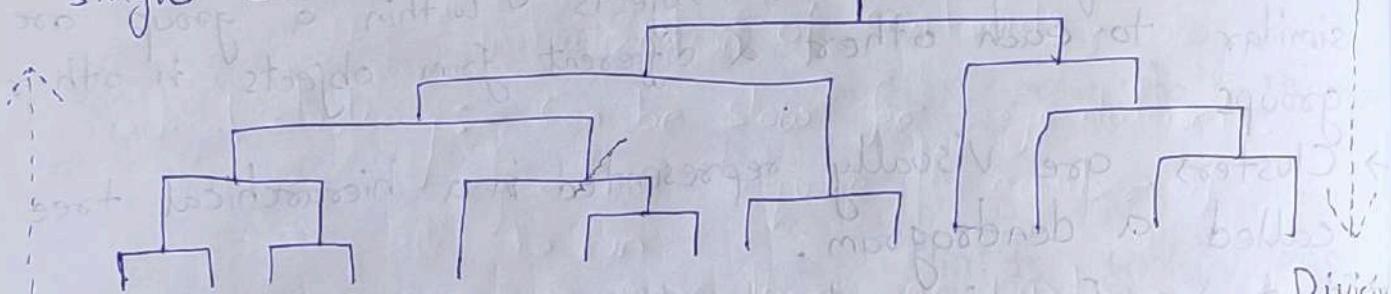
Divisive Hierarchical clustering : It works just opposite to the agglomerative clustering.

- ✓ ↳ It starts by considering all the data points into a big single cluster & later on splitting them into smaller heterogeneous clusters continuously until all data points are in their own cluster.
- ✗ ↳ ~~Hierarchical clustering~~ is a connectivity based clustering model that groups the data points together that are close to each other based on the measure of similarity or distance.
- ✓ ↳ It is also known as top-down approach.
Top-down clustering requires a method for splitting a cluster that contains the whole data and proceeds by splitting clusters recursively until individual data have been split into singleton clusters.

2024/2/22 11:26

Agglomerative Hierarchical clustering :

- It is also known as bottom-up approach or Hierarchical agglomerative clustering (HAC).
- Bottom-up algorithms treat each data as a singleton cluster at the outset & then successively agglomerate pairs of clusters until all clusters have been merged into a single cluster that contains all data.



Agglomerative

Agglomerative algorithm :

- These algorithms produce a sequence of clustering of decreasing number of clusters, M, at each step.

Step 1: Create 'n' clusters, one for each data point

Step 2: Compute the proximity Matrix

Step 3: Repeat - Merge two closest clusters
- Update the proximity matrix

Step 4: Until only a single cluster remains -

- Dataset to Dissimilarity matrix :

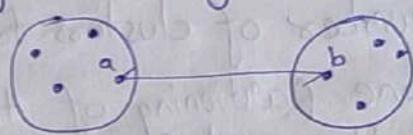
Dissimilarity matrix - also called distance matrix.

It describes pairwise distinction between M objects.

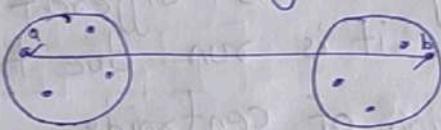
It is a square symmetrical $M \times M$ matrix with the ij^{th} element equal to the value of a chosen measure of distinction between the i^{th} & j^{th} object.

Distance between clusters :

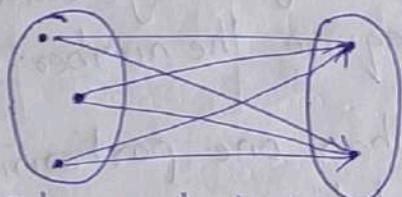
↳ Single-Linkage Clustering → It is a minimum distance b/w clusters



↳ Complete-Linkage clustering → It is a maximum distance b/w clusters



↳ Average-Linkage clustering → It is a average distance b/w clusters.



↳ Centroid Linkage clustering → It is a distance b/w cluster centroids



Advantages & Dis-Advantages of Hierarchical clustering :

↳ Advantages : - It doesn't require number of clusters to be specified
- It is easy to implement
- Produces a dendrogram which helps with understanding the data.

↳ Dis-Advantages : - Can never undo any previous steps thought the algorithm.
- Generally has long runtimes
- Sometimes difficult to identify the number of clusters by the dendrogram.

