Unit:1 Introduction to machine Learning

• Concept of machine Learning:-

It was Arthur Samuel; an early American leader in the field of computer gaming and Artifical Intelligence coined the term "machine Learning" in 1959. He defined machine Learning as "the field of study that gives computers the ability to learn without being explicity programmed".

• Definition:-

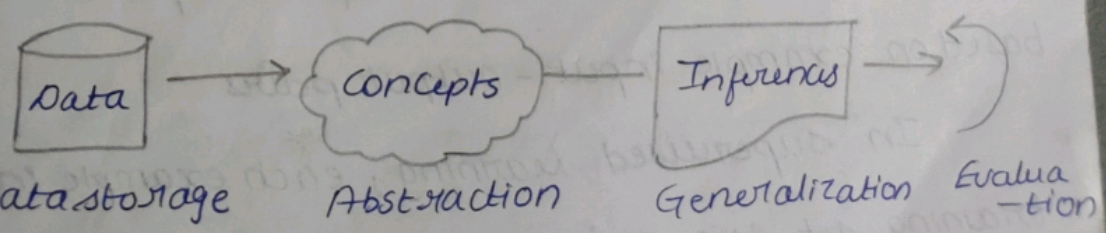
machine Learning (ML) is a branch of artifical intelligence that focuses on developing algorithms and models that enables computers to learn form and make decision based on data; without being explicity programmed.

It involves traning a model on a dataset; identifying patterns and using those patterns to make predictions or decision on newdata.

How machine Learns:-

Basic components of Learning process

this Learning process; wheather by a human or a machine, can be divided into four components; namely; data storage; abstraction; generalization and evaluation



Unit:1 Introduction to Machine Learning

• Concept of machine Learning:-

It was Arthur Samuel; an early American Leader in the field of computer gaming and Artifical Intelligence coined the term "machine Learning" in 1959. He defined machine Learning as "The field of study that gives computers the ability to learn without being explicity programmed".

• Definition:-

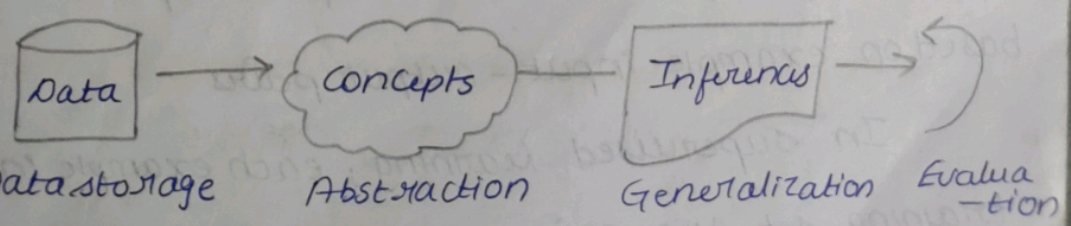
machine Learning (ML) is a branch of artifical intelligence that focuses on developing algorithms and models that enables computers to learn form and make decision based on data; without being explicity programmed.

It involves traning a model on a dataset; identifying patterns and using those patterns to make predictions or decision on newdata.

How machine Learns:-

Basic components of Learning process

this Learning process; wheather by a human or a machine, can be divided into four components, namely; data storage; abstraction; generalization and evaluation



1. **Data storage:** which facilities for storing and retrieving huge amounts of data are an important components of learning process.

2. **Abstraction:-** The second components of the learning process is known as abstraction. Abstraction is the process of extracting knowledge about stored data. The data is transformed into an abstract from the summarizes the original information.

3. **Generalization:-** The third component of learning process is known as generalization. The term Generalization describes the process of turning the knowledge about stored data into form of future action.

4. **Evaluation:-** It is the last component of learning process known as Evaluation. It is the process of giving feedback.

2. Different types of learning:-

In general, machine learning algorithms can be classified into three types.

1. **Supervised learning:-**

Supervised learning is the machine learning tasks of learning a function that maps an input to an output based on example input - output pairs

In supervised learning, each example in the training set is a pair consisting of an input object

1. **Data storage:** which facilities for storing and retrieving huge amounts of data are an important components of learning process.

2. **Abstraction:-** The second components of the learning process is known as abstraction. Abstraction is the process of extracting knowledge about stored data. The data is transformed into an abstract from the summarizes the original information.

3. **Generalization:-** The third component of learning process is known as generalization. The term Generalization describes the process of turning the knowledge about stored data into form of future action.

4. **Evaluation:-** It is the last component of learning process known as Evaluation. It is the process of giving feedback.

---

2. Different types of learning:-

In general, machine learning algorithms can be classified into three types.

1. **Supervised learning:-**

Supervised learning is the machine learning tasks of learning a function that maps an input to an output based on example input-output pairs

In supervised learning, each example in the training set is a pair consisting of an input object

Storing and retrieving
rtant components of

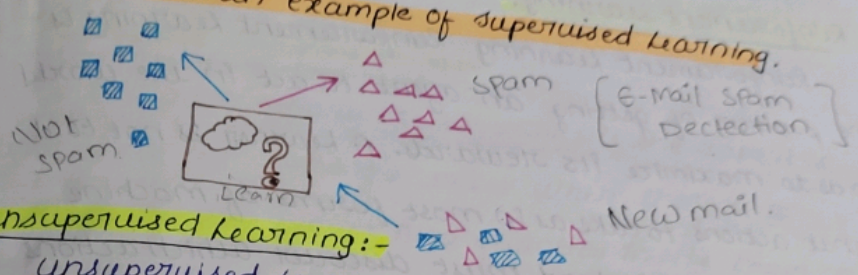
s of the learning
bstraction is the
ut stored data.
abstract from the
on.

of learning process
Generalization
owledge about
on.
t of learning
process of

can be

and an output value. A supervised learning algorithm analyzes the training data and produces a function, which can be used for mapping new examples. Both classification and regression problems are supervised learning problems.

Eg:- Consider an example of supervised learning.



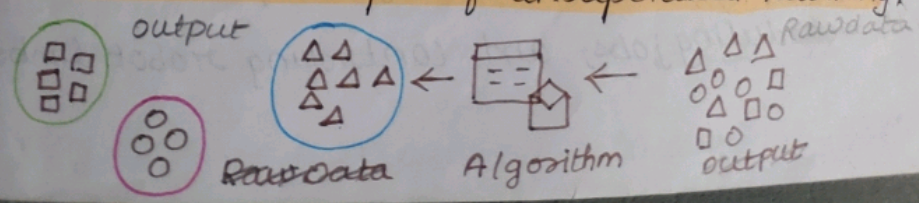
2. **Unsupervised learning:-**

Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses. In unsupervised learning algorithms; a classification or categorization is not included in the observations. There are no output values and so there is no estimation of function.

Since; the examples given to the learner are unlabeled; the accuracy of the structure that is output by the algorithm cannot be evaluated. The most common unsupervised learning method is cluster analysis; which is used for exploratory data analysis to find hidden patterns or grouping in data.

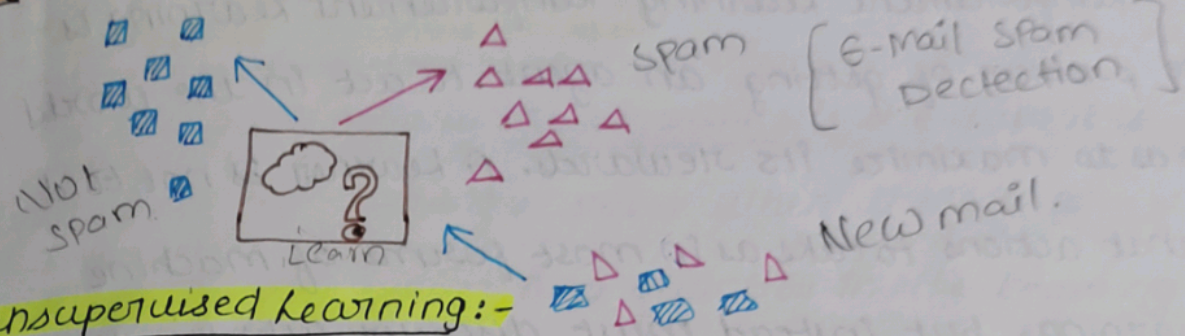
Eg:- Consider an example of unsupervised learning.

1.



and an output value. A supervised learning algorithm analyzes the training data and produces a function, which can be used for mapping new examples. Both classification and regression problems are supervised learning problems.

Eg:- Consider an example of supervised learning.



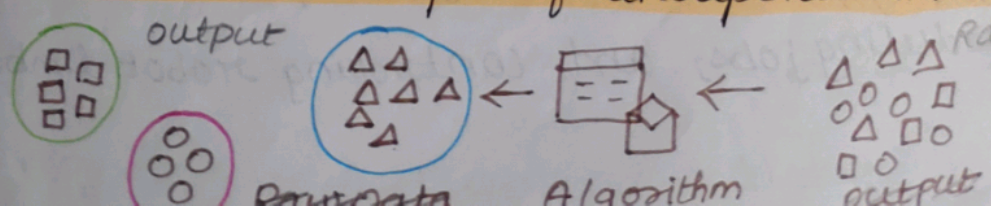
2. Unsupervised Learning:-

Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses. In unsupervised learning algorithms; a classification or categorization is not included in the observations. There are no output values and so there is no estimation of function.

Since; the examples given to the learner are unlabeled; the accuracy of the structure that is output by the algorithm cannot be evaluated. The most common unsupervised learning method is cluster analysis; which is used for exploratory data analysis to find hidden patterns or grouping in data.

Eg:- Consider an example of unsupervised learning.

1.



Eg:-2.

Gender	age
M	48
M	67
F	53
M	49
F	34

3. Reinforcement Learning:-

Reinforcement Learning Reinforcement Learning is the problem of getting an agent to act in the world so as to maximize its rewards. A learner is not told what actions to take as in most forms of machine learning; but instead must discover which actions yield the most reward by trying them.

In the most interesting and challenging cases; actions may affect not only the immediate rewards but also the next situations and, through that all subsequent rewards.

Eg:- Consider an example of Reinforcement Learning.

Consider teaching a dog a new trick: we cannot tell it what to do; but we can reward/punish it if it does the right/wrong thing. It has to find out what it did that made it get the reward/punishment. We can use a similar method to train computers to do many tasks; such as playing backgammon or chess; scheduling jobs; and controlling robot limbs.

3. Challenges of mach

Machine Learning ha
by enabling data-dri
crucial to acknowled
Professionals face wh
application from scr

1. Inadequate Traini

The back bone of a
trained on. The chal
shortage of both qua
dataset. Noisy; in
significantly impac

NOTE:- Noisy data is
or meaningless inf
that machines can

2. Poor Quality of D

Data quality is a
and inaccurate da
classification and
data is essential fo
a meticulous app

3. Non-Representa

The representa
the generalization
training data fa

Eg:-2.

Gender	age
M	48
M	67
F	53
M	49
F	34

3. Reinforcement Learning:-

Reinforcement Learning Reinforcement Learning is the problem of getting an agent to act in the world so as to maximize its rewards. A learner is not told what actions to take as in most forms of machine learning; but instead must discover which actions yield the most reward by trying them.

In the most interesting and challenging cases; actions may affect not only the immediate rewards but also the next situations and, through that all subsequent rewards.

Eg:- Consider an example of Reinforcement Learning.

Consider teaching a dog a new trick: we cannot tell it what to do; but we can reward/punish it if it does the right/wrong thing. It has to find out what it did that made it get the reward/punishment. We can use a similar method to train computers to do many tasks; such as playing backgammon or chess; scheduling jobs; and controlling robot limbs.

3. Challenges of machine Learning:-

Machine Learning has undoubtedly transformed industries by enabling data-driven decision-making. It is a crucial to acknowledge the practical challenges that professionals face while honing ML skills and developing application from scratch.

1. Inadequate Training Data:-

The back bone of any ML algorithm is the data it is trained on. The challenges arises when there is a shortage of both quality and quantity in the training dataset. Noisy, incorrect, or unclean data can significantly impact the efficiencies of ML algorithms.

NOTE:- Noisy data is data that contains extra, irrelevant or meaningless information. It can also refer to data that machines can't interpret correctly.

2. Poor Quality of Data:-

Data quality is a recurring issue, with noisy, incomplete and inaccurate data undermining the accuracy of classification and overall results. Achieving high-quality data is essential for the success of ML models, necessitating a meticulous approach to data preparation.

3. Non-Representative Training Data:-

The representative of training data directly influences the generalization capability of ML models. If training data fails to cover all relevant classes, the

3. Challenges of machine Learning:-

Machine Learning has undoubtedly transformed industries by enabling data-driven decision-making. It is a crucial to acknowledge the practical challenges that professionals face while honing ML skills and developing application from scratch.

1. Inadequate Training Data:-

The back bone of any ML algorithm is the data it is trained on. The challenges arises when there is a shortage of both quality and quantity in the training dataset. Noisy, incorrect, or unclean data can significantly impact the efficiencies of ML algorithms.

NOTE:- Noisy data is data that contains extra, irrelevant or meaningless information. It can also refer to data that machines can't interpret correctly.

2. Poor Quality of Data:-

Data quality is a recurring issue, with noisy, incomplete and inaccurate data undermining the accuracy of classification and overall results. Achieving high-quality data is essential for the success of ML models, necessitating a meticulous approach to data preparation.

3. Non-Representative Training Data:-

The representative of training data directly influences the generalization capability of ML models. If training data fails to cover all relevant classes, the

model may produce less accurate predictions; leading to bias against specific classes or groups. Using representative data in training mitigates biases and enhances predictions accuracy.

4. Overfitting and underfitting:-

Overfitting occurs when a model captures noise and inaccuracies from a large dataset; adversely affecting it's performance. This can be mitigated by employing linear and parametric algorithm; increasing training data; or reducing model complexity.

Conversely; underfitting arises from a model being too simple for the data resulting is incomplete and inaccurate predictions.

5. monitoring and maintenance:-

Regular monitoring and maintenance are essential to ensure the continued effectiveness of ML models. Changes in data or user exa expectations may necessitate code adjustments and resource updates; emphasizing the needs for ongoing vigilance.

6. Getting Bad Recommendations:-

ML models operating in a specific context may provide outdated or irrelevant recommendations; known as data drift. Regularly updating and monitoring data helps mitigate this issue; ensuring recommend

-ations align wi

7. Customer segme

Accurate custo

ML algorithms.

customer behavio

-tions based on

personalized use

4. Applications of

Applications of

databases is calle

volume of data

model with val

1. In retail busi

consumer beha

2. In finance; ban

models to we

and the stock

3. In manfactu

optimization;

4. In medicine

medical diagn

5. In telecommu

for network o

quality of serv

model may produce less accurate predictions; leading to bias against specific classes or groups. Using representative data in training mitigates biases and enhances predictions accuracy.

4. Overfitting and underfitting:-

Overfitting occurs when a model captures noise and inaccuracies from a large dataset; adversely affecting it's performance. This can be mitigated by employing linear and parametric algorithm; increasing training data; or reducing model complexity.

Conversely; underfitting arises from a model being too simple for the data resulting is incomplete and inaccurate predictions.

5. monitoring and maintenance:-

Regular monitoring and maintenance are essential to ensure the continued effectiveness of ML models. Changes in data or user ~~exa~~ expectations may necessitate code adjustments and resource updates; emphasizing the needs for ongoing vigilance.

6. Getting Bad Recommendations:-

ML models operating in a specific context may provide outdated or irrelevant recommendations; known as data drift. Regularly updating and monitoring data helps mitigate this issue; ensuring recommend

-ations align with current user expectations.

7. Customer segmentation:-

Accurate customer segmentation is crucial for effective ML algorithms. Developing algorithms that recognize customer behavior and trigger relevant recommenda-tions based on past experiences is essential for personalized user interactions.

