



NAAC A⁺
Accredited

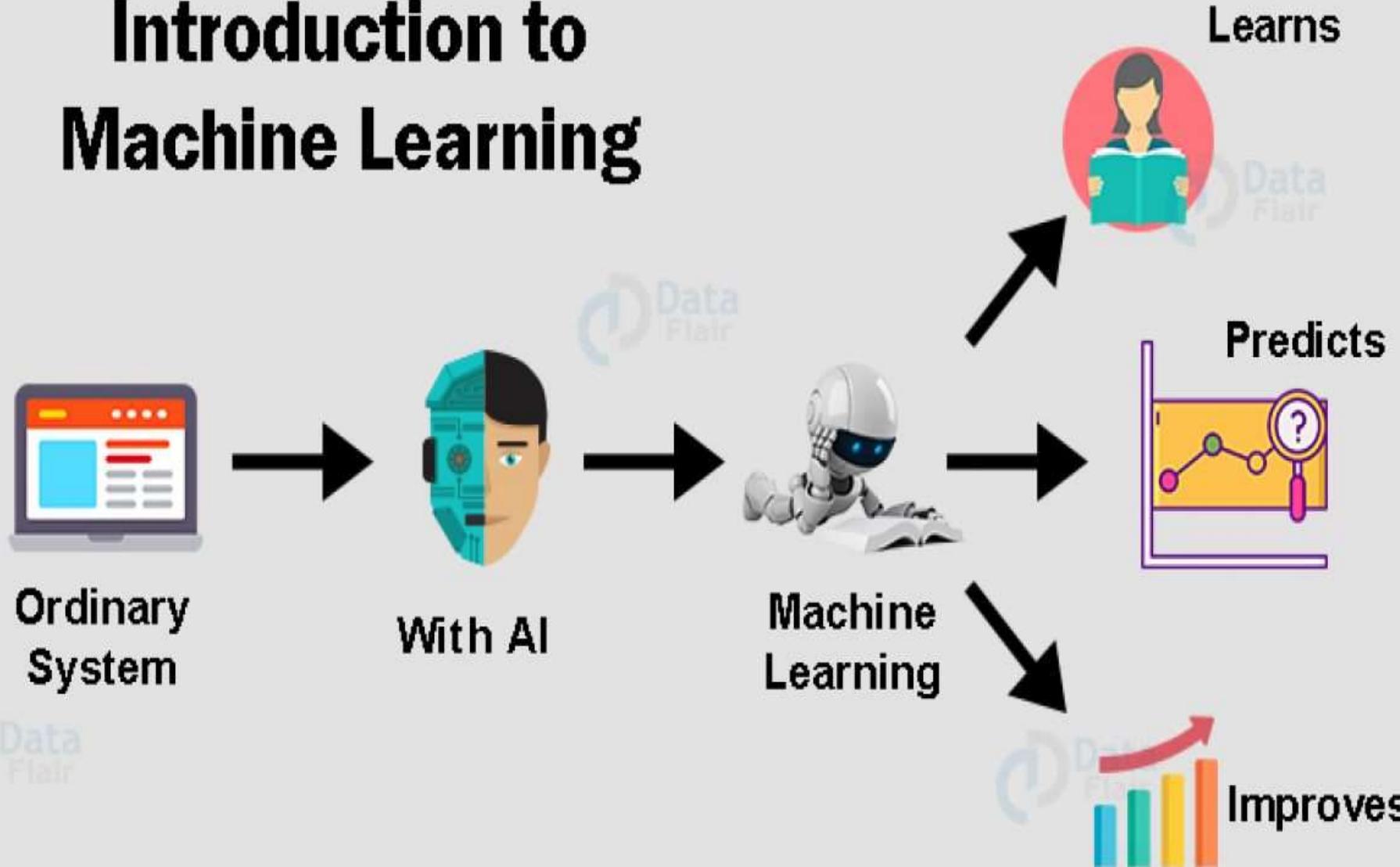
Unit 5

Learning

Topics to be Covered:

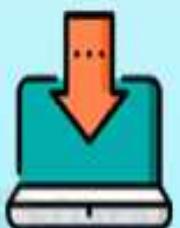
- Forms of Learning
 - Supervised Learning
 - Unsupervised Learning
 - Reinforcement Learning
- Ensemble Learning
- Learning Decision Trees
- Artificial Neural Networks
- Expert Systems
- Machine Learning
- Natural Language Processing

Introduction to Machine Learning



How does Machine Learning Work?

Input Data → Analyze Data → Find Patterns → Prediction → Stores the Feedback



Types of Machine Learning

Machine Learning

Supervised

Task Driven
(Predict next value)



Unsupervised

Data Driven
(Identify Clusters)