Hierarchical clustering v/s k-means clustering

↳ k-means clustering :-

- Much more efficient
- Requires number of clusters to be specified
- Gives only one partitioning of the data based on the predefined number of clusters
- Potentially returns different clusters each time it is run due to random initialization of centroids.

↳ Hierarchical clustering :-

- Can be slow for large datasets.
- Does not require the number of clusters to run
- Gives more than one partitioning depending on the resolution.
- Always generates the same clusters.

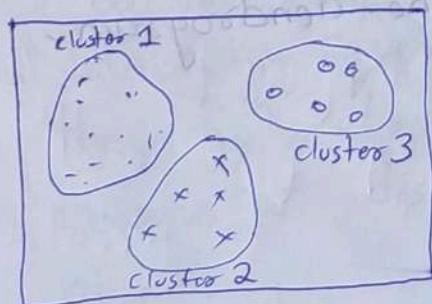
Clustering & K-Means clustering & v/s classification :-

↳ Clustering is the act of organizing similar objects into groups within a ML (Machine Learning) algorithm.

↳ Clustering has many uses in data science, like image processing, knowledge discovery in data, unsupervised learning & various other applications.

↳ Grouping unlabelled data is called clustering.

↳ Cluster is a group of objects that are similar to the other objects in the cluster & dissimilar to data points in other clusters.



↳ Classification : It is a supervised ML (Machine Learning) method where the model tries to predict the correct label of a given input data.

The model is fully trained using the training data & then it is evaluated on test data before being used to perform prediction on new unseen data.

Clustering Applications :

- ↳ Retail / Marketing —
 - Identifying buying patterns of customers
 - Recommending new books or movies to new customers
- ↳ Banking —
 - Fraud detection in credit card use
 - Identifying clusters of customers (e.g.: loyal)
- ↳ Insurance —
 - Fraud detection in claims analysis
 - To evaluate insurance risk of customers
- ↳ Publication —
 - Auto categorizing news based on their content
 - Recommending similar news articles
- ↳ Medicine — Characterizing patient behaviour
- ↳ Biology — Clustering genetic markers to identify family ties.

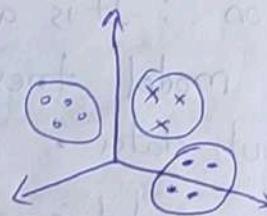
Why Clus.

- We use clustering for following purposes:
 - Exploratory data analysis
 - Summary generation
 - Outlier detection
 - Finding duplicates
 - Pre processing step

Clustering algorithms

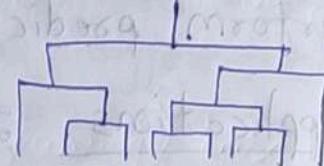
↳ Partitioned based clustering :

- Relatively efficient
- k-Means, k-Median, Fuzzy c-means



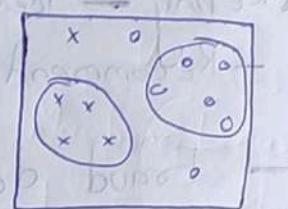
↳ Hierarchical clustering :

- Produces trees of clusters
- ex: Agglomerative, Divisive



↳ Density-Based clustering :

- Produces arbitrary shaped clusters
- ex:- DBSCAN



K-Means Clustering :-

↳ It is an unsupervised Learning algorithm which groups the unlabelled dataset into different clusters.

k- defines number of pre-defined clusters that need to be created in the process.

↳ I

K-Means algorithms :-

↳ Partitioning clustering → K-Means divides the data into non-overlapping subsets (clusters) without any cluster-interaction structure.

examples within a cluster are very similar

examples across different clusters are very different.

↳ K-means is a centroid - based clustering algorithm , where we calculate the distance b/w each data point and a centroid to assign it to a cluster.

It tries to group similar kinds of items in form of clusters.

Step 1: Initialize 'k'

Step 2: Calculate the distance

Step 3: Assign to centroid, Assign each point to the closest centroid

SSE = The sum of the squared differences b/w
(centroids) each point and its centroid

$$SSE = \sum_{i=1}^n (x_i - c_j)^2$$

Step 4: Compute new centroids - for each cluster

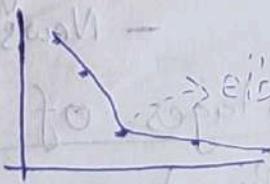
Step 5: Repeat until there are no more changes. Step 2-4

K-Means clustering algorithm :

- Randomly placing k centroids, one for each cluster.
- Calculate the distance of each point from each centroid
- Assign each data point object to its closest centroid, creating a cluster.
- Recalculate the position of the k-centroids
- Repeat the steps 2-4, until the centroids no longer move.

