KOSHINDIA

KOUSTUBH SHARMA

Machine Learning

Page No.:
Date:

Types

(1) Supervised Learning:

→ Where we have input variables x & output variable y & we use an algorithm to map the input function to the output function.

→ Goal is to approximate the mapping function so well that when we have new x, we could predict y.

Popular Algorithms:

→ Linear Regression
→ Random Forest
→ Support Vector Machine

(2) Unsupervised Learning.

→ Where we only have input data x & no corresponding output variable.

→ Goal is to model the underlying structure or distribution in the data in order to learn more about data.

→ Hence, we have an input but not output. The algorithm detects patterns based on innate characteristics of input data.

Popular Algorithms
→ Apriori Algorithm
→ K-Means
→ Hierarchical Clustering.

KOSH

Page No.:

Date:

(3) Reinforcement Learning

→ It allows software agents and machine to automatically determine the ideal behaviour within a specific context to optimize its performance.

→ It is about interaction b/w environment & learning agent. Learning agent leverages two mechanism, exploration & exploitation.

Exploration = learning agent acts on trial & error basis.

exploitation = when it acts based on the knowledge gained from the environment it is refered to as exploitation.

Environment rewards the learning agent for correct actions, which is reinforcement signal & based on the reinforcement signal, agent improves its knowledge to select the next action.

KOSH

Page No.:

Date:

Algorithms

- Classification Algorithm

A Supervised learning approach in which the computer program learns from the input given to it & uses this learning to classify new observations

- Anomaly Detection Algorithm.

It is the identification of unusual/outliers in the dataset.

- Clustering Algorithm

Grouping the data based on some conditions

- Regression Algorithm.

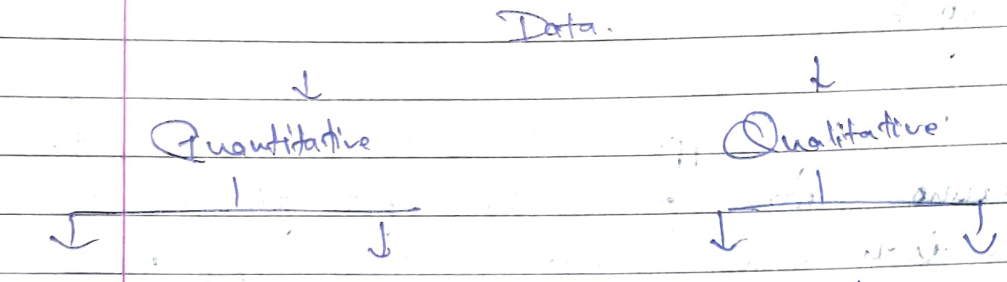
Data itself is predicted. Allows making prediction by learning the relationship between the features of your data & observed continuous valued response.

KOSH

Page No.:

Date:

Statistics.



Qualitative Data.

→ Deals with characteristics & descriptors that cannot be easily measured but can be observed effectively.

Nominal → Data with no inherent order

Ordinal → Data with an ordered series.

Quantitative Data

→ Deals with nos & things that could be measured objectively

Discrete = AKA Categorical data, holds finite no of possible values. (No of students in class)

Continuous = Data can hold infinite no of possible values.
(Weighing scale.

KOSH

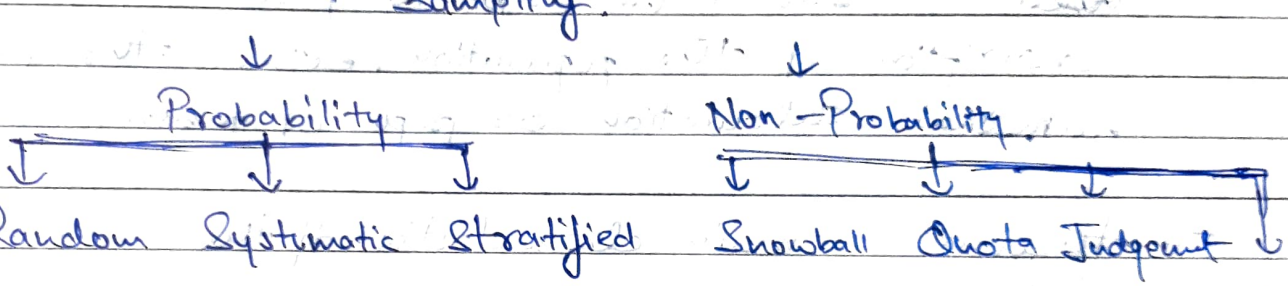
Page No.:

Date:

Population - Collection or set of individuals or objects or events whose properties are to be analysed.

Sample - A subjet of population

Sampling - Is a technique that deals with the selection of individual observations within a population. Used to infer statistical knowledge about population.



Sampling : A sampling is a technique in which samples from a large population are chosen using the theory of probability

Random Sampling : Each member of the population has equal chance being selected in the sample.

Systematic Sampling : Every nth record is choosen to be a part of the sample.

Stratified Sampling : A Stratum is used to form sample. A stratum is a subset of the population that shares at least one common characteristics.

KOSH

Page No.:

Date:

Random sampling is used to select a sufficient no. of subjects from each stratum.

Types of Statistics

① Descriptive Statistics

→ used to summarise the data to provide description of the population, either through numerical calculation or graph or tables.

→ focuses on main characteristics of the data.

② Inferential Statistics.

→ It makes inferences & predictions about a population based on a sample of data taken from the population in question.

→ It generalizes the large dataset & applies probability to draw conclusion

KOSH

Page No.:
Date:

Discriptive Statistics

→ It is a method to describe & understand the features of a specific dataset by giving short summaries about the sample & measure the data.

