MALLA REDDY COLLEGE OF ENGINEERING
(Approved by AICTE-New Delhi, Affiliated to JNTUH-Hyderabad)
Recognised under Section 2(f) & 12(B) of the UGC Act 1956,
An ISO 9001:2015 Certified Institution.
Maisammaguda, Dhulapally, post via Kompally, Secunderabad - 500100

R22 B.Tech. CSE (AI and ML) Syllabus

JNTU Hyderabad

SUBJECTE CODE:AM502PC

MACHINE LEARNING

B.Tech. III Year I Sem.

MALLA REDDY COLLEGE OF ENGINEERING

Course Objectives:

- To introduce students to the basic concepts and techniques of Machine Learning.
- To have a thorough understanding of the Supervised and Unsupervised learning techniques
- To study the various probability-based learning techniques

Course Outcomes:

- Distinguish between, supervised, unsupervised and semi-supervised learning
- Understand algorithms for building classifiers applied on datasets of non-linearly separable classes
- Understand the principles of evolutionary computing algorithms
- Design an ensemble to increase the classification accuracy

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

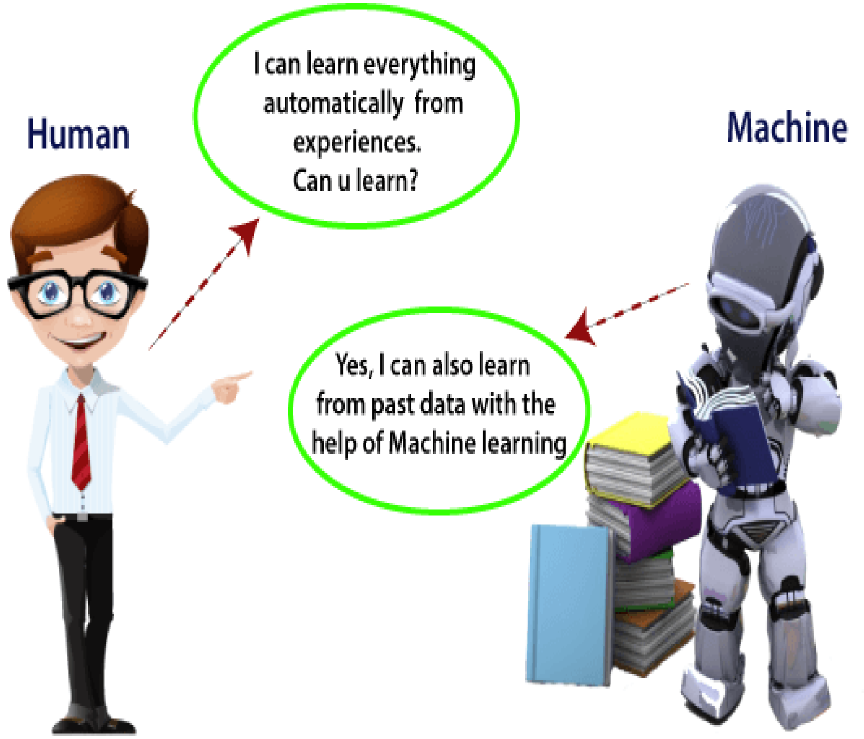
Learning – Types of Machine Learning – Supervised Learning – The Brain and the Neuron – Design a Learning System – Perspectives and Issues in Machine Learning – Concept Learning Task – Concept Learning as Search – Finding a Maximally Specific Hypothesis – Version Spaces and the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Learning –

A rapidly developing field of technology, machine learning allows computers to automatically learn from previous data. For building mathematical models and making predictions based on historical data or information, machine learning employs a variety of algorithms. It is currently being used for a variety of tasks, including speech recognition, email filtering, auto-tagging on Facebook, a recommender system, and image recognition.



.

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Learning

Machine learning contains a set of algorithms that work on a huge amount of data. Data is fed to these algorithms to train them, and on the basis of training, they build the model & perform a specific task.

These ML algorithms help to solve different business problems like **Regression, Classification, Forecasting, Clustering, and Associations, etc.**

Based on the methods and way of learning, machine learning is divided into mainly four types, which are:

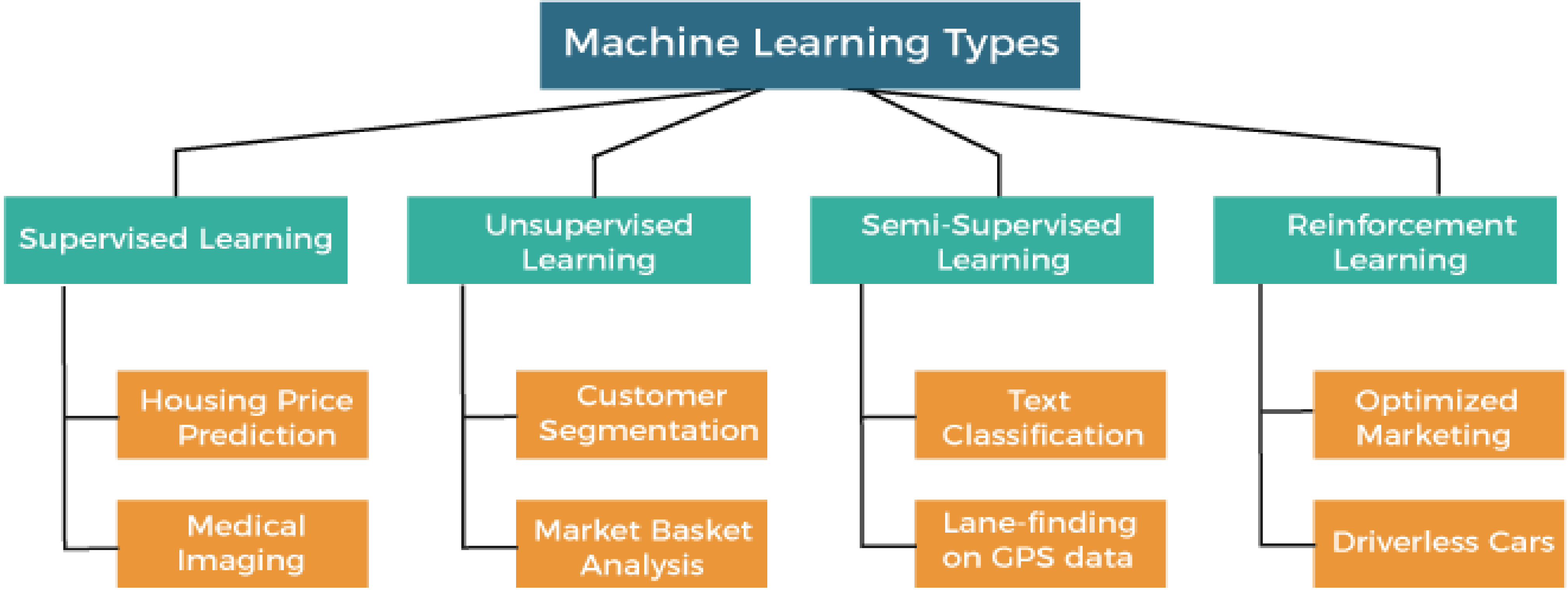
1. Supervised Machine Learning
2. Unsupervised Machine Learning
3. Semi-Supervised Machine Learning
4. Reinforcement Learning

.

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Learning

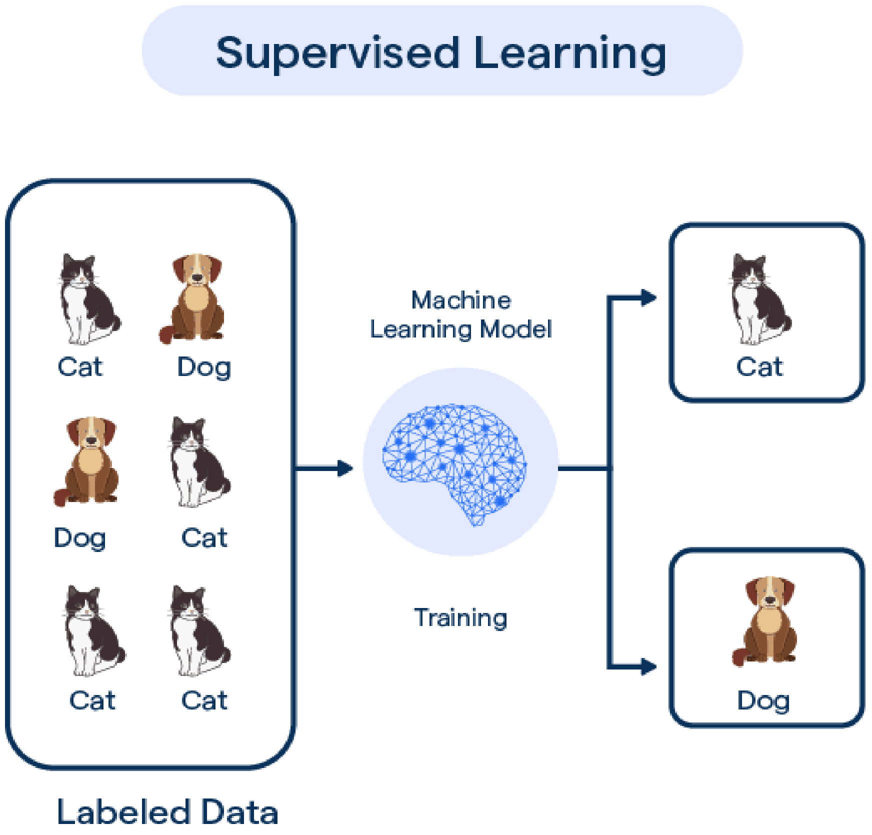


MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Learning

Supervised Machine Learning - Supervised machine learning is based on supervision. It means in the supervised learning technique, we train the machines using the "labelled" dataset, and based on the training, the machine predicts the output. Here, the labelled data specifies that some of the inputs are already mapped to the output.



