

Unit 54 - NLP and ANN. Learning Tools, Techniques and Applications.

* Introduction to Learning:-

- Learning is defined as follows:-

1. "Learning denotes changes in a system that enables a system to do some task more efficiently the next time" — Herbert Simon.
2. "Learning is constructing or modifying representations of what is being experienced" — Ryszard Michalski.
3. "Learning is making useful changes in our minds" — Marvin Minsky.

. Machine Learning:-

It refers to a system capable of the autonomous acquisition and integration of knowledge".

- Learning is essential for unknown environments.
- Learning modifies the agent's decision mechanism to improve the performance.
- Learning is important for agents to deal with unknown environments and changes.

* Machine Learning -

- It is used for building and exploring methods for programming computers to make them learn.
- Machine learning has a knowledge base and set of rules (Prediction rules) which helps in making a decision.

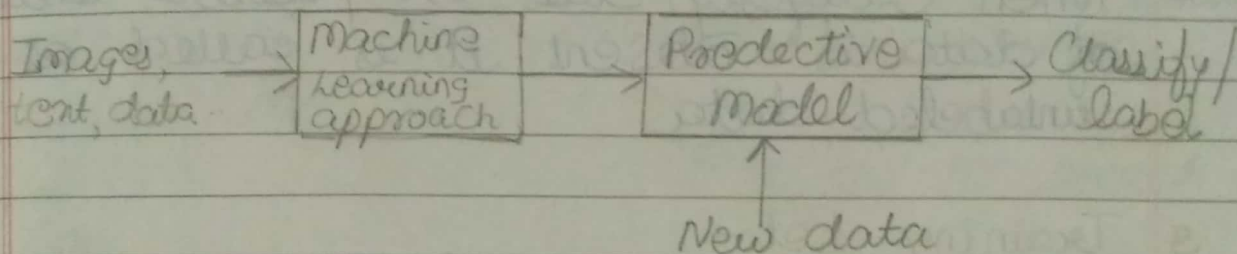


Fig. Machine Learning approach.

* Learning Concepts:-

1. Labeled Data.

- Given a specific set of inputs we are aware of the output.
- The data can be sampled into categories. The samples are labeled which can be positive or negative.

• Positive Sample:-

It provides information that belongs to a particular category.

• Negative Sample:-

- It is the information that does not

belong to that category.

Eg- Student xyz is in pune university.
So xyz becomes positive sample for
category pune university whereas
xyz becomes negative sample for
Mumbai university.

2. Unlabeled Data-

- When category/class or possible outcome of data is absent it is called as unlabeled data.

3. Training Set:-

It is made up of positive and negative samples.

4. Test Set:-

- It measures the accuracy of the classifier or learning approach.

* Supervised Learning:-

- The classification problem consist of full entities:-

1. Labeled Data Set.

Some available documents and their classes is known.

2. Classifier-

A program that maps input documents to appropriate classes.

- Full fig. shows how labeled data is classified into two different classes:-

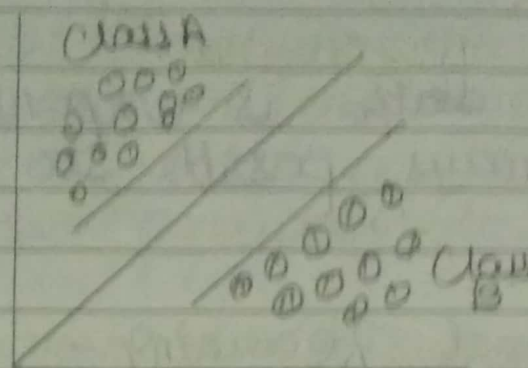


Fig.- Supervised Learning

- Supervised learning is a learning with labeled data.
- It is like a supervisor telling an agent whether categorization is correct or not.
- The aim of this type of learning is to find a function that maps input to output.
- Classifier and decision making engine should minimize false positive and false negatives.
- False Positive :-
Stands for result yes i.e. classified in particular group wrongly.
- Confusion / Error matrix is used here. It represents the performance of an classifier.

| | | | |
|--------------|-------------|----------------|----------------|
| Actual Class | Student | True Positive | False Positive |
| | Not Student | False Negative | True Negative |

Fig.- Confusion Matrix

- Challenges:-
Supervised data is expensive and it is not always possible to get it.

* Unsupervised Learning:-

- The data may not always be properly labeled.
- Huge unlabeled data is present that should be used during training.
- Learning based on unlabeled data is called as unsupervised learning.