→ Discriptive statistics are broken down into two categories
- Measure of central tendency (rep. summary of Dataset)
- Measure of variability (spread)

Central tendency
- Mean (Avg of all value)
- Mode (Most recurrent)
- Median (Measure of Centre value)

Measure of Spread
- Raye
- IQR
- Variance & Standard Dev (Difference of each elmnt from mean actual value)

Variance is measure of how much a variable differes from
Tell us about the spread of dataset by breaking it in quartile
How spread apart the value in a dataset are.

Raye = Max(x_i) − Min(x_i)

Quartile
- $Q_1 \Rightarrow 25\%^{th}$ value
- $Q_2 \Rightarrow 50\%^{th}$ value
- $Q_3 \Rightarrow 75\%^{th}$ value.

IQR = $Q_3 - Q_1$

Standard Deviation = $(x_i - M)$, M = Mean.

Variance. $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$

x = idividual data point
n = No of data points
$\bar{x}$ = mean.

μ = Mean of Population N = no of DP in Population.

$\bar{x}$ = Mean of sample n = no of DP in sample.

Page No.:

Date:

KOSH

⇒ Population Variance is avg of Squared Deviation

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} = (x_i - \mu)^2$$

⇒ Sample variance is is the avg of squared differences from the mean.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{N} = (x_i - \bar{x})^2$$

Standard Deviation is measure of despersion of a set of data from its mean.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$$

KOSH

Page No.:
Date:

Information Gain & Entropy.

Entropy : Measure of uncertainity in data.

$$H(S) = -\sum_{i=1}^{N} p_i \log_2 p_i$$

S = Set of all instances in the dataset.
N = Number of distinct class values
p_i = Probability event.

Information Gain : How much information a particular feature / variable gives us about final outcome.

$$\text{Gain}(A,S) = H(S) - \sum_{j=1}^{V} \frac{|S_j|}{|S|} \cdot H(S_j)$$

$$= H(S) - H(A,S)$$

$H(S)$ — Entropy of whole dataset S
$|S_j|$ — No of instances with j value of a attribute A
$|S|$ — Total no of instances in dataset A.
V — Set of distinct values of an attribute A.
$H(S_j)$ - entropy a subset of instance for an attribute A.
$H(A,S)$ - Entropy of an attribute A.

KOSH

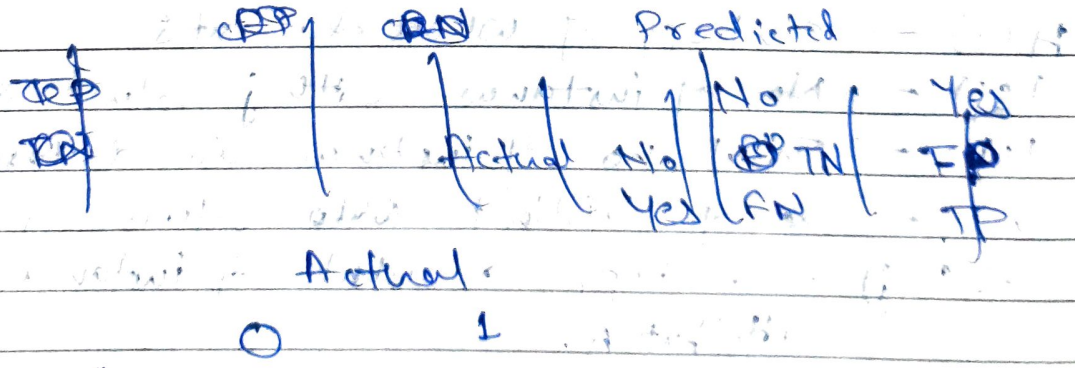
Page No.:
Date:

Use case – Decision Tree.

⇒ Confusion Matrix

- It is matrix used to describe the performance of a classification model or classifier on a set of test data for which the true values are known.
- Represents a tabular representation of Actual vs Predicted values

$$\frac{TP + TN}{TP + TN + FP + FN}$$



Actual

	0	1
Predicted 0	TN	FN
1	FP	TP

KOSH

Page No.:

Date:

Probability

$$P(x) = \frac{\text{Desired}}{\text{Total}}$$

Terminologies:

→ Random Experiment: Experiment for which the outcome cannot be predicted with certainity.

→ Sample Space: The entire possible set of outcomes of a random experiment is the sample space (S) of that experiment.

→ Event: Out comes of an experiment

Event Type:

→ Disjoint: Don't have a common outcome.

→ Non Disjoint: can have common outcome

2:05:45

KOSH

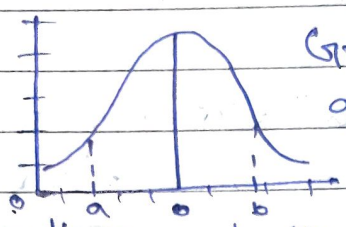
Page No.:
Date:

Probability Distribution

① Probability Density function.

→ It is the equation describing a continuous probability distribution

Properties:



Graph is also known as bell curve.

① Graph of PDF is continuous in range
② Area bounded by the curve of density function & the x axis is equal to 1 (area below the curve)
③ Probability that a random variable assumes a value b/w a & b is equal to the area under PDF bounded by a & b.

