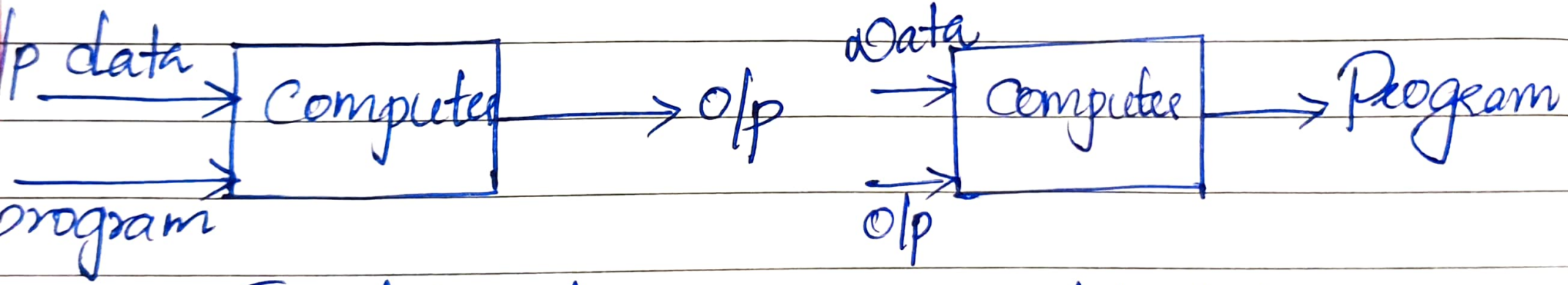
MODULE - 1

Introduction

The field of study that gives computers the ability to learn without being explicitly programmed



Traditional Programming

Machine learning

Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience or instructions, in order to look for patterns in data and make better decesions in the future based on the examples that we provide.

Machine leaeming works on the a simple concept

Facebook :- continuously notices

↳ friends that you connect with
↳ people that you visit
↳ your interest etc.

on the basis of continuous learning, a list of facebook users are suggested that you can become friends with.

Tag friends :-

When you upload a picture of you with a friend, Facebook instantly recognizes that friend. This is possible with the help of machine learning.

Ads Recommendation

When you shop any product online, after some days you keep receiving notifications for shopping suggestions.

The shopping website or the app recommends you some items that somehow matches with your interest.

This is possible with the help of machine learning on the basis of your behaviours with the website/app, past purchases, items liked or added to cart etc the product recommendations are made.

The primary aim of machine learning :-

is to allow the computers learn automatically without any human interventions.

* A computer program is said to learn from experience 'E' with respect to some class of tasks 'T' and performance measure 'P', if its performance at tasks T, as measured by P, improves with experience E.

eg:

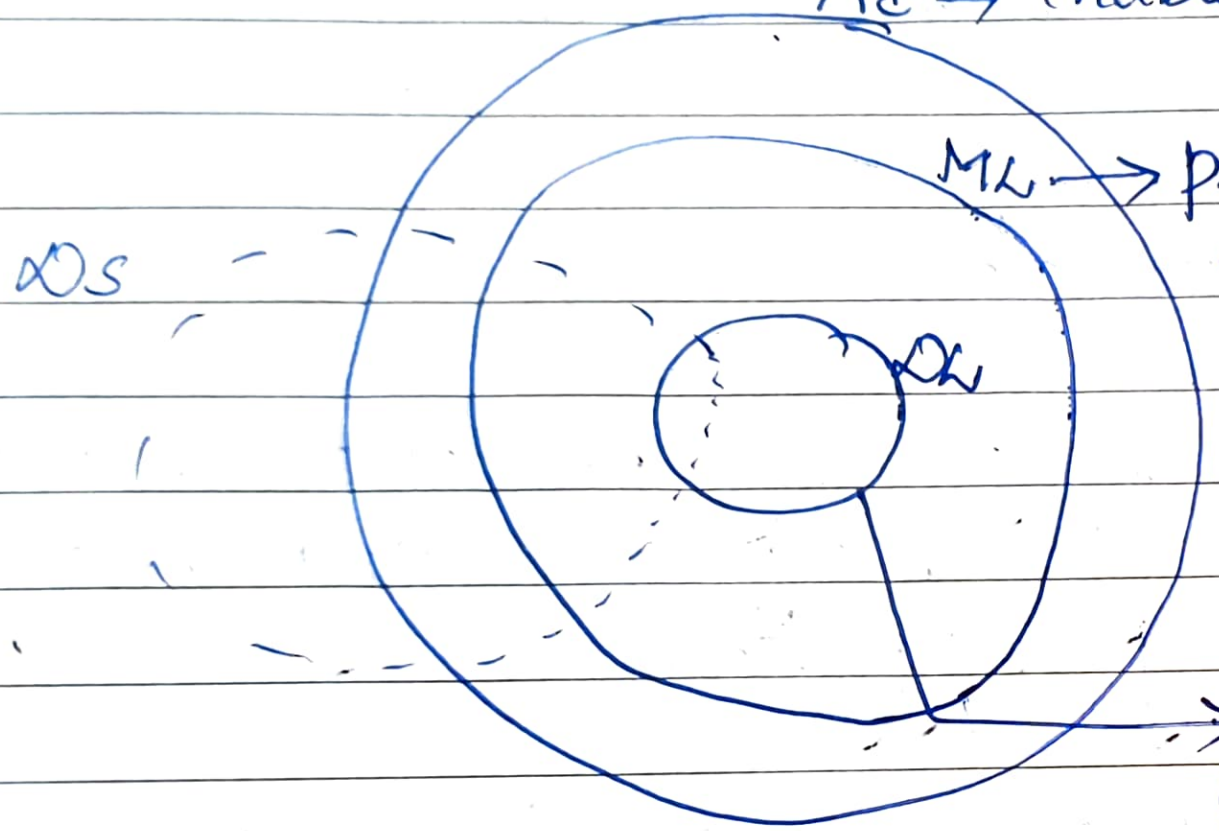
Handwriting recognition learning problem

- Task T: Recognising and classifying handwritten words within images
- Performance P: Percentage of words correctly classified

- Training Experience E:
  A dataset of handwritten words with given classifications.

eg:- A robot driving learning problem

- Task T: Driving on highways using vision sensors.
- Performance Measure P: Average distance traveled before an error
- Training Experience: A sequence of images and steering commands recorded while observing a human drives



AI → Enable the machine to think
goal: takes its own decision

ML → provides the statistical tools to explore and analyse the data. or understanding the data.

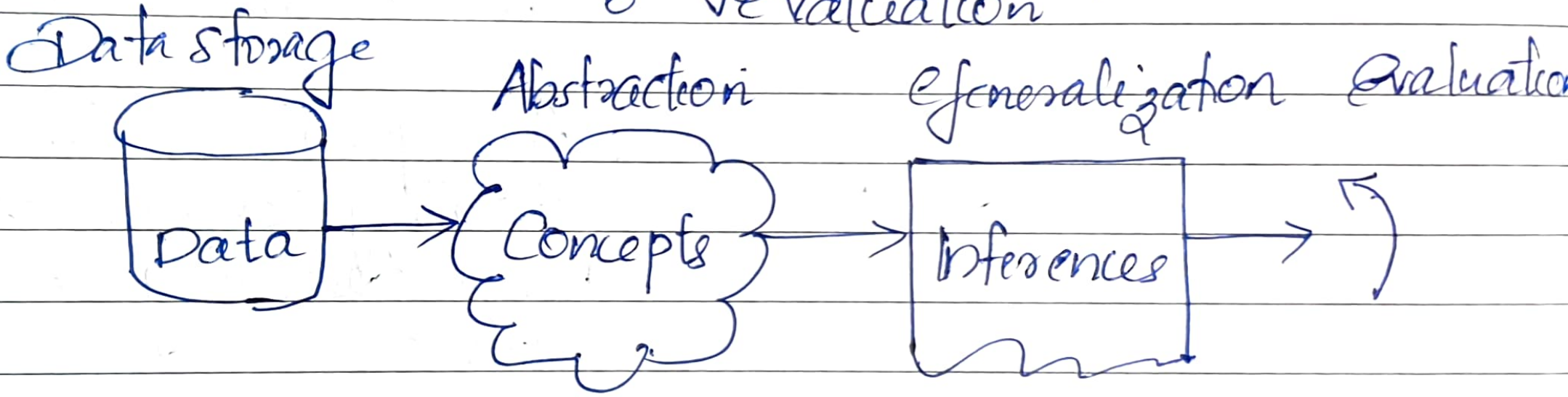
→ Can a machine able to learn? how a human learn?
(we are creating architecture called multineural n/w)