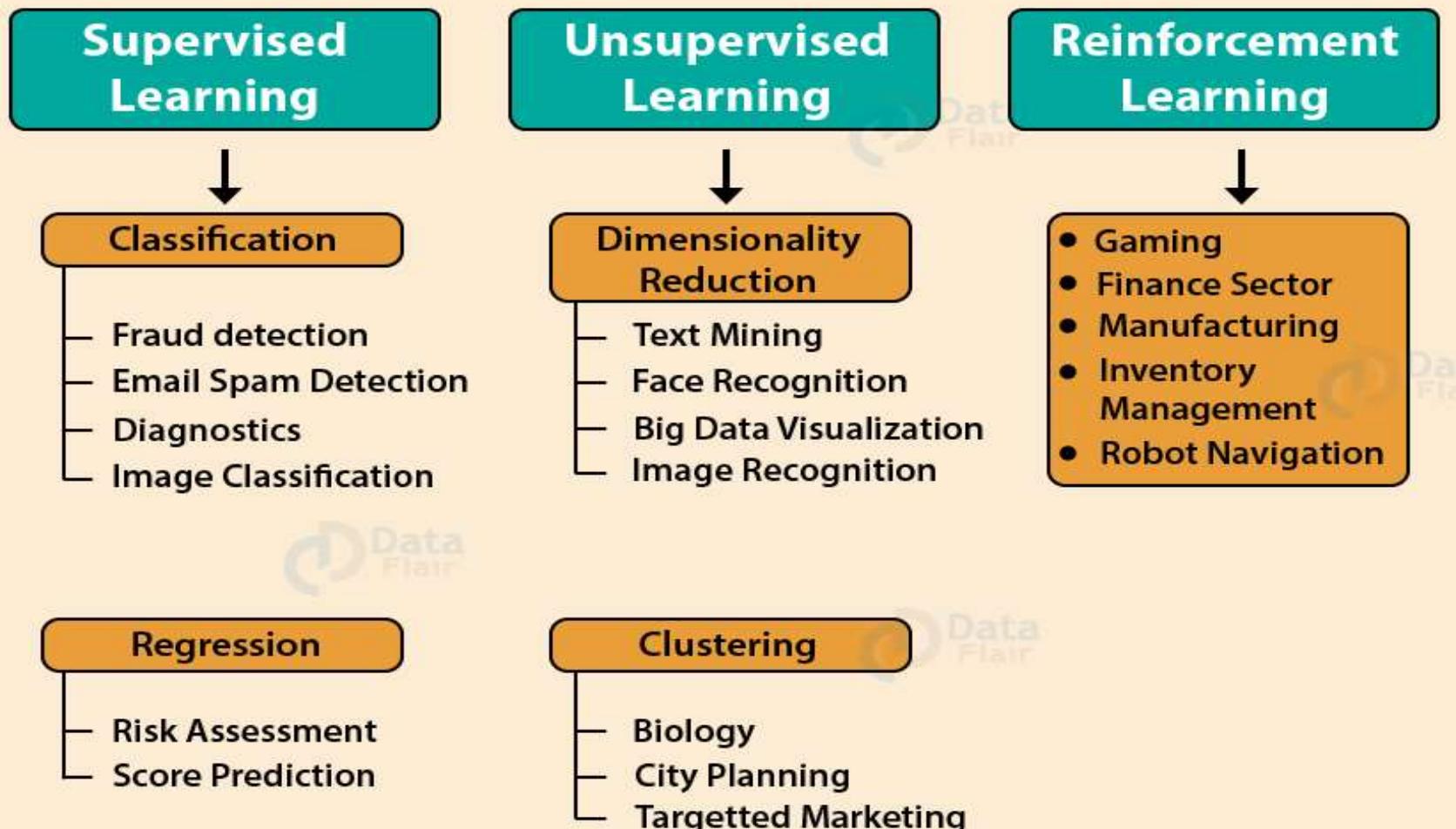


Reinforcement

Learn from
Mistakes



Types of Machine Learning



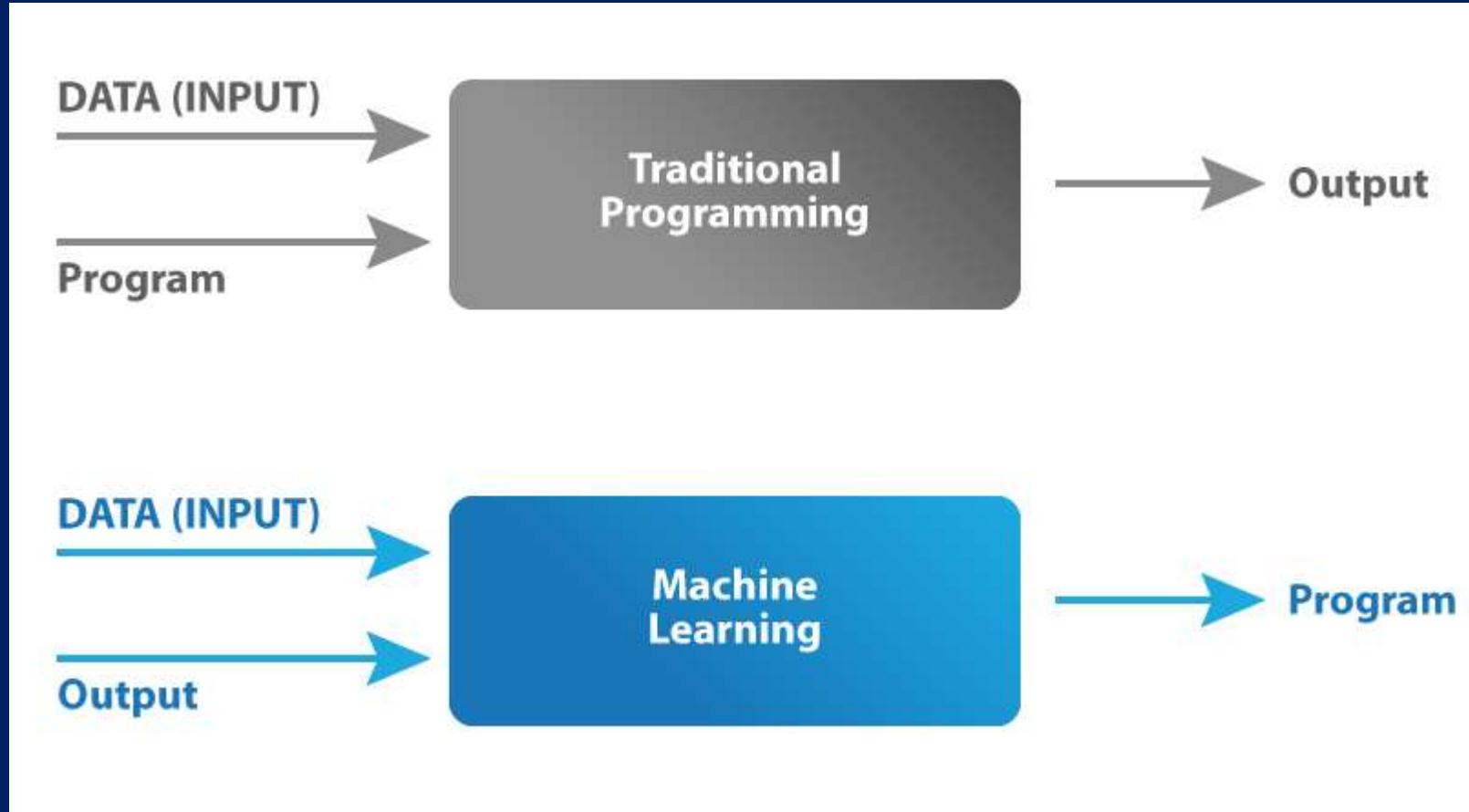
What is Machine Learning?

- Arthur Samuel coined the term Machine Learning in the year 1959.
- He was a pioneer in Artificial Intelligence and computer gaming, and defined Machine Learning as

“Field of study that gives computers the capability to learn without being explicitly programmed”.

- A machine is said to be learning from **past Experiences** (data feed in) with respect to some class of **Tasks**, if its **Performance** in a given Task improves with the Experience.

How it is different from traditional programming?



Some Terminology of Machine Learning

• Model:

- Also known as “hypothesis”, a machine learning model is the mathematical representation of a real-world process.
- A machine learning algorithm along with the training data builds a machine learning model.

• Feature:

- A feature is a measurable property or parameter of the data-set.

• Feature Vector:

- It is a set of multiple numeric features. We use it as an input to the machine learning model for training and prediction purposes.

• Training:

- An algorithm takes a set of data known as “training data” as input. The learning algorithm finds patterns in the input data and trains the model for expected results (target). The output of the training process is the machine learning model.

Some Terminology of Machine Learning

•Prediction:

- Once the machine learning model is ready, it can be fed with input data to provide a predicted output.

•Target (Label):

- The value that the machine learning model has to predict is called the target or label.