• Understanding and Visualization:-

- Here learning is based on similarities and differences which are visible.
- Eg:- A child is able to separate cars from boats or ships.
- Child can differentiate that cars have wheels whereas boats do not have wheels.
- This is called as dissimilarity measure.
- So, based on visualization, objects can be categorized.
- Unsupervised learning performs hierarchical groupings (clustering).
- Clustering mechanism results in grouping of the objects based on similarities and differences.
- So, it supports 2 types of clustering:-

1. Hierarchical Clustering:-

Here objects are arranged into hierarchical structure in such a way that similar objects are grouped together.

2. Non-hierarchical Clustering:-

- It partitions the data set into disjoint clusters.

- Relationships between clusters is undetermined.

- Fall fig. shows clustering of unlabeled data:

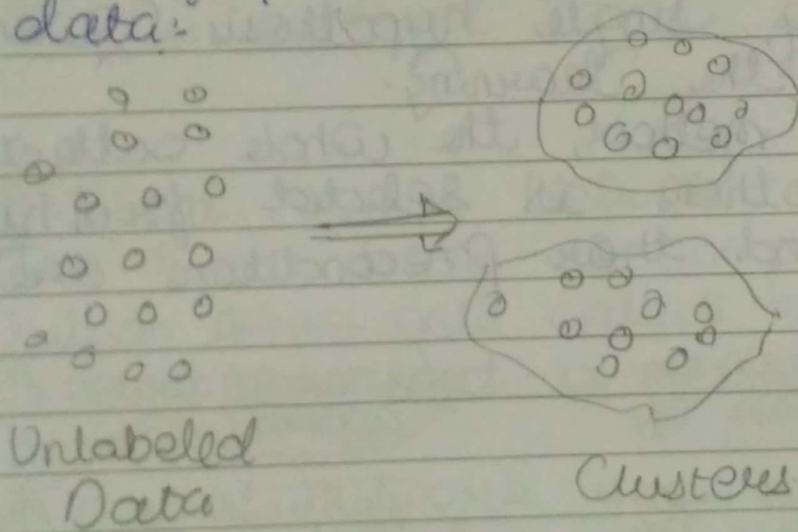


Fig:- Clustering of data.

* Semi-Supervised Learning:-

- In real life scenarios, we try to learn from both supervised and unsupervised learning methods.

- Semi-Supervised learning tries to learn from both labeled and unlabeled data.

- Let U be set of unlabeled data
 L be set of labeled data
- The learning approach tries to identify unlabeled data U with reference to labeled data L and keeps on labeling the unlabeled data.

5.9 * Ensemble Learning:-

- In this learning method several multiple learners/ learning algorithms are trained.
- Generally, single hypothesis (given fact) drives the learning.
- In this method, the whole collection of hypothesis is selected from hypothesis space and their preconditions are combined.