4. Applications of machine learning:-

Applications of machine learning methods to large databases is called datamining. In data mining; a large volume of data is processed to construct a simple model with valuable use.

1. In retail business; machine learning is used to study consumer behavior.
2. In finance; banks analyze their past data to build models to use in credit applications, fraud detection and the stock market.
3. In manufacturing; learning models are used for optimization; control and troubleshooting.
4. In medicine; learning programs are used for medical diagnosis.
5. In telecommunications; call patterns are analyzed for network optimization and maximizing the quality of service.

-ations align with current user expectations.

7. Customer segmentation:-

Accurate customer segmentation is crucial for effective ML algorithms. Developing algorithms that recognize customer behavior and trigger relevant recommenda-tions based on past experiences is essential for personalized user interactions.

---

4. Applications of machine Learning:-

Applications of machine Learning methods to large databases is called datamining. In data mining; a large volume of data is processed to construct a simple model with valuable use.

1. In retail bussiness; machine Learning is used to study consumer behavior.
2. In finance; banks analyze their past data to build models to use in credit applications, fraud detection and the stock market.
3. In manfacturing; Learning models are used for optimization; control and trouble shooting.
4. In medicine; Learning programs are used for medical diagnosis.
5. In telecommunications; call patterns are analyzed for network optimization and maxmizing the quality of service.

6. In science; large amounts of data in physics, astronomy and biology can only be analyzed fast enough by computers. The world wide web is huge; it is constantly growing and searching for relevant information cannot be done manually.

7. In artifical Intelligence, it is used to teach a system to learn and adapt to change so that the system designer need not foresee and provide solutions for all possible situations.

8. It is used to find solutions to many problems in vision; speech Recognition and robotics.

Imp. 5. Baye's Theorem :-

Bayes theorem is a fundamental concept in Probability theory that describes how to update the Probability of a hypothesis (event) based on new evidence. It provides a way to calculate the posterior probability of a hypothesis; given prior knowledge and observed data.

The theorem is mathematically expressed as:

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

Where:

- $P(A|B)$ is the posterior probability: The probability of event of A occuring given that event B has occured.
- $P(B|A)$ is the likelihoo event B given that
- $P(A)$ is the prior prob event A before observ
- $P(B)$ is the marginal Probability of event possible hypothesis.

Applications of Bayes

1. Spam Email Filtering: as spam or not based
2. Medical Diagnosis:- disease based on test -ge.
3. Recommendation sy what movies or pr choices and prefere
4. Wheather Forecast data with real-tir Weather events.
5. Credit Scoring:- of a borrower s history and beha
6. Fraud Detection by updating the p spending patterns

6. In science; large amounts of data in physics, astronomy and biology can only be analyzed fast enough by computers. The world wide web is huge; it is constantly growing and searching for relevant information cannot be done manually.

7. In artifical Intelligence; it is used to teach a system to learn and adapt to change so that the system designer need not foresee and provide solutions for all possible situations.

8. It is used to find solutions to many problems in vision; speech Recognition and robotics.

Imp.

5. Baye's Theorem:-

Bayes theorem is a fundamental concept in Probability theory that describes how to update the Probability of a hypothesis (event) based on new evidence. It provides a way to calculate the posterior probability of a hypothesis; given prior knowledge and observed data.

The theorem is mathematically expressed as:

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

Where:

- $P(A|B)$ is the posterior probability: The probability of event of A occuring given that event B has occured.

- $P(B|A)$ is the Likelihood: The probability of observing event B given that A is true.
- $P(A)$ is the prior probability: The initial probability of event A before observing event B.
- $P(B)$ is the marginal likelihood or evidence: The total probability of event B occuring, considering all possible hypothesis.

Applications of Bayes theorem:-

1. Spam Email Filtering:- Bayes theorem helps classify emails as spam or not based on word patterns and probabilities.
2. Medical Diagnosis:- It is updated the probability of a disease based on test results and prior health knowledge.
3. Recommendation Systems:- Bayes theorem helps predict what movies or product you'll like based on past choices and preferences.
4. Wheather Forecasting:- It combines prior weather data with real-time conditions to predict future weather events.
5. Credit Scoring:- Bayes theorem assesses the likehood of a borrower repaying a loan based on financial history and behavior.
6. Fraud Detection:- It detects fraudulent transcations by updating the probability of fraud based on spending patterns.

- $P(B|A)$ is the Likelihood: The probability of observing event B given that A is true.
- $P(A)$ is the prior probability: The initial probability of event A before observing event B.
- $P(B)$ is the marginal likelihood or evidence: The total probability of event B occuring, considering all possible hypothesis.

Applications of Bayes theorem:-

1. Spam Email Filtering:- Bayes theorem helps classify emails as spam or not based on word patterns and probabilities.
2. Medical Diagnosis:- It is updated the probability of a disease based on test results and prior health knowledge.
3. Recommendation Systems:- Bayes theorem helps predict what movies or product you'll like based on past choices and preferences.
4. Wheather Forecasting:- It combines prior weather data with real-time conditions to predict future weather events.
5. Credit Scoring:- Bayes theorem assesses the likehood of a borrower repaying a loan based on financial history and behavior.
6. Fraud Detection:- It detects fraudulent transcations by updating the probability of fraud based on spending patterns.

Challenges of Bayes Theorem:

1. **Data Quality:-** Bayes theorem relies on accurate data, so if the information you have is incomplete or wrong, the results won't be reliable.

2. **Choosing the Right prior:-** The prior probability can strongly affect the results and choosing the wrong prior can lead to biased conclusion.

3. **Independence Assumption:-** In some applications, it assumes that all features are independent, which is often not in case in real-world data.

4. **Complexity with Large Data:-** As the amount of data grows, calculating probabilities for all possible outcomes can become very complex and computationally expensive.

5. **Overfitting:-** If the prior data is too specific, the model might become too focused on past data and fail to generalize well to new situation.

6. **Lack of Data:-** If you don't have enough data for some events. Bayes theorem might not give reliable results.

It gives probability of an event based on prior knowledge of conditions.

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

$P(A|B) \rightarrow$ Hypothesis

$P(B|A) \rightarrow$ Likelyhood

$P(A) \rightarrow$ Prior knowledge

$P(B) \rightarrow$ Marginal.

Proof:- Let us

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

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

from ① $P(A$

from ② $P(B$

$P(A|B) \cdot$

$P(A$

fro

→ A is called as

→ $P(A|B)$ is fi

Probability o

→ $P(B|A)$ is fi

with proba

→ $P(A)$ is prob

given data.

→ $P(B)$ is prob

Example:-

$P(\text{King}|A$

Challenges of Bayes Theorem:

1. **Data Quality:-** Bayes theorem relies on accurate data, so if the information you have is incomplete or wrong. the results won't be reliable.

2. **Choosing the Right prior:-** The prior probability can strongly affect the results and choosing the wrong prior can lead to biased conclusion.

3. **Independence Assumption:-** In some applications; it assumes that all features are independent; which is often not in case in real-world data.

4. **Complexity with Large Data:-** As the amount of data grows; calculating probabilities for all possible outcomes can become very complex and computationally expensive.

5. **Overfitting:-** If the prior data is too specific; the model might become too focused on past data and fail to generalize well to new situation.

6. **Lack of Data:-** If you don't have enough data for some events. Bayes theorem might not give reliable results.

It gives probability of an event based on prior knowledge of conditions.

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

$P(A|B) \rightarrow$ Hypothesis

$P(B|A) \rightarrow$ Likelyhood

$P(A) \rightarrow$ prior knowledge

$P(B) \rightarrow$ Marginal.

n accurate data, so
mplete or wrong.

robability can
ng the wrong prior

lications; it assumes
h is often not in

mount of data
possible outcomes
nally expensive.

ific; the model
nd fail to

data for some
rable results.

on prior

Proof:- Let us take two events A and B

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \longrightarrow (1)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \longrightarrow (2)$$

from (1) $P(A \cap B) = P(A|B) \cdot P(B) \longrightarrow (i)$

from (2) $P(B \cap A) = P(B|A) \cdot P(A) \longrightarrow (ii)$

from (i) and (ii)

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \longrightarrow (iii)$$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \longrightarrow (iv)$$

from equation (4).

→ A is called as hypothesis and B is called as given data.

→ P(A|B) is finding probability of hypothesis, when probability of trained examples are given.

→ P(B|A) is find probability of given data provided with probability of hypothesis that is true.

→ P(A) is probability of hypothesis before observing the given data.

→ P(B) is probability of hypothesis for given data.

Example:-

$$P(\text{King}|\text{face}) = \frac{P(\text{face}|\text{King}) \cdot P(\text{King})}{P(\text{face})}$$

$$= \frac{\frac{\cancel{4}^{1}}{\cancel{4}_{1}} \cdot \frac{\cancel{4}^{1}}{\cancel{52}_{13}}}{\frac{\cancel{12}^{3}}{\cancel{52}_{13}}}$$

$$= \frac{1 \cdot \frac{1}{13}}{\frac{3}{13}} = 1 \times \frac{1}{13} \times \frac{13}{3} = \boxed{\frac{1}{3}}$$

P(face) = 12

P(face/King)

$\frac{4}{4}$

Black - spade;
Red - Heart;
Total - 26.

Proof:- Let us take two events A and B

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \longrightarrow (1)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \longrightarrow (2)$$

from (1) $p(A \cap B) = P(A|B) \cdot p(B) \longrightarrow (i)$

from (2) $P(B \cap A) = P(B|A) \cdot p(A) \longrightarrow (ii)$

from (i) and (ii)

$$P(A|B) \cdot p(B) = P(B|A) \cdot p(A) \longrightarrow (iii)$$

$$P(A|B) = \frac{p(B|A) \cdot P(A)}{P(B)} \longrightarrow (iv)$$

from equation (4).

→ A is called as hypothesis and B is called as given data.

→ P(A|B) is finding probability of hypothesis; when probability of trained examples are given.

→ p(B|A) is find probability of given data provided with probability of hypothesis that is true.

→ p(A) is probability of hypothesis before observing the given data.

→ p(B) is probability of hypothesis for given data.

Example:-

P(face) = 12

P(face|King) $\frac{4}{4}$

$$P(King|Face) = \frac{p(Face|King) \cdot p(King)}{p(Face)}$$

$$= \frac{\frac{\cancel{4}^{1}}{\cancel{4}_{1}} \cdot \frac{\cancel{4}^{1}}{\cancel{52}_{13}}}{\frac{\cancel{12}^{3}}{\cancel{52}_{13}}}$$

$$= \frac{1 \cdot \frac{1}{13}}{\frac{3}{13}} = 1 \times \frac{1}{13} \times \frac{13}{3} = \boxed{\frac{1}{3}}$$

6. The Naive Bayes classifier:-

The Naive Bayes classifiers is a probabilistic machine learning model based on Bayes theorem. It is particularly useful for classification tasks; such as spam filtering; sentiments analysis; and medical diagnosis.

Types of Naive Bayes classifier:-

There are three types of Naive bayes classifier.

1. **Gaussian Naive Bayes:-**

It used when features are continous and assumed to follow a normal distribution.

2. **Multinomial Naive Bayes:-**

It used for discrete data.

For eg:- Text classification.

3. **Bernoulli Naive Bayes:-**

It is used for binary features.

For eg:- spam detection.

Advantages:-

1. Fast and efficient; even with large datasets.
2. Works well with high-dimensional data.
3. Requires less training data compared to other models.

Disadvantages:-

1. Assumption of independence between features is often unrealistic.

2. struggles with correlat
3. performs poorly if d

Eg:- Consider an examp

Day	Outlook	Temper
D_1	Sunny	Hot
D_2	Sunny	Hot
D_3	Overcast	Hot
D_4	Rain	mi
D_5	Rain	coo
D_6	Rain	Coo
D_7	Overcast	co
D_8	Sunny	m
D_9	Sunny	
D_{10}	Rain	
D_{11}	Sunny	
D_{12}	Overcast	
D_{13}	Overcast	
D_{14}	Rain	

Outlo	es
Sunny	
rcast	
n	

6. The Naive Bayes classifier:-

The Naive Bayes classifiers is a probabilistic machine learning model based on Bayes theorem. It is Particularly useful for classification tasks; such as spam filtering; sentiments analysis; and medical diagnosis.

Types of Naive Bayes Classifier:-

There are three types of Naive bayes classifier.

1. **Gaussian Naive Bayes:-**

It used when features are continous and assumed to follow a normal distribution.

2. **Multinomial Naive Bayes:-**

It used for discrete data.

For eg:- Text classification.

3. **Bernoulli Naive Bayes:-**

It is used for binary features.

For eg:- spam detection.

Advantages:-

1. Fast and efficient; even with large datasets.
2. Works well with high-dimensional data.
3. Requires less training data compared to other models.

Disadvantages:-

1. Assumption of independence between features is often unrealistic.

2. struggles with correlated features.
3. performs poorly if data is not normally distributed.

Eg:- consider an example of Naive bayes classifier.

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D_1	Sunny	Hot	High	Weak	NO
D_2	Sunny	Hot	High	Strong	NO
D_3	Overcast	Hot	High	Weak	Yes
D_4	Rain	mild	High	Weak	Yes
D_5	Rain	cool	Normal	Weak	Yes
D_6	Rain	Cool	Normal	Strong	NO
D_7	Overcast	cool	Normal	Strong	Yes
D_8	Sunny	mild	High	Weak	NO
D_9	Sunny	cool	Normal	Weak	Yes
D_{10}	Rain	mild	Normal	Weak	Yes
D_{11}	Sunny	mild	Normal	Strong	Yes
D_{12}	Overcast	mild	High	Strong	Yes
D_{13}	Overcast	Hot	Normal	Weak	Yes
D_{14}	Rain	mild	High	Strong	NO

Outlook	Yes	NO
Sunny	2/9	3/5
Overcast	4/9	0
rain	3/9	2/5

Temperature	Yes	NO
Hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humidity	Yes	NO
High	3/9	4/5
Normal	6/9	1/5

Windy	Yes	NO
Strong	3/9	3/5
Weak	6/9	3/5

2. struggles with correlated features.

3. performs poorly if data is not normally distributed.

Eg:- Consider an example of Naive bayes classifier.

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D_1	Sunny	Hot	High	Weak	NO
D_2	Sunny	Hot	High	Strong	NO
D_3	Overcast	Hot	High	Weak	Yes
D_4	Rain	mild	High	Weak	Yes
D_5	Rain	Cool	Normal	Weak	Yes
D_6	Rain	Cool	Normal	Strong	NO
D_7	Overcast	Cool	Normal	Strong	Yes
D_8	Sunny	mild	High	Weak	NO
D_9	Sunny	Cool	Normal	Weak	Yes
D_{10}	Rain	mild	Normal	Weak	Yes
D_{11}	Sunny	mild	Normal	Strong	Yes
D_{12}	Overcast	mild	High	Strong	Yes
D_{13}	Overcast	Hot	Normal	Weak	Yes
D_{14}	Rain	mild	High	Strong	NO

Outlook	Yes	NO
Sunny	2/9	3/5
Overcast	4/9	0
Rain	3/9	2/5

Temperature	Yes	NO
Hot	2/9	2/5
mild	4/9	2/5
Cool	3/9	1/5

Humidity	Yes	NO
High	3/9	4/5
Normal	6/9	1/5

Windy	Yes	NO
Strong	3/9	3/5
Weak	6/9	3/5

$P(\text{play Tennis} = \text{Yes}) = \frac{9}{14} = 0.64$

$P(\text{play Tennis} = \text{NO}) = \frac{5}{14} = 0.36$

Formula:

$$\boxed{\vartheta_{NB} = \underbrace{\text{argmax}}_{\text{Naive bayes.}} \; \underbrace{P(v_j)}_{\text{yes or NO values}} * \underbrace{P(d_i|v_j)}_{\text{attributes}}}$$

Consider:-

outlook = Sunny

Temperature = cool

Humidity = High

Windy = Strong.