The main goal of the supervised learning technique is to map the input variable(x) with the output variable(y). Some real-world applications of supervised learning are **Risk Assessment, Fraud Detection, Spam filtering**

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Learning

Supervised Machine Learning

- **Classification**
- **Regression**

a) Classification

Classification algorithms are used to solve the classification problems in which the output variable is categorical, such as "**Yes" or No, Male or Female, Red or Blue, etc**. The classification algorithms predict the categories present in the dataset. Some real-world examples of classification algorithms are **Spam Detection, Email filtering, etc.**

Some popular classification algorithms are given below:

- **Random Forest Algorithm**
- **Decision Tree Algorithm**
- **Logistic Regression Algorithm**
- **Support Vector Machine Algorithm**

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Learning

Supervised Machine Learning

- **Regression**

Regression algorithms are used to solve regression problems in which there is a linear relationship between input and output variables. These are used to predict continuous output variables, such as market trends, weather prediction, etc.

Some popular Regression algorithms are given below:

- **Simple Linear Regression Algorithm**
- **Multivariate Regression Algorithm**
- **Decision Tree Algorithm**
- **Lasso Regression**

Types of Machine Learning

Advantages and Disadvantages of Supervised Learning

Advantages:

- Since supervised learning work with the labelled dataset so we can have an exact idea about the classes of objects.
- These algorithms are helpful in predicting the output on the basis of prior experience.

Disadvantages:

- These algorithms are not able to solve complex tasks.
- It may predict the wrong output if the test data is different from the training data.
- It requires lots of computational time to train the algorithm.

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

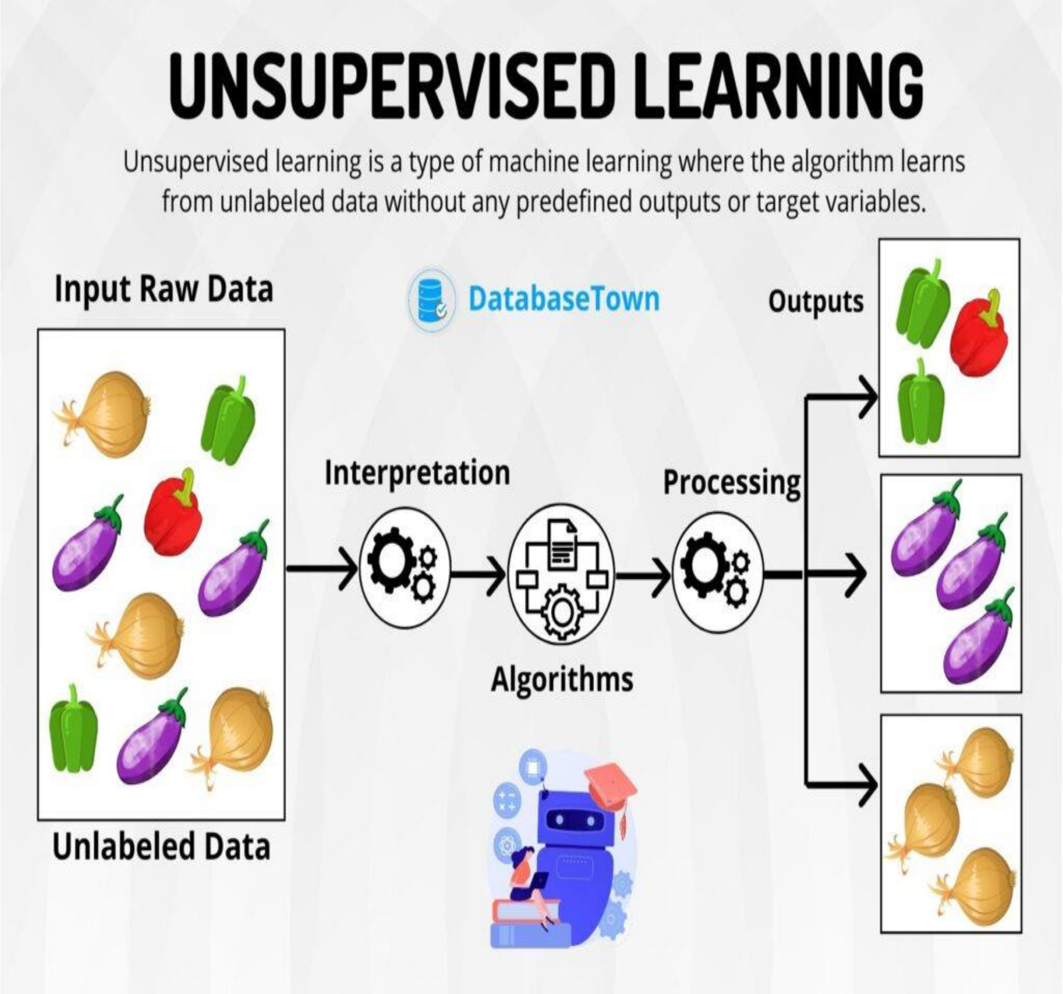
Types of Machine Learning

Unsupervised Machine Learning

Unsupervised learning is different from the Supervised learning technique; as its name suggests, there is no need for supervision. It means, **in unsupervised machine learning**, the machine is trained using the unlabeled dataset, and the machine predicts the output without any supervision.

In unsupervised learning, the models are trained with the data that is neither classified nor labelled, and the model acts on that data without any supervision.

The main aim of the unsupervised learning algorithm is to group or categories the unsorted dataset according to the similarities, patterns, and differences. Machines are instructed to find the hidden patterns from the input dataset.



MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Learning

Unsupervised Machine Learning

Categories of Unsupervised Machine Learning

Unsupervised Learning can be further classified into two types, which are given below:

- **Clustering**
- **Association**

1) Clustering

The clustering technique is used when we want to find the inherent groups from the data. It is a way to group the objects into a cluster such that the objects with the most similarities remain in one group and have fewer or no similarities with the objects of other groups. An example of the clustering algorithm is grouping the customers by their purchasing behaviour.

Some of the popular clustering algorithms are given below:

- **K-Means Clustering algorithm,   Mean-shift algorithm**
- **DBSCAN Algorithm**
- **Principal Component Analysis**
- **Independent Component Analysis**

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Learning

Unsupervised Machine Learning

Categories of Unsupervised Machine Learning

- **Association**

Association rule learning is an unsupervised learning technique, which finds interesting relations among variables within a large dataset. The main aim of this learning algorithm is to find the dependency of one data item on another data item and map those variables accordingly so that it can generate maximum profit. This algorithm is mainly applied in **Market Basket analysis, Web usage mining, continuous production**, etc.

Some popular algorithms of Association rule learning are **Apriori Algorithm, Eclat, FP-growth algorithm.**

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Learning- Unsupervised Machine Learning

Advantages and Disadvantages of Unsupervised Learning Algorithm

Advantages:

- These algorithms can be used for complicated tasks compared to the supervised ones because these algorithms work on the unlabeled dataset.
- Unsupervised algorithms are preferable for various tasks as getting the unlabeled dataset is easier as compared to the labelled dataset.

Disadvantages:

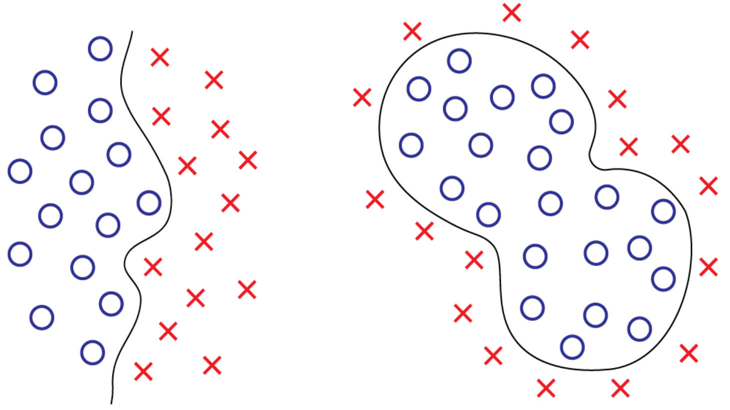
- The output of an unsupervised algorithm can be **less accurate as the dataset is not labelled**, and algorithms are not trained with the exact output in prior.
- Working with Unsupervised learning is more difficult as it works with the unlabelled dataset that does not map with the output.

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Learning- Unsupervised Machine Learning

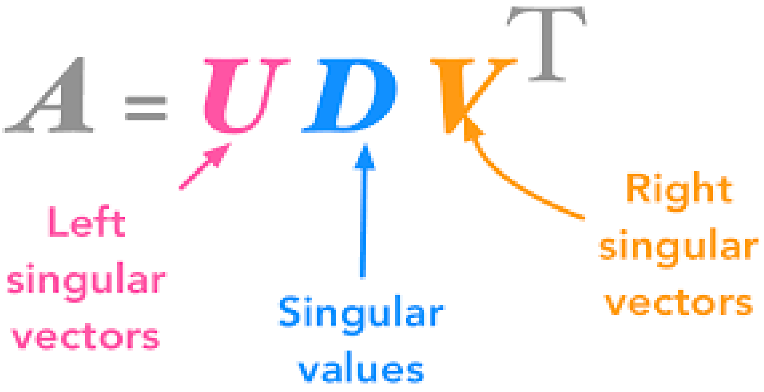
Applications of Unsupervised Learning



Classification

Anomaly Detection

- **Network Analysis:** Unsupervised learning is used for ident plagiarism and copyright in document network analysis of text da scholarly articles.
- **Recommendation Systems:** Recommendation systems widely unsupervised learning techniques for building recommendation applications for different web applications and e-commerce websites.
- **Anomaly Detection:** Anomaly detection is a popular application (unsupervised learning, which can identify unusual data points within th dataset. It is used to discover fraudulent transactions.
- **Singular Value Decomposition:** Singular Value Decomposition or SV(is used to extract particular information from the database. For exampl extracting information of each user located at a particular location.



MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Semi-Supervised Learning

Semi-Supervised learning is a type of Machine Learning algorithm that lies between Supervised and Unsupervised machine learning. It represents the intermediate ground between Supervised (With Labelled training data) and Unsupervised learning (with no labelled training data) algorithms and uses the combination of labelled and unlabeled datasets during the training period.

To overcome the drawbacks of supervised learning and unsupervised learning algorithms, the concept of Semi-supervised learning is introduced. The main aim of semi-supervised learning is to effectively use all the available data, rather than only labelled data like in supervised learning

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Semi-Supervised Learning

Advantages and disadvantages of Semi-supervised Learning

Advantages:

- It is simple and easy to understand the algorithm.
- It is highly efficient.
- It is used to solve drawbacks of Supervised and Unsupervised Learning algorithms.

Disadvantages:

- Iterations results may not be stable.
- We cannot apply these algorithms to network-level data.
- Accuracy is low.

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Reinforcement Learning

Reinforcement learning works on a feedback-based process, in which an AI agent (A software component) automatically explore its surrounding by hitting & trail, taking action, learning from experiences, and improving its performance

The reinforcement learning process is similar to a human being;

for example, a child learns various things by experiences in his day-to-day life. An example of reinforcement learning is to play a game, where the Game is the environment, moves of an agent at each step define states, and the goal of the agent is to get a high score. Agent receives feedback in terms of punishment and rewards.

Due to its way of working, reinforcement learning is employed in different fields such as **Game theory, Operation Research, Information theory, multi-agent systems.**

A reinforcement learning problem can be formalized using **Markov Decision Process(MDP).** In MDP, the agent constantly interacts with the environment and performs actions; at each action, the environment responds and generates a new state.

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

Types of Machine Reinforcement Learning

A reinforcement learning problem can be formalized using **Markov Decision Process(MDP).** In MDP, the agent constantly interacts with the environment and performs actions; at each action, the environment responds and generates a new state.

Categories of Reinforcement Learning

Reinforcement learning is categorized mainly into two types of methods/algorithms:

- **Positive Reinforcement Learning:** Positive reinforcement learning specifies increasing the tendency that the required behaviour would occur again by adding something. It enhances the strength of the behaviour of the agent and positively impacts it.
- **Negative Reinforcement Learning:** Negative reinforcement learning works exactly opposite to the positive RL. It increases the tendency that the specific behaviour would occur again by avoiding the negative condition.

Types of Machine Reinforcement Learning

Real-world Use cases of Reinforcement Learning

- **Video Games:**
  RL algorithms are much popular in gaming applications. It is used to gain super-human performance. Some popular games that use RL algorithms are **AlphaGO** and **AlphaGO Zero**.

- **Resource Management:**
  The "Resource Management with Deep Reinforcement Learning" paper showed that how to use RL in computer to automatically learn and schedule resources to wait for different jobs in order to minimize average job slowdown.

- **Robotics:**
  RL is widely being used in Robotics applications. Robots are used in the industrial and manufacturing area, and these robots are made more powerful with reinforcement learning. There are different industries that have their vision of building intelligent robots using AI and Machine learning technology.

- **Text Mining**
  Text-mining, one of the great applications of NLP, is now being implemented with the help of Reinforcement Learning by Salesforce company.

Types of Machine Reinforcement Learning

Advantages and Disadvantages of Reinforcement Learning

Advantages

- It helps in solving complex real-world problems which are difficult to be solved by general techniques.
- The learning model of RL is similar to the learning of human beings; hence most accurate results can be found.
- Helps in achieving long term results.

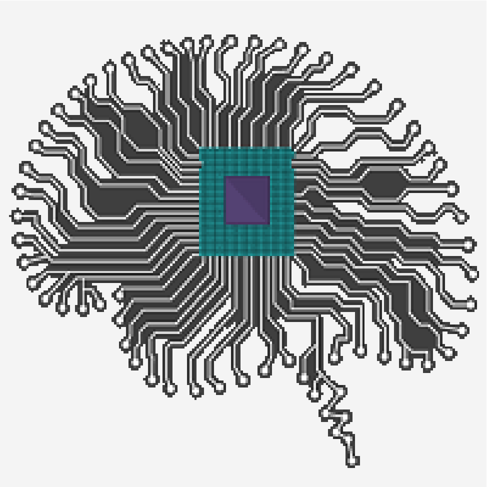
Disadvantage

- RL algorithms are not preferred for simple problems.
- RL algorithms require huge data and computations.
- Too much reinforcement learning can lead to an overload of states which can weaken the results.

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

The Brain and the Neuron



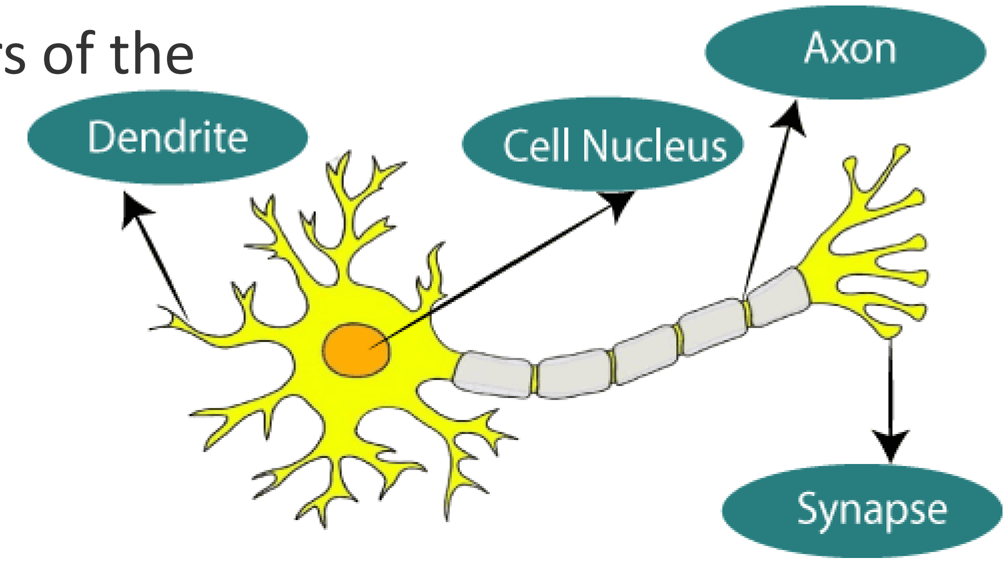
The term "Artificial neural network" refers to a biologically inspired sub-field of artificial intelligence modeled after the brain. An Artificial neural network is usually a computational network based on biological neural networks that construct the structure of the human brain. Similar to a human brain has neurons interconnected to each other, artificial neural networks also have neurons that are linked to each other in various layers of the networks. These neurons are known as nodes.

MALLA REDDY COLLEGE OF ENGINEERING

UNIT - I

The Brain and the Neuron

Artificial Neural Network" is derived from Biological neural networks that develop the structure of a human brain. Similar to the human brain that has neurons interconnected to one another, artificial neural networks also have neurons that are interconnected to one another in various layers of the networks. These neurons are known as nodes.

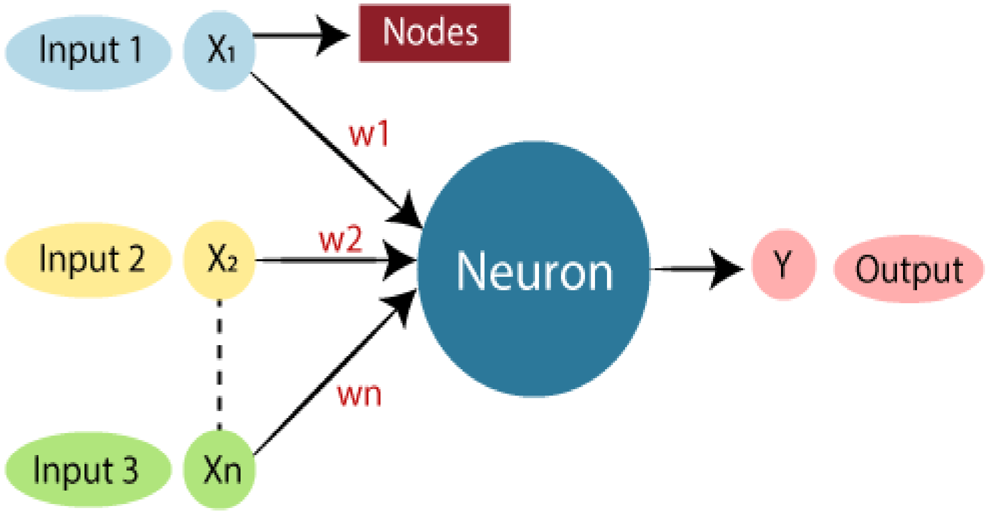


MALLA REDDY COLLEGE OF ENGINEERING

The Brain and the Neuron

Dendrites from Biological Neural Network represent inputs in Artificial Neural Networks, cell nucleus represents Nodes, synapse represents Weights, and Axon represents Output.