•Overfitting:

- When a massive amount of data trains a machine learning model, it tends to learn from the noise and inaccurate data entries. Here, the model fails to characterise the data correctly.

•Underfitting:

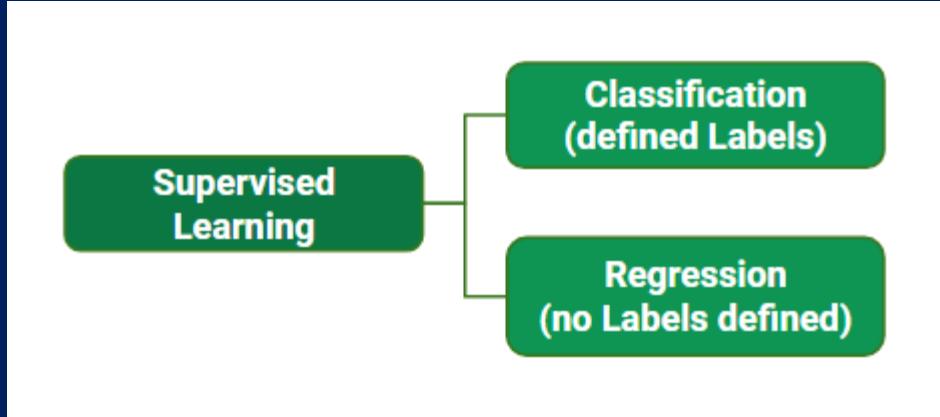
- It is the scenario when the model fails to decipher the underlying trend in the input data. It destroys the accuracy of the machine learning model. In simple terms, the model or the algorithm does not fit the data well enough.

Supervised Learning

- Supervised learning is when the model is getting trained on a labelled dataset.
- Labelled dataset is one which have both input and output parameters.
- In this type of learning both training and validation datasets are labelled.
- The supervised learning model has a set of input variables (x), and an output variable (y).
- An algorithm identifies the mapping function between the input and output variables. The relationship is $y = f(x)$.
- The learning is monitored or supervised in the sense that we already know the output and the algorithm are corrected each time to optimise its results.
- The algorithm is trained over the data set and amended until it achieves an acceptable level of performance.

Types of Supervised Learning

We can group the supervised learning problems as:



- Classification Problems – Various labels train the algorithm to identify items within a specific category. E.g., Disease or no disease, Apple or an orange, Beer or wine.
- Regression Problems – Used to predict future values and the model is trained with the historical data.
E.g., Predicting the future price of a product.

Classification

- Classification is the task of “classifying things” into sub-categories.
- In Machine Learning and Statistics, Classification is the problem of identifying to which of a set of categories (sub populations), a new observation belongs to, on the basis of a training set of data containing observations and whose categories membership is known.

Classification..

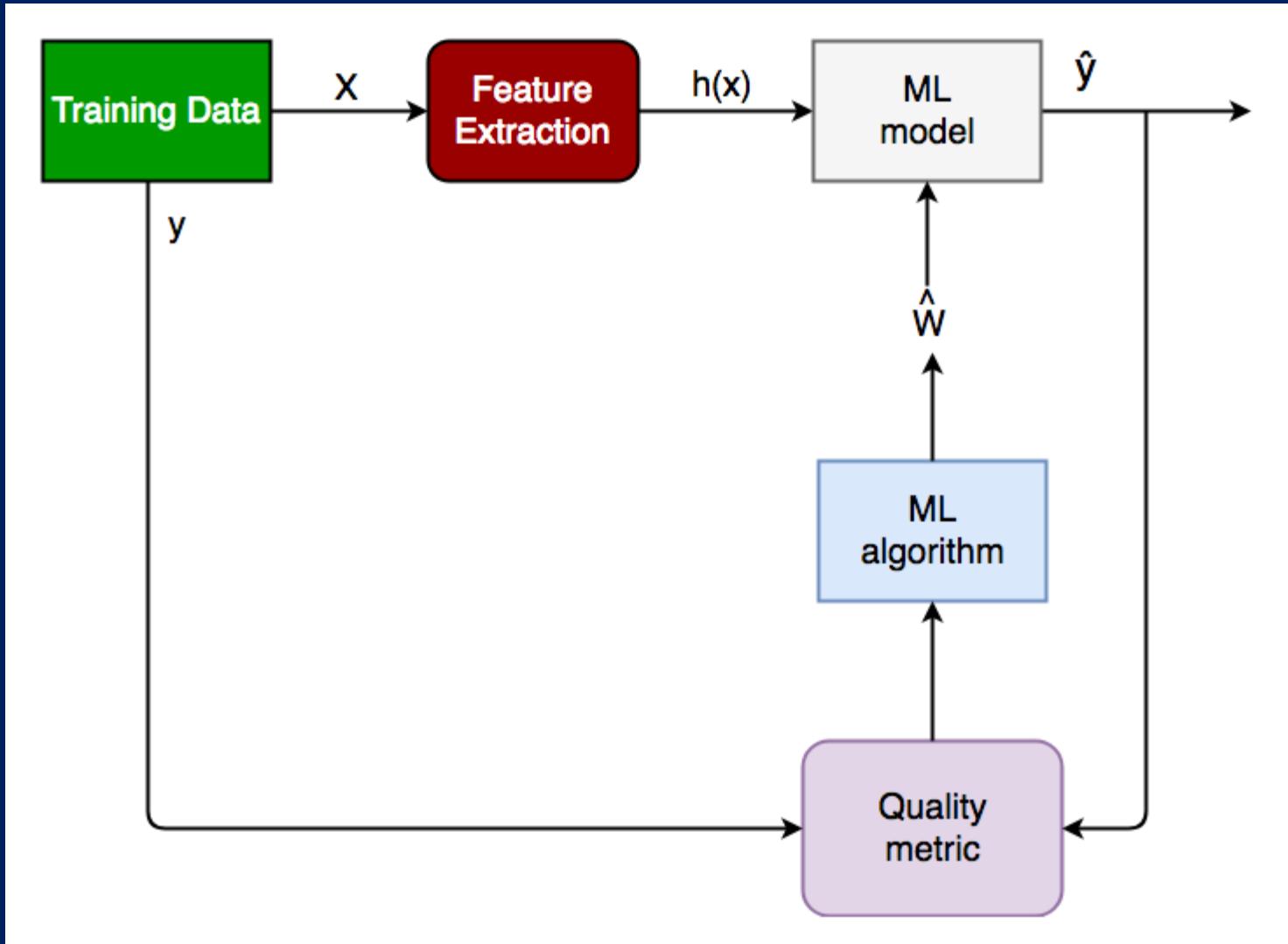


Figure 1: Generalized Classification Block Diagram.