$P(\text{outlook} = \text{sunny}|v_j)\; P(\text{Temperature} = \text{cool}|v_j)$

$P(\text{Humidity} = \text{High}|v_j)\; P(\text{windy} = \text{strong}|v_j)$

$\vartheta_{NB}(\text{Yes}) = P(\text{Yes}).\, P(\text{sunny}|\text{yes}).\, P(\text{cool}|\text{Yes}).\, P(\text{High}|\text{yes}).\, P(\text{strong}|\text{yes}) \longrightarrow (1)$

$\vartheta_{NB}(\text{NO}) = P(\text{NO}).\, P(\text{sunny}|\text{NO}).\, P(\text{cool}|\text{NO}).\, P(\text{High}|\text{NO}).\, P(\text{strong}|\text{NO}) \longrightarrow (2)$

from; (1)

$0.64 \times \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9}$

$= \boxed{0.005}$

from; (2)

$0.36 \times \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5}$

$= \boxed{0.020}$

$$\boxed{\vartheta_{NB} = \frac{\vartheta_{NB}(\text{Yes})}{\vartheta_{NB}(\text{Yes}) + \vartheta_{NB}(\text{NO})}} = \frac{0.005}{0.005 + 0.020} = \boxed{0.2}$$

$$\boxed{\vartheta_{NB} = \frac{\vartheta_{NB}(\text{NO})}{\vartheta_{NB}(\text{Yes}) + \vartheta_{NB}(\text{NO})}} = \frac{0.020}{0.005 + 0.020} =$$

$P(\text{play Tennis} = \text{yes}) = \frac{9}{14} = 0.64$

$P(\text{play Tennis} = \text{NO}) = \frac{5}{14} = 0.36$

Formula:

$$\boxed{v_{NB} = \text{argmax}\ p(v_j) * P(a_i|v_j)}$$

yes or NO values

attributes.

Naive bayes.

Consider:-

outlook = sunny

Temperature = cool

Humidity = High

windy = Strong.

$P(\text{outlook} = \text{sunny}|v_j)\ p(\text{Temperature} = \text{cool}|v_j)$

$P(\text{Humidity} = \text{High}|v_j)\ p(\text{windy} = \text{strong}|v_j)$

$v_{NB}(\text{Yes}) = P(\text{Yes}).\ p(\text{sunny}|\text{yes}).\ p(\text{cool}|\text{yes}).\ p(\text{High}|\text{yes}).\ P(\text{strong}|\text{yes}) \longrightarrow (1)$

$v_{NB}(\text{NO}) = P(\text{NO}).\ P(\text{sunny}|\text{NO}).\ P(\text{cool}|\text{NO}).\ P(\text{High}|\text{NO}).\ P(\text{strong}|\text{NO}) \longrightarrow (2)$

from; (1)

$0.64 \times \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9}$

$= \boxed{0.005}$

from; (2)

$0.36 \times \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5}$

$= \boxed{0.020}$

$$\boxed{v_{NB} = \frac{v_{NB}(\text{Yes})}{v_{NB}(\text{Yes}) + v_{NB}(\text{NO})}} = \frac{0.005}{0.005 + 0.020} = \boxed{0.2}$$

$$\boxed{\hat{y}_{NB} = \frac{\hat{y}_{NB}(NO)}{\hat{y}_{NB}(YES) + \hat{y}_{NB}(NO)}} = \frac{0.020}{0.005 + 0.020} = \boxed{0.8}$$

Unit: 2

Regression

→ Linear Regression.

→ Logistic Regression.

1. Regression:-

It is a process of determining a relationship between one or more independent variables and one dependent variable output variable.

Examples:-

1. Predict the height of a person given the age of the person.

   Independent - Age

   Dependent - Height.

2. predict the price of a car given the car model year of manufacturing; milage; engine capacity; interior etc.

   Independent - milage; year of manufacturing

   dependent - car.

3. predict the student age admission in college based on college requirements percentage graduation.

- **Types of Regression:-**

There are four types of Regression, they are:

1. Simple Regression.
2. multi Regression
3. polynomial Regression.
4. Logistic Regression.

1. Simple Regression:-

The relationship
and the dependent
relation is called a

y = a+b

- Example:-

Week	
1	
2	
3	
4	
5	

Week — Indepen
Sales in Thousar

y=
a
a
c

a1 =

a0 =

Here; there

	x
	1
	2
	3

Unit: 2

Regression

→ Linear Regression.

→ Logistic Regression.

---

1. Regression:-

It is a process of determining a relationship between one or more independent variables and one dependent variable output variable.

Examples:-

1. Predict the height of a person given the age of the Person.

   Independent - Age

   Dependent - Height.

2. predict the price of a car given the car model year of manufacturing; milage; engine capacity; interior etc.

   independent - milage; year of manufacturing

   dependent - car.

3. predict the student age admission in college based on college requirements percentage graduation.

- **Types of Regression:-**

There are four types of Regression, they are:

1. Simple Regression.
2. multi Regression
3. polynomial Regression.
4. Logistic Regression.

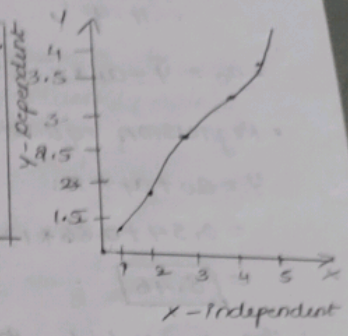
1. **Simple Regression:-**

The relationship between independent variable x and the dependent variable y is modeled by a relation is called as simple Regression.

$$y = a + bx.$$

• **Example:-**

Week	Sales in Thousands.
1	1.2
2	1.8
3	2.6
4	3.2
5	3.8



Week — Independent

Sales in Thousands - Dependent.

$$y = a_0 + a_1 * x + e$$

$a_1 \rightarrow$ co-efficient

$a_0 \rightarrow$ intercept

$e \rightarrow$ error.

$$a_1 = \frac{(\overline{xy}) - (\bar{x})(\bar{y})}{\overline{x}^2 - \overline{x}^2}$$

$$a_0 = \bar{y} - a_1 \times \bar{x}$$

Here, there are 5 items i.e; $i = 1, 2, 3, 4, 5$

	x_i (Week)	y_i (Sales)	x_i^2	$x_i * y_i$
	1	1.2	1	1.2
	2	1.8	4	3.6
	3	2.6	9	7.8
	4	3.2	16	12.8
	5	3.8	25	19
Sum	15	12.6	55	44.4
Average	$x = 3$	$y = 2.52$	$\bar{x}^2 = 11$	$\overline{xy} = 8.88$

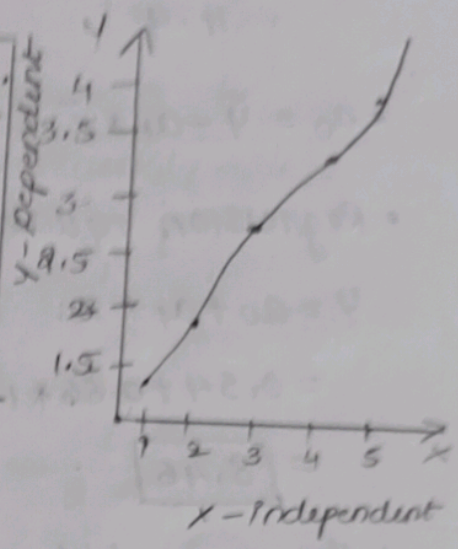
1. Simple Regression:-

The relationship between independent variable x and the dependent variable y is modeled by a relation is called as simple Regression.

$$y = a + bx.$$

Example:-

Week	Sales in Thousands.
1	1.2
2	1.8
3	2.6
4	3.2
5	3.8



Week – Independent

Sales in Thousands – Dependent.

$$y = a_0 + a_1 * x + e$$

$a_1 \rightarrow$ co-efficient

$a_0 \rightarrow$ intercept

$e \rightarrow$ error.

$$a_1 = \frac{(\overline{xy}) - (\bar{x})(\bar{y})}{\overline{x}^2 - \bar{x}^2}$$

$$a_0 = \bar{y} - a_1 \times \bar{x}$$

Here; there are 5 items i.e; $i = 1, 2, 3, 4, 5$

	x_i (week)	y_i (sales)	x_i^2	$x_i * y_i$
	1	1.2	1	1.2
	2	1.8	4	3.6
	3	2.6	9	7.8
	4	3.2	16	12.8
	5	3.8	25	19
Sum	15	12.6	55	44.4
Average	$x = 3$	$y = 2.52$	$\overline{x}^2 = 11$	$\overline{xy} = 8.88$

$\bar{x} = 3 \quad \bar{y} = 2.52 \quad \overline{x^2} = 11 \quad \overline{xy} = 8.88$

- $a_1 = \dfrac{(\overline{xy}) - (\bar{x})(\bar{y})}{\overline{x^2} - \bar{x}^2} = \dfrac{8.88 - 3 \times 2.52}{11 - 3^2}$

$= \dfrac{8.88 - 3 \times 2.52}{11 - 9} = \dfrac{1.32}{2} = \boxed{0.66}$

- $a_0 = \bar{y} - a_1 * \bar{x} = 2.52 - 0.66 * 3 = \boxed{0.54}$

- Regression equation is.

$y = a_0 + a_1 * x.$

$= 0.54 + 0.66 * 12$

$= \boxed{8.46}$

$\therefore$ The predicted 12th week sale is $x = 8.46$.

$y = a_0 + a_1 * x.$

$= 0.54 + 0.66 * 7$

$= \boxed{5.16}$

$\therefore$ The predicted 7th week sale is $x = 5.16$.

---

2. **Multi Regression:-**

multiple regression is also called as **multiple linear regression.** is a statistical technique used to model the relationship between a dependent variable and two or more independent variables. It extends simple linear regression by multiple predictions to better explain variations in the dependent variable

- Let us consider two variables x_1 and x_2

$y = F(x_1, x_2)$

$y = a_0 + a_1 x_1 + a_2 x_2$

→ (n) independent variables as:

$y = F(x_1, x_2 \cdots x_n)$

$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots$

→ x_1, x_2, x_n → predictor

→ y is dependent

→ a_0, a_1, a_2 → coefficients

Example:-

(x_1) Product 1 sale	(x_2) Product 2 sale
1	4
2	5
3	8
4	2

Consider: $x = \begin{bmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{bmatrix}$ ar

→ The coefficient of a mu

$a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$

→ The regression coeffic

Calculated the same

$a^1 = ((x^T x)$

$\bar{x} = 3 \quad \bar{y} = 2.52 \quad \overline{x^2} = 11 \quad \overline{xy} = 8.88$

- $a_1 = \dfrac{(\overline{xy}) - (\bar{x})(\bar{y})}{\overline{x^2} - \bar{x}^2} = \dfrac{8.88 - 3 \times 2.52}{11 - 3^2}$

$= \dfrac{8.88 - 3 \times 2.52}{11 - 9} = \dfrac{1.32}{2} = \boxed{0.66}$

- $a_0 = \bar{y} - a_1 * \bar{x} = 2.52 - 0.66 * 3 = \boxed{0.54}$

- Regression equation is.

$y = a_0 + a_1 * x.$

$= 0.54 + 0.66 * 12$

$= \boxed{8.46}$

$\therefore$ The predicted 12th week sale is $x = 8.46$.

$y = a_0 + a_1 * x.$

$= 0.54 + 0.66 * 7$

$= \boxed{5.16.}$

$\therefore$ The predicted 7th week sale is $x = 5.16$.

---

2. **Multi Regression:-**

multiple regression is also called as **multiple linear regression.** is a statistical technique used to model the relationship between a dependent variable and two or more independent variables. It extends simple linear regression by multiple predictions to better explain variations in the dependent variable

- Let us consider two variables x_1 and x_2

$y = F(x_1, x_2)$

$y = a_0 + a_1 x_1 + a_2 x_2$

→ (n) independent variables as:

$$y = f(x_1, x_2 \cdots x_n)$$

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots\cdots + a_n x_n + \varepsilon.$$

→ x_1, x_2, x_n → predictors.

→ y is dependent

→ a_0, a_1, a_2 → coefficients

Example:-

(x_1) Product 1 sale	(x_2) Product 2 sale	(x_3) Y (weekly sale). product
1	4	1
2	5	6
3	8	8
4	2	12

Consider: $x = \begin{bmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \\ 6 \\ 8 \\ 12 \end{bmatrix}$

→ The coefficient of a multiple regression equation is:

$$a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

→ The regression coefficient for multiple regression is calculated the same way as linear regression.

$$a^{\wedge} = ((x^T x)^{-1} x^T) Y.$$

1. $X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{bmatrix}$

→ (n) independent variables as:

$$y = f(x_1, x_2 \dots x_n)$$

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots\dots + a_n x_n + \varepsilon.$$

→ x_1, x_2, x_n → predictor

→ y is dependent

→ a_0, a_1, a_2 → coefficients

Example:-

(x_1) product 1 sale	(x_2) product 2 sale	(x_3) Y (weekly sale). product
1	4	1
2	5	6
3	8	8
4	2	12

Consider: $x = \begin{bmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \\ 6 \\ 8 \\ 12 \end{bmatrix}$

→ The coefficient of a multiple regression equation is:

$$a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

→ The regression coefficient for multiple regression is calculated the same way as linear regression.

$$a^{\wedge} = ((x^T x)^{-1} x^T) y.$$

1. $X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{bmatrix}$

$$\Rightarrow \begin{bmatrix} 1+1+1+1 & 1+2+3+4 & 4+5+8+2 \\ 1+2+3+4 & 1+4+9+16 & 4+10+24+8 \\ 4+5+8+2 & 4+10+24+8 & 16+25+64+4 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 4 & 10 & 19 \\ 10 & 30 & 46 \\ 19 & 46 & 109 \end{bmatrix}$$

$$\Rightarrow (x^T x)^{-1} = \begin{bmatrix} 4 & 10 & 19 \\ 10 & 30 & 46 \\ 19 & 46 & 109 \end{bmatrix}$$

$$\Rightarrow (x^T x)^{-1} = \frac{1}{\det(A)} \times adj(A).$$

$$\det A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \Rightarrow \begin{matrix} a = 4 & d = 10 & g = 19 \\ b = 10 & e = 30 & h = 46 \\ c = 9 & f = 46 & i = 109. \end{matrix}$$