- Main idea is to mimic human brain

OR Elements of Machine Learning.

Basic Components of learning Process

The learning process, whether by a human or a machine, can be divided into four components, namely:
- Data storage
- Abstraction
- Generalization
- Evaluation.



Data storage :-

- Facilities for storing and retrieving huge amounts of data

- Humans and Computers alike utilize data storage as a foundation for advanced reasoning

  ↳ In a human being, the data is stored in the brain and data is retrieved using electrochemical signals.

↳ Computers use hard disk drives, flash mly, random access memory and similar devices to store data and use cables and other technology to retrieve data.

Abstraction :-

* Abstraction is the process of extracting knowledge about stored data

* Creating general concepts about the data as a whole

* Creation of knowledge involves application of known models and creation of new models

* The process of fitting a model to a dataset is known as training. When the model has been trained, the data is transformed into an abstract form that summarizes the original information

Generalization

* The term generalization describes the process of turning the knowledge about stored data into a form that can be utilized for future action.

Evaluation

It is the process of giving feedback to the user to measure the utility of the learned knowledge. This feedback is then utilized to effect improvements in the whole learning process.

Applications of Machine learning :-

Application of machine learning methods to large database is called Data Mining

In retail business, machine learning is used to study consumer behaviour

In finance, banks analyze their past data to build models to use in credit applications, fraud detection, and the stock market

In manufacturing, learning models are used for optimization, control and troubleshooting

In medicine, learning programs are used for medical diagnosis

Evaluation

It is the process of giving feedback to the user to measure the utility of the learned knowledge. This feedback is then utilized to effect improvements in the whole learning process.

Applications of Machine learning :-

Application of machine learning methods to large database is called Data Mining

In retail business, machine learning is used to study consumer behaviour

In finance, banks analyze their past data to build models to use in credit applications, fraud detection, and the stock market

In manufacturing, learning models are used for optimization, control and troubleshooting

In medicine, learning programs are used for medical diagnosis

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

6) Machine learning methods are applied in the design of computer-controlled vehicles to steer correctly when driving on a variety of roads.

7) Machine learning methods have been used to develop programmes for playing games such as chess, backgammon and Go.

Evaluation

It is the process of giving feedback to the user to measure the utility of the learned knowledge. This feedback is then utilized to effect improvements in the whole learning process.

Applications of Machine learning :-

Application of machine learning methods to large database is called Data Mining

In retail business, machine learning is used to study consumer behaviour

In finance, banks analyze their past data to build models to use in credit applications, fraud detection, and the stock market

In manufacturing, learning models are used for optimization, control and troubleshooting

In medicine, learning programs are used for medical diagnosis

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

General Classes of Machine learning Problems :-

1) Learning Associations

(a) :- Association rule learning

Association rule learning is a machine learning method for discovering interesting relations, called "association rules", between variables in large databases using

some measure of "interestingness".

(b) Example

Consider a supermarket chain. The management of the chain is interested in knowing whether there are any patterns in the purchase of products by customers like the following

"If a customer buys onions and potatoes together, then he/she is likely to also buy hamburger".