Classification steps

1. X : pre-classified data, in the form of a $N \times M$ matrix. N is the no. of observations and M is the number of features
2. y : An $N \times d$ vector corresponding to predicted classes for each of the N observations.
3. Feature Extraction : Extracting valuable information from input X using a series of transforms.
4. ML Model : The “Classifier” we’ll train.
5. y' : Labels predicted by the Classifier.
6. Quality Metric : Metric used for measuring the performance of the model.
7. ML Algorithm : The algorithm that is used to update weights w' , which update the model and “learns” iteratively.

Types of Classification

• Binary Classification :

- When we have to categorize given data into 2 distinct classes.
- Example – On the basis of given health conditions of a person, we have to determine whether the person has a certain disease or not.

• Multiclass Classification :

- The number of classes is more than 2.
- For Example – On the basis of data about different species of flowers, we have to determine which species does our observation belong to.

Types of Classifiers (algorithms)

There are various types of classifiers. Some of them are :

- Linear Classifiers : Logistic Regression
- Tree Based Classifiers : Decision Tree Classifier
- Support Vector Machines
- Artificial Neural Networks
- Bayesian Regression
- Gaussian Naive Bayes Classifiers
- Stochastic Gradient Descent (SGD) Classifier
- Ensemble Methods : Random Forests, AdaBoost, Bagging Classifier, Voting Classifier, ExtraTrees Classifier

Practical Applications of Classification

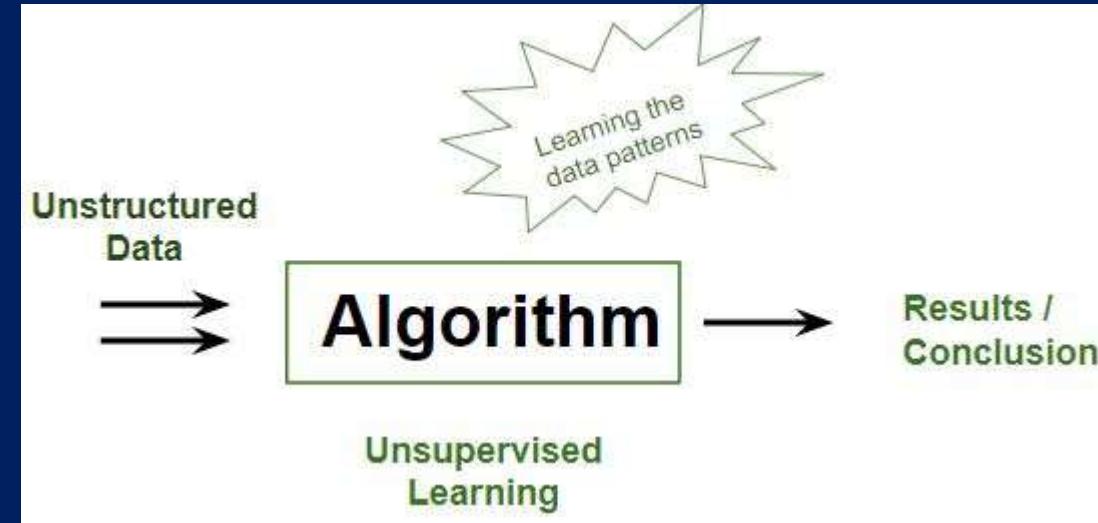
1. Google's self driving car uses deep learning enabled classification techniques which enables it to detect and classify obstacles.
2. Spam E-mail filtering is one of the most widespread and well recognized uses of Classification techniques.
3. Detecting Health Problems, Facial Recognition, Speech Recognition, Object Detection, Sentiment Analysis.

All use Classification at their core.

Unsupervised Learning

- This approach is the one where the output is unknown, and we have only the input variable at hand.
- The algorithm learns by itself and discovers an impressive structure in the data.
- The goal is to decipher the underlying distribution in the data to gain more knowledge about the data.

- **Unstructured data:** May contain noisy(meaningless) data, missing values or unknown data.
- **Unlabelled data:** Data only contains value for input parameters, there is no targeted value(output).



Types of Unsupervised Learning

We can group the unsupervised learning problems as:



1. **Clustering:** This means bundling the input variables with the same characteristics together.

E.g., grouping users based on search history

2. **Association:** Here, we discover the rules that govern meaningful associations among the data set.
E.g., People who watch ‘X’ will also watch ‘Y.’

Unsupervised Learning based Algorithms

- K-Means Clustering
- K-Medoids Clustering
- Partition Around Medoids (PAM)
- DBSCAN – Density-Based Spatial Clustering of Applications with Noise
- BIRCH – Balanced Iterative Reducing and Clustering using Hierarchies
- Hierarchical Clustering

Semi-supervised Learning

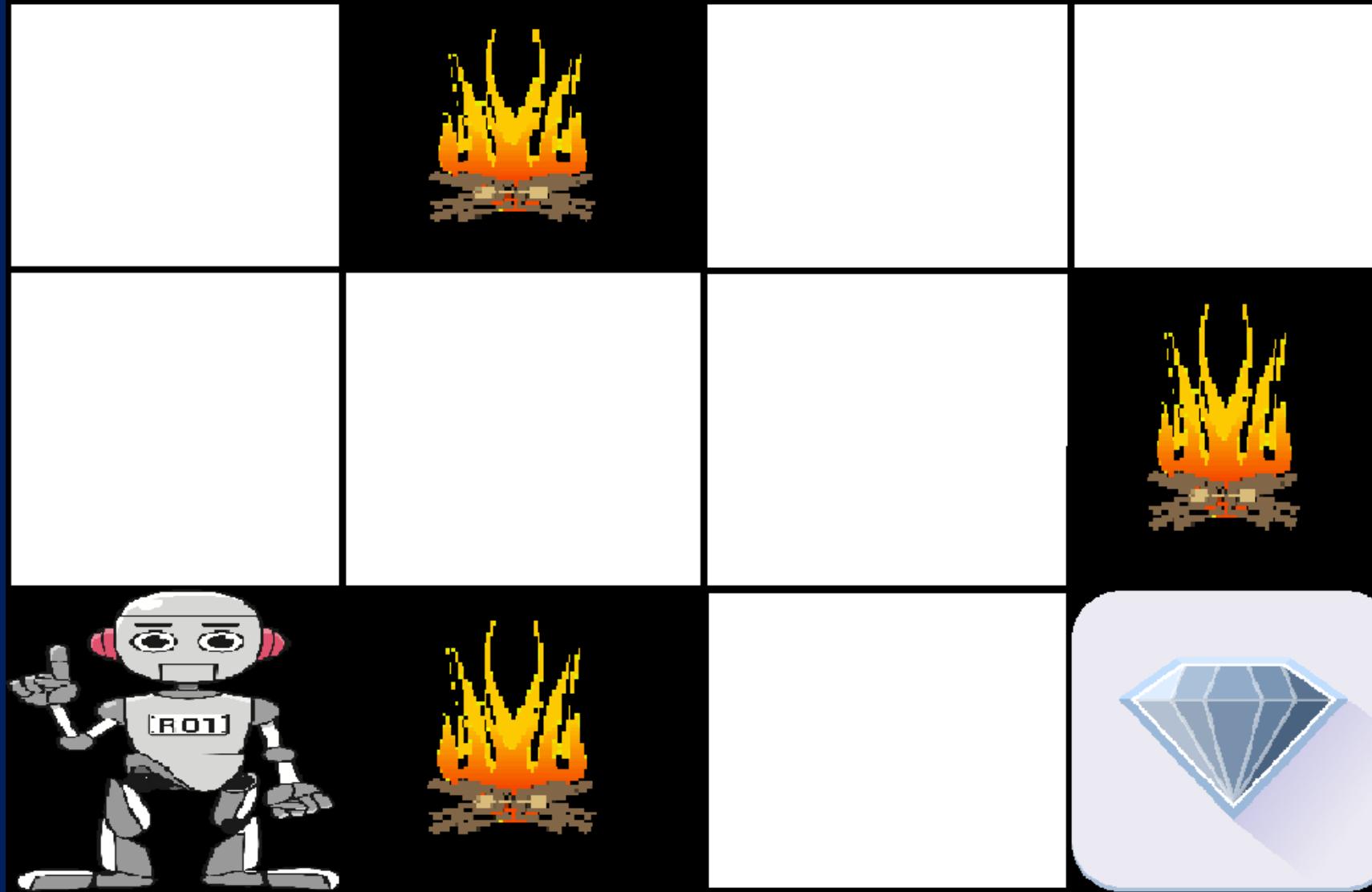
- In semi-supervised learning, data scientists train model with a minimal amount of labelled data and a large amount of unlabelled data.
- Usually, the first step is to cluster similar data with the help of an unsupervised machine learning algorithm.
- The next step is to label the unlabelled data using the characteristics of the limited labelled data available.
- After labelling the complete data, one can use supervised learning algorithms to solve the problem.
- This technique is mostly applicable in case of image data-sets where usually all images are not labelled.