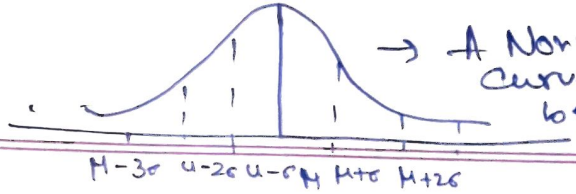
② Normal Distribution (Gaussian Distribution)

→ Normal Distribution is a probability distribution that associates the normal random variable X with cumulative properties

→ It denotes the symmetric properties of the mean. Idea behind is that the data near the mean occurs more frequently than the data away from the mean, or data near the mean represents the entire dataset

→ Graph of Normal Distribution depends on two factors: the Mean & the Standard Deviation.

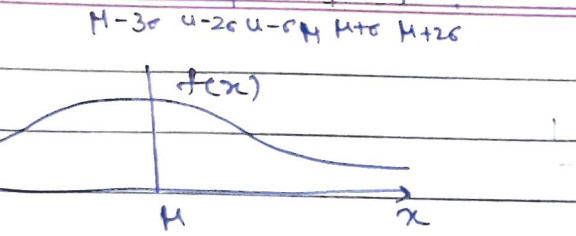
Mean: determines location of center of the graph
Standard deviation: Determines height of the graph.

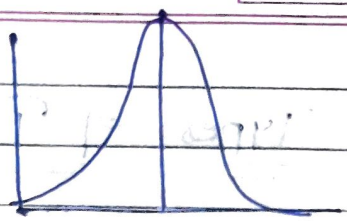


→ A Normal Distribution Curve is also a bell curve.

Page No.:
Date:

KOSH





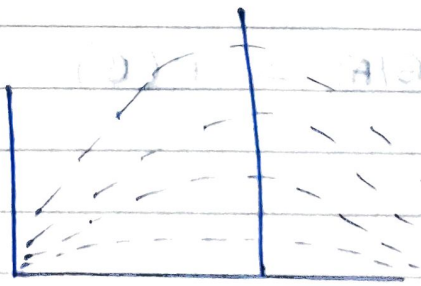
If Standard Deviation is large, curve is short & wide

if Standard Deviation is small, curve is tall & narrow

⇒ **Central Limit Theorem** (only true for large dataset as small dataset has a lot of variations due to scaling factor)

- It states that the sampling distribution of the mean of any independent random variable will be normal or nearly normal, if the sample size is large enough.

- If we have a large population & we divided it into many samples, then the mean of all the samples of the population will be almost equal to the mean of the entire population. Hence, mean is normally distributed. Holds true only for a large data set



Mean of each Sample is almost along the same line.

The accuracy or resemblance depends on two factors:
① No. of sample points considered ② Shape of underlying population ← shape depends on
① Standard Deviation
② Mean of sample.

⇒ Types of Probability

① Marginal Probability.
→ Probability of occurence of single event.

8 Hearts in a deck of card = $\frac{13}{52}$.

② Joint Probability.
→ Measure of two events happening at same time.

Ace of hearts = $\frac{1}{52}$

③ Conditional Probability.
→ Probability of an event based on occurance of a previous event.

An event B will occur given that A has already occured.

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$ (In case of A & B are dependent)

$$P(B/A) = P(B)$$ (In case A & B are Independent)

KOSH

Page No.:
Date:

⇒ Baye's Theorem

- Shows the relationship between one conditional probability & its inverse.
- It is the probability of an event occuring based on prior knowledge of conditions that might be related to the same event.

$$P(A|B) = P(B|A) \cdot P(A) / P(B)$$

P(A|B) is refered to as posterior which means the probability of occurence of A given B.

P(B|A) is refered to as likelihood ratio which measures the probability of occurence of B given A.

P(A) is refered as Prior which represent the actual probability distribution of A.

Eg: we have 3 bowls.

Bowl A = 2 blue 4 red

Bowl B = 8 blue 4 red

Bowl C = 1 blue 3 red.

1 ball is drawn from each. Probability of drawing a blue from A given 2 blue is drawn from Bowl. are drawn in total.

Sol

Event 1 = Picking a blue ball from bowl A. ⇒ A

Event 2 = Picking 2 blue balls. ⇒ X.

$$P(A/X) = \frac{P(A \cap X)}{P(X)} \qquad P(A \cap X) = \frac{{}^{6}C_2 + {}^{12}C_2}{{}^{18}C_2}$$

$$P(X) = \frac{({}^{6}C_1 \times {}^{12}C_1) + ({}^{12}C_1 \times {}^{4}C_1) + ({}^{4}C_1 \times {}^{6}C_1)}{{}^{22}C_2}$$

Page No.:
Date:

⇒ Inferential Statistics & Point Estimation

Point Estimation is concerned with the use of the sample data to measure the single value that that serves as an approximate value or the best estimate of an unknown population parameter.

Population —Random→ Sample.

Sample —Mean→ $\bar{x}$

$\bar{x}$ —used to→ μ

μ : Median / Mode / Mean etc. of Population

Estimator = function of the sample used to find out the estimate.
Eg: Sample mean.

Estimate = calculated value from the function
Eg: population mean.

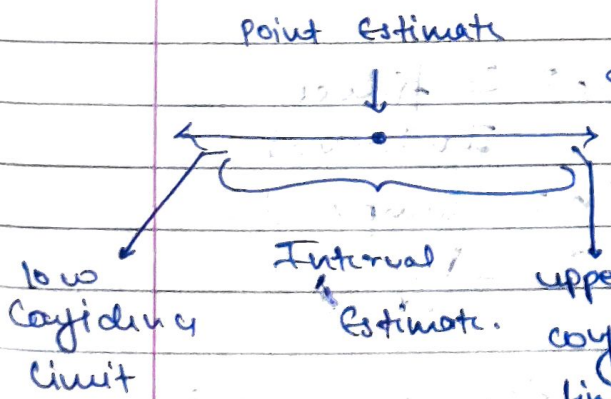
KOSH

Page No.:
Date:

Method to find Estimate.