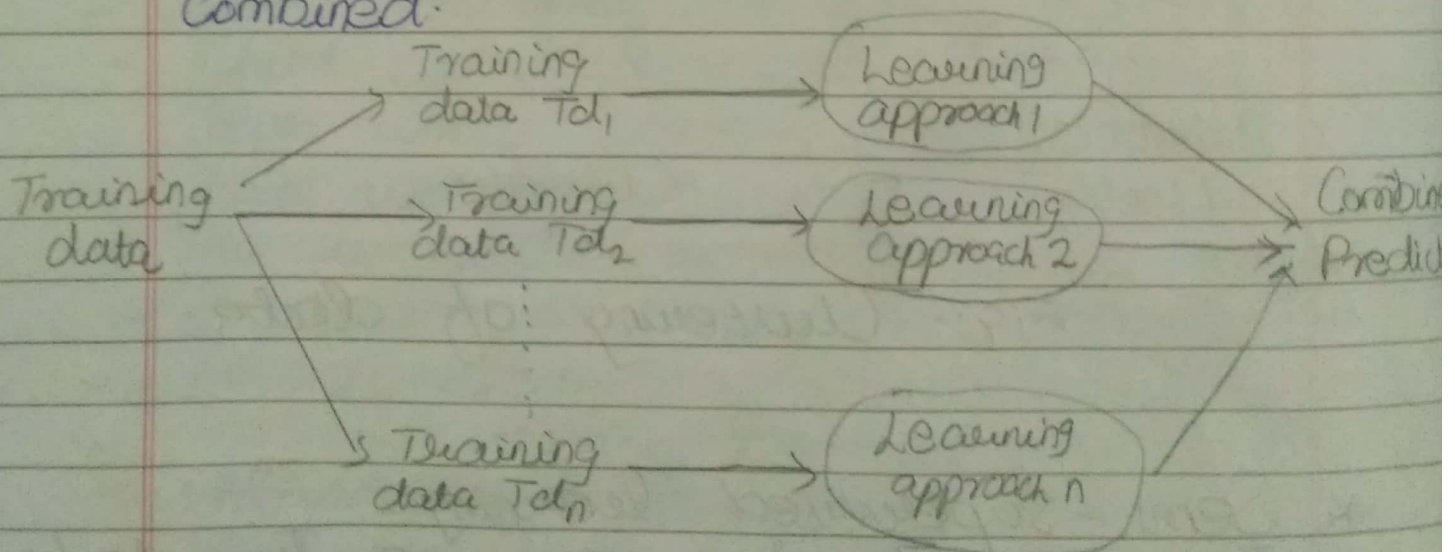


Fig:- Ensemble method: The concept.

* Reinforcement Learning:-

- RL is a machine learning approaches where agent explores an environment.
- The agent perceives its current state and takes actions.
- The environment in return provides a reward as positive or negative.
- The algorithm tries to find a policy for maximizing reward for the agent.

- Definition:-

RL is a computational approach to learning where an agent tries to maximize the total amount of reward it receives when interacting with complex uncertain environment.

• RL Problem:-

- It is a goal directed problem and agent ~~to~~ uses its decision making capabilities.
- It is a supervised learning approach.
- RL emphasizes learning from interactⁿ.

with its environment.

- RL uses a formal framework for defining interaction between a learning agent and its environment in terms of states, actions and rewards.

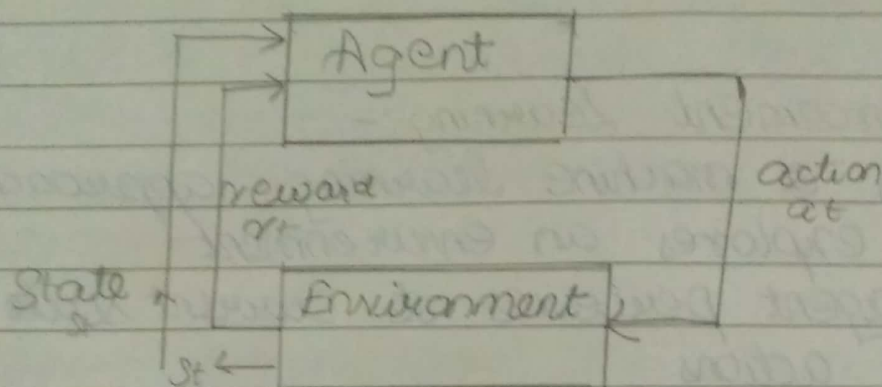


Fig:- Agent-Environment Interaction.

- For every interaction, agent gets some input i .
- This i gives brief detail about the current state.
- Based on this input agent takes an action a .
- Every action when executed results in the transition to a new state and a value associated with its transition is given to the agent by means of a signal known as reinforcement signal.
- The goal is to choose the actions that maximize the final summation of x values.

* Features of RL:-

- The learner is not told what actions to take, instead it finds out what to do by trial - and - error search.
- The environment is stochastic; i.e., the behaviour is non-deterministic means a "state" does not fully determine its next "state".
- The learner has to balance between the need to explore its environment and the need to exploit its current knowledge.

6 * Artificial Neural Network:- (ANN)

- The human brain is composed of billion nerve cells called as neurons.
- A neuron can send a message to other neuron.
- A neural network is an artificial representation of human brain that tries to simulate its learning process.
- An ANN is often called as neural n/w.
- + A neural networks model a brain which is called as learning by example.
- A neural network typically take a vector of input values and produce a vector of output values.
- + Inside they train weights of "neuron".

- ANN is composed of multiple nodes which represents neurons.
- The neurons are connected by links and they interact with each other.
- The nodes can take input data and perform simple operations on the data.
- The output of each node is called its activation or node value.
- Each link is associated with weights.
- ANNs are capable of learning, which takes place by altering weight values.
- The following illustration shows a simple ANN.