Reinforcement Learning

- In this approach, machine learning models are trained to make a series of decisions based on the rewards and feedback they receive for their actions.
- The machine learns to achieve a goal in complex and uncertain situations and is rewarded each time it achieves it during the learning period.
- Reinforcement learning is different from supervised learning in the sense that there is no answer available, so the reinforcement agent decides the steps to perform a task.
- The machine learns from its own experiences when there is no training data set present.



Example



Main points in Reinforcement learning

- Input:
 - The input should be an initial state from which the model will start
- Output:
 - There are many possible output as there are variety of solution to a particular problem
- Training:
 - The training is based upon the input, The model will return a state and the user will decide to reward or punish the model based on its output.
- The model keeps continues to learn.
- The best solution is decided based on the maximum reward.

Types of Reinforcement:

1. Positive –

Positive Reinforcement is defined as when an event, occurs due to a particular behaviour, increases the strength and the frequency of the behaviour. In other words, it has a positive effect on behaviour.

- Advantages of reinforcement learning are:
 1. Maximizes Performance
 2. Sustain Change for a long period of time
- Disadvantages of reinforcement learning:
 1. Too much Reinforcement can lead to overload of states which can diminish the results.

Types of Reinforcement..:

2. Negative –

Negative Reinforcement is defined as strengthening of a behaviour because a negative condition is stopped or avoided.

- Advantages of reinforcement learning:
 1. Increases Behaviour
 2. Provide defiance to minimum standard of performance
- Disadvantages of reinforcement learning:
 1. It only provides enough to meet up the minimum behaviour.

Reinforcement Learning Algorithms

- Temporal Difference (TD)
- Q-Learning
- Deep Adversarial Networks

Various Practical Applications of Reinforcement Learning

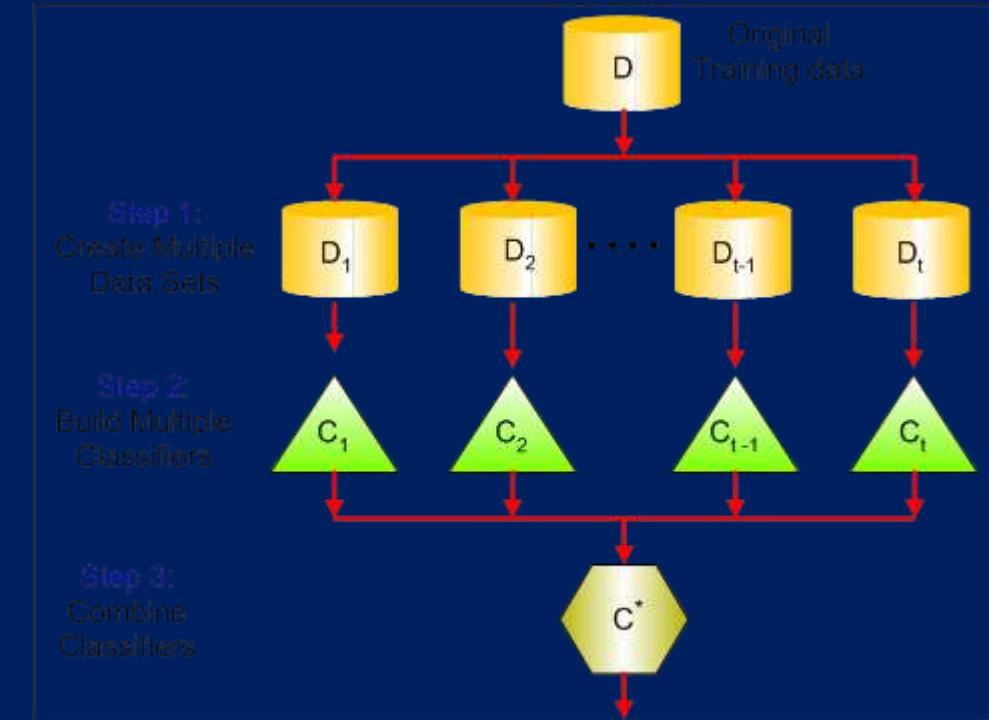
- RL can be used in robotics for industrial automation.
- RL can be used in machine learning and data processing.
- RL can be used to create training systems that provide custom instruction and materials according to the requirement of students.

Ensemble Learning

- Ensemble Learning is a technique that create multiple models and then combine them to produce improved results.
- Ensemble learning usually produces more accurate solutions than a single model would.
- Ensemble learning methods is applied to regression as well as classification.
- Ensemble learning for regression creates multiple regressors i.e. multiple regression models such as linear, polynomial, etc.
- Ensemble learning for classification creates multiple classifiers i.e. multiple classification models such as logistic, decision tress, KNN, SVM, etc.

Advantage : Improvement in predictive accuracy.

Disadvantage : It is difficult to understand an ensemble of classifiers.