1. Method of Moments: form an equation in the sample data set & analyze that similar Equation in the population dataset as well.
   - we take the known facts about the Population sample & extend them to sample & using sample then predict the known parameters such as mean, mode etc.
2. Maximum of Likelihood:
3. Baye's Estimator
4. Best Unbiased Estimators.
5. Interval Estimation Method (Imp)

Interval Estimate: An interval, a range of values used to estimate a population parameter.



: If we want to find a value (Mean mode etc) of a population, we set a range & at that rate the value lies.

So, instead of predicting a point we predict a range. in which a value may occur.

Kosh

Page No.:
Date:

→ Difference b/w point estimate & actual population parameter value is Sampling Error

→ When μ is estimated, sampling error $= \mu - \bar{x}$

Confidence Interval

→ It is the measure of your confidence that the interval estimated contains the population parameter

→ It is used to describe the amount of uncertainity associated with a sample estimate of a population parameter.

→ Arange of values so constructed that there is specified probability of including the true value of a parameter within it.

Margin of error - E

→ For a given level of confidence it is the greatest possible distance b/w the point estimate & the value of parameter it is estimating.

$$\left[E = Z_c \frac{\sigma}{\sqrt{n}} \right]$$

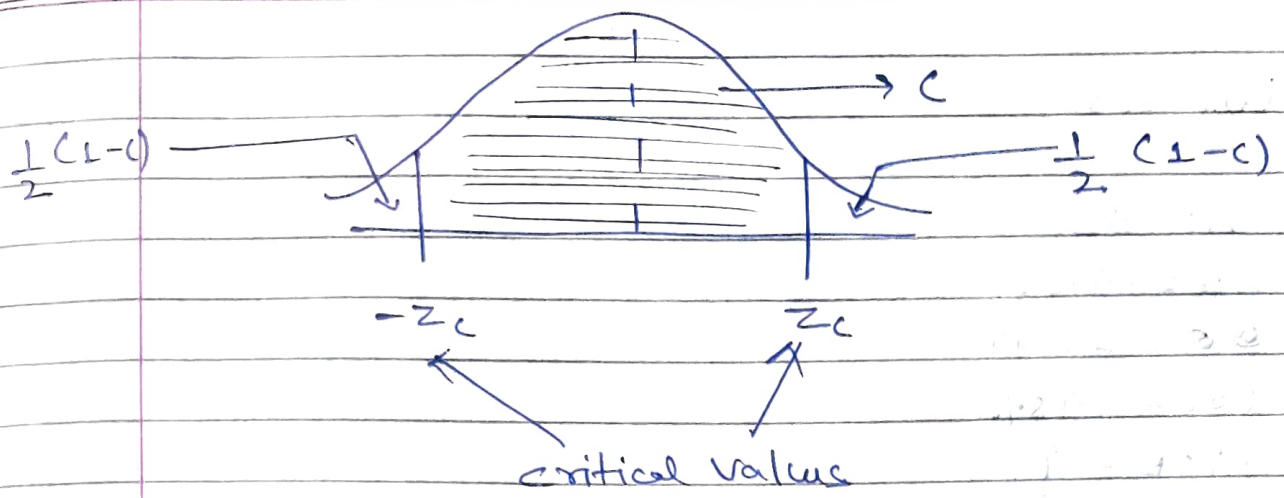
Z_c = Confidence Interval

n = Sample size

σ = S.D.

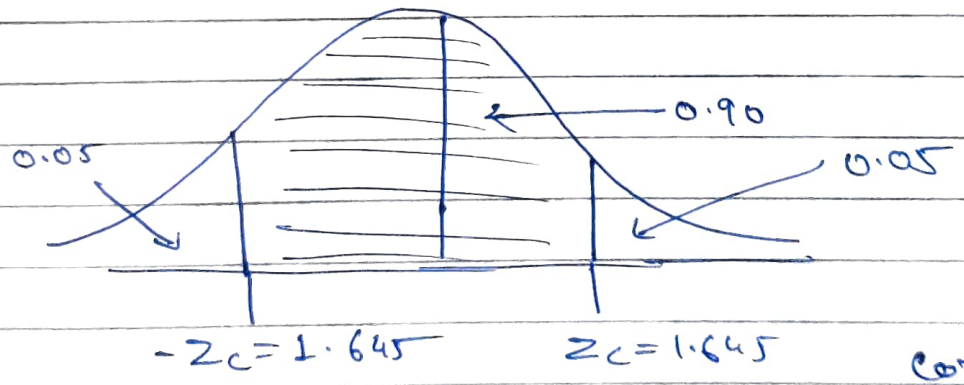
The level of confidence c is the probability that the interval estimate contains the population parameter.

Page No.:
Date:



c is the area beneath the normal curve b/w the critical values. Z-score can be calculated using standard normal table.

↳ level of confidence = 90% i.e. 90% confidence that interval contains population parameter.



Correspondiy
Z score = ± 1.645

Sequence :

1. Sample statistic (used to calculate population parameter)
2. Select confidence level.
3. Margin of error
4. Specify confidence interval

Use Case

$n = 32$

$\bar{x} = 74.22$

$\sigma = 23.44$

(z) LoC = 95%.

(E) MoE = ?

$$\Rightarrow E = z_c \frac{\sigma}{\sqrt{n}}$$

$$E = 1.96 \frac{23.44}{\sqrt{32}} \approx 8.12$$

KOSH

Page No.:
Date:

Hypothesis Testing

→ Used to check whether hypothesis is accepted or rejected.