K-means accuracy :

- ↳ External approach - Compare the clusters with the ground truth, if it is available.
- ↳ Internal approach - Average the distance b/w data points within a cluster.
- The elbow method is a graphical representation of finding the optimal 'k' in k-means clustering



28/7/23

5. Recommendation Engines

Recommender Systems :

- ↳ Recommender systems that captures the pattern of people's behaviour and use it to predict what else they might want or like.
- ↳ These are the type of ML (Machine Learning) based systems that are used to predict the ratings or preferences of items for a given user.
- ↳ These are a subclass of Machine Learning which generally deal with ranking or rating products/users.
- ↳ Matrix factorization algorithms are probably the most popular & effective Collaborative filtering methods for recommender systems.
- ↳ Netflix, Amazon & Spotify are well known examples of robust recommendation systems.

Applications of recommender systems

- ↳ Netflix, Amazon, LinkedIn, Facebook, Twitter are major applications.
- ↳ What to buy?
 - E-commerce, books, movies, beers, shoes
- ↳ Where to eat?
- ↳ Which job to apply?
- ↳ Who you should be friends with?
 - LinkedIn, Facebook
- ↳ Personalize your experience on the web
 - News platform, new personalization

Advantages of Recommender Systems :

- ↳ Broader exposure
- ↳ Possibility of continual usage or purchase of products
- ↳ Provides better experience

↳ They are efficient ML (Machine Learning) solutions that can help increase user retention & lead to significant increase in your business revenues.

↳ 5 benefits that businesses can achieve using recommendation engines : Revenue, Customer satisfaction, Personalization, Discovery, Provide Reports

Types of recommendation Systems :

↳ There are 6 types of recommendation systems

- Collaborative-Recommender System
- Content based recommender system
- Demographic based recommender system
- Utility based recommender system
- Knowledge based recommender system
- Hybrid recommender system

↳ Collaborative Recommender System — It recommends items based on similarity measures between users and/or items. The basic assumption behind the algorithm is that users with similar interests have common preferences. It predicts which item an user will like based on the item preferences of other similar users.

↳ Content based recommendation System — Content based system recommends items to the customer similar to previously high rated items by the customer. It uses the features & properties of the item.

It is a supervised ML used to induce a classifier to discriminate between interesting & uninteresting items for the users.

↳ Demographic based recommender system — It categorize the users based on their attributes & recommends the movies by utilizing their demographic data.

- ↳ Utility based recommender system — An utility based recommender system is proposed based on IUV-GA which can classify item data based on central item attributes and calculate utility of items with user preferring attributes and recommend top -N utility items to user.
- ↳ Knowledge based recommender system — It is a specific type of recommender system that are based on explicit knowledge about the item assortment, user preferences & recommendation criteria. It doesn't have a damping up problem since its recommendations do not depend on a base of user rating.

- ↳ Hybrid Recommender System — It is a specific type of system that used data of both collaborative data and content based data simultaneously which helps to suggest a similar or close item to the user.

- ### Implementing recommender System :
- ↳ Memory Based
 - Uses the entire user-item dataset to generate a recommendation.
 - Uses statistical techniques to approximate users or items eg: Pearson correlation, cosine similarity, Euclidean Distance etc..
 - ↳ Model Based
 - Develops a model of users in an attempt to their preferences
 - Models can be created using Machine Learning techniques like regression, clustering, classification etc..

29/7/23

→ Collaborative filtering has basic 2 approaches:

- (i) User based collaborative filtering
- (ii) Item based collaborative filtering

(i) User based collaborative filtering — It is a technique used to predict the items that ~~a~~ user might like on the basis of rating given to that item by other users who have similar taste with that of the target user.

(ii) Item based collaborative filtering — It is a technique or method to predict a user's taste & find the items that ~~a~~ user might prefer on the basis of information collected from various other users having similar tastes or preferences.

We find the missing rating with the help of the rating given to the other items by the user.

Challenges of collaborative filtering :

↳ Data Sparsity — Users in general rate only a limited number of items.

— The term used to describe the phenomenon of not observing enough data in a dataset. To improve database and data processing performance.

2024/2/22 11:27

↳ Cold start — Difficulty in recommendation to new users or new items.

— It describes the difficulty of making recommendation when the users or the items are new.

↳ Scalability — Increase in number of users or items.

— It is a key factor when determining which type of recommender system to use.