* Consider two sets, {onion, potato} and {burger}

We can define an association between the set of products {onion, potato} and the set {burger}. This association is represented in the form of a rule as follows

{onion, potato} ⟹ {burger}

* Conditional probability = Measure of how likely a customer, who has bought onion and potato, to buy burger also is given by

P({onion, potato} | {burger})

In finding an association rule $X \Rightarrow Y$, we are using a conditional probability of the form $P(Y|X)$ where 'Y' is the product the customer may buy and 'X' is the product or set of

products the customer has already purchased

$P(Y|X, \Theta)$ → to make a distinction among customers, we may estimate $P(Y|X, \Theta)$, where 'Θ' is a set of customers attributes, like gender, age, marital status and so on.

Algorithms -

There are several algorithms for generating association rules.

eg:- a) Apriori algorithm

b) Eclat algorithm

c) Frequency pattern algorithm

Classification :-

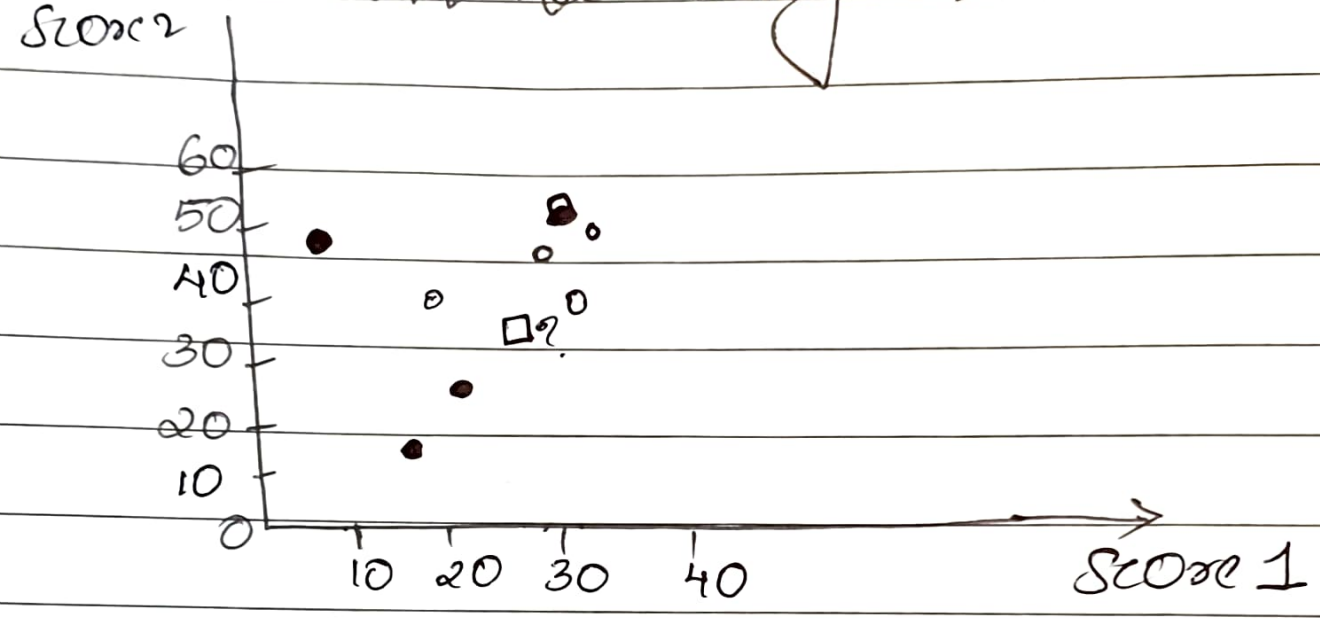
Classification is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known.

eg :- Consider the following data

Score 1	29	22	10	31	17	33	32	20
Score 2	43	29	47	55	18	54	40	41
Result	pass	Fail	Fail	pass	Fail	pass	pass	pass.

↳ Given a training set of data.
↳ Two attributes Score1 and Score2.
↳ Class label = Result
↳ Class label has 2 possible values "Pass" and "Fail"
↳ The data can be divided into two categories
(1) Set of data for which the class label is pass
(2) Set of data for which the class label is fail.