Steps :

① State the hypothesis : Stating null & alternative hypothesis.

② Formulate an analysis plan : This stage involve construction of analysis plan.

③ Analyse Sample Data : Calculation & interpretation of the test statistic as described in the analysis plan.

④ Interpret Result : The application of the decision rule described in the analysis plan.

NULL HYPOTHESIS : Result is no different from assumption (H0)

ALTERNATIVE HYPOTHESIS : Result disapproves the assumption.

Eg: 4 boys Nick, John, Bob, Harry are picked randomly each day. What is the probability that John is not cheating.

Sol P(John picked for a day) = 3/4 = 0.75

P(John picked for 3 days) = 3/4 × 3/4 × 3/4 = 0.42.

P(John —"— 12 days) = $(3/4)^{12}$ = 0.032.

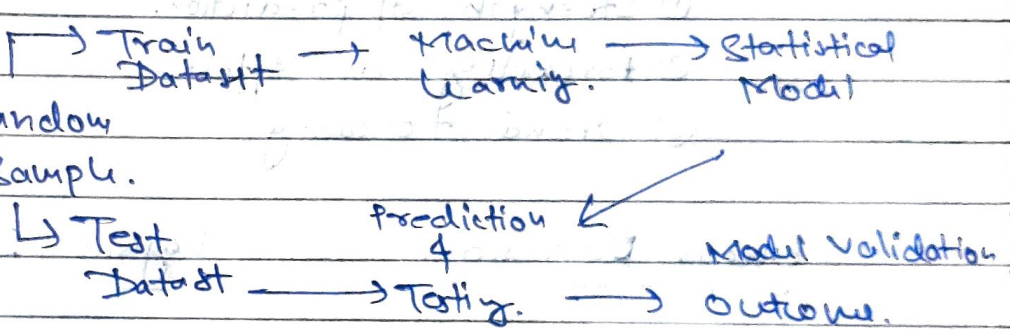
As, P(Event) < 5% (Threshold); John is cheating.

Kosh

Page No.:
Date:

⇒ Super Vised Learning

→ where we have input variable (x) & output variable (y) & we use the algorithm to map the function from the input to output.



New Data → Model → Prediction.

Algorithm

1. Linear Regression (estimate real value)
2. Logistic Regression (estimate discrete / binary etc values)
3. Decision Tree (Classification)
4. Random forest (Ensemble of DT)
5. Naive Bayes Classification (Classification)

Regression

→ It is a form of predictive modelling technique which investigates relationship between a dependent & independent variable.

Uses:

① Strength of prediction.
② Forcasting an effect.
③ Trend forcasting.

Linear Regression	Logistic Regression
→ Data is modelled using straight line	– The probability of some obtained event is represented as a linear function of a combination of predictor variable.
→ $y = mx + c$, we find coorelation b/w x & y.	– Data is modelled using a sigmoid function.
→ for continuous Variable	– Categorical Variable.
→ Value of variable as an output	– Probability of occurance of an event as an output.
→ Accuracy is measured by loss, R squared, Adjusted R squared etc.	– Accuracy, precision, Recall, f1 score, ROC curve, confusion Matric. etc.
→ Maps continuous x to continuous y	– Maps continuous x to categorical y

Kush

Page No.:
Date:

Linear Regression is used in:

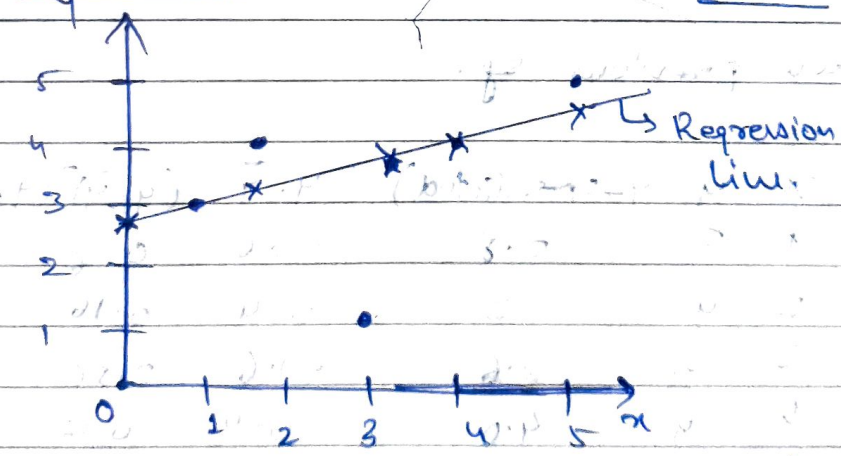
- Evaluating Trends & sales Estimate.
- Analyzing effect of price changes
- Assessment of risk in financial servius & insurance domain.

y axis = Dependent variable
x axis = independent variable.

Eg

X	Y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	3	-2	-0.6	4	1.2
2	4	-1	0.4	1	-0.4
3	2	0	-1.6	0	0
4	4	1	0.4	1	0.4
5	5	2	1.4	4	2.8
3	3.6			10	4

Mean



Linear regression eq ⇒ $y = mx + c$. (slope, y intercept)

$$\left[m = \frac{\Sigma (x-\bar{x}) \cdot (y-\bar{y})}{\Sigma (x-\bar{x})^2} \right]$$

$x - \bar{x}$ = difference (distance) of point x mean of x

$y - \bar{y}$ = similar to x.

∴ $m = 4/10$

∴ $3.6 = 0.4 \times 3 + c$

$c = 2.4$

∴ Predicted values for when $x = \{1, 2, 3, 4, 5\}$

$y = 0.4 \times 1 + 2.4 = 2.8$

$y = 0.4 \times 2 + 2.4 = 3.2$

etc.