$\Rightarrow 30 \times 109 = 3270.$

$46 \times 46 = 2116$

$10 \times 109 = 1090$

$46 \times 19 = 874$

$10 \times 46 = 460$

$30 \times 19 = 570.$

$$\det(A) = 4(3270 - 2116) - 10(1090 - 874) + 19(460 - 570)$$

$$= 4(1154) - 10(216) + 19(-110)$$

$$= 4616 - 2160 - 2090$$

$$\det A = 366.$$

3. <u>Cofactor:-</u>

$$M_1 = \begin{vmatrix} 30 & 46 \\ 46 & 109 \end{vmatrix} = (30)(109) - 46(46)$$

$$= 3270 - 2116.$$

$$= 1154$$

$$M_{12} = \begin{vmatrix} 10 & 46 \\ 19 & 109 \end{vmatrix} = (10)(109) - (19)4$$

$$= 216$$

$$M_{13} = \begin{vmatrix} 10 & 30 \\ 19 & 46 \end{vmatrix} = 10(46) - 19(30$$

$$= -110.$$

$$M_{21} = \begin{vmatrix} 10 & 19 \\ 46 & 109 \end{vmatrix} = 10(109) - 19($$

$$= 216$$

$$M_{22} = \begin{vmatrix} 4 & 19 \\ 19 & 109 \end{vmatrix} = 4(109) -$$

$$= 75$$

$$M_{23} = \begin{vmatrix} 4 & 10 \\ 19 & 46 \end{vmatrix} = 4(46) - 1$$

$$= -6.$$

$$M_{31} = \begin{vmatrix} 10 & 19 \\ 30 & 46 \end{vmatrix} = 10(46) - 19$$

$$= -110.$$

$$M_{32} = \begin{vmatrix} 4 & 19 \\ 10 & 46 \end{vmatrix} = 4(46) -$$

$$= -6$$

$$M_{33} = \begin{vmatrix} 4 & 10 \\ 10 & 30 \end{vmatrix} = 4(30)$$

$$= 20$$

$$\text{cofactor } A = \begin{bmatrix} 1154 & -2 \\ -216 & 7 \\ -110 & \end{bmatrix}$$

$$A^{-1} = \frac{1}{\det A} \times adj(A)$$

$$= \frac{1}{366} \times \begin{bmatrix} 1154 & -21 \\ -216 & 75 \\ -110 & 6 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 3.1530 \\ -0.5902 \\ -0.3005 \end{bmatrix}$$

$$(x^T x)^{-1} x^T = \begin{bmatrix} 3.15 \\ -0.5 \\ -0. \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1+1+1+1 & 1+2+3+4 & 4+5+8+2 \\ 1+2+3+4 & 1+4+9+16 & 4+10+24+8 \\ 4+5+8+2 & 4+10+24+8 & 16+25+64+4 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 4 & 10 & 19 \\ 10 & 30 & 46 \\ 19 & 46 & 109 \end{bmatrix}$$

$$\Rightarrow (x^T x)^{-1} = \begin{bmatrix} 4 & 10 & 19 \\ 10 & 30 & 46 \\ 19 & 46 & 109 \end{bmatrix}$$

$$\Rightarrow x^T x^{-1} = \frac{1}{\det(A)} \times adj(A).$$

$$\det A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \Rightarrow \begin{matrix} a = 4 & d = 10 & g = 19 \\ b = 10 & e = 30 & h = 46 \\ c = 9 & f = 46 & i = 109. \end{matrix}$$

$\Rightarrow 30 \times 109 = 3270.$

$46 \times 46 = 2116$

$10 \times 109 = 1090$

$46 \times 19 = 874$

$10 \times 46 = 460$

$30 \times 19 = 570.$

$$\det(A) = 4(3270 - 2116) - 10(1090 - 874) + 19(460 - 570)$$

$$= 4(1154) - 10(216) + 19(-110)$$

$$= 4616 - 2160 - 2090$$

$$\det A = 366.$$

3. <u>Cofactors:-</u>

$$M_1 = \begin{vmatrix} 30 & 46 \\ 46 & 109 \end{vmatrix} = (30)(109) - 46(46)$$

$$= 3270 - 2116.$$

$$= 1154$$

$$m_{12} = \begin{vmatrix} 10 & 46 \\ 19 & 109 \end{vmatrix} = (10)(109) - (19)46 = 216$$

$$m_{13} = \begin{vmatrix} 10 & 30 \\ 19 & 46 \end{vmatrix} = 10(46) - 19(30) = -110$$

$$m_{21} = \begin{vmatrix} 10 & 19 \\ 46 & 109 \end{vmatrix} = 10(109) - 19(46) = 216$$

$$m_{22} = \begin{vmatrix} 4 & 19 \\ 19 & 109 \end{vmatrix} = 4(109) - 19(19) = 75$$

$$m_{23} = \begin{vmatrix} 4 & 10 \\ 19 & 46 \end{vmatrix} = 4(46) - 19(10) = -6$$

$$m_{31} = \begin{vmatrix} 10 & 19 \\ 30 & 46 \end{vmatrix} = 10(46) - 19(30) = -110$$

$$m_{32} = \begin{vmatrix} 4 & 19 \\ 10 & 46 \end{vmatrix} = 4(46) - 19(10) = -6$$

$$m_{33} = \begin{vmatrix} 4 & 10 \\ 10 & 30 \end{vmatrix} = 4(30) - 10(10) = 20$$

$$\text{cofactor } A = \begin{bmatrix} 1154 & -216 & -110 \\ -216 & 75 & 6 \\ -110 & 6 & 20 \end{bmatrix}$$

$$A^{-1} = \frac{1}{\det A} \times \text{adj}(A)$$

$$= \frac{1}{366} \times \begin{bmatrix} 1154 & -216 & -110 \\ -216 & 75 & 6 \\ -110 & 6 & 20 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 3.1530 & -0.5902 & -0.3005 \\ -0.5902 & 0.2049 & 0.0164 \\ -0.3005 & 0.0164 & 0.0546 \end{bmatrix}$$

$$(X^T X)^{-1} X^T = \begin{bmatrix} 3.15 & -0.59 & -0.30 \\ -0.59 & 0.20 & 0.016 \\ -0.30 & 0.016 & 0.054 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{bmatrix}$$

$$m_{12} = \begin{vmatrix} 10 & 46 \\ 19 & 109 \end{vmatrix} = (10)(109) - (19)46.$$
$$= 216$$

$$m_{13} = \begin{vmatrix} 10 & 30 \\ 19 & 46 \end{vmatrix} = 10(46) - 19(30)$$
$$= -110.$$

$$m_{21} = \begin{vmatrix} 10 & 19 \\ 46 & 109 \end{vmatrix} = 10(109) - 19(46)$$
$$= 216$$

$$m_{22} = \begin{vmatrix} 4 & 19 \\ 19 & 109 \end{vmatrix} = 4(109) - 19(19)$$
$$= 75$$

$$m_{23} = \begin{vmatrix} 4 & 10 \\ 19 & 46 \end{vmatrix} = 4(46) - 19(10).$$
$$= -6.$$

$$m_{31} = \begin{vmatrix} 10 & 19 \\ 30 & 46 \end{vmatrix} = 10(46) - 19(30)$$
$$= -110.$$

$$m_{32} = \begin{vmatrix} 4 & 19 \\ 10 & 46 \end{vmatrix} = 4(46) - 19(10)$$
$$= -6$$

$$m_{33} = \begin{vmatrix} 4 & 10 \\ 10 & 30 \end{vmatrix} = 4(30) - 10(10)$$
$$= 20.$$

$$\text{cofactor } A = \begin{bmatrix} 1154 & -216 & -110 \\ -216 & 75 & 6 \\ -110 & 6 & 20 \end{bmatrix}$$

$$A^{-1} = \frac{1}{\det A} \times adj(A)$$

$$= \frac{1}{366} \times \begin{bmatrix} 1154 & -216 & -110 \\ -216 & 75 & 6 \\ -110 & 6 & 20 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 3.1530 & -0.5902 & -0.3005 \\ -0.5902 & 0.2049 & 0.0164 \\ -0.3005 & 0.0164 & 0.0546 \end{bmatrix}$$

$$(X^T X)^{-1} X^T = \begin{bmatrix} 3.15 & -0.59 & -0.30 \\ -0.59 & 0.20 & 0.016 \\ -0.30 & 0.016 & 0.054 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1.36 & 0.47 & -1.02 & 0.19 \\ -0.32 & -0.11 & 0.138 & 0.242 \\ -0.068 & 0.002 & 0.18 & -0.128 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1.36 & 0.47 & -1.02 & 0.19 \\ -0.32 & -0.11 & 0.138 & 0.24 \\ -0.068 & 0.002 & 0.18 & -0.128 \end{bmatrix} \times \begin{bmatrix} 1 \\ 6 \\ 8 \\ 12 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} -1.7 \\ 3.028 \\ -0.152 \end{bmatrix}$$

$a_0 = -1.7$

$a_1 = 3.028$

$a_2 = -0.152.$

3. **Polynomial Regression:-**

Polynomial regression is a type of linear regression where the relationship between the dependent variable (y) and the independent variable (x) is modeled as an n-degree polynomial rather than a straight line.

It is useful when data shows a non-linear relationship.

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3.$$

Generally; Polynomial of maximum degree four are used as higher order polynomial we take some strange change and make the curve more flexible. It's leads to situation of overfitting and hence it is avoided.

Construct the polynomial 2nd degree equation:

$$y = a_0 + a_1 x + a_2 x^2$$

$y \rightarrow$ dependent

$x, x^2 \rightarrow$ independent

$a_0, a_1, a_2 \rightarrow$ coefficie

$$\boxed{A = X^{-1} B}$$

$$X = \begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x \end{bmatrix}$$

$$B = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i, y_i \\ \Sigma x_i^2, y_i \end{bmatrix}$$

Example:-

x	y
1	1
2	4
3	9
4	15

n=4

x_i	y_i	$x_i y$
1	1	1
2	4	8
3	9	27
4	15	60
10	29	9

$$X = \begin{bmatrix} 4 & 1 \\ 10 & \\ 30 & \end{bmatrix}$$

$A =$

$X = \begin{bmatrix} 4 \\ \\ \end{bmatrix}$

$$\Rightarrow \begin{bmatrix} 1.36 & 0.47 & -1.02 & 0.19 \\ -0.32 & -0.11 & 0.138 & 0.242 \\ -0.068 & 0.002 & 0.18 & -0.128 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1.36 & 0.47 & -1.02 & 0.19 \\ -0.32 & -0.11 & 0.138 & 0.24 \\ -0.068 & 0.002 & 0.18 & -0.128 \end{bmatrix} \times \begin{bmatrix} 1 \\ 6 \\ 8 \\ 12 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} -1.7 \\ 3.028 \\ -0.152 \end{bmatrix}$$

$a_0 = -1.7$

$a_1 = 3.028$

$a_2 = -0.152.$

3. **Polynomial Regression:-**

Polynomial regression is a type of linear regression where the relationship between the dependent variable (y) and the independent variable (x) is modeled as an n-degree polynomial rather than a straight line.

It is useful when data shows a non-linear relationship.

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3.$$

Generally; Polynomial of maximum degree four are used as higher order polynomial we take some strange change and make the curve more flexible. It's leads to situation of overfitting and hence it is avoided.

Construct the polynomial 2nd degree equation:

$$y = a_0 + a_1 x + a_2 x^2.$$

$y \rightarrow$ dependent

$x, x^2 \rightarrow$ Independent

$a_0, a_1, a_2 \rightarrow$ coefficients {This are calculated by using formula}

$$\boxed{A = X^{-1}B}$$

$$X = \begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix}$$

$$B = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{bmatrix}$$

Example:-

X	Y
1	1
2	4
3	9
4	15

$n=4$

x_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y$	x_i^3	x_i^4
1	1	1	1	1	1	1
2	4	8	4	16	8	16
3	9	27	9	81	27	81
4	15	60	16	240	64	261
10	29	96	30	338	100	354

$$X = \begin{bmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix} \quad B = \begin{bmatrix} 29 \\ 96 \\ 338 \end{bmatrix}$$

$$A = X^{-1}B.$$

$$X = \begin{bmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix}$$

$x, x^2 \rightarrow$ Independent

$a_0, a_1, a_2 \rightarrow$ coefficients {This are calculated by using formula}

$$\boxed{A = X^{-1}B}$$

$$X = \begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix}$$

$$B = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i, y_i \\ \Sigma x_i^2, y_i \end{bmatrix}$$

Example:-

X	Y
1	1
2	4
3	9
4	15

$n=4$	x_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y$	x_i^3	x_i^4
	1	1	1	1	1	1	1
	2	4	8	4	16	8	16
	3	9	27	9	81	27	81
	4	15	60	16	240	64	261
	10	29	96	30	338	100	354

$$X = \begin{bmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix} \quad B = \begin{bmatrix} 29 \\ 96 \\ 338 \end{bmatrix}$$

$$A = X^{-1}B.$$

$$X = \begin{bmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix}$$

$a=4 \quad b=10 \quad c=30$

$d=10 \quad e=30 \quad f=100$

$g=30 \quad h=100 \quad i=354$

$(30)(354)-(100)(100)$

$=10620-10000$

$=620.$

$(10)(354)-(30)(100)$

$=3540-3000$

$=540.$

$10(100)-(30)(30)$

$=1000-900$

$=100$

$\det(x)=4(620)-10(540)+30(100)$

$=2480-5400+3000$

$=80.$

Cofactor:-

$m_1=(30)(354)-100(100)=10620-10000=620\ +$

$m_{12}=(10)(354)-100(30)=+(3540)-3000)=-540$

$m_{13}=10(100)-(30)(30)=1000-900=100\ +$

$m_{21}=(10)(354)-(30)(100)=3540-3000=540\ -$

$m_{22}=4(354)-(30)(30)=1416-900=516\ +$

$m_{23}=4(100)-(10)(30)=400-300=100\ -$

$m_{31}=(10)(100)-(30)(30)=1000-900=100\ +$

$m_{32}=4(100)-(30)(10)=400-300=100\ -$

$m_{33}=4(30)-10(10)=120-100=20\ +$

$$C=\begin{bmatrix}620 & -540 & 100\\ -540 & 516 & -100\\ 100 & -100 & 20\end{bmatrix}$$

$$adj(x)=C^T=\begin{bmatrix}620 & -540 & 100\\ -540 & 516 & -100\\ 100 & -100 & 20\end{bmatrix}$$

$$x^{-1}=\frac{1}{\det(A)}\times Adj(x).$$

$$=\frac{1}{80}\times\begin{bmatrix}620 & -540 & 100\\ -540 & 516 & -100\\ 100 & -100 & 20\end{bmatrix}$$

$$x^{-1}=\begin{bmatrix}7.75 & -6.75 & 1.25\\ -6.75 & 6.45 & -1.25\\ 1.25 & -1.25 & 0.25\end{bmatrix}$$

$x^{-1}B.$

$$\Rightarrow\begin{bmatrix}7.75 & -6.75 & 1.25\\ -6.75 & 6.45 & -1.25\\ 1.25 & -1.25 & 0.25\end{bmatrix}$$

$$\Rightarrow\begin{bmatrix}7.75\times29-6.75\times96\\ -6.75\times29+6.45\times9\\ 1.25\times29+(-1.25\times$$

$$\Rightarrow\begin{bmatrix}-0.75\\ 0.95\\ 0.75\end{bmatrix}$$

$a_0=-0.75$

$a_1=0.95$

$a_2=0.75$

- features scaling | sele

1. feature scaling:-

Feature scaling

standardizing numer

are on a similar sca

of machine learning

$a=4 \quad b=10 \quad c=30$

$d=10 \quad e=30 \quad f=100$

$g=30 \quad h=100 \quad i=354$

$(30)(354)-(100)(100)$

$=10620-10000$

$=620.$

$(10)(354)-(30)(100)$

$=3540-3000$

$=540.$

$10(100)-(30)(30)$

$=1000-900$

$=100$

$\det(x)=4(620)-10(540)+30(100)$

$=2480-5400+3000$

$=80.$

<u>Cofactor:-</u>

$m_1=(30)(354)-100(100)=10620-10000=620+$

$m_{12}=(10)(354)-100(30)=+(3540)-3000)=-540$

$m_{13}=10(100)-(30)(30)=1000-900=100+$

$m_{21}=(10)(354)-(30)(100)=3540-3000=540-$

$m_{22}=4(354)-(30)(30)=1416-900=516+$

$m_{23}=4(100)-(10)(30)=400-300=100-$

$m_{31}=(10)(100)-(30)(30)=1000-900=100+$

$m_{32}=4(100)-(30)(10)=400-300=100-$

$m_{33}=4(30)-10(10)=120-100=20+$

$$C=\begin{bmatrix} 620 & -540 & 100 \\ -540 & 516 & -100 \\ 100 & -100 & 20 \end{bmatrix}$$

$$adj(x)^T = \begin{bmatrix} 620 & -540 & 100 \\ -540 & 516 & -100 \\ 100 & -100 & 20 \end{bmatrix}$$

$$x^{-1} = \frac{1}{det(A)} \times Adj(x).$$

$$= \frac{1}{80} \times \begin{bmatrix} 620 & -540 & 100 \\ -540 & 516 & -100 \\ 100 & -100 & 20 \end{bmatrix}$$

$$x^{-1} = \begin{bmatrix} 7.75 & -6.75 & 1.25 \\ -6.75 & 6.45 & -1.25 \\ 1.25 & -1.25 & 0.25 \end{bmatrix}$$