* If we have some new data, for eg:- Score1 = 25 and Score 2 = 36, wha value should be assigned to "Result



↳ to find out the above question using given data alone we need to find the or the formula or method

* The problem of finding this rule or formula or method is the Classificat problem

Real life Examples:-

a) Optical character Recognition
(b) Face Regonition
(c) Speech Recognition
(d) Medical Diagnosis
(e) Knowledge Extraction
(f) Compression

Discriminant:-

A discriminant of a classification problem is a rule or a function that is used to assign labels to new observations

eg:- If score 1 + score 2 $\geq$ 60, THEN "pass" E
"fail"

IF score 1 $\geq$ 20 AND score 2 $\geq$ 40 THE
"pass" ELSE "FAIL".

Algorithms:-

machine learning

There are several algorithms for classification

a) Logistic Regression
b) Naive Bayes algorithm
c) k-NN algorithm
d) Decision tree algorithm
e) Support vector machine algorithm
f) Random forest algorithm

* A classification problem requires that examples be classified into one of two or more classes

* A classification can have real-valued or discrete input variables

* A problem with two classes is often a two class or binary classification

* A problem with more than two classes often called a multi-class classification problem.

Regression :-

* Regression problem is the problem of predicting the value of a numeric variable based on observed values of the variable.

* The value of the o/p variable may be a number, such as an integer or a floating point value. Input variables may be discrete or real-valued

→ We are required to estimate the price of a car aged 25 years with distance 53240 KM and weight 1200 pounds.

* This is an example of a regression problem because we have to predict the value of the numeric variable "price".

* Let 'x' denote the set of input variable
y ⇒ output variable

* General approach to regression is to assume a model i.e a mathematical relation b/w 'x' and 'y' involving some parameter 'θ'.

$$y = f(x, \theta)$$

* The function $f(x, \theta)$ is called regression function.

Aim :- Estimates the values of the dependent variable 'y' are as close as possible to the correct values given in the training set

eg:-

$$y = f(x, \theta)$$

$$price = a_0 + a_1(Age) + a_2 * (Distance) + a_3 * (weight)$$

* Based on three different aspects,
  - the number and type of independent variables
  - type of dependent variables
  - shape of regression line

Regression models can be classified as below.

- Simple linear Regression ($y = a + bx$)
- Multivariate linear Regression ($y = a_0 + a_1x_1 + \cdots + a_nx_n$)
- Polynomial Regression ($y = a_0 + a_1x + \cdots a_nx^n$)
- Logistic Regression (0/1 (probability))