Biological Neural Network	Artificial Neural Network
Dendrites	Inputs
Cell nucleus	Nodes
Synapse	Weights
Axon	Output



MALLA REDDY COLLEGE OF ENGINEERING

The Brain and the Neuron

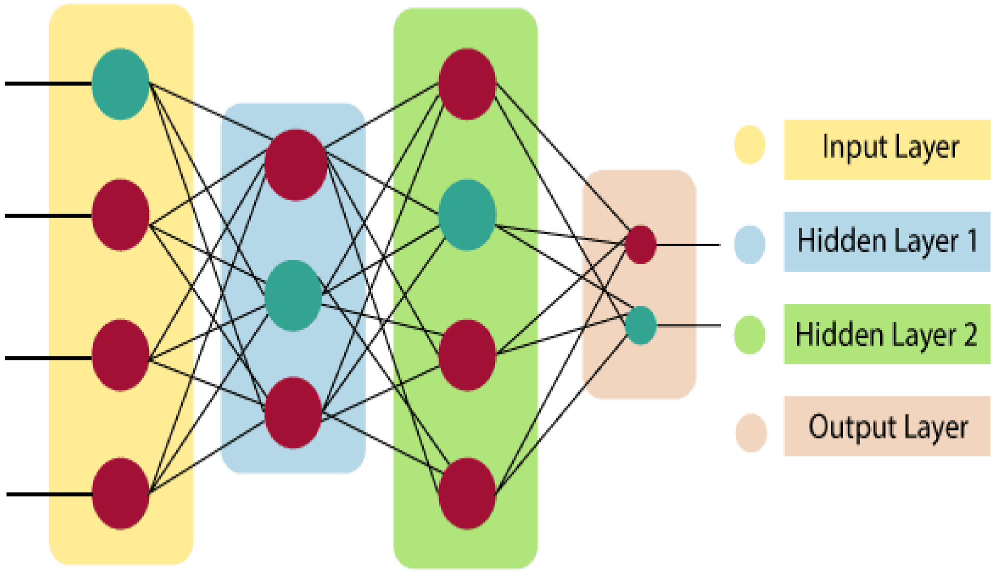
We can understand the artificial neural network with an example, consider an example of a digital logic gate that takes an input and gives an output. "OR" gate, which takes two inputs. If one or both the inputs are "On," then we get "On" in output. If both the inputs are "Off," then we get "Off" in output. Here the output depends upon input. Our brain does not perform the same task. The outputs to inputs relationship keep changing because of the neurons in our brain, which are "learning."

MALLA REDDY COLLEGE OF ENGINEERING

The Brain and the Neuron

The architecture of an artificial neural network:

To understand the concept of the architecture of an artificial neural network, we have to understand what a neural network consists of. In order to define a neural network that consists of a large number of artificial neurons, which are termed units arranged in a sequence of layers. Lets us look at various types of layers available in an artificial neural network.



MALLA REDDY COLLEGE OF ENGINEERING

The Brain and the Neuron

Artificial Neural Network primarily consists of three layers:

Input Layer:

As the name suggests, it accepts inputs in several different formats provided by the programmer.

Hidden Layer:

The hidden layer presents in-between input and output layers. It performs all the calculations to find hidden features and patterns.

Output Layer:

The input goes through a series of transformations using the hidden layer, which finally results in output that is conveyed using this layer.

The artificial neural network takes input and computes the weighted sum of the inputs and includes a bias. This computation is represented in the form of a transfer function.

MALLA REDDY COLLEGE OF ENGINEERING

The Brain and the Neuron

It determines weighted total is passed as an input to an activation function to produce the output. Activation functions choose whether a node should fire or not. Only those who are fired make it to the output layer. There are distinctive activation functions available that can be applied upon the sort of task we are performing.

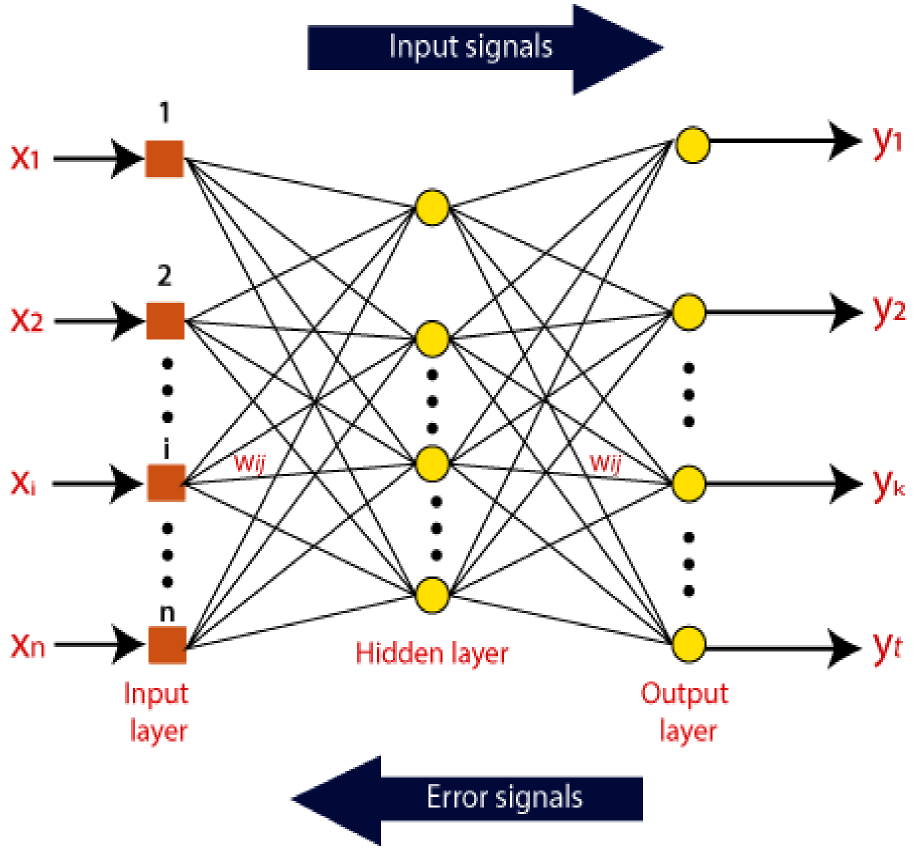
$$\sum_{i=1}^{n} Wi * Xi + b$$

MALLA REDDY COLLEGE OF ENGINEERING

The Brain and the Neuron

How do artificial neural networks work?

Artificial Neural Network can be best represented as a weighted directed graph, where the artificial neurons form the nodes. The association between the neurons outputs and neuron inputs can be viewed as the directed edges with weights. The Artificial Neural Network receives the input signal from the external source in the form of a pattern and image in the form of a vector. These inputs are then mathematically assigned by the notations x(n) for every n number of inputs.



The Brain and the Neuron

How do artificial neural networks work?

Afterward, each of the input is multiplied by its corresponding weights (these weights are the details utilized by the artificial neural networks to solve a specific problem). In general terms, these weights normally represent the strength of the interconnection between neurons inside the artificial neural network. All the weighted inputs are summarized inside the computing unit.

If the weighted sum is equal to zero, then bias is added to make the output non-zero or something else to scale up to the system's response. Bias has the same input, and weight equals to 1. Here the total of weighted inputs can be in the range of 0 to positive infinity. Here, to keep the response in the limits of the desired value, a certain maximum value is benchmarked, and the total of weighted inputs is passed through the activation function.

The activation function refers to the set of transfer functions used to achieve the desired output. There is a different kind of the activation function, but primarily either linear or non-linear sets of functions.

MALLA REDDY COLLEGE OF ENGINEERING

The Brain and the Neuron

Advantages of Artificial Neural Network (ANN)

Parallel processing capability:

Artificial neural networks have a numerical value that can perform more than one task simultaneously.

Storing data on the entire network:

Data that is used in traditional programming is stored on the whole network, not on a database. The disappearance of a couple of pieces of data in one place doesn't prevent the network from working.

Capability to work with incomplete knowledge:

After ANN training, the information may produce output even with inadequate data. The loss of performance here relies upon the significance of missing data.

Having a memory distribution: For ANN is to be able to adapt, it is important to determine the examples and to encourage the network according to the desired output by demonstrating these examples to the network. The succession of the network is directly proportional to the chosen instances, and if the event can't appear to the network in all its aspects, it can produce false output.

Having fault tolerance:

Extortion of one or more cells of ANN does not prohibit it from generating output, and this feature makes the network fault-tolerance.

MALLA REDDY COLLEGE OF ENGINEERING

The Brain and the Neuron

Disadvantages of Artificial Neural Network (ANN)

Assurance of proper network structure:
There is no particular guideline for determining the structure of artificial neural networks. The appropriate network structure is accomplished through experience, trial, and error.

Unrecognized behavior of the network:
It is the most significant issue of ANN. When ANN produces a testing solution, it does not provide insight concerning why and how. It decreases trust in the network.

Hardware dependence:
Artificial neural networks need processors with parallel processing power, as per their structure. Therefore, the realization of the equipment is dependent.

Difficulty of showing the issue to the network:
ANNs can work with numerical data. Problems must be converted into numerical values before being introduced to ANN. The presentation mechanism to be resolved here will directly impact the performance of the network. It relies on the user's abilities.

The duration of the network is unknown:
The network is reduced to a specific value of the error, and this value does not give us optimum results.

MALLA REDDY COLLEGE OF ENGINEERING

Design a Learning System

In order to illustrate some of the basic design issues and approaches to machine learning, let us consider designing a program to learn to play checkers, with the goal of entering it in the world checkers tournament.

Choosing the Training Experience:

The first design choice we face is to choose the type of training experience from which our system will learn. The type of training experience available can have a significant impact on success or failure of the learner. One key attribute is whether the training experience provides direct or indirect feedback regarding the choices made by the performance system.

For example,

in learning to play checkers, the system might learn from direct training examples consisting of individual checkers board states and the correct move for each.

In order to complete the design of the learning system, we must now choose

1. the exact type of knowledge to be learned
2. a representation for this target knowledge
3. a learning mechanism

MALLA REDDY COLLEGE OF ENGINEERING

Design a Learning System

Choosing the Target Function:
The next design choice is to determine exactly what type of knowledge will be learned and how this will be used by the performance program. Let us begin with a checkers-playing program that can generate the legal moves from any board state. The program needs only to learn how to choose the best move from among these legal moves.