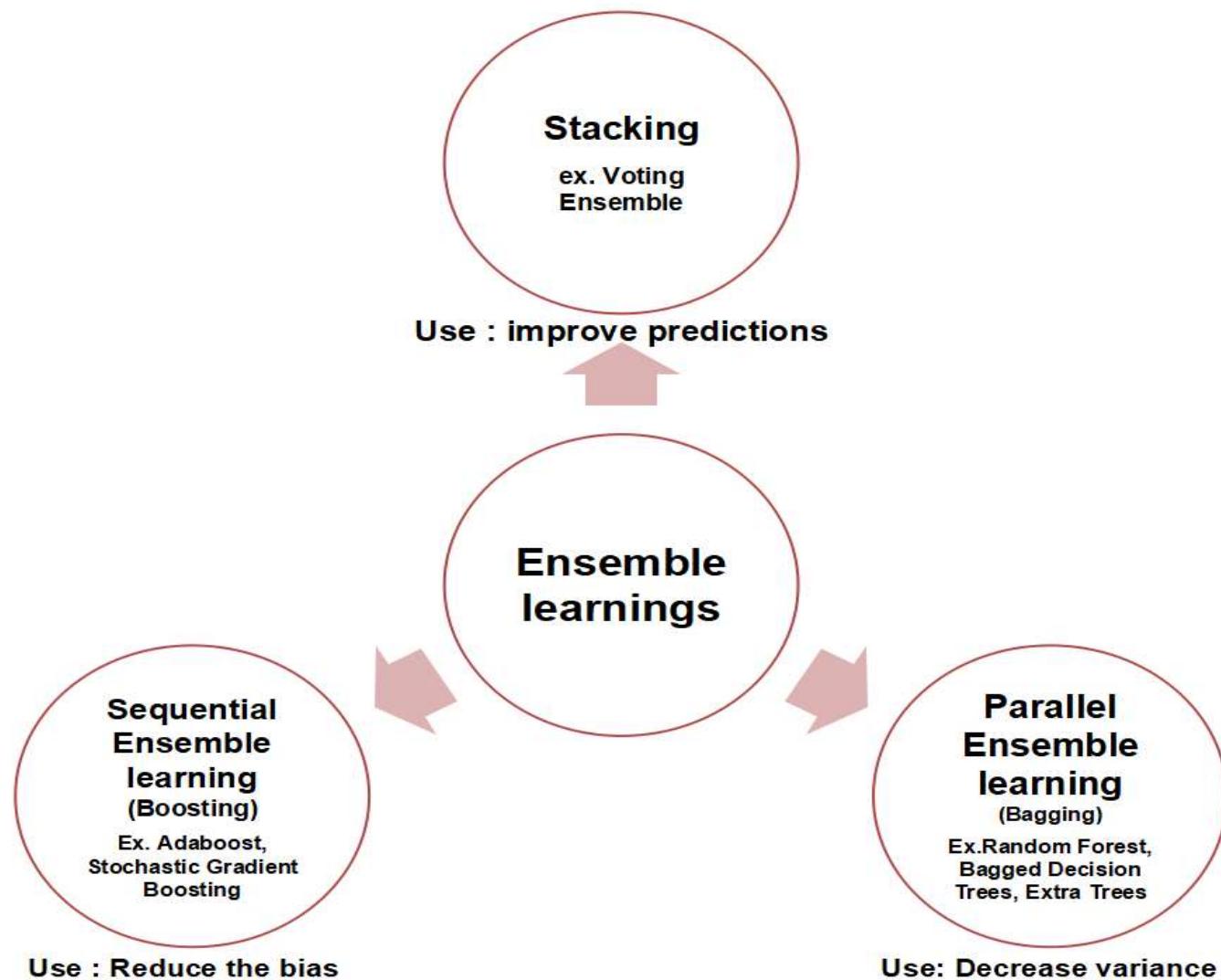




Ensemble Learning...

There are two steps in ensemble learning:

- Multiples machine learning models were generated using same or different machine learning algorithm. These are called “base models”.
- The prediction perform on the basis of base models.



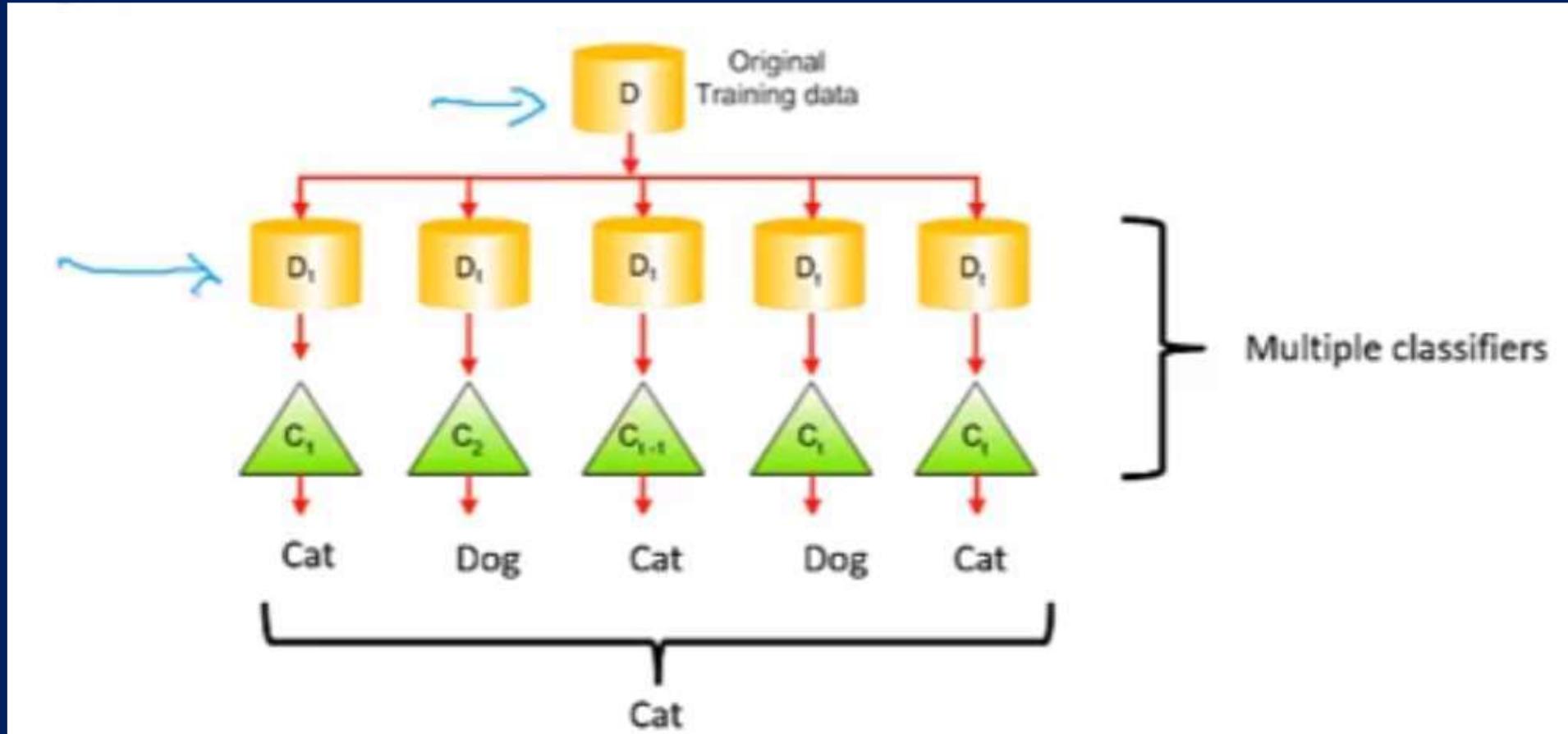
Ensemble Learning Techniques/Methods...