→ Now, we we calculate the distance b/w actual & predicted value. And reduce the error / distance.

→ The line with least error, will be best fit line

To check the goodness of fit we use R squared Method.

→ R squared value is a statistical measure of how close the data are to the fitted regression line.

→ Also called coefficient of determination or coefficient of multiple determination

Eg: From previous Eg.

x	y	y-(predicted)	$y-\bar{y}$	$(y-\bar{y})^2$	$(y_p-\bar{y})$	$(y_p-\bar{y})^2$
1	3	2.8	−0.6	0.36	−0.8	0.64
2	4	3.2	0.4	0.16	−0.4	0.16
3	2	3.6	−1.6	2.56	0	0
4	4	4.0	0.4	0.16	0.4	0.16
5	5	4.4	1.4	1.96	0.8	0.64
	3.6			5.2		1.6

Predicted value

$$R^2 = \frac{\sum (y_p - \bar{y})^2}{\sum (y - \bar{y})^2}$$

actual value

$$R^2 = \frac{1.6}{5.2}$$

$R^2 \approx 0.3$

It is not a good fit.

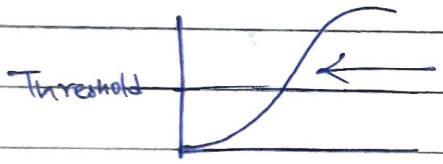
Logistic Regression.

→ It produces results in a binary format which is used to the outcome of a categorical dependent variable.

→ Outcome should be discrete / categorical such as:
0 / 1 | High / Low | T / F. | Y / N

→ Since the value of Y will be b/w 0 & 1, linear line has to be clipped b/w 0 & 1.

→ A Sigmoid function Caps all the values from $-\infty$ to ∞ to 0 & 1



→ for values on curve, we are threshhold value to either say they are 1 or 0.

Eg:
If curve gives value 0.8, it is > 0.5 hence, will be given as 1 or rounded to 1.

→ Logistic Reg eq is derivied from stright line eq.

Eq of st line
$$Y = C + B_1X_1 + B_2X_2 \ldots \longrightarrow \text{Rang } -\infty \text{ to } \infty.$$

Since, the Logistic regression has Y value b/w 0 & 1. we will 1-Y.

$$\therefore \quad \frac{Y}{1-Y}$$

KOSH

Page No.:
Date:

Use cases

1. Whether Prediction (rain or not)
2. Determine illness

⇒ Classification

→ Process of identifying & dividing the dataset into different categories or group by adding label

Algorithms:

Decision Tree
Random forest
Naive Bayes
KNN.
Logistic Regression
Linear Regression.
SVM

KOSH

Page No.:

Date:

Decision Tree: graphical representation of all the possible solution to a decision.

Random forest :- Builds multiple decision tree & merges them togetur.
- More accurate & stable
- Corrects decision trees mistake of overfitting
- Trained with bagging method.

Naive Bayes:- Based on Bayes Theorm.
- Assumes that the presence of a perticular feature in a class is is unrelated to the presence of any other featur.

KNN - Stores all the available cases & classifies new cases based on a similarity measure.
- While k in KNN algorithm we ~~exits~~ which to take vote from.

Page No.:
Date:

Decision Tree using CART Algorithm

1. Root Node - Base node (Parent Node)
2. Leaf Node - End of the tree
3. Splitting - Dividing
4. Pruning - Removing sub nodes

→ Entropy : Measure of impurity

$$Entropy(S) = -P(yes)\log_2 P(yes) - P(no)\log_2 P(no)$$

S = No of sample space.

P(S) = 0.5 then Entropy = 1

P(S) = 0 or 1 then Entropy = 0

→ Information Gain: Measure the reduction in entropy
Decides which attribute should be selected at the decision node

$$Information\ Gain = Entropy(S) - [(Weighted\ Avg) \times Entropy\ (each\ feature)]$$

KOSH

Page No.:
Date:

Eg → Temp, outlook dataset.

14 instances, 9 yes, 5 no.

$E(S) = -P(Yes)\log_2 P(Yes) - P(NO)\log_2 P(NO)$

$E(S) = -(9/14)\log_2(9/14) - (5/14)\log_2(5/14)$

$E(S) = 0.94$

Selection of Root Node:

lets take Column Outlook:

Sunny → 2Y 3N Overcast → 4Y Rainy → 3Y 2N

$E(Sunny) = -2/5\log_2(2/5) - 3/5\log_2(3/5) = 0.971$

$E(Overcast) = 0$

$E(Rainy) = -3/5\log_2(3/5) - 2/5\log_2(2/5) = 0.971$

Information Gain from outlook $= 5/14 * 0.971 + 4/14 \times 0 + 5/14 * 0.971 = 0.693$

(weighted Average.)

Information Gain $= 0.94 - 0.693 = 0.247$

lets take Column Temp:

hot → 2Y 2N Mild → 4Y 2N Cool → 3Y 1N.

$E(hot) = -2/4\log_2(2/4) - 2/4\log_2(2/4) =$

{ Similarly, performed for all the attributes.

⇒ Random forest

- Capable of performing regression & classification
- Randomly select 'n' variables out of 'N' variables in the dataset.
- Create decision trees.
- Create different decision trees with different combination of variables.
- Use majority voting to get final result

KOSH

Page No.:
Date:

⇒ KNN Algorithm (Lazy Learner)

- It stores all the available cases & classifies the new data or case based on similarity measure
- k denotes no of voting numbers / nearest members to to checked.