$x^{-1}B.$

$$\Rightarrow \begin{bmatrix} 7.75 & -6.75 & 1.25 \\ -6.75 & 6.45 & -1.25 \\ 1.25 & -1.25 & 0.25 \end{bmatrix} \times \begin{bmatrix} 29 \\ 96 \\ 338 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 7.75 \times 29 - 6.75 \times 96 + 1.25 \times 338 \\ -6.75 \times 29 + 6.45 \times 96 - 1.25 \times 338 \\ 1.25 \times 29 + (-1.25 \times 96) + 0.25 \times 338 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} -0.75 \\ 0.95 \\ 0.75 \end{bmatrix}$$

$a_0 = -0.75$

$a_1 = 0.95$

$a_2 = 0.75$

- **features scaling / selection:-**

1. **feature scaling:-**

Feature scaling is the process of normalizing or standardizing numerical input features to ensure they are on a similar scale. This helps improve the performance of machine learning models, especially those that use

$$adj(x) = C^T = \begin{bmatrix} 620 & -540 & 100 \\ -540 & 516 & -100 \\ 100 & -100 & 20 \end{bmatrix}$$

$$x^{-1} = \frac{1}{det(A)} \times Adj(x)$$

$$= \frac{1}{80} \times \begin{bmatrix} 620 & -540 & 100 \\ -540 & 516 & -100 \\ 100 & -100 & 20 \end{bmatrix}$$

$$x^{-1} = \begin{bmatrix} 7.75 & -6.75 & 1.25 \\ -6.75 & 6.45 & -1.25 \\ 1.25 & -1.25 & 0.25 \end{bmatrix}$$

$x^{-1}B$.

$$\Rightarrow \begin{bmatrix} 7.75 & -6.75 & 1.25 \\ -6.75 & 6.45 & -1.25 \\ 1.25 & -1.25 & 0.25 \end{bmatrix} \times \begin{bmatrix} 29 \\ 96 \\ 338 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 7.75 \times 29 - 6.75 \times 96 + 1.25 \times 338 \\ -6.75 \times 29 + 6.45 \times 96 - 1.25 \times 338 \\ 1.25 \times 29 + (-1.25 \times 96) + 0.25 \times 338 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} -0.75 \\ 0.95 \\ 0.75 \end{bmatrix}$$

$a_0 = -0.75$

$a_1 = 0.95$

$a_2 = 0.75$

- **Features scaling / selection:-**

1. **Feature scaling:-**

Feature scaling is the process of normalizing or standardizing numerical input features to ensure they are on a similar scale. This helps improve the performance of machine learning models, especially those that use

distance based algorithms.

Eg:- Logistic Regression.

Types of features scaling:-

1. Normalization (min-max scaling):-

Normalization is a features scaling technique that transforms data into a specific range; typically [0,1] or [-1,1]. This ensures that all features contribute equally to a machine learning model; preventing bias towards larger-valued features.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$x' \rightarrow$ scaled feature.

$x \rightarrow$ original feature value.

$x_{min} \rightarrow$ minimum value of the feature.

$x_{max} \rightarrow$ maximum value of the feature.

2. Standardization (z-score Normalization).

Standardization in a feature scaling technique that transforms data so that it has a mean of 0 and a standard deviation of 1. This ensures that features contribute equally to the model; making it especially useful for algorithm that assume normally distributed data.

$$x' = \frac{x - \mu}{\sigma}$$

x' = standardized value

x = original feature value

μ = mean of the feature

σ = standard deviation of th

2. Feature selection:-

Feature selection is the process of relevant features from a dataset to Performance, reduce overfitting and -ational cost.

Types of feature selection methods

1. Filter methods:-

Select features based on statisti the model.

Eg's:- 1. pearson correlation

2. chi-square test.

2. Wrapper methods:-

use a model to evaluate th subsets.

Eg's:- Forward selection; Ba

3. Embedded Methods:-

Features selection is bu Process.

Eg's:- Random Forest.

4. Dimensionality Reduction:

Transform Features

Eg:- principal component

distance based algorithms.

Eg:- Logistic Regression.

Types of features scaling:-

1. Normalization (min-max scaling):-

Normalization is a features scaling technique that transforms data into a specific range; typically [0,1] or [-1,1]. This ensures that all features contribute equally to a machine learning model; preventing bias towards larger-valued features.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$x' \rightarrow$ scaled feature.

$x \rightarrow$ original feature value.

$x_{min} \rightarrow$ minimum value of the feature.

$x_{max} \rightarrow$ maximum value of the feature.

2. Standardization (Z-score Normalization).

Standardization in a feature scaling technique that transforms data so that it has a mean of 0 and a standard deviation of 1. This ensures that features contribute equally to the model; making it especially useful for algorithm that assume normally distributed data.

$$x' = \frac{x - \mu}{\sigma}$$

x' = standardized value.
x = original feature value
μ = mean of the feature
σ = standard deviation of the feature.

2. **Feature Selection:-**

Feature selection is the process of selecting the most relevant features from a dataset to improve model performance, reduce overfitting and decreases computational cost.

Types of feature selection methods:-

1. **Filter methods:-**

Select features based on statistical tests independent of the model.

Eg's:- 1. pearson correlation
2. chi-square test.

2. **Wrapper methods:-**

use a model to evaluate the importance of feature subsets.

Eg's:- Forward selection; Backward elimination.

3. **Embedded Methods:-**

Features selection is built into the model training process.

Eg's:- Random forest.

4. **Dimensionality Reduction:-**

Transform features into a lower-dimensional space.

Eg:- principal component Analysis (PCA).

x' = standardized value.

x = original feature value

μ = mean of the feature

σ = standard deviation of the feature.

2. Feature Selection:-

Feature selection is the process of selecting the most relevant features from a dataset to improve model performance, reduce overfitting and decreases computational cost.

Types of feature selection methods:-

1. **Filter methods:-**

Select features based on statistical tests independent of the model.

Eg's:- 1. pearson correlation
2. chi-square test.

2. **Wrapper methods:-**

use a model to evaluate the importance of feature subsets.

Eg's:- Forward selection, Backward elimination.

3. **Embedded Methods:-**

Features selection is built into the model training process.

Eg's:- Random Forest.

4. **Dimensionality Reduction:-**

Transform features into a lower-dimensional space.

Eg:- Principal component Analysis (PCA).

• Classification:-
Imp

It is a supervised learning method which can be used to predict categorical values, or class label of new observations based on past data. The goal is to learn from labeled training data so that the model can accurately assign categories to new data points

→ Classification is divided into two types:-

1. Binary classification
2. multi~~ple~~ class classification

1. Binary classification:-

Binary classification is a type of supervised learning where the model predicts one of two possible classes.

- Spam detection (Spam or Not Spam)
- Diseases Diagnosis (positive or Negative)
- Sentiment analysis (positive or Negative)
- Fraud detection (Fraud or Not Fraud),

• Key concepts:-

1. Labels / classes:-

Usually represented as 0 and 1 (or -1 and 1)

2. Algorithms used:-

- Logistic Regression
- Support vector machines (SVM)
- Decision Trees
- K-nearest neighbours (K-NN)

2. multi class classification:-

multiclass classification is a
where the goal is to classify inpu
two distinct classes.

• Examples:-

- classifying types of animals (ca
- Handwritten digit recognition
- Emotion detection (Happy, s

Algorithms for multi class:-

- Logistic Regression (One-vs
- Decision Trees.
- Naive Bayes
- SVM (One-vs-one or o

Evaluation:-

- Accuracy.
- confusion matrix.

• Key Differences from Bi

Aspect:-	Bina classi
1. output classes	2
2. output format	Singl (0

Classification:-

Imp

It is a supervised learning method which can be used to predict categorical values, or class label of new observations based on past data. The goal is to learn from labeled training data so that the model can accurately assign categories to new data points

→ Classification is divided into two types:-

1. Binary classification
2. multi~~ple~~ class classification

1. Binary classification:-

Binary classification is a type of supervised learning where the model predicts one of two possible classes.

- spam detection (spam or Not spam)
- Diseases Diagnosis (positive or Negative)
- Sentiment analysis (positive or Negative)
- Fraud detection (Fraud or Not Fraud),

- Key concepts:-

1. Labels / classes:-

usually represented as 0 and 1 (or -1 and 1)

2. Algorithms used:-

- Logistic Regression
- Support vector machines (svm)
- Decision Trees

- K-nearest neighbours (K-NN)

2. **Multi class classification:-**

multi class classification is a machine learning tasks where the goal is to classify inputs in to more than two distinct classes.

Examples:-

- classifying types of animals (cat, Dog, Horse etc).
- Handwritten digit recognition (Digits 0 through 9)
- Emotion detection (Happy, sad, Angry, Neutral).

Algorithms for multi class:-

- Logistic Regression (One-vs-Rest or softmax)
- Decision Tress.
- Naive Bayes
- SVM (one-vs-one or one-vs-Rest strategy)

Evaluation:-

- Accuracy.
- confusion matrix.

- **Key Differences from Binary classification:-**

Aspect:-	Binary classification	multi class classification.
1. output classes	2	3 or more
2. output format	single label (0 or 1)	One label from multiple (eg 0 to 9)

- K-nearest neighbours (K-NN)

2. **Multi class classification:-**

multiclass classification is a machine learning tasks where the goal is to classify inputs in to more than two distinct classes.

Examples:-

- classifying types of animals (Cat, Dog, Horse etc).
- Handwritten digit recognition (Digits 0 through 9)
- Emotion detection (Happy, sad, Angry, Neutral).

Algorithms for multi class:-

- Logistic Regression (One-vs-Rest or softmax)
- Decision Trees.
- Naive Bayes
- SVM (One-vs-one or one-vs-Rest strategy)

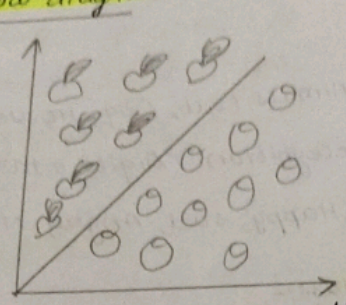
Evaluation:-

- Accuracy.
- confusion matrix.
- **Key Differences from Binary classification:-**

Aspect:-	Binary classification	multi class classification.
1. Output classes	2	3 or more
2. Output format	Single label (0 or 1)	One label from multiple (eg 0 to 9)

3. Evaluation	Precision Recall, F1	Averaged metrics (macro, micro, weighted)

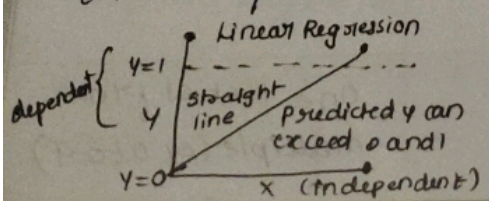
- Classification algorithms can be better understood using below diagram.

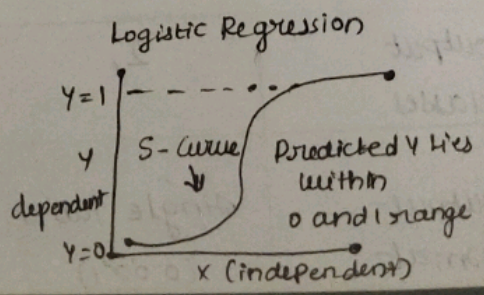


In the diagram, there are two diagram (apple and orange) there are two classes these features that are similar to each other and disimilar to other classes. The algorithm which implements the classification on a dataset is known as a classifier.

- Logistic Regression:-

Logistic Regression is a statistical method uses for binary classification problems, where the output (dependent variable) is categorical and typically has two possible outcomes; such as yes/no; 0/1 or true/false.





Let us consider an example of Logistic

Logistic Regression model predict a aged 40 will purchase the predict formula.

Examples:-

Age (x)	purchase (y)
25	0
30	0
45	1
50	1

Step:1 - Logistic Regression equatio

Probability of (y) = p(y=

$= \frac{1}{1+e^{-}}$ (

Here β_0 = Intercept.

β_1 = coefficient.

x - Independent v

NOTE:- This values usally com Likelyhood estimation.

$\beta_0 = -10$

$\beta_1 = 0.25.$

$= \frac{1}{1+e^{-}}(-10+0.25$

$= 0.5.$

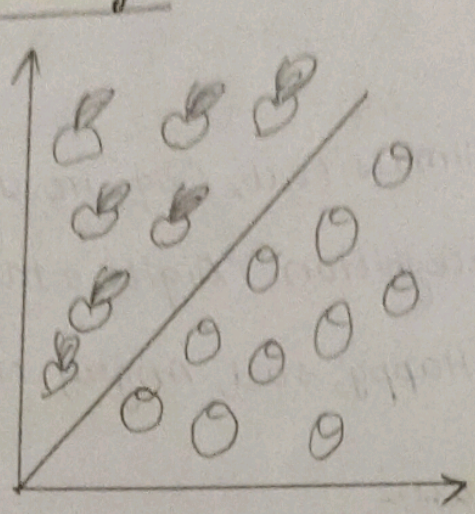
If $p >= 0.5$ predi

$p < 0.5$ pred

- Difference between Logis Regression.

3. Evaluation	precision Recall, F1	averaged metrics (macro, micro, weighted)

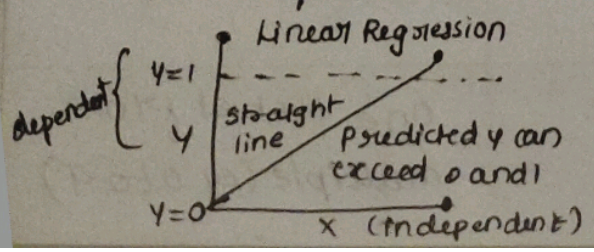
- Classification algorithms can be better understood using below diagram.

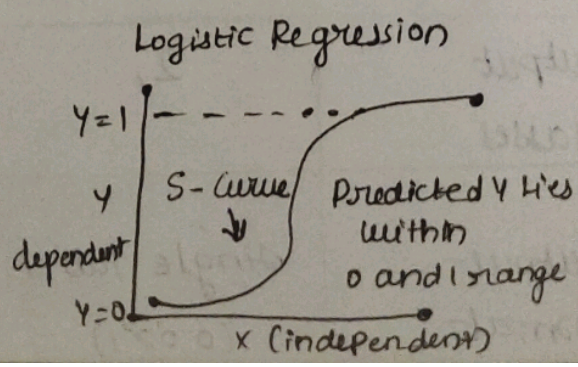


In the diagram; there are two diagram (apple and orange) there are two classes these features that are similar to each other and disimilar to other classes. The algorithm which implements the classification on a dataset is known as a classifier.

- **Logistic Regression:-**

Logistic Regression is a statistical method uses for binary classification problems; where the output (dependent variable) is categorical and typically has two possible outcomes; such as yes/no; 0/1 or true/false.





Let us consider an example of Logistic Regression:

Logistic Regression model predict weather a person aged 40 will purchase the product using Regression formula.

Example:-

Age (x)	purchase (y)
25	0
30	0
45	1
50	1

Step:1 - Logistic Regression equation.

Probability of (y) = $p(y=1/x)$.

$$= \frac{1}{1+e^{-}(\beta_0+\beta_1 \cdot x)}$$

Here β_0 = Intercept.

β_1 = coefficient.

x = Independent variable.

NOTE:- This values usally computated maximum Likelyhood estimation.

$\beta_0 = -10$

$\beta_1 = 0.25$.

$$= \frac{1}{1+e^{-}(-10+0.25 \times 40)}.$$

$= 0.5$.