- Voting and Averaging
 - Voting used in classification
 - Averaging used in regression
- Stacking
- Bootstrap Aggregating/Bagging
- Boosting

Ensemble Learning Techniques/Methods...

Voting and Averaging

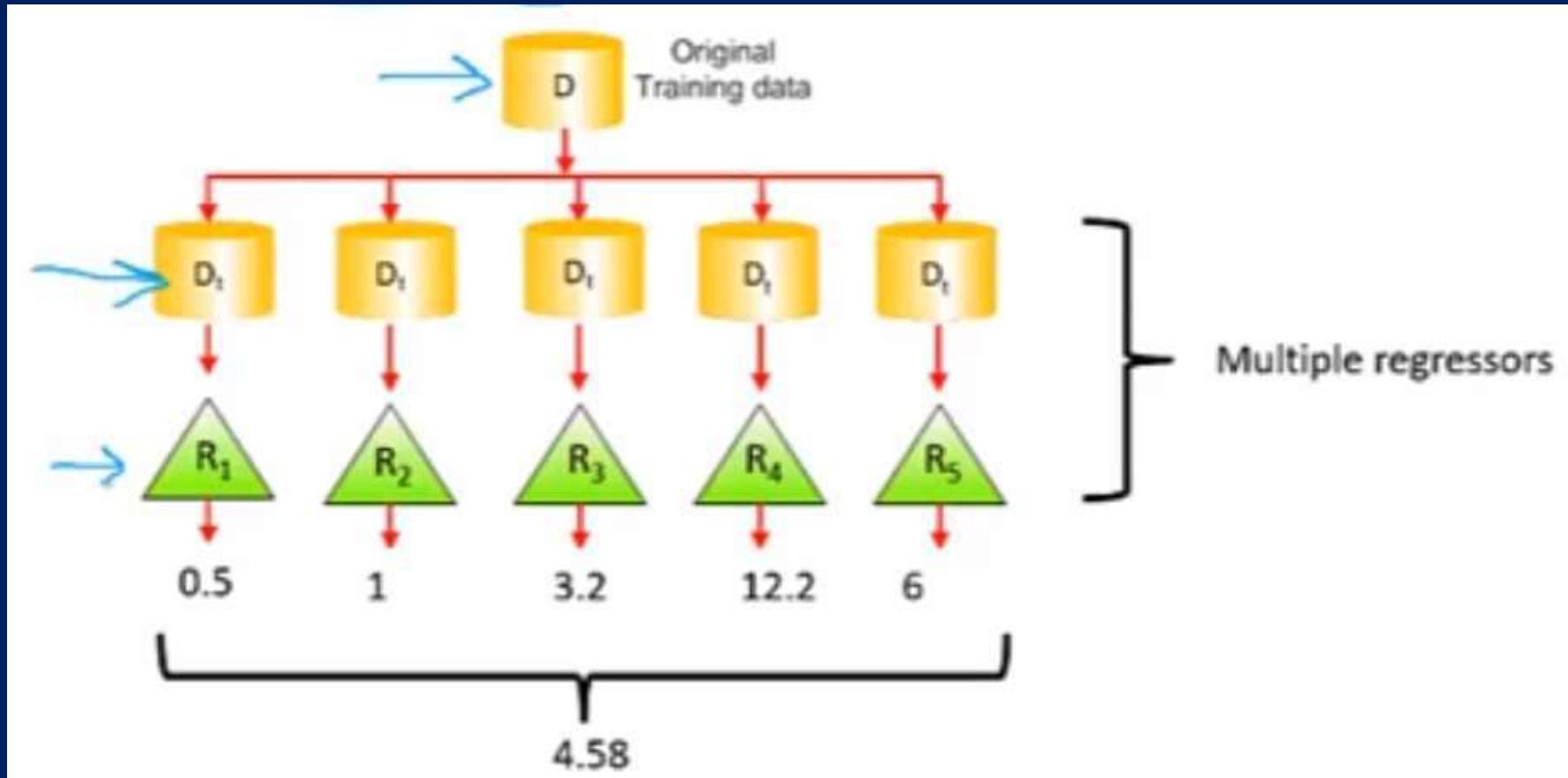
- Voting used in classification



Ensemble Learning Techniques/Methods...