Choosing a Representation for the Target Function:
X1: the number of black pieces on the board
x2: the number of red pieces on the board
x3: the number of black kings on the board
x4: the number of red kings on the board
x5: the number of black pieces threatened by red (i.e., which can be captured on red's next turn)
X6: the number of red pieces threatened by black
Thus, our learning program will represent V(b) as a linear function of the form
V(b)=w0+w1x1+w2x2+w3x3+w4x4+w5x5+w6x6

MALLA REDDY COLLEGE OF ENGINEERING

Design a Learning System

Partial design of a checkers learning program:

Task T: playing checkers
Performance measure P: percent of games won in the world tournament
Training experience E: games played against itself
Target function: V:Board → R
Target function representation
V(b)=w0+w1x1+w2x2+w3x3+w4x4+w5x5+w6x6

Choosing a Function Approximation Algorithm
In order to learn the target function f we require a set of training examples, each describing a specific board state b and the training value Vtrain(b) for b

MALLA REDDY COLLEGE OF ENGINEERING

Design a Learning System

ESTIMATING TRAINING VALUES

Rule for estimating training values.

Vtrain (b) ← V(Successor(b))
Adjusting the weights

$$E \equiv \sum_{\langle b, V_{train}(b)\rangle \in \ training\ examples} (V_{train}(b) - \hat{V}(b))^2$$

MALLA REDDY COLLEGE OF ENGINEERING

Design a Learning System

The Performance System
is the module that must solve the given per case playing checkers, by using the learned target function(s). It takes an instance of a new problem (new game) as input and produces a trace of its solution (game history) as output.
The **Critic** takes as input the history or trace of the game and produces as output a set of training examples of the target function**. As shown in the diagram**, each training example in this case corresponds to some game state in the trace, along with an estimate Vtrain of the target function value for this example.

The **Generalizer** takes as input the training examples and produces an output hypothesis that is its estimate of the target function. It generalizes from the specific training examples, hypothesizing a general function that covers these examples and other cases beyond the training examples.

The **Experiment Generator** takes as input the current hypothesis (currently learned function) and outputs a new problem (i.e., initial board state) for the Performance System to explore. Its role is to pick new practice overall system.

MALLA REDDY COLLEGE OF ENGINEERING

Design a Learning System

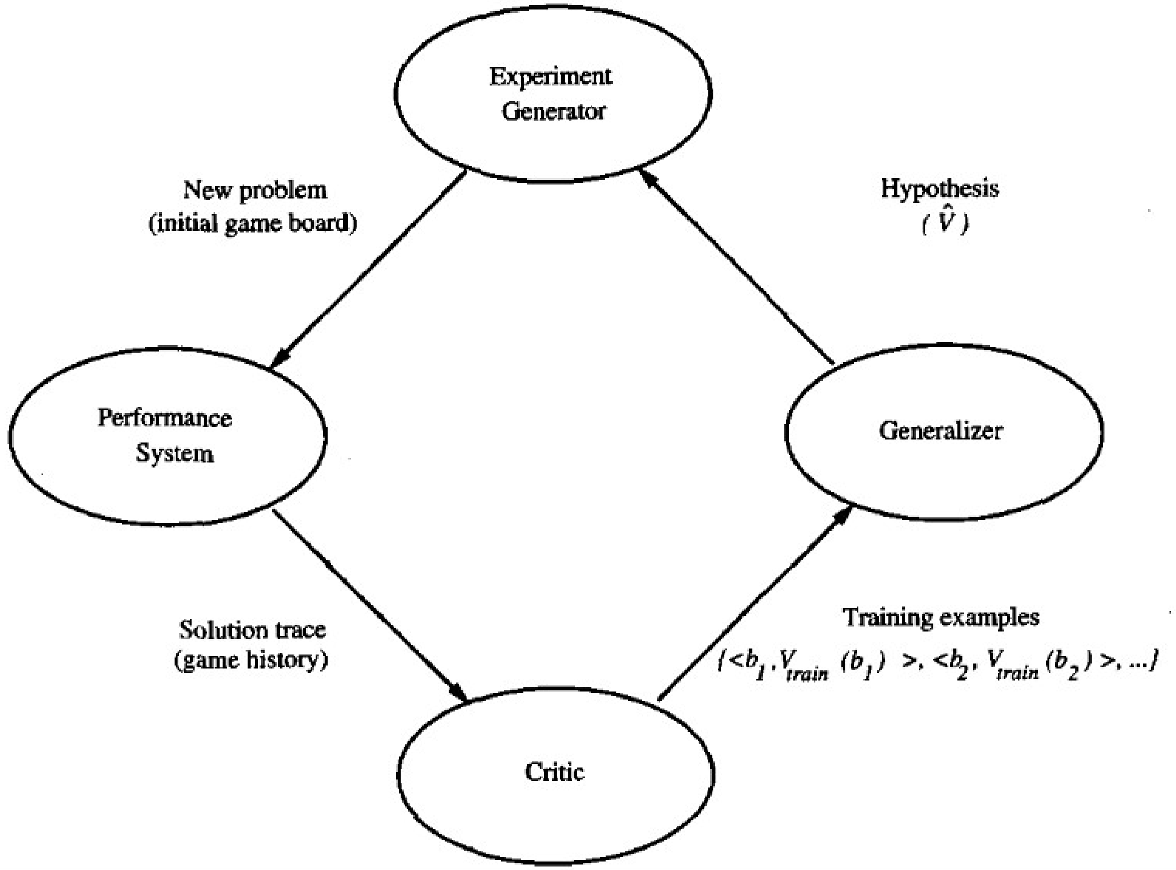


Fig: Final Design of Checkers Learning Problem

MALLA REDDY COLLEGE OF ENGINEERING

Design a Learning System

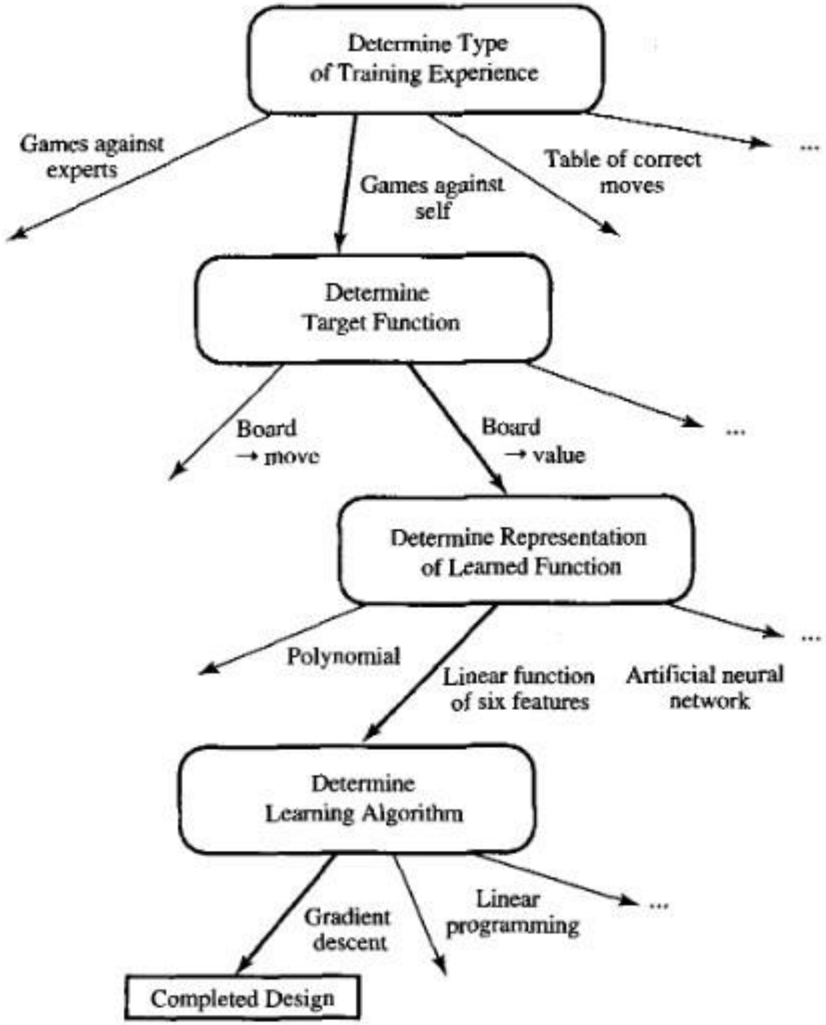


Fig: Final Design of Checkers Learning Problem

MALLA REDDY COLLEGE OF ENGINEERING

Perspectives and Issues in Machine Learning

One useful perspective on machine learning is that it involves searching a very large space of possible hypotheses to determine one that best fits the observed data and any prior knowledge held by the learner

For example,
consider the space of hypotheses that could in principle be output by the above checkers learner. This hypothesis space consists of all evaluation functions that can be represented by some choice of values for the **weights W0 through w6.**

The learner's task is thus to search through this vast space to locate the hypothesis that is most consistent with the available training examples

MALLA REDDY COLLEGE OF ENGINEERING

Perspectives and Issues in Machine Learning

What algorithms exist for learning general target functions from specific training examples? In what settings will particular algorithms converge to the desired function, given sufficient training data? Which algorithms perform best for which types of problems and representations?

- How much training data is sufficient? What general bounds can be found to relate the confidence in learned hypotheses to the amount of training experience and the character of the learner's hypothesis space?

- When and how can prior knowledge held by the learner guide the process of generalizing from examples?
- Can prior knowledge be helpful even when it is only approximately correct?

- What is the best strategy for choosing a useful next training experience, and how does the choice of this strategy alter the complexity of the learning problem?

- What is the best way to reduce the learning task to one or more function approximation problems? Put another way, what specific functions should the system attempt to learn? Can this process itself be automated?

- How can the learner automatically alter its representation to improve its ability to represent and learn the target function?

MALLA REDDY COLLEGE OF ENGINEERING

Concept Learning as Search –

Concept learning:

Inferring a boolean-valued function from training examples of its input and output.