Use Case:

Recomendation System.
Fil Sepration.

Distance b/w the new data point & the neighbour is calculated by:

1. Euclidean Method
2. Manhattan Method.

Steps:

1. Handle Data
2. Similarity (calculating distance b/w two data instance)
3. Neighbours (locate k most similar data instance)
4. Response (Generate a response from a set of data instance)
5. Accuracy (summaggize the accuracy of predictions.
6. Main (tie all together)

Page No.:
Date:

Kash

⇒ Naive Bayes

- Based on Bayes theorem.

$$P(A|B) = \frac{P(B|A)\,(P(A)}{P(B)}$$

Use cases
- ~~Nlp~~ Text classification
- Spam Filtering.
- Medical Diagnostic.

3 Models on ~~Keras~~ Scikit learn.
1. Gaussian (used in classification, assumes feature follow Normal Dist.)
2. Multinomial (used for discrete count)
3. Bernoulli (used for feature vectors are binary)

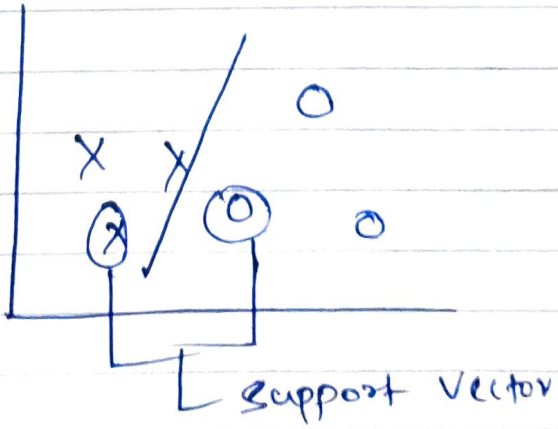
KOSH

Page No.:

Date:

⇒ SVM

- Supervised classification method that seperates data usy hyperplane.
- Classification & Regression Algorithm.
- Can be used to seperate non linear data with the help of kernel function
- Aim is to draw a hyper plane to seperate 2 classes.
- first we draw a random ~~data~~ hyperplane.
- Then we check the distance with the nearest data points aka, support vectors.
- An optimum hyperplane will have maximum distance from each of these support vectors.
- Distance b/w hyperplane & support vector is called Margin.



KOSH

Page No.:

Date:

In case of non linear data ?

- In this case transform the 2D space into 3D space.
- Use case
  - Cancer classification

KOSH

Page No.:
Date:

⇒ Unsupervised learning:

- It is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses.
- Training data is collection of information without any labels label

⇒ Clustering:

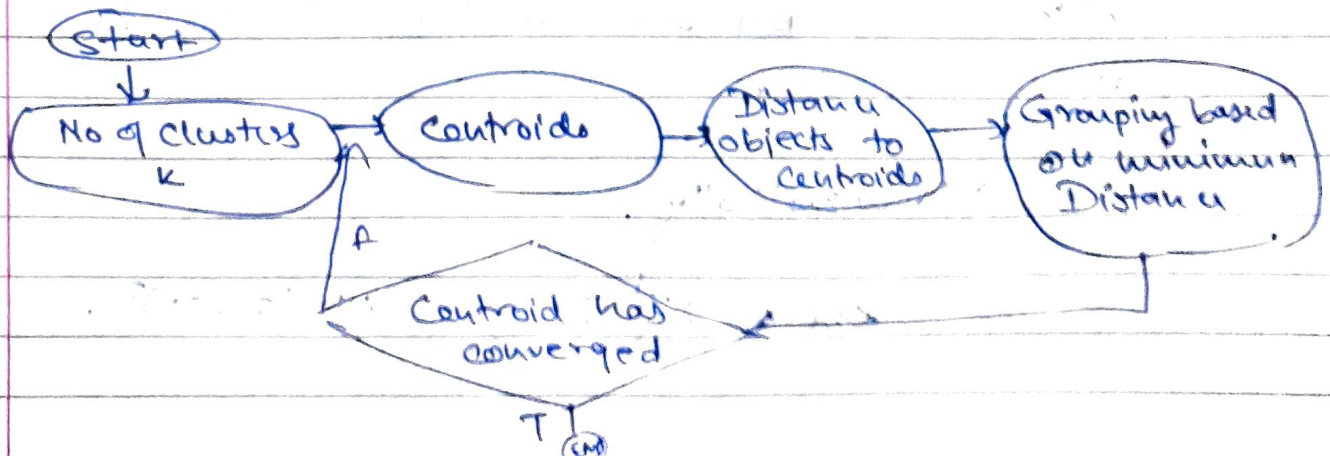
- Dividing the dataset into groups of similar datapoints without prior knowledge.

Types:

- Exclusive Clustering → eg: K means clustering
- Overlapping Clustering → eg: C means Clustering.
- Hierarchical Clustering.

⇒ K Means Clustering:

- Algorithm that groups similar datapoints.



KOSH

Page No.:
Date:

Steps for calculation:

→ Decide k: No of clusters to be made
→ Then we provide centroids of all the clusters.
→ Algo calculate Euclidian Distance of the points from each centroid & assigns the point to the closet cluster.
→ Next the centroids are calculated again, when we have are new cluster.
→ The distance of the points from the center of the clusters are calculated again & points are assigned to the closet cluster.
→ Again new centroid of the cluster is calculated.
→ These steps are repeated until we have a repetition in centroid or new centroids are very clos to the previous ones.

How to decide noCluster(k).

① The Elbow Method.

- Compute the sum of squared error (SSE) for some values of k. The SSE is defined as the sum of the squared distance b/w each member of the cluster & its centroid.