If $p >= 0.5$ predict as 1 (will purchase).

$p < 0.5$ predict as 0 (will not purchase).

- **Difference between Logistic Regression and Linear Regression.**

Let us consider an example of Logistic Regression:

Logistic Regression model predict weather a person aged 40 will purchase the product using Regression formula.

Example:-

Age (x)	purchase (Y)
25	0
30	0
45	1
50	1

Step:1 - Logistic Regression equation.

Probability of (Y) = $p(Y=1/x)$.

$$= \frac{1}{1+e^{-}(\beta_0+\beta_1 \cdot x)}$$

Here β_0 = intercept.

β_1 = coefficient.

x = Independent variable.

NOTE:- This values usally computated maximum likelyhood estimation.

$\beta_0 = -10$

$\beta_1 = 0.25$.

$$= \frac{1}{1+e^{-}(-10+0.25 \times 40)}.$$

$= 0.5$.

If $p >= 0.5$ predict as 1 (will purchase).

$P < 0.5$ predict as 0 (will not purchase).

- **Difference between Logistic Regression and Linear Regression.**

Logistic Regression	Linear Regression.
1. predicts a ~~cont~~ categorical output (typically binary)	1. predicts a continuous output (real values).
2. Between 0 and 1 (as it represents probability)	2. Any real number ($-\infty$ to $+\infty$).
3. classification (algorithum type).	3. Regression (Algorithm type).
4. $y = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n$ (eq^n is used)	$P(Y=1)$ equation is used.
5. Sigmoid (Logistic function).	5. Linear function.
6. Log Loss cross-Entropy (Error measurement).	6. mean squared Error (MSE).
7. Spam detection diseases prediction. fraud detection. (application examples)	7. predicting house prices; temperature sales etc. (application Examples.
8. Same diagrams as Before.	

• Logistic Regression with a variable.

(I) Logistic Regression with o

The Logistic Regr

called as univariate

classification algorith

predict a binary ou

single input featur

Logistic

statistical model

function to mode

belongs to a part

• The mathemat

where; $h_\theta(x)$ is

output

→ β_0 is t

→ β_1 is t

→ e is th

• Same exam

(II) Logistic

also calle

extension

are

Logistic Regression	Linear Regression.
1. predicts a cont categorical output (typically binary)	1. predicts a continuous output (real value).
2. Between 0 and 1 (as it represents probability)	2. Any real number ($-\infty$ to $+\infty$).
3. classification (algorithum type).	3. Regression (Algorithm type).
4. $y = w_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n$ (eq^n is used)	$P(Y=1)$ equation is used.
5. Sigmoid (Logistic function).	5. Linear function.
6. Log Loss cross-Entropy (Error measurement).	6. mean squared Error (MSE).
7. spam detection diseases prediction. fraud detection. (application examples)	7. predicting house prices; temperature sales etc. (application Examples.
8. Same diagrams as Before.	

- Logistic Regression with one variable and with multi variable.

(i) Logistic Regression with one variable:-

The Logistic Regression with one variable also called as univariate logistic Regression is a classification algorithm used when we want to predict a binary outcome (0 or 1) based on a single input feature.

Logistic Regression with one variable is a statistical model their uses the logistic (sigmoid) function to model the probability that a given input belongs to a particular class.

- The mathematical formula is; $h_0(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 \times x)}}$

where; $h_0(x)$ is the predicted probability that the output is 1.

→ β_0 is the bias (intercept)

→ β_1 is the weight for the input feature x.

→ e is the base of the natural logarithm.

- Same example of Logistic Regression.

(ii) Logistic Regression with multi variable:-

Logistic Regression with multiple variable also called multivariate logistic Regression is an extension of logistic regression used when there are two or more input features.

- Logistic Regression with one variable and with multi variable.

(i) Logistic Regression with one variable:-

The Logistic Regression with one variable also called as univariate Logistic Regression is a classification algorithm used when we want to predict a binary outcome (0 or 1) based on a single input feature.

Logistic Regression with one variable is a statistical model their uses the Logistic (sigmoid) function to model the probability that a given input belongs to a particular class.

- The mathematical formula is; $h_\theta(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 \times x)}}$

where; $h_\theta(x)$ is the predicted probability that the output is 1.

→ β_0 is the bias (intercept)

→ β_1 is the weight for the input feature x.

→ e is the base of the natural logarithm.

- Same example of Logistic Regression.

(ii) Logistic Regression with multi variable:-

Logistic Regression with multiple variable also called multivariate Logistic Regression is an extension of Logistic regression used when there are two or more input features.

Multivariate Logistic regression is a classification algorithm that models the probability of a binary outcome (0 or 1) based on multiple input features using the Logistic (sigmoid) function.

Mathematical form:-

$$h_\theta(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\dots+\beta_n x_n)}}$$

or in vector form:-

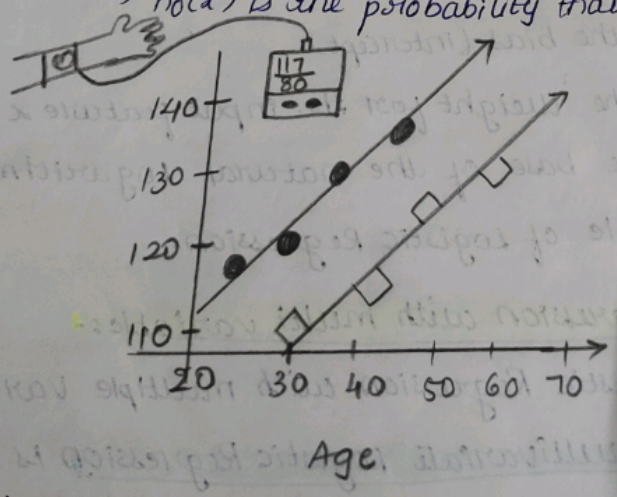
$$h_\theta(x) = \frac{1}{1+e^{-\beta^T x}}.$$

Where:- → $x_1, x_2, \dots, x_n$ are given feature inputs that model can predict.

→ x is the vector of input features (including $x_0 = 1$ for bias).

→ β is the vector of parameters (weights)

→ $h_\theta(x)$ is the probability that the outputs



Sample problem:-

Multivariate Logistic regression is a classification algorithm that models the probability of a binary outcome (0 or 1) based on multiple input features using the Logistic (sigmoid) function.

Mathematical form:-

$$h_0(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}}$$

or in vector form:-

$$h_0(x) = \frac{1}{1+e^{-\beta^T x}}.$$

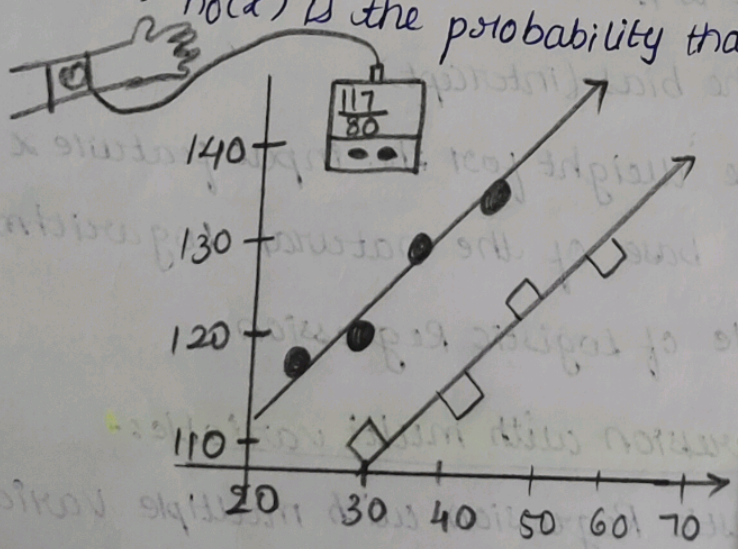
Where:-

→ $x_1, x_2, \ldots, x_n$ are given feature inputs that model can predict.

→ x is the vector of input features (including $x_0 = 1$ for bias).

→ β is the vector of parameters (weights)

→ $h_0(x)$ is the probability that the output is



Sample problem:-

Unit: 5 Unsupervised Learning.

• Unsupervised Learning:-

→ Unsupervised Learning is a machine learning concept where the unlabelled and unclassified information is analyzed to discover hidden knowledge.

→ The algorithms work on the data without any prior training.

• Example:-

→ movie promotions to the correct group of people.

→ Earlier times: same set of movie to all the visitors of the page.

→ Now: based on their interest; understand what type of movie is liked by what segment of people.

• Clustering:-

→ Cluster analysis find the commonalities, similarities between the data objects and categories them as per the presence and absence of those similarities. Clustering which helps in segmentations of the set of objects in to groups of similar objects.

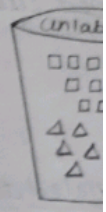
→ An association rule is an unsupervised learning method which is used for finding the relationship between variables / objects in the large database (dataset)

• Clustering:-

clustering is d

Learning task th

data into cluste



• Different

The m

1, K-me

2, Hie

3, Den

1, K-

bo

Unit: 5

Unsupervised Learning.

• Unsupervised Learning:-

→ Unsupervised Learning is a machine learning concept where the unlabelled and unclassified information is analyzed to discover hidden knowledge.

→ The algorithms work on the data without any prior training.

• Example:-

→ movie promotions to the correct group of people.

→ Earlier times: same set of movie to all the visitors of the page.

→ Now: based on their interest; understand what type of movie is liked by what segment of people.

• Clustering:-

→ Cluster analysis find the commonalities, similarities between the data objects and categories them as per the presence and absence of those similarities. Clustering which helps in segmentations of the set of objects in to groups of similar objects.

→ An association rule is an unsupervised learning method which is used for finding the relationship between variables/objects in the large database (dataset)

- **Clustering:-**

clustering is defined as an unsupervised machine learning task that automatically divides the data into clusters of groups of similar items.

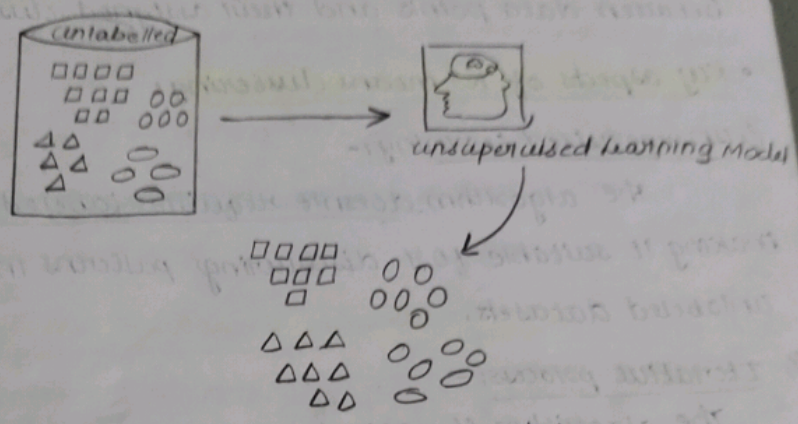


figure.

- **Different types of clustering techniques:-**

The major clustering techniques are:-

1, K-means (~~partitioning~~ ~~methods~~)

2, Hierarchical method.

3, Density based clustering method.

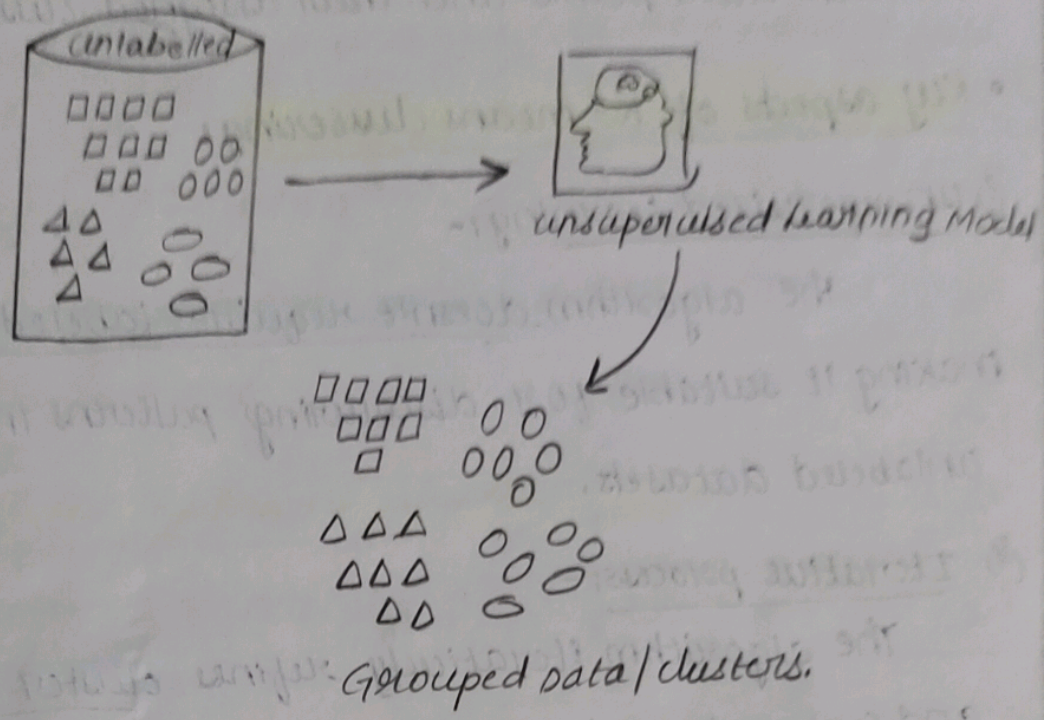
---

1. **K-means clustering:-**

K-means clustering is an unsupervised learning algorithm used to group data points into clusters based on their similarity. It aims to partition data into K distinct clusters, where each data point

- **Clustering:-**

Clustering is defined as an unsupervised machine learning task that automatically divides the data into clusters of groups of similar items.



figure

- **Different types of clustering techniques:-**

The major clustering techniques are:-

1, K-means (~~partitioning methods~~)

2, Hierarchical method.

3, Density based clustering method.

1, **K-means clustering:-**

K-means clustering is an unsupervised learning algorithm used to group data points into clusters based on their similarity. It aims to partition data into K distinct clusters, where each data point

belongs to the cluster with the nearest mean, also known as the centroid. The algorithm iteratively adjusts the cluster centroids to minimize the distance between data points and their assigned clusters.

- **Key aspects of K-means clustering:**

(i) Unsupervised Learning:-

The algorithm doesn't require labeled data, making it suitable for discovering patterns in unlabeled datasets.

(ii) Iterative process:-

The algorithm iteratively refines cluster assignments and centroids until convergence; meaning the cluster assignments stabilize.

(iii) Centroid-based:-

Each cluster is represented by a centroid, which is the average position of all data points within that cluster.

(iv) Distance metric:-

The algorithm uses a distance metric typically Euclidean distance to measure the similarity between data points and centroids.

v) Choosing the number of clusters (K):-

The user needs to specify the number of

clusters (K) to
and techniques
determine as o

- Applications:-

K-means cl
like:
1, Customer se
2, Image seg
3, Anomaly

- Steps of th

1, Initializat
Centroids
2, Assignme
Centroid
3, Update:
the data
Iterati

belongs to the cluster with the nearest mean; also known as the centroid. The algorithm iteratively adjusts the cluster centroids to minimize the distance between data points and their assigned clusters.

- **Key aspects of K-means clustering:**

(i) **Unsupervised learning:-**

The algorithm doesn't require labeled data, making it suitable for discovering patterns in unlabeled datasets.

(ii) **Iterative process:-**

The algorithm iteratively refines cluster assignment and centroids until convergence; meaning the cluster assignments stabilize.

(iii) **Centroid-based:-**

Each cluster is represented by a centroid; which is the average position of all data points within that cluster.