Classification
Regression
Ranking

→ Clustering
→ Association
→ Segmentation
→ Dimension Reduction

→ Decision Process
→ Reward system
→ Recommendation System

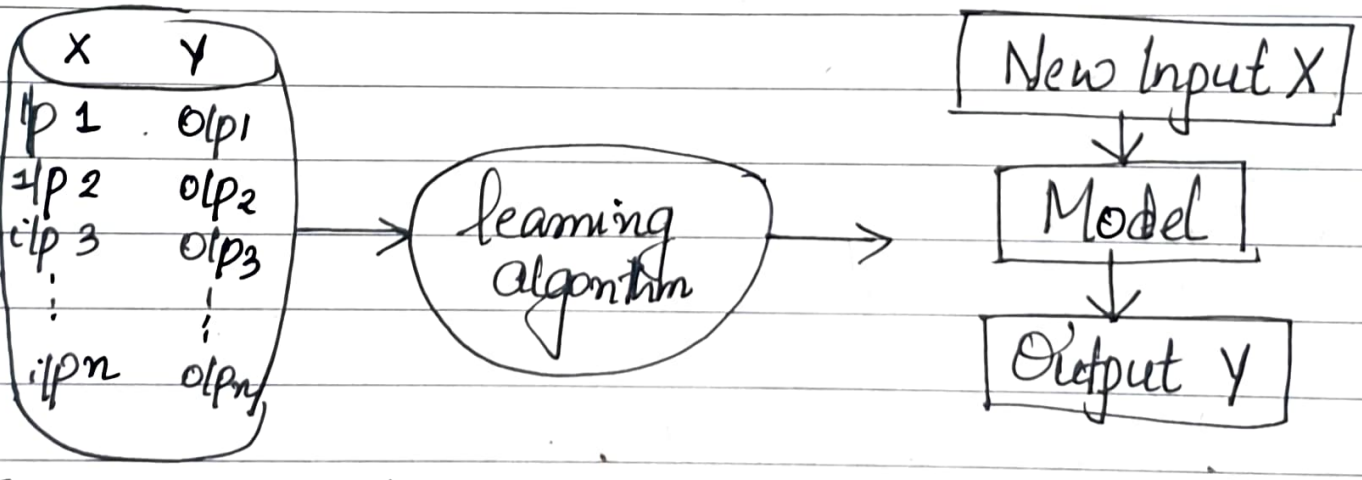
Supervised learning :-

→ For learning we need data. Every data has i/p and o/p for every data instance we have input : x and output : y

→ From this the machine learning system will build a model so that given an observation, 'x', it will try to find out what is the corresponding 'y'.

This is called supervised learning because for every instance we tell what is the output. This is called as labelled data

* For supervised learning
  - X, Y (pre-classified training example)
  - Given an observation X, what is the best label for Y.



Training instances :-

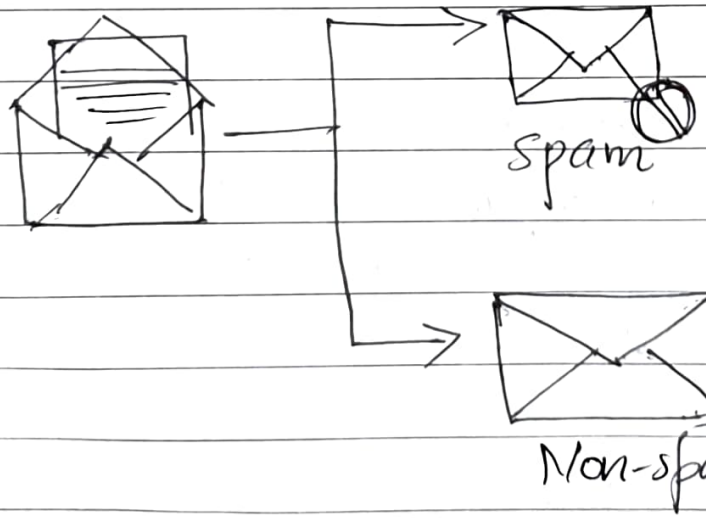
Given all the training instances the learning algorithm will come up with a model, and this model can be used to classify or to find the corresponding 'y' value for a new observation X

Classification

Classification is the task of predicting a discrete class label

In a classification pblm data is classified into one of two or more classes

A classification problem with two classes is called binary, more than two classes is called a multi-class classification

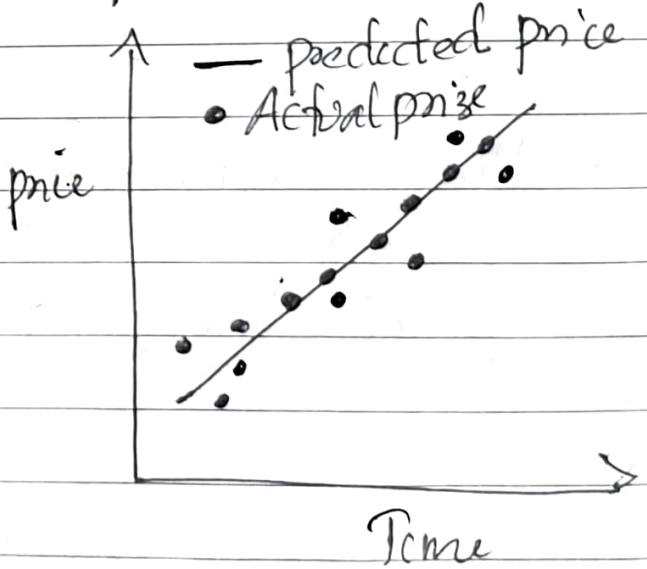


Regression

* Regression is the task of predicting a continuous quantity

* A regression pblm requires the prediction of a quantity

* A regression pblm with multiple i/p variables is called a multivariate regression pblm