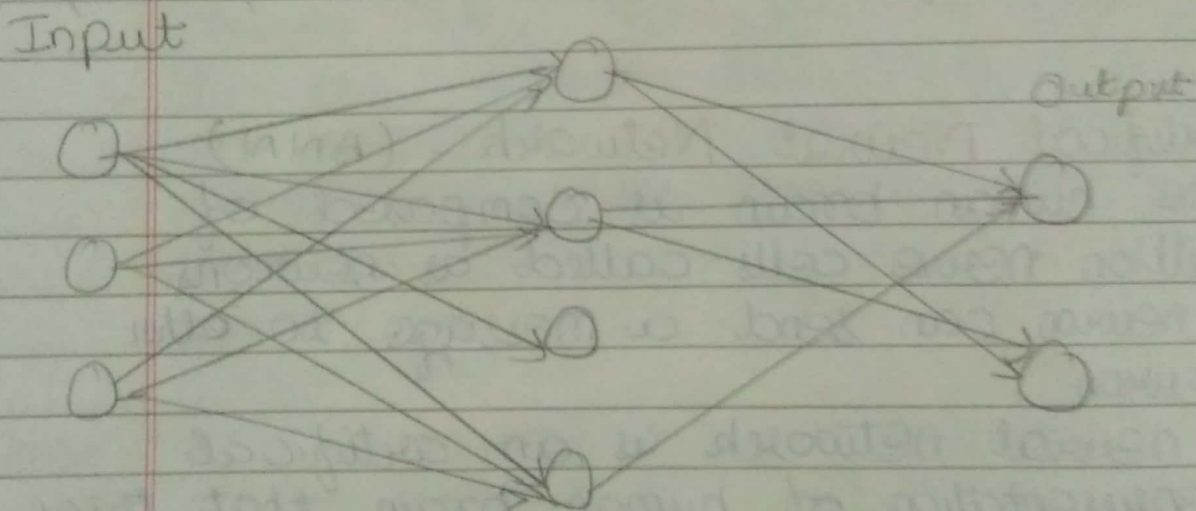


Fig:- ANN model

- ANN has different inputs which are weighted as per the strength of a signal.
- ANNs are mathematical functions which determine the activation of the neuron.

- ANN's combine artificial neurons to process information.
- ANN - Back Propagation Algorithm:-
 - Here artificial neurons send their signals forward and then errors are propagated backwards.
 - This algo. is supervised and hence error can be calculated.
 - The user provide the algo examples of input and output they want and the difference betⁿ actual o/p and expected o/p is calculated. This is error.
 - The goal is to adjust the weight in order to minimize the errors.

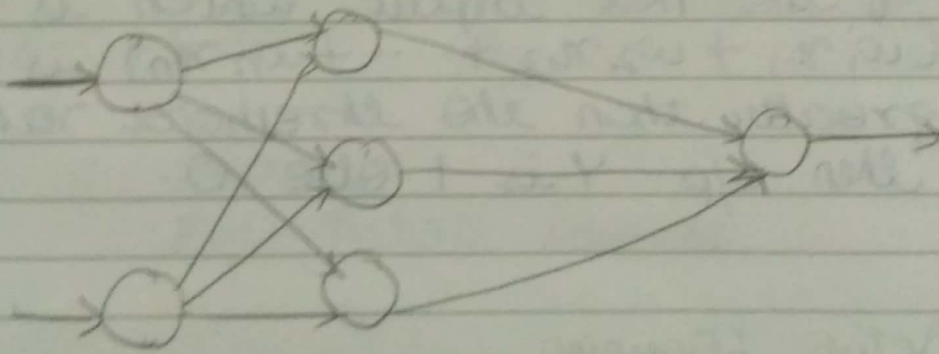


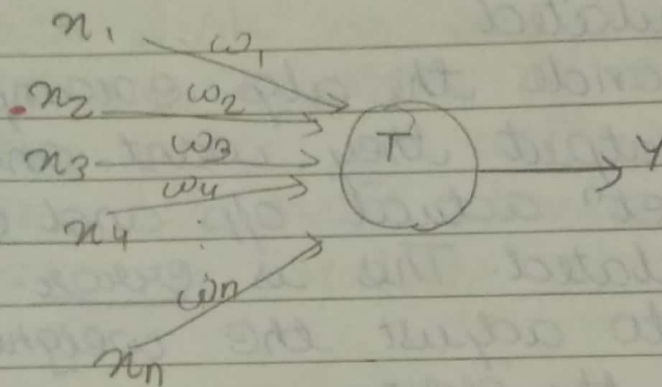
Fig:- ANN Backpropagation algo.

- This algo has 2 passes:-
 1. Forward Pass:-
 - Here hidden layer uses the input values which computes the neurons output.
 - These outputs are input to the next hidden layer.
 - Finally the output value is calculated.

2. Backward Pass:-

- In the output layer error is calculated for each neuron.
- The error is propagated and weights are adjusted to get desired value.

- A perceptron is a single model of trainable 'neuron'.



- If the net input which is $(w_1 n_1 + w_2 n_2 + \dots + w_n n_n)$ is greater than the threshold value T then o/p Y is 1 else 0.