A CONCEPT LEARNING TASK:

What hypothesis representation shall we provide to the learner in this case? Let us begin by considering a simple representation in which each hypothesis consists of a conjunction of constraints on the instance attributes.

Gadgets -- Tablet, Smart phone

Features of Both(Binary Value Attributes) – Size, color, Screentype, Shape (X1,X2,X3,X4)

No of possible instances – 2^d

here d = 4 (attributes) therefore 2 power4 = 16

Possible count = 2 (2^{d}) 2power 16 = 65536 (approximately)

MALLA REDDY COLLEGE OF ENGINEERING

Concept Learning as Search –

In particular, let each hypothesis be a vector of six constraints, specifying -**sports learning task**, enjoy the values of the six attributes

Sky, AirTemp, Humidity, Wind, Water, and Forecast.

3 values (rainy, cloudy, sunny)

Different instances possible – 3*2*2*2*2*2 = 96 (here 2 is values of each attribute)

Now, Syntactically Distinct Hypothesis :-

(Additionally 2 more values) i.e (Ø,?)

(Ø,rainy,sunny,cloudy,?) = 5*4*4*4*4*4 = 5120

Semantically Distinct Hypothesis = (?,rainy,cloudy,sunny) (Ø – Null Value)

=1+(4*3*3*3*3*3) = 973

For each attribute, the hypothesis will either indicate by a "?’

that any value is acceptable for this attribute,

specify a single required value (e.g., Warm) for the attribute, or indicate by a "θ" that no value is acceptable.

MALLA REDDY COLLEGE OF ENGINEERING

CONCEPT LEARNING AS SEARCH Concept learning can be viewed as the task of searching through a large space of hypotheses implicitly defined by the hypothesis representation. The goal of this search is to find the hypothesis that best fits the training examples.

It is important to note that by selecting a hypothesis representation, the designer of the learning algorithm implicitly defines the space of all hypotheses that the program can ever represent and therefore can ever learn.

General-to-Specific Ordering of Hypotheses Many algorithms for concept learning organize the search through the hypothesis space by relying on a very useful structure that exists for any concept learning problem :

a general-to-specific ordering of hypotheses. By taking advantage of this naturally occurring structure over the hypothesis space, we can design learning algorithms that exhaustively search even infinite hypothesis spaces without
explicitly enumerating every hypothesis.

MALLA REDDY COLLEGE OF ENGINEERING

Concept Learning as Search –

Concept Learning as Search –

To illustrate the general-to-specific ordering, consider the two hypotheses

h1 = (Sunny, ?, ?, Strong, ?, ?)

h2 = (Sunny, ?, ?, ?, ?, ?)

Now consider the sets of instances that are classified positive by hl and by h2. Because h2 imposes fewer constraints on the instance, it classifies more instances as positive. In fact, any instance classified positive by h1 will also be classified positive by h2. Therefore, we say that h2 is more general than h1.

MALLA REDDY COLLEGE OF ENGINEERING

Finding a Maximally Specific Hypothesis –

Find S Algorithm

Finding a Maximally Specific Hypothesis:

This Algorithm considers only Positive examples most specific hypothesis

Representation :

-> Most Specific Hypothesis => Ø

-> Most General Hypothesis => ?

Finding a Maximally Specific Hypothesis – Version Spaces and the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Find S Algorithm

Step 1: Initialise with most specific hypothesis (Ø)

h_0 = <Ø, Ø , Ø , Ø , Ø > = 5 attributes

step 2 : for each +ve sample

for each attribute,

if(value = hypothesis value) => ignore

else

replace with the most general hypothesis (?)

MALLA REDDY COLLEGE OF ENGINEERING

Finding a Maximally Specific Hypothesis – Version Spaces and the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Find S Algorithm

S no	Origin	Manufacturer	Color	Year	Type	class
1	Japan	Honda	Blue	1980	Eco	+
2	Japan	Toyota	Green	1970	Sport	-
3	Japan	Toyota	Blue	1990	Eco	+
4	USA	Audi	Red	1980	Eco	-
5	Japan	Honda	White	1980	Eco	+
6	Japan	Toyota	Green	1980	Eco	+
7	Japan	Honda	Red	1980	Eco	-

MALLA REDDY COLLEGE OF ENGINEERING

Finding a Maximally Specific Hypothesis – Version Spaces and the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Find S Algorithm

ho = <Ø,Ø,Ø,Ø,Ø>

h1 = <Japan, Honda, Blue,1980,Eco>

h2=h1

h3 = < Japan, ?, Blue,?,Eco>

h4 = h3

h5 = < Japan, ?, ?,?, Eco >

h6 = h7

Disadvantages:

- Consider only +ve Values
- h6 may not be sole hypothesis that fits the complete data

MALLA REDDY COLLEGE OF ENGINEERING

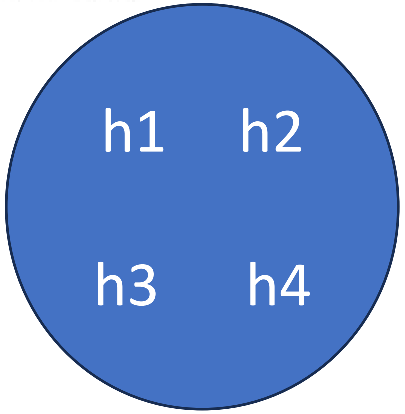
Version Spaces and the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Version Spaces and the Candidate Elimination Algorithm:

subset of hypothesis (H) consistent with the training examples

Definition: The **version space**, denoted $VS_{H,D}$, with respect to hypothesis space H and training examples D, is the subset of hypotheses from H consistent with the training examples in D.

$$VS_{H,D} \equiv \{h \in H \mid Consistent(h, D)\}$$



consistent h(x) = c(x)

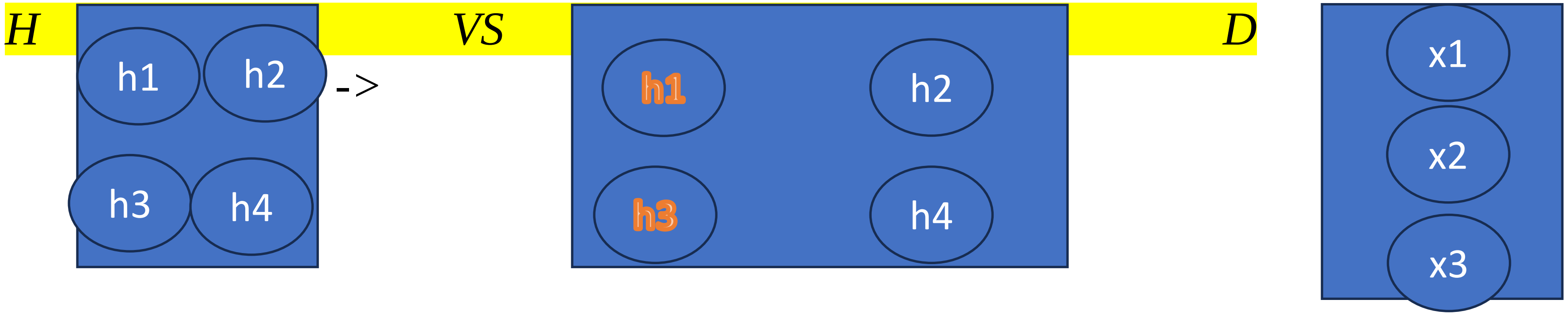
MALLA REDDY COLLEGE OF ENGINEERING

Version Spaces and the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

the list then eliminate algorithm:

- **Version Space ← list containing every hypothesis in H**
- **from this step we keep on removing inconsistent hypothesis from version space**
- **for each training example**
- **<x,c(x)> remove any hypothesis i.e $h(x) \neq c(x)$**
- **output the list of hypothesis into version space after checking for all training examples**

subset of hypothesis (H) consistent with the training examples **consistent** $\mathbf{h(x) = c(x)}$



MALLA REDDY COLLEGE OF ENGINEERING

Version Spaces and the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Initialize G to the set of maximally general hypotheses in H
Initialize S to the set of maximally specific hypotheses in H
For each training example d, do

- If d is a positive example
  - Remove from G any hypothesis inconsistent with d
  - For each hypothesis s in S that is not consistent with d
    - Remove s from S
    - Add to S all minimal generalizations h of s such that
      - h is consistent with d, and some member of G is more general than h
    - Remove from S any hypothesis that is more general than another hypothesis in S
- If d is a negative example
  - Remove from S any hypothesis inconsistent with d
  - For each hypothesis g in G that is not consistent with d
    - Remove g from G
    - Add to G all minimal specializations h of g such that
      - h is consistent with d, and some member of S is more specific than h
    - Remove from G any hypothesis that is less general than another hypothesis in G

the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

The Candidate Elimination Algorithm:

- ✓ use the concept of version space
- ✓ It considers both +ve and –ve values (Samples) – Yes or No
- ✓ Both specific and General Hypothesis

for Positive Samples , move from specific to General

For Negative Samples , move from general to Specific

S = {Ø, Ø , Ø , Ø , Ø } +

G = {? ? ? ? ? } -

MALLA REDDY COLLEGE OF ENGINEERING

the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

The Candidate Elimination Algorithm:

✓ Algorithm

Step 1: initialise General to Specific

S = {Ø, Ø , Ø , Ø , ………Ø }
G = {? ? ? ? ………..? }

Depends on no of attributes

Step 2: for each example ,
if example is positive
make specific to general
else
example is negative
make general to specific

MALLA REDDY COLLEGE OF ENGINEERING

the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

The Candidate Elimination Algorithm:

Example:

Enjoysport

S = {Ø, Ø , Ø , Ø , Ø , Ø} G = {? ? ? ? ? ? }

Dataset:

S No	Sky	Temperature	Humidity	Wind	Water	Forecast	Enjoysport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes (+)
2	Sunny	Warm	High	Strong	Warm	Same	Yes (+)
3	Rainy	Cold	High	Strong	Warm	Change	No (-)
4	Sunny	Warm	High	Strong	Cool	Change	Yes (+)

the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

The Candidate Elimination Algorithm:

Example:

Enjoysport

S = {Ø, Ø , Ø , Ø , Ø } G = {? ? ? ? ? }

Dataset:

S No	Sky	Temperature	Humidity	Wind	Water	Forecast	Enjoysport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes (+)
2	Sunny	Warm	High	Strong	Warm	Same	Yes (+)

S1 = {Sunny,Warm,Normal,Strong,Warm,Same}

G1 = {?,?,?,?,?,?}

S2 = {Sunny,Warm,?,Strong,Warm,Same}

G2 = {?,?,?,?,?,?}

the Candidate Elimination Algorithm – Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

The Candidate Elimination Algorithm:

Example:

Enjoysport

S = {Ø, Ø , Ø , Ø , Ø }   G = {? ? ? ? ? }

Dataset:

S No	Sky	Temperature	Humidity	Wind	Water	Forecast	Enjoysport
3	Rainy	Cold	High	Strong	Warm	Change	No (-)
4	Sunny	Warm	High	Strong	Cool	Change	Yes (+)

S3 = {Sunny,Warm,?,Strong,Warm,Same}

G3 = {<Sunny,?,?,?,?,?>, {<?,Warm,?,?,?,?>, <?,?,?,?,?same, >}

S4 = {Sunny,Warm,?,Strong,?,?}

G4 = {<Sunny,?,?,?,?,?>, {<?,Warm,?,?,?,?>,}

there for S4 and G4 => final Hypothesis

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Perceptron

What is the Perceptron model in Machine Learning?

Perceptron is Machine Learning algorithm for supervised learning of various binary classification tasks. Further, ***Perceptron is also understood as an Artificial Neuron or neural network unit that helps to detect certain input data computations in business intelligence***.

Perceptron model is also treated as one of the best and simplest types of Artificial Neural networks. However, it is a supervised learning algorithm of binary classifiers.

Hence, we can consider it as a single-layer neural network with four main parameters, i.e., **input values, weights and Bias, net sum, and an activation function.**

MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Perceptron

Takes real valued input,

calculate linear combination

of these inputs and generates output

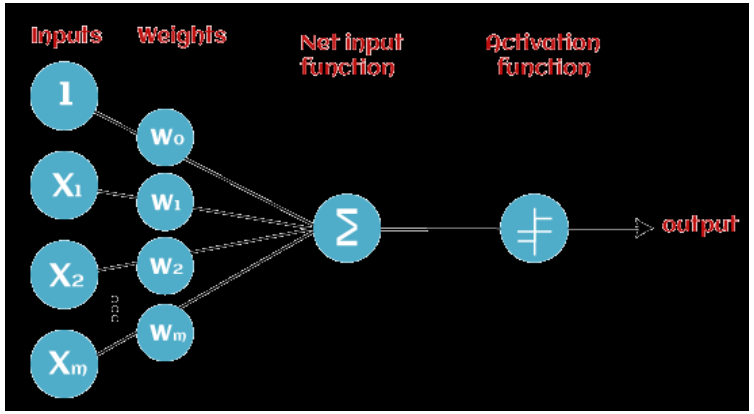
Output = 1 if result > threshold

= -1 otherwise (i.e not greater than threshold)

O(x1,x2,………xn) = { 1 if w0+w1x1+w2x2+……………………wnxn > O

-1 otherwise

(Here w0,w1,w2……………………wn weights of inputs)



MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Perceptron

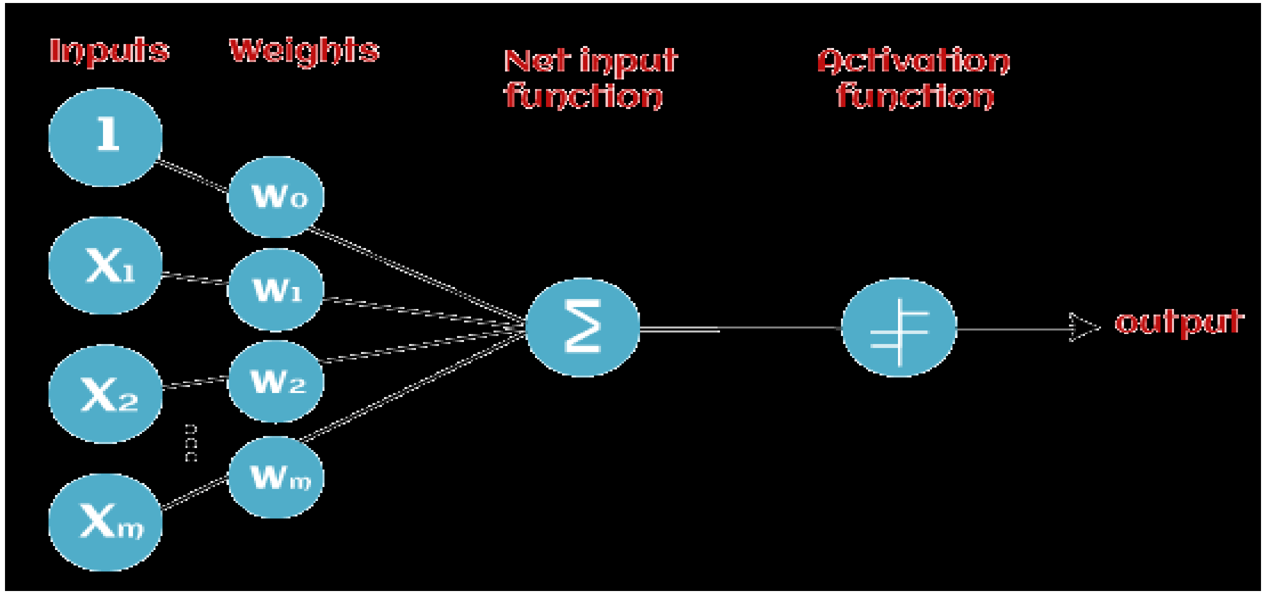
Output = 1 if result > threshold

= -1 otherwise (i.e not greater than threshold)

O(x1,x2,………xn) = { 1 if w0+w1x1+w2x2+……………………wnxn > O

-1 otherwise

(Here w0,w1,w2……………………wn weights of inputs)



MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

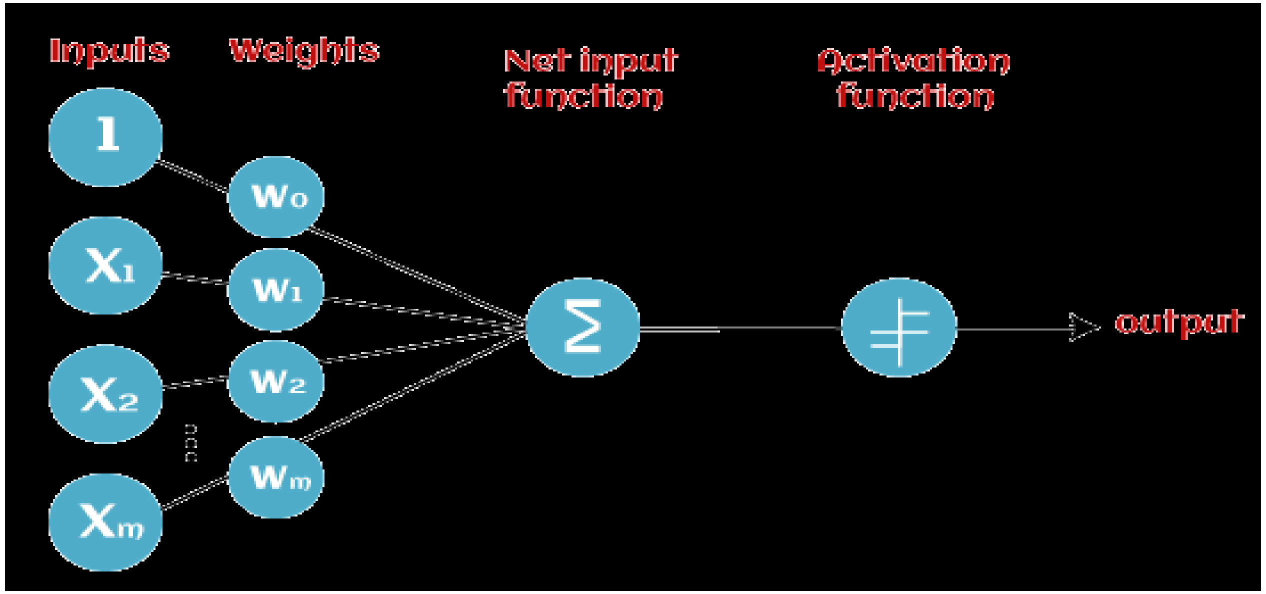
Perceptron

Calculating weights

Formula to change weights

$W_i \leftarrow w_i + \Delta w_i$

$\Delta w_i = n(t-o)x_i$



MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Linear Separability :

Linear separability is an important concept in **machine learning**, particularly in the field of **supervised learning**. It refers to the ability of a set of data points to be separated into distinct categories using a linear decision boundary. In other words, if there exists a straight line that can cleanly divide the data into two classes, then the data is said to be linearly separable.

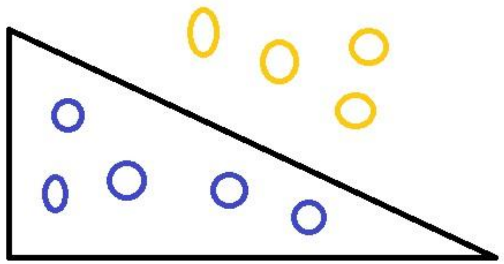
Linear separability is a concept in machine learning that refers to the ability to separate data points in binary classification problems using a linear decision boundary. If the data points can be separated using a line, linear function, or flat hyperplane, they are considered linearly separable. Linear separability is an important concept in neural networks, and it is introduced in the context of linear algebra and optimization theory.