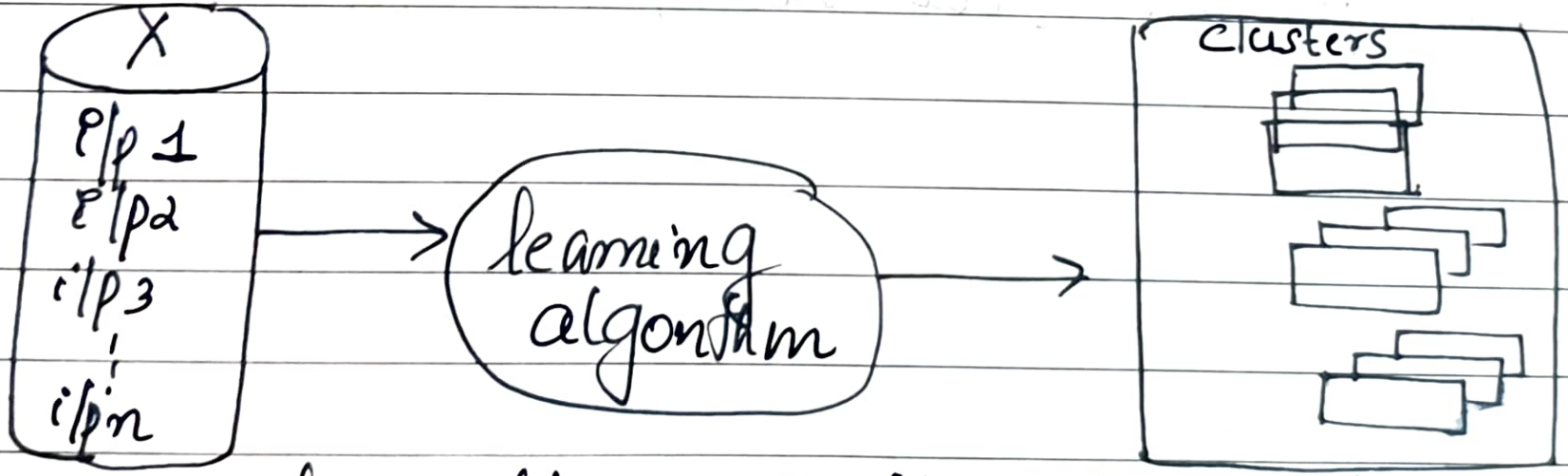


Unsupervised Learning :-

- X

* Given a set of X's, cluster or summarize them. Here you don't have label to the data. Given the different data points, you cluster them, summarize them, or find

Some common pattern between them ..