(iv) **Distance metric:-**

The algorithm uses a distance metric typically Euclidean distance to measure the similarity between data points and centroids.

v) **Choosing the number of clusters (K):-**

The user needs to specify the number of

th the nearest mean,
id. The algorithm iteratively
to minimize the distance
heir assigned clusters.

tering:

quire labeled data,
ing patterns in

ines cluster assignment
neaning the cluster

centroid, which
points within

tric typically
similarity

nber of

clusters (k) to be formed. This can be challenging, and techniques like the elbow method are used to determine an optimal k.

- **Applications:-**

K-means clustering is widely used in various field like:

1, Customer segmentation
2, Image segmentation
3, Anomaly detection.

- **Steps of the K-means algorithm:-**

1, Initialization:- Randomly select k initial cluster centroids.

2, Assignment:- Assign each data point to the nearest centroid forming clusters.

3, Update:- Recalculate the centroids by averaging the data points within each cluster.

4, Iteration:- Repeat steps 2 and 3 until cluster assignment stabilize or a maximum numbers of iteration is reached.

- **Example problem.**

Let us consider an example of K-means clustering.

clusters (K) to be formed. This can be challenging, and techniques like the elbow method are used to determine an optimal k.

- **Applications:-**

K-means clustering is widely used in various field like:

1, Customer segmentation

2, Image segmentation

3, Anomaly detection.

- **Steps of the K-means algorithm:-**

1, Initialization:- Randomly select K initial cluster centroids.

2, Assignment:- Assign each data point to the nearest centroid forming clusters.

3, Update:- Recalculate the centroids by averaging the data points within each cluster.

4, Iteration:- Repeat steps 2 and 3 until cluster assignment stabilize or a maximum numbers of iteration is reached.

- **Example problem.**

Let us consider an example of K-means clustering.

Data points	(x_1)	(y_1)	Distance to (x_2) 2 / (y_2) 10	5 / 8	1 / 2	Cluster	New Cluster
A_1	2	10	0.00	3.61	8.06	1	
A_2	2	5	5.00	4.24	3.16	3	
A_3	8	4	8.49	5.00	7.28	2	
B_1	5	8	3.61	0.00	7.21	2	
B_2	7	5	7.07	3.61	6.71	2	
B_3	6	4	7.21	4.12	5.39	2	
C_1	1	2	8.06	7.12	0.00	3	
C_2	4	9	2.24	1.41	7.62	2	

1. Initial Centroids:

A_1: (2,10)

B_1: (5,8)

C_1: (1,2)

The formula is : $d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

2. Current Centroids:-

A_1: (2,10)

B_1: (6,6)

C_1: (1.5,3.5).

Data points	(x_1)	(y_1)	Distance to 2 / 10	6 / 6	1.5 / 3.5	cluster	New cluster
A_1	2	10	0.00	5.66	6.52	1	1
A_2	2	5	5.00	4.12	1.58	3	3
A_3	8	4	8.49	2.83			
B_1	5	8	3.61	2.24			
B_2	7	5	7.07	1.41			
B_3	6	4	7.21	2.00			
C_1	1	2	8.06	6.40			
C_2	4	9	2.24	3.61			

3. New Centroids:

A_1: (3,9.5)

B_1: (6.5, 5.25)

C_1: (1.5,3.5)

Data points	(x_1)	(y_1)	Distance 3 / 9.5
A_1	2	10	1.12
A_2	2	5	4.61
A_3	8	4	7.43
B_1	5	8	2.50
B_2	7	5	6.02
B_3	6	4	6.26
C_1	1	2	7.76
C_2	4	9	1.12

4. New Centroids:

A_1: (3.67, 9)

B_1: (7, 4.33)

C_1: (1.5, 3.5)

Datapoints	(x_1)	(y_1)	Distance to (x_2) (y_2) 2 | 10	5 | 8	1 | 2	Cluster	New Cluster
A_1	2	10	0.00	3.61	8.06	1	
A_2	2	5	5.00	4.24	3.16	3	
A_3	8	4	8.49	5.00	7.28	2	
B_1	5	8	3.61	0.00	7.21	2	
B_2	7	5	7.07	3.61	6.71	2	
B_3	6	4	7.21	4.12	5.39	2	
C_1	1	2	8.06	7.12	0.00	3	
C_2	4	9	2.24	1.41	7.62	2	

1. Initial Centroids:

A_1: (2,10)

B_1: (5,8)

C_1: (1,2).

The formula is : $\boxed{d(p_1,p_2)=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}}$

2. Current Centroids:-

A_1: (2,10)

B_1: (6,6)

C_1: (1.5,3.5).

Data points	(x_1)	(y_1)	Distance to 2 | 10	6 | 6	1.5 | 3.5	cluster	New cluster
A_1	2	10	0.00	5.66	6.52	1	1
A_2	2	5	5.00	4.12	1.58	3	3

A3	8	4	8.49	2.83	6.52	2	2
B1	5	8	3.61	2.24	~~1.58~~ 5.70	2	2
B2	7	5	7.07	1.41	~~6.52~~ 5.70	2	2
B3	6	4	7.21	2.00	~~5.70~~ 4.53	2	2
C1	1	2	8.06	6.40	~~5~~ 1.58	3	3
C2.	4	9	2.24	3.61	6.04	2	1

3. New Centroids:

A_1: (3, 9.5)

~~B1~~ : (6.5, 5.25)

C_1 : (1.5, 3.5)

Data points			Distance to			cluster	New cluster.
	(x_1)	(y_1)	3 | 9.5	6.5 | 5.25	1.5 | 3.5		
A1	2	10	1.12	6.54	6.52	1	1
A2	2	5	4.61	4.51	1.58	3	3
A3	8	4	7.43	1.95	6.52	2	2
B1	5	8	2.50	3.13	5.70	2	1
B2	7	5	6.02	0.56	5.70	2	2
B3	6	4	6.26	1.35	4.53	2	2
C1	1	2	7.76	6.39	1.58	3	3
C2	4	9	1.12	4.51	6.04	1	1

4. New Centroids:

A_1: (3.67, 9)

B_1: (7, 4.33)

C_1: (1.5, 3.5)

A_3	8	4	8.49	2.83	6.52	2	2
B_1	5	8	3.61	2.24	~~1.58~~ 5.70	2	2
B_2	7	5	7.07	1.41	~~6.52~~ 5.70	2	2
B_3	6	4	7.21	2.00	~~5.70~~ 4.53	2	2
C_1	1	2	8.06	6.40	~~5~~ 1.58	3	3
C_2	4	9	2.24	3.61	6.04	2	1

3. New Centroids:

A_1: (3, 9.5)

B_1: (6.5, 5.25)

C_1: (1.5, 3.5)

Data points			Distance to			Cluster	New Cluster
	(x_1)	(y_1)	3 | 9.5	6.5 | 5.25	1.5 | 3.5		
A_1	2	10	1.12	6.54	6.52	1	1
A_2	2	5	4.61	4.51	1.58	3	3
A_3	8	4	7.43	1.95	6.52	2	2
B_1	5	8	2.50	3.13	5.70	2	1
B_2	7	5	6.02	0.56	5.70	2	2
B_3	6	4	6.26	1.35	4.53	2	2
C_1	1	2	7.76	6.39	1.58	3	3
C_2	4	9	1.12	4.51	6.04	1	1

4. New Centroids:

A_1: (3.67, 9)

B_1: (7, 4.33)

C_1: (1.5, 3.5)

Data points	(x_i)	(y_i)	Distance to 3.67 / 9	7 / 4.33	1.5 / 3.5	Cluster	New Cluster
A_1	2	10	1.94	7.56	6.52	1	1
A_2	2	5	4.33	5.04	1.58	3	3
A_3	8	4	6.62	1.05	6.52	2	2
B_1	5	8	1.67	4.18	5.70	1	1
B_2	7	5	5.21	0.67	5.70 / 4.58	2	2
B_3	6	4	5.52	1.05	4.53	2	2
C_1	1	2	7.49	6.44	1.58	3	3
C_2	4	9	0.33	5.55	6.04	1	1

2, Hierarchical clustering:-

→ Hierarchical clustering determines cluster assignments by a building a hierarchy. This is implemented bottom-up or a top-down approach.

→ Agglomerative clustering is the bottom-up approach. It merges the two points that are the points have been merged into a single cluster.

→ Divisive clustering is the top-down approach. It starts with all points are one cluster and splits the most similar cluster at each step until only single data points remain.

→ These methods produce a t of points called a dendrogram

→ Hierarchical clustering is meaning cluster assignment you run an algorithm tu data.

- Steps of Hierarchical clu

1. Initialization:- Each data as a separate cluster.

2. Distance calculation:- A indicating the distance

3. Merging / splitting:-

- Agglomerative:- The merged, Creatin
- Divisive:- A clust two smaller

4. Iteration:-

Step 2 and 3 a level of clustering i remains.

5. Visualization:-

The final as a dendogram

Data points			Distance to					Cluster	
	(x_1)	(y_1)	3.67	9	7	4.33	1.5 / 3.5		
A_1	2	10	1.94		7.56		6.52		
A_2	2	5	4.33		5.04		1.58	1	
A_3	8	4	6.62		1.05		6.52	3	1
B_1	5	8	1.67		4.18		5.70	2	3
B_2	7	5	5.21		0.67		5.70 4.58	1	2
B_3	6	4	5.52		1.05		4.53	2	1
C_1	1	2	7.49		6.44		1.58	2	2
C_2	4	9	0.33		5.55		6.04	3	2
								1	3
									1

2) **Hierarchical clustering:-**

→ Hierarchical clustering determines cluster assignments by a building a hierarchy. This is implemented bottom-up or a top-down approach.

→ Agglomerative clustering is the bottom-up approach. It merges the two points that are the points have been merged into a single cluster.

→ Divisive clustering is the top-down approach. It starts with all points are one cluster and splits the least similar cluster at each step until only

nce to

			Cluster	New Clus.
/4.33	1.5 / 3.5			
1.56	6.52		1	1
.04	1.58		3	3
05	6.52		2	2
18	5.70		1	1
57	5.70 4.58		2	2
5	4.53		2	2
4	1.58		3	3
	6.04		1	1

ermines cluster
erarchy. This is
op-down approach.
he bottom-up approach
rt are the points
e cluster.
down approach
cluster and split
tep until only

single data points remain.

→ These methods produce a tree-based hierarchy of points called a dendrogram.

→ Hierarchical clustering is a deterministic process; meaning cluster assignment won't change when you run an algorithm twice on the same input data.

- Steps of Hierarchical clustering:-

1. Initialization:- Each data point is initially treated as a separate cluster.

2. Distance calculation:- A distance matrix is computed, indicating the distance between each pair of cluster.

3. Merging / splitting:-
   - Agglomerative:- The two closest clusters are merged, creating a new cluster.
   - Divisive:- A cluster is choosen and split into two smaller clusters.

4. Iteration:-
   Step 2 and 3 are repeated until a desired level of clustering is reached or a single cluster remains.

5. Visualization:-
   The final cluster hierarchy is visualized as a dendogram

single data points remain.

→ These methods produce a tree-based hierarchy of points called a dendrogram.

→ Hierarchical clustering is a deterministic process; meaning cluster assignment won't change when you run an algorithm twice on the same input data.

- Steps of Hierarchical clustering:-

1. Initialization:- Each data point is initially treated as a separate cluster.

2. Distance calculation:- A distance matrix is computed, indicating the distance between each pair of cluster.

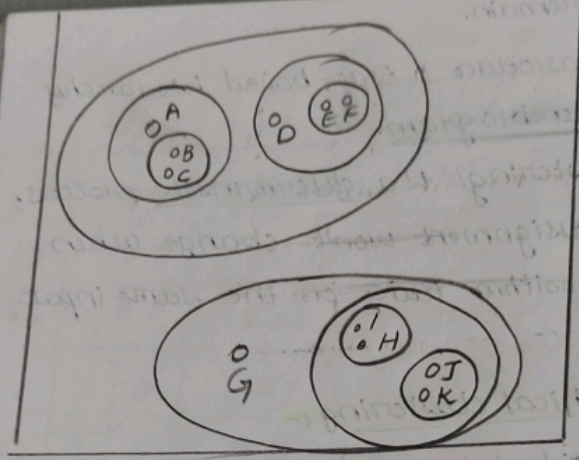
3. Merging / splitting:-
   - Agglomerative:- The two closest clusters are merged, creating a new cluster.
   - Divisive:- A cluster is choosen and split into two smaller clusters.

4. Iteration:-

   Step 2 and 3 are repeated until a desired level of clustering is reached or a single cluster remains.

5. Visualization:-

   The final cluster hierarchy is visualized as a dendogram



- **Example problem:-**

Let us consider an example for hierarchical clustering method.

→ consider the following set of 6 one dimensional data points:

→ 18, 22, 25, 27, 42, 43.

→ Apply the agglomerative hierarchical clustering algorithm to build the hierarchical clustering dendogram

→ merge the cluster using min distance and update the proximity matrix accordingly.

→ clearly show the proximity matrix corresponding to each iteration of the algorithm.

Consider steps:-

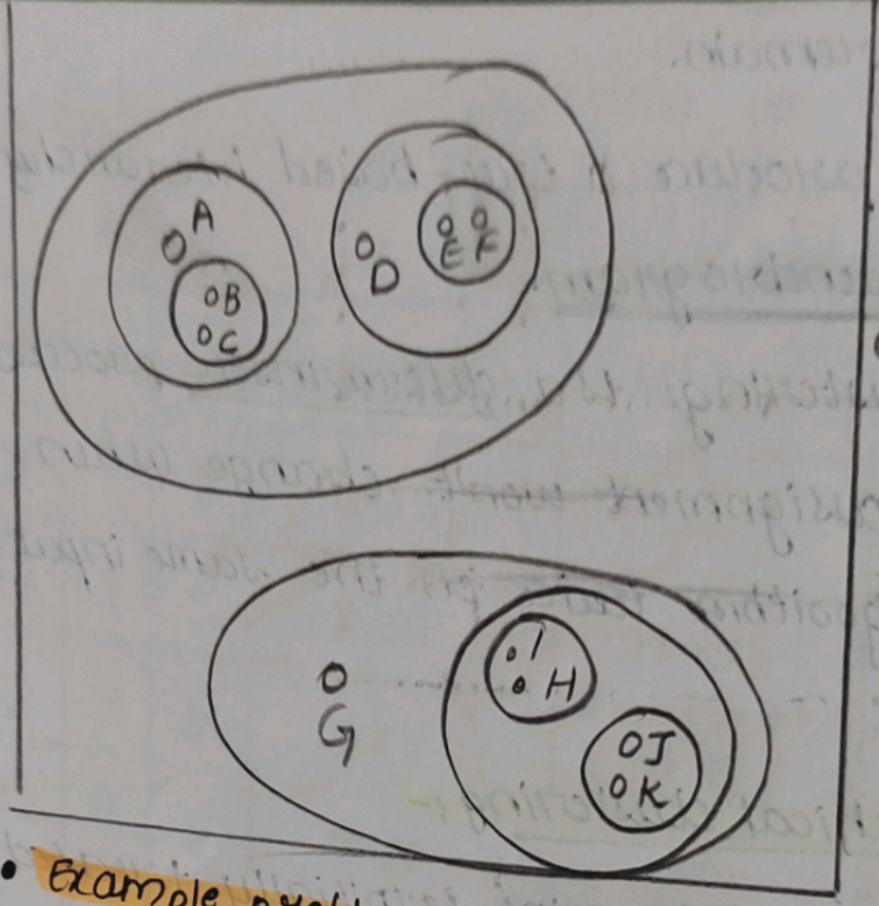
Step: 1.