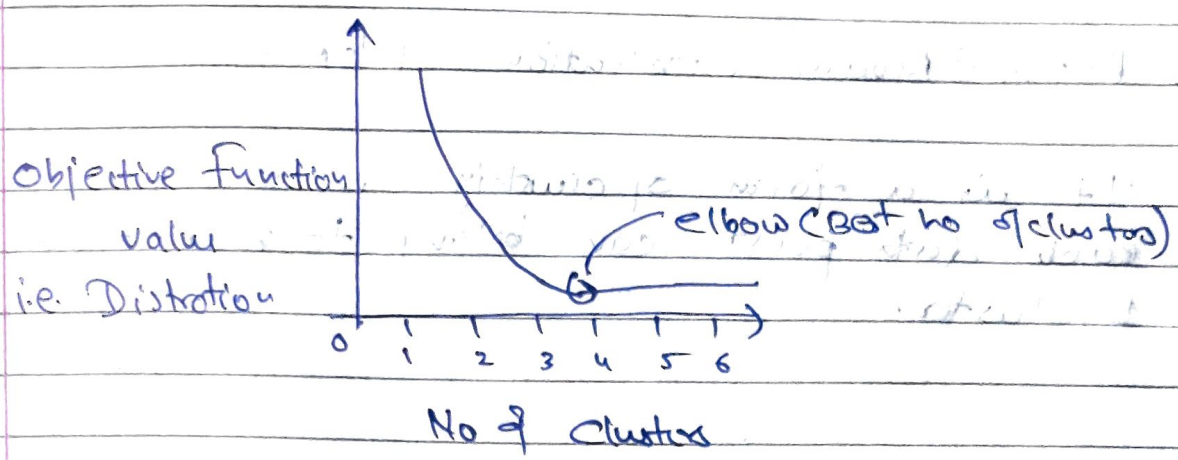
Mathematically,

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} dist(x, c_i)^2.$$

KOSH

Page No.:

Date:



Cons of K Means:

→ Can't handle ~~noise~~ & noisy data & outliers.

→ All items are forced to clusters.

⇒ Fuzzy C Means (extention of k Means)

- It is a form of clustering in which each data point can belong to more than 1 cluster.

Drawback

- Need to define c
- Need to define membership cutoff value.
- These are non deterministic.

⇒ Hivarchical Clustering.

- Alternative approach which builds a hivarchy from the bottom up, & doesnot require to specify no of clusters.
- 2 types:

Alomrative	& Division
↓	↓
Builds from bottom level	has all points initially in one cluster.

Disadvantage

- Too slow on big data.

KOSH

Page No.:
Date:

⇒ Market basket Analysis

- Two algorithms:

→ ① Association Rule Mining.
→ Shows how items are associated with each other.

3 ways to measure association:

ⓐ Support: Gives fraction of transactions which contains the items A & B.

- Support = $\frac{\text{freq}(A,B)}{N}$

ⓑ Confidence: Gives how often the item A&B occurs together, given no of times A occurs.

Confidence = $\frac{\text{freq}(A,B)}{\text{freq}(A)}$

ⓒ Lift: It indicates the strength of a rule over the random occurrence of A & B

Lift = $\frac{\text{Support}}{\text{Supp}(A) \times \text{Supp}(B)}$

→ ② Apriori Algorithm: It uses frequent itemsets to generate association rules. It is based on the concept that a subset of a frequent itemset must also be frequent itemset.

KOSH

Page No.:
Date:

⇒ Reinforcement Learning

- Reinforcement Learning is a type of ML where a agent learns to behave in a environment by performing actions & seeing the results.

- 4 components:
  - → Agent (reinforcement Algo) – Learns from Trial & Error.
  - → Environment: (The world.
  - → Action (A) – All possible states Agent can take
  - → State (S) – Current condition returned by the environment
  - → Reward (R) – An instant return from the environment to appraise the last action.
  - → Policy (π) – Approach the agent determines the next action based on the current state.
  - → Value (V) – Expected long term return with discount as opposed to short term return (R)
  - → Action Value (Q) – Similar to Value, except, it takes an extra parameter, the current action (A)

- Reward Maximization theory states that, a RL Agent must be trained in such a way that he takes the best action so that the reward is maximum.

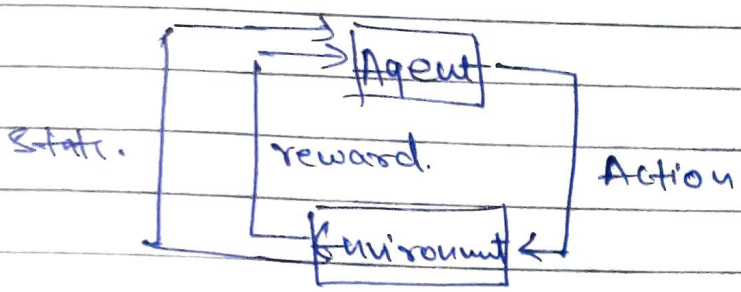
- Exploration & Exploitation → It is about exploring & capturing more information about an environment.
  - ↓
  - It is about using the already known exploited info to highlight the rewards

KOSH

Page No.:

Date:

Markov Decision Process : Approach for mapping a solution in reinforcement learning



- Uses Bellman's Equation for decision making.

$$V(s) = \max_{a} \left(R(s,a) + \gamma V(s') \right)$$

- $V(s)$: value of being in a particular state
- s : Particular State (room)
- a : action.
- γ : Discount factor
- s' : State at which robot goes from s
- $R(s,a)$: Reward function, outputs a reward value.