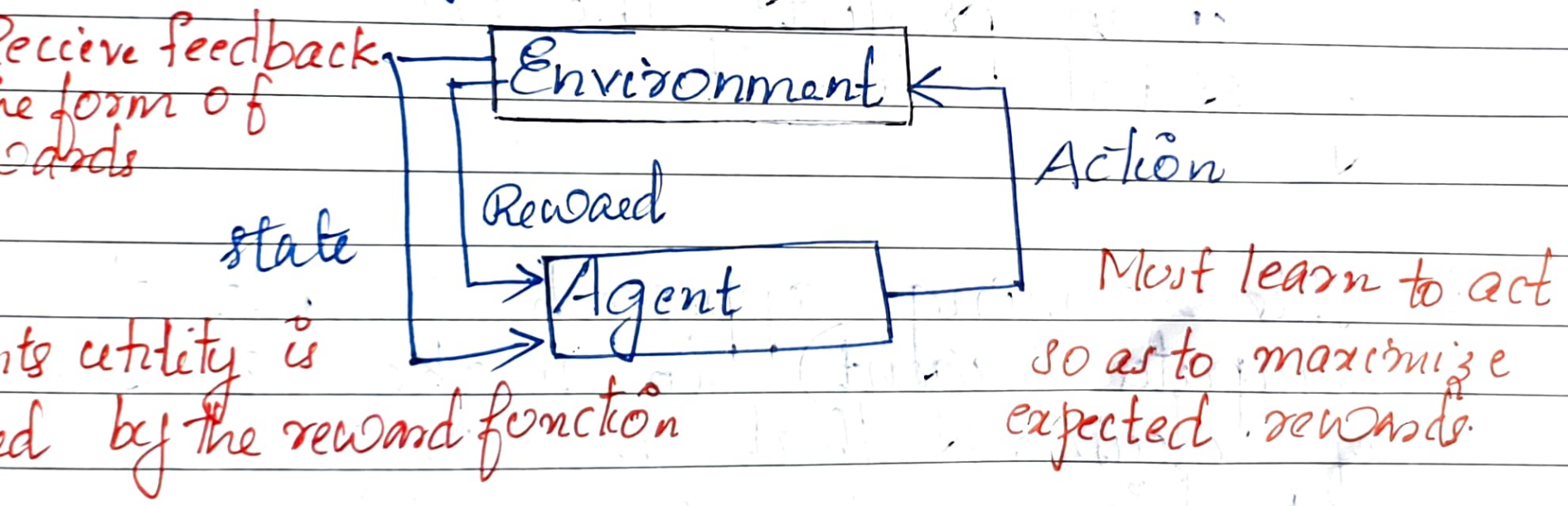
Based on the similarity of the data item we can find out certain groups among the data

→ Mainly used algorithms are

- → k-means clustering
- → Fuzzy C-means clustering
- → Hierarchical clustering
- → Apriori Algorithm
- → Eclat Algorithm
- → Frequency Pattern Algorithm etc.

Reinforcement learning :-

It is a type of machine learning in which a computer learns to perform a task through repeated trial and error interactions with a dynamic environment.



Receive feedback
he form of
ards

nts utility is
ed by the reward function

Must learn to act
so as to maximize
expected rewards.

It is different from supervised and unsupervised machine learning algorithms that require

large datasets for training purposes.

* Reinforcement algorithms are designed to continuously learn from experience rather than only from data

* In Reinforcement learning, an agent will have to interact with environment and findout what is the best outcome.

* Agent follows the concept of hit and trial method, agent is awarded and penalized with a point for a correct or a wrong answer.

On the basis of the positive rewards points gained and the model trains itself.

Once the agent gets trained, it is ready to predict the new data presented to it.

This learning approach enables the computer (agent) to make a series of decisions that maximize a reward metric for the task without any human intervention and without being explicitly programmed to achieve the task.

Basic Elements of a Reinforcement learning :-

1) Environment :- Physical world in which the agent operates

2) State : Current situation of the agent

3) Policy : Method to map agent's state to actions

4) Reward : Feedback from the environment

5) Value : Future reward that an agent would receive by taking an action in a particular state.

Applications of Reinforcement Learning :-

→ Robotics for industrial automation
→ Business strategy planning
→ Aircraft control and Robot motion control
→ Gaming

Supervised Learning :-

Input Representation :-

* Considers the problem of assigning the label "family car" or "not family car" to cars.
* Let us assume that the features that depersate a family car from other cars are the
  * price
  * Engine power

  } ⇒ Constitute the i/p representation for the problem
* People look the cars and label them that, the cars in the family cars are positive egs. and others are negative egs.
* The price and Engine powers are 2 attributes of cars, which is known as class Recognizer

we can make decision by using these particular input representation, we ignores all other attributes as irrelevant, like seating capacity, colours etc.

In supervised classification :-

- Learning the class 'C' of a family car from examples
  - ↳ Prediction :- Is car 'x' a family car?
  - ↳ Knowledge Extraction :- What do people expect from a family car?

- Output :- (labels)
  - ↳ positive (+) and
  - ↳ negative (-)

- Input Representation :-

x_1 :- Price ; x_2 :- Engine power.