In the context of machine learning, linear separability is an important property because it makes classification problems easier to solve. If the data is linearly separable, we can use a linear classifier, such as logistic regression or support vector machines (SVMs), to accurately classify new instances of data.

MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

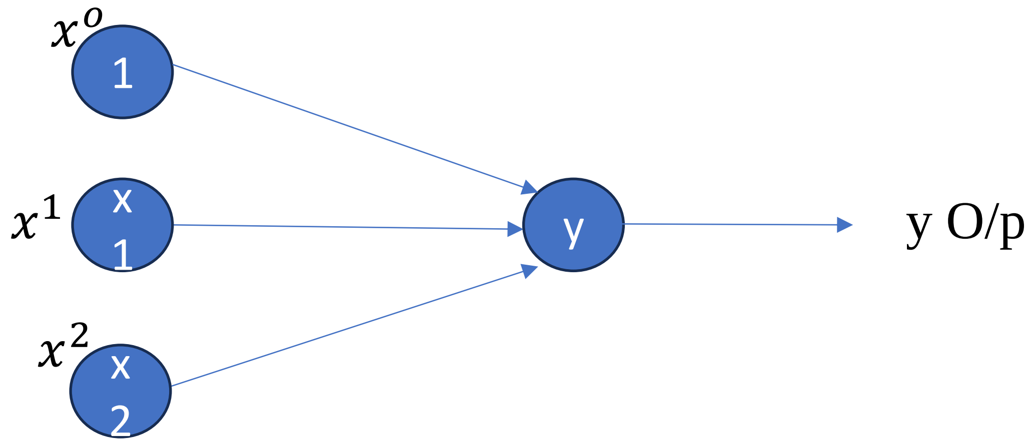
Linear Separability :



A Decision line is drawn to separate possible and negative responses

$$y_{in} = b + \sum_{i=1}^{n} xiwi$$

$$b + \sum_{i=1}^{n} xiwi \quad = 0$$



MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Linear Separability :

Requirement for +ve responses of the net is

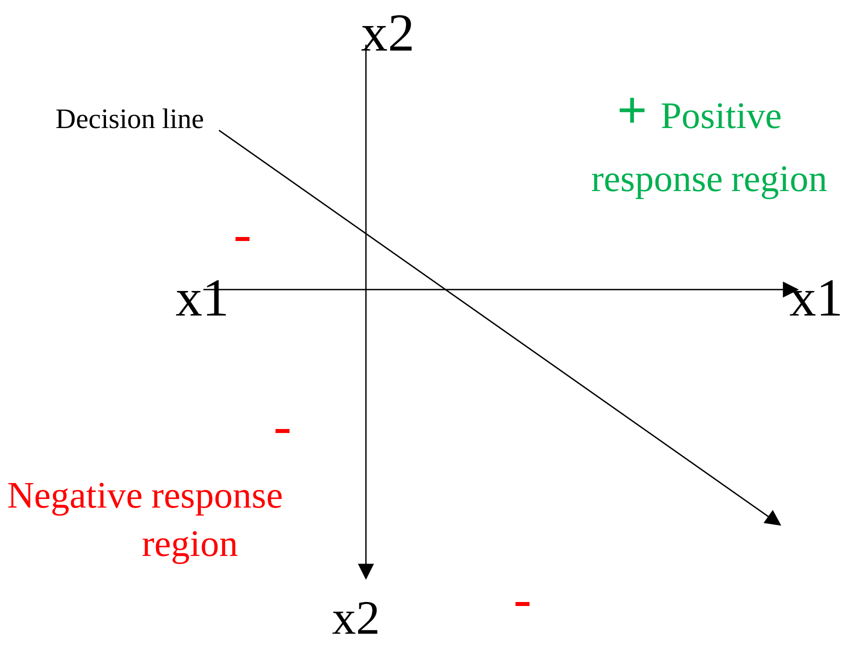
$$\boxed{b + \sum_{i=1}^{n} xiwi \quad > 0}$$

b+x1w1+x2w2 = 0

$$x2= -\frac{w1}{w2}x1- \frac{b}{w2}$$

example with AND Gate

X1	X2	Y
0	0	0
0	1	0
1	0	0
1	1	1



MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Linear **Regression :**

$$y = a0+a1x+e$$

$$\textit{Where } a1 = \frac{\overline{(xy)} - \overline{(x)}\overline{(y)}}{\overline{x^2} - \overline{x}^{-2}} \qquad a0 = \overline{y} - a_1 + \overline{x}$$

Week	Sale in thousands
1	1.2
2	1.8
3	2.6
4	3.2
5	3.8

Week (x1)	Sale in thousands (yi)	X1 power of 2	X1+yi
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

MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Linear **Regression :**

Week (x1)	Sale in thousands (yi)	X1 power of 2	X1+yi
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
X=3	12.6	55	44.4
Mean	2.52	11	8.88

$$a1 = \overline{(\boldsymbol{xy})} - \frac{\overline{(\boldsymbol{x})(\boldsymbol{y})}}{x^2 - x^{-2}}$$

$$= \frac{\boldsymbol{8.88\text{-}3{*}2.52}}{\boldsymbol{11\ \text{-}3^2}} = \boldsymbol{0.66}$$

$$\boldsymbol{a0} = \overline{\boldsymbol{y}} - \boldsymbol{a_1} + \overline{\boldsymbol{x}}$$

ao= 2.52-(0.66+3) = 0.54

If x = 7

Y = 0.54*0.66+ 7 = **5.16**

If x = 12

Y = 0.54*0.66+ 12 = **8.46**

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Linear Regression ***:***

Linear regression is one of the easiest and most popular Machine Learning algorithms.

It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as **sales, salary, age, product price,** etc.

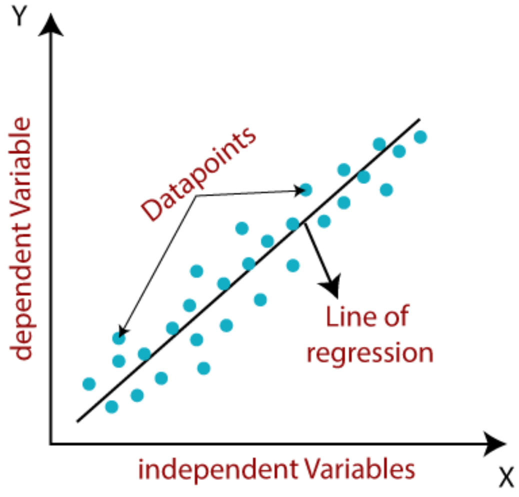
Linear regression algorithm shows a linear relationship between a **dependent (y)** and **one or more independent (y) variables**, hence called as linear regression.

Since linear regression shows the linear relationship, which means *it finds how the value of the dependent variable is changing according to the value of the independent variable*.

MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Linear Regression **:**



Mathematically, we can represent a linear regression as:

y= a0+a1x+ ε

Here,

Y= Dependent Variable (Target Variable)

X= Independent Variable (predictor Variable)

a0= intercept of the line (Gives an additional degree of freedom)

a1 = Linear regression coefficient (scale factor to each input value).

ε = random error

The values for x and y variables are training datasets for Linear Regression model representation.

Types of Linear Regression

Linear regression can be further divided into two types of the algorithm

Simple Linear Regression:

If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression

MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Linear Regression ***:***

Types of Linear Regression

Linear regression can be further divided into two types of the algorithm

Simple Linear Regression:

If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression

Multiple Linear regression:

If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

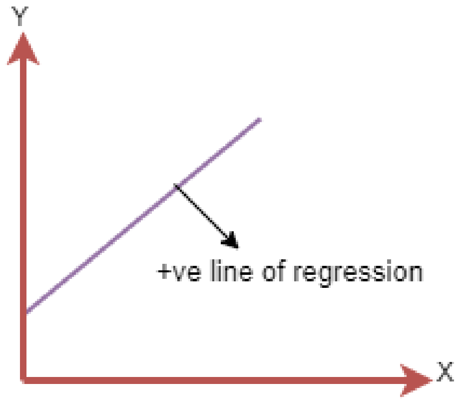
MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Linear **Regression :**

Linear Regression Line

A linear line showing the relationship between the dependent and independent variables is called a **regression line**.
A regression line can show two types of relationship:



The line equation will be: **$Y= a_0+a_1x$**

- **Positive Linear Relationship:**

If the dependent variable increases on the Y-axis and independent variable increases on X-axis, then such a relationship is termed as a Positive linear relationship.

MALLA REDDY COLLEGE OF ENGINEERING

Linear Discriminants: – Perceptron – Linear Separability – Linear Regression

Linear Regression **:**

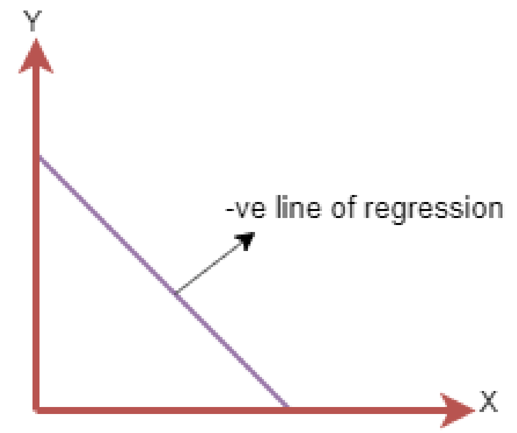
Linear Regression Line

A linear line showing the relationship between the dependent and independent variables is called a **regression line**.

A regression line can show two types of relationship:

Negative Linear Relationship:

If the dependent variable decreases on the Y-axis and independent variable increases
on the X-axis, then such a relationship is called a negative linear relationship



The line of equation will be: $\mathbf{Y = -a_0 + a_1 x}$