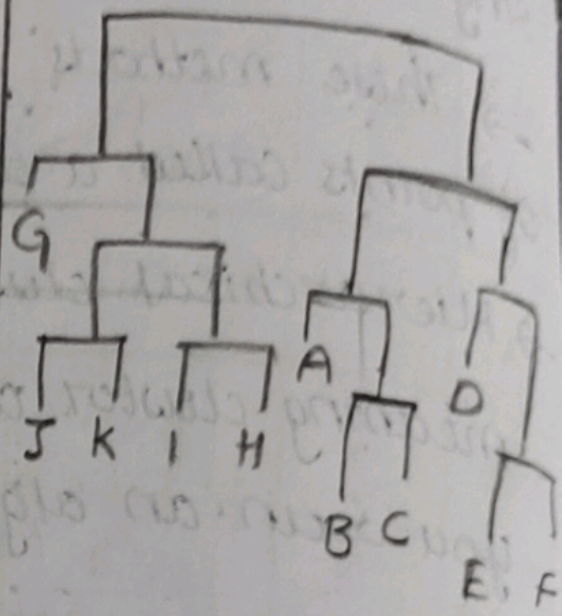
	18	22
18	0	4
22	4	0
25	7	3
27	9	5
42	24	2
43	25	2

Step: 2

	18
18	0
22	4
25	
27	
(42,43)	

Step: 3





- **Example problem:-**

Let us consider an example for hierarchical clustering method.

→ consider the following set of 6 one dimensional data points:

→ 18, 22, 25, 27, 42, 43.

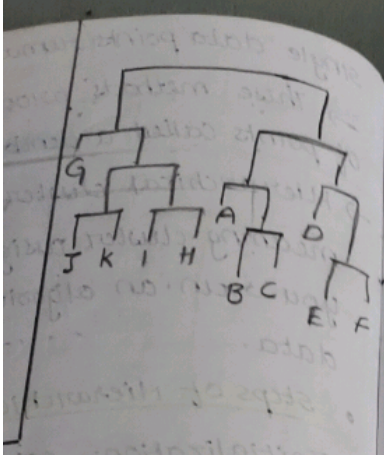
→ Apply the agglomerative hierarchical clustering algorithm to build the hierarchical clustering dendogram

→ merge the cluster using min distance and update the proximity matrix accordingly.

→ clearly show the proximity matrix corresponding to each iteration of the algorithm.

Consider steps:-

Step: 1.



for hierarchical

e dimensional

chical clustering

clustering dendogram

nce and update

x corresponding

	18	22	25	27	42	43
18	0	4	7	9	24	25
22	4	0	3	5	20	21
25	7	3	0	2	17	18
27	9	5	2	0	15	16
42	24	20	17	15	0	1
43	25	21	18	16	1	0

(42,43)

step:2

	18	22	25	27	(42,43)
18	0	4	7	9	24
22	4	0	3	5	20
25	7	3	0	2	17
27	9	5	2	0	15
(42,43)	24	20	17	15	0

(42,43) (25,27).

step:3

	18	22	(25,27)	(42,43)
18	0	4	7	24
22	4	0	3	20
(25,27)	7	3	0	15
(42,43)	24	20	15	0

(42,43),(25,27),22)

	18	22	25	27	42	43
18	0	4	7	9	24	25
22	4	0	3	5	20	21
25	7	3	0	2	17	18
27	9	5	2	0	15	16
42	24	20	17	15	0	1
43	25	21	18	16	1	0

(42,43)

step:2

	18	22	25	27	(42,43)
18	0	4	7	9	24
22	4	0	3	5	20
25	7	3	0	2	17
27	9	5	2	0	15
(42,43)	24	20	17	15	0

(42,43) (25,27).

step:3

	18	22	(25,27)	(42,43)
18	0	4	7	24
22	4	0	3	20
(25,27)	7	3	0	15
(42,43)	24	20	15	0

(42,43),((25,27),22)

Step:4 –

	18	22, 25, 27	42, 43
18	0	4	24
22, 25, 27	4	0	15
42, 43	24	15	0

(42,43),((25,27),22),18)

Step:5

	18, 22, 25, 27	42, 43
18, 22, 25, 27	0	15
42, 43	15	0

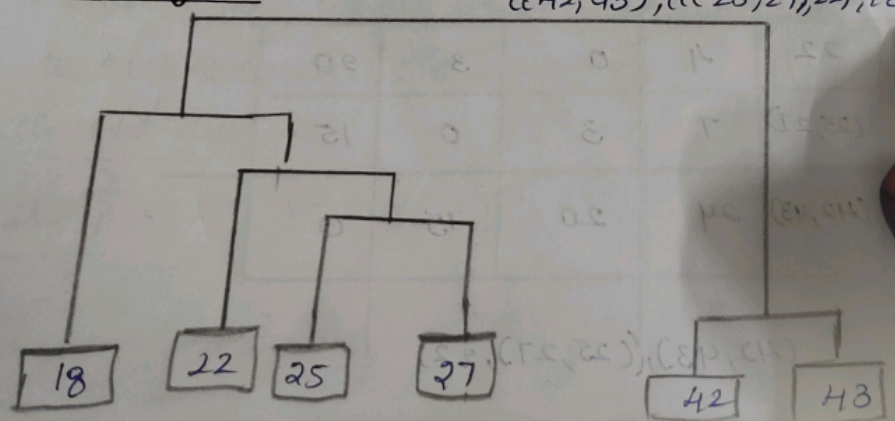
((42,43),(((25,27),22),18))

Step:6

	18, 22, 25, 27, 42, 43
18, 22, 25, 27, 42, 43	0.

• Dendrogram:-

((42,43),(((25,27),22),18))



3, Density Based clustering Metho

Density-based clustering is an technique that identifies clusters points that are closely packed to outliers as noise.

It works by finding separated by regions as out

The most well-know category is DBSCAN (Densit -ring of Applications with

• Key concepts:-

1. Core points – A point wit number of neighboring p specified radius (Epsilo

2. Border points – points tha neighborhood of a core enough neighbors to be

3. Noise (outliers) – point points reachable

DBSCAN

Step:4 -

	18	22, 25, 27	42, 43
18	0	4	24
22, 25, 27	4	0	15
42, 43	24	15	0

(42,43),((25,27),22),18)

Step:5

	18, 22, 25, 27	42, 43
18, 22, 25, 27	0	15
42, 43	15	0

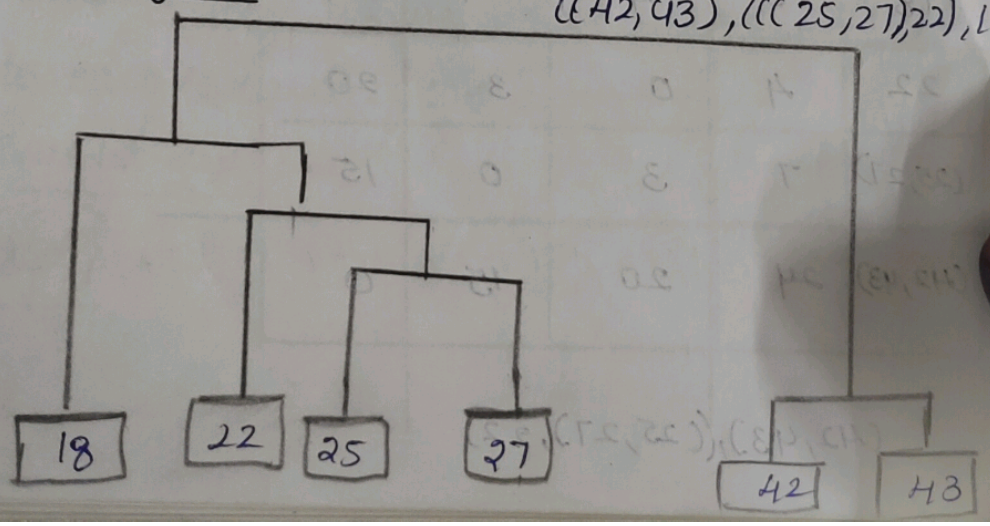
((42,43),(((25,27),22),18))

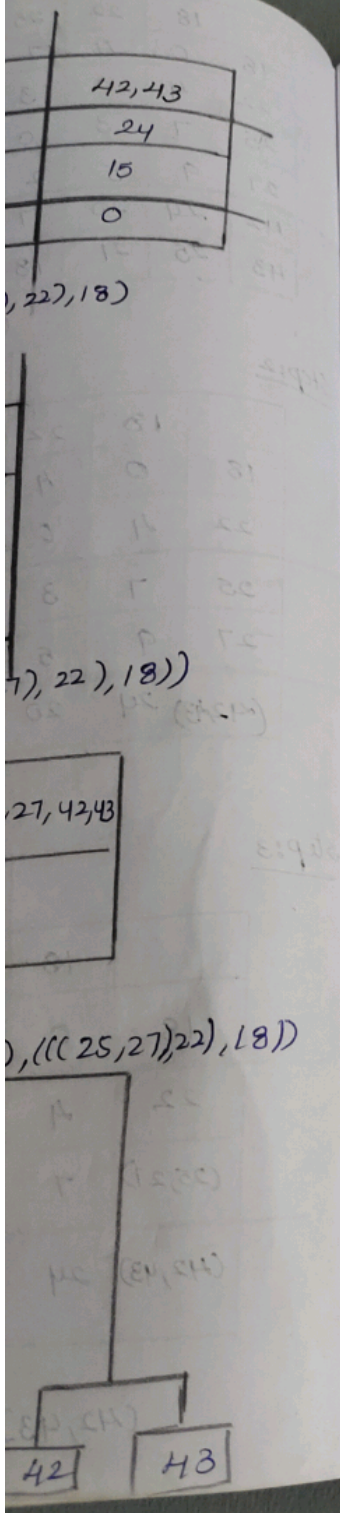
Step:6

	18, 22, 25, 27, 42, 43
18, 22, 25, 27, 42, 43	0.

- Dendrogram:-

((42,43),(((25,27),22),18))





3, Density Based clustering Methods:-

Density-based clustering is an unsupervised learning technique that identifies clusters by grouping data points that are closely packed together, and marking outliers as noise.

It works by finding regions of high density separated by regions as outliers.

The most well-known algorithm in this category is DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

- Key concepts:-

1. Core points - A point with at least a minimum number of neighboring points (minpts) with in a specified radius (Epsilon).

2. Border points - points that are within the neighborhood of a core point but do not have enough neighbors to be core points themselves.

3. Noise (outliers) - points that are neither core points nor reachable from any core point.

DBSCAN Algorithm:-

Steps:-

1. pick a point at random.

2. If it's a core point, create a cluster around it.

3. Expand the cluster by including all points density

3) Density Based clustering Method:-

Density-based clustering is an unsupervised learning technique that identifies clusters by grouping data points that are closely packed together, and marking outliers as noise.

It works by finding regions of high density separated by regions as outliers.

The most well-known algorithm in this category is DBSCAN (Density-Based Spatial clustering of Applications with Noise).

- Key concepts:-

1. Core points - A point with at least a minimum number of neighboring points (minpts) with in a specified radius (Epsilon).

2. Border points - points that are within the neighborhood of a core point but do not have enough neighbors to be core points themselves.

3. Noise (outliers) - points that are neither core points nor reachable from any core point.

DBSCAN Algorithm:-

Steps:-

1. Pick a point at random.

2. If it's a core point, create a cluster around it.

3. Expand the cluster by including all points density

reachable from it.

4. Repeat until all points are processed.

- Advantages:-
- can find arbitrarily shaped clusters.
- Robust to outliers.
- No need to specify the no. of. clusters beforehand (unlike k-means)
- Limitations:-
- struggles with varying density clusters.
- performance can degrade in high-dimensional spaces.
- sensitive to the choice of ε and minpts.
- Popular Alternatives:-
- OPTICS - (ordering points to identify the clustering structure):
- HDBSCAN - (Hierarchical DBSCAN), which improves performance on datasets with variable densities.
- Consider an example:-

→ Apply the DBSCAN algorithm to the given data points

→ create the clusters with minpts = 4

epsilon (ε) = 1.9

Data points:-

P_1: (3,7) P_2: (4,6)

P_3: (5,5) P_4: (6,4)

P_5: (7,3) P_6: (6,2)

P_7: (7,2) P_8: (8,4)

P_9: (3,3) P_{10}: (2,6)

P_{11}: (3,5) P_{12}: (2,4)

- use Euclidian distance a

between each points.

Distance (A(x_1,y_1); B(x_2

	P_1	P_2	P_3	P_4	P_5
P1	0				
P2	1.41	0			
P3	2.83	1.41	0		
P4	4.24	2.83	1.41	0	
P5	5.66	4.24	2.83	1.41	
P6	5.83	4.47	3.16	2.00	
P7	6.40	5.[illegible]	[illegible].61	2.24	
	5.83	4.[illegible]		2.00	
		3.16			
			3.[illegible]		

reachable from it.

4. Repeat until all points are processed.

- **Advantages:-**
- can find arbitrarily shaped clusters.
- Robust to outliers.
- No need to specify the no. of. clusters beforehand (unlike k-means)
- **Limitations:-**
- struggles with varying density clusters.
- performance can degrade in high-dimensional spaces.
- sensitive to the choice of ε and minpts.
- **Popular Alternatives:-**
- OPTICS - (ordering points to identify the clustering structure):
- HDBSCAN - (Hierarchical DBSCAN), which improves performance on datasets with variable densities.
- **Consider an example:-**

→ Apply the DBSCAN algorithm to the given data points

→ create the clusters with minpts = 4

epsilon (ε) = 1.9

Data points:-

P_1: (3,7) P_2: (4,6)

P_3: (5,5) P_4: (6,4)

P_5: (7,3) P_6: (6,2)

P_7: (7,2) P_8: (8,4)

P_9: (2,3) P_{10}: (2,6)

P_{11}: (3,5) P_{12}: (2,4)

- use Euclidian distance and calculate the distance between each points.

Distance $(A(x_1, y_1); B(x_2, y_2)) = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
P1	0											
P2	1.41	0										
P3	2.83	1.41	0									
P4	4.24	2.83	1.41	0								
P5	5.66	4.24	2.83	1.41	0							
P6	5.83	4.47	3.16	2.00	1.41	0						
P7	6.40	5.00	3.61	2.24	1.00	1.00	0					
P8	5.83	4.47	3.16	2.00	1.41	2.83	2.24	0				
P9	4.00	3.16	2.83	3.16	4.00	3.16	4.12	5.10	0			
P10	1.41	2.00	3.16	4.47	5.83	5.66	6.40	6.32	3.16	0		
P11	2.00	1.41	2.00	3.16	4.47	4.24	5.00	5.10	2.00	1.41	0	
P12	3.16	2.83	3.16	4.00	5.10	4.47	5.39	6.00	1.41	2.00	1.41	0

Data points:-

P_1: (3, 7) P_2: (4, 6)

P_3: (5, 5) P_4: (6, 4)

P_5: (7, 3) P_6: (6, 2)

P_7: (7, 2) P_8: (8, 4)

P_9: (3, 3) P_{10}: (2, 6)

P_{11}: (3, 5) P_{12}: (2, 4)

- use Euclidian distance and calculate the distance between each points.

Distance $(A(x_1, y_1), B(x_2, y_2)) = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
P1	0											
P2	1.41	0										
P3	2.83	1.41	0									
P4	4.24	2.83	1.41	0								
P5	5.66	4.24	2.83	1.41	0							
P6	5.83	4.47	3.16	2.00	1.41	0						
P7	6.40	5.00	3.61	2.24	1.00	1.00	0					
P8	5.83	4.47	3.16	2.00	1.41	2.83	2.24	0				
P9	4.00	3.16	2.83	3.16	4.00	3.16	4.12	5.10	0			
P10	1.41	2.00	3.16	4.47	5.83	5.66	6.40	6.32	3.16	0		
P11	2.00	1.41	2.00	3.16	4.47	4.24	5.00	5.10	2.00	1.41	0	
P12	3.16	2.83	3.16	4.00	5.10	4.47	5.39	6.00	1.41	2.00	1.41	0