Voting and Averaging

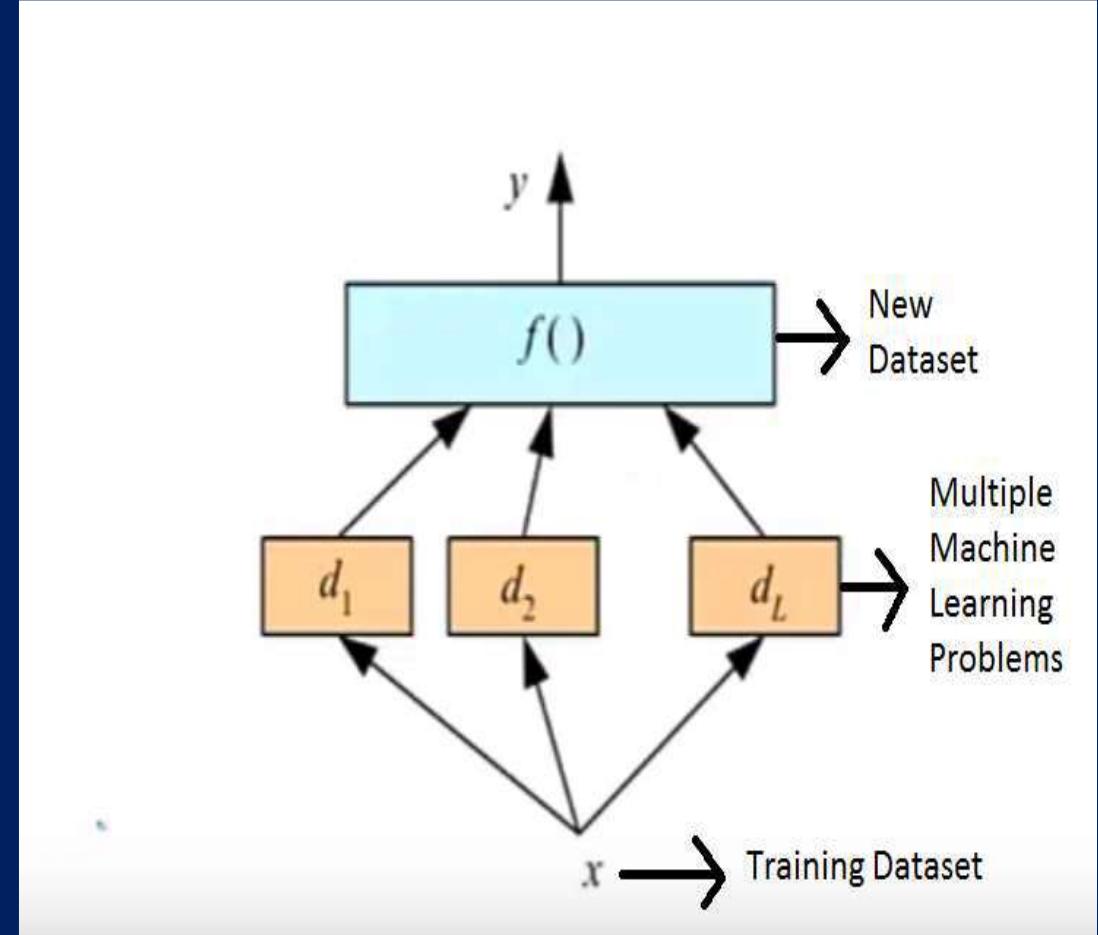
- Averaging used in regression



Ensemble Learning Techniques/Methods...

Stacking

- Also known as stacked generalization.
- Multiple machine learning models are combined using another (combiner) machine learning algorithm.
- The basic idea is to train machine learning algorithms with training dataset and then generate a new dataset with these models.
- Then this new dataset is used as input for the combiner machine learning algorithm.



Ensemble Learning Techniques/Methods..

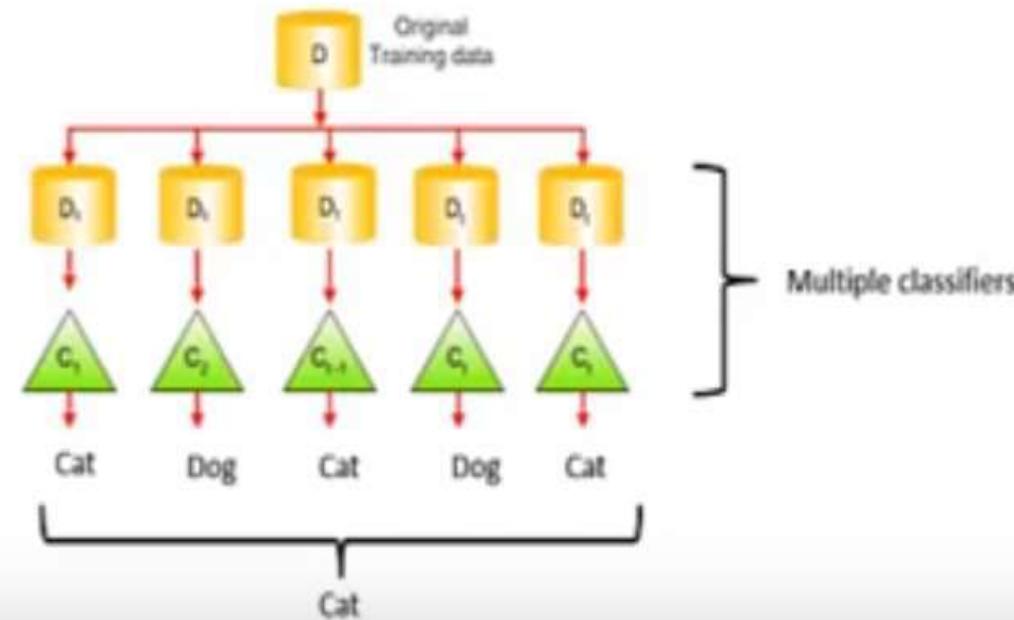
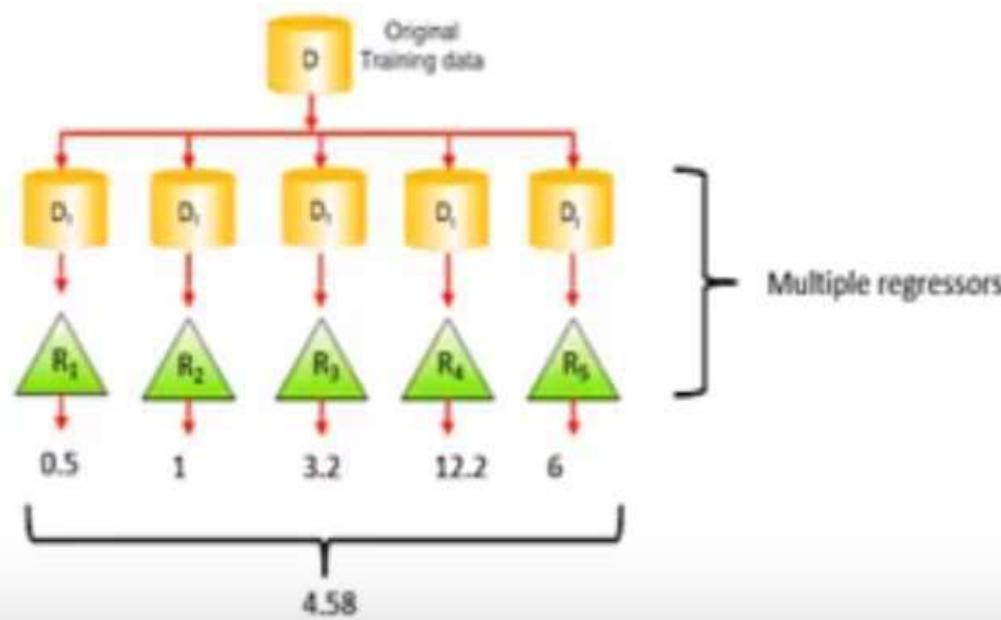
Bootstrap Aggregating / Bagging

- Again at first step multiple machine learning models are generated. These models are generated using the same machine learning algorithm with n random observations of m sub-samples of the original dataset using bootstrap sampling method.
- The second step is aggregating the result generated from these models. Well known methods, such as voting and averaging, are used for this purpose.
- For example, Random Forest algorithm uses the bagging technique.

Ensemble Learning Techniques/Methods..

Bootstrap Aggregating / Bagging

- Voting and averaging is used in second step for final prediction



Ensemble Learning Techniques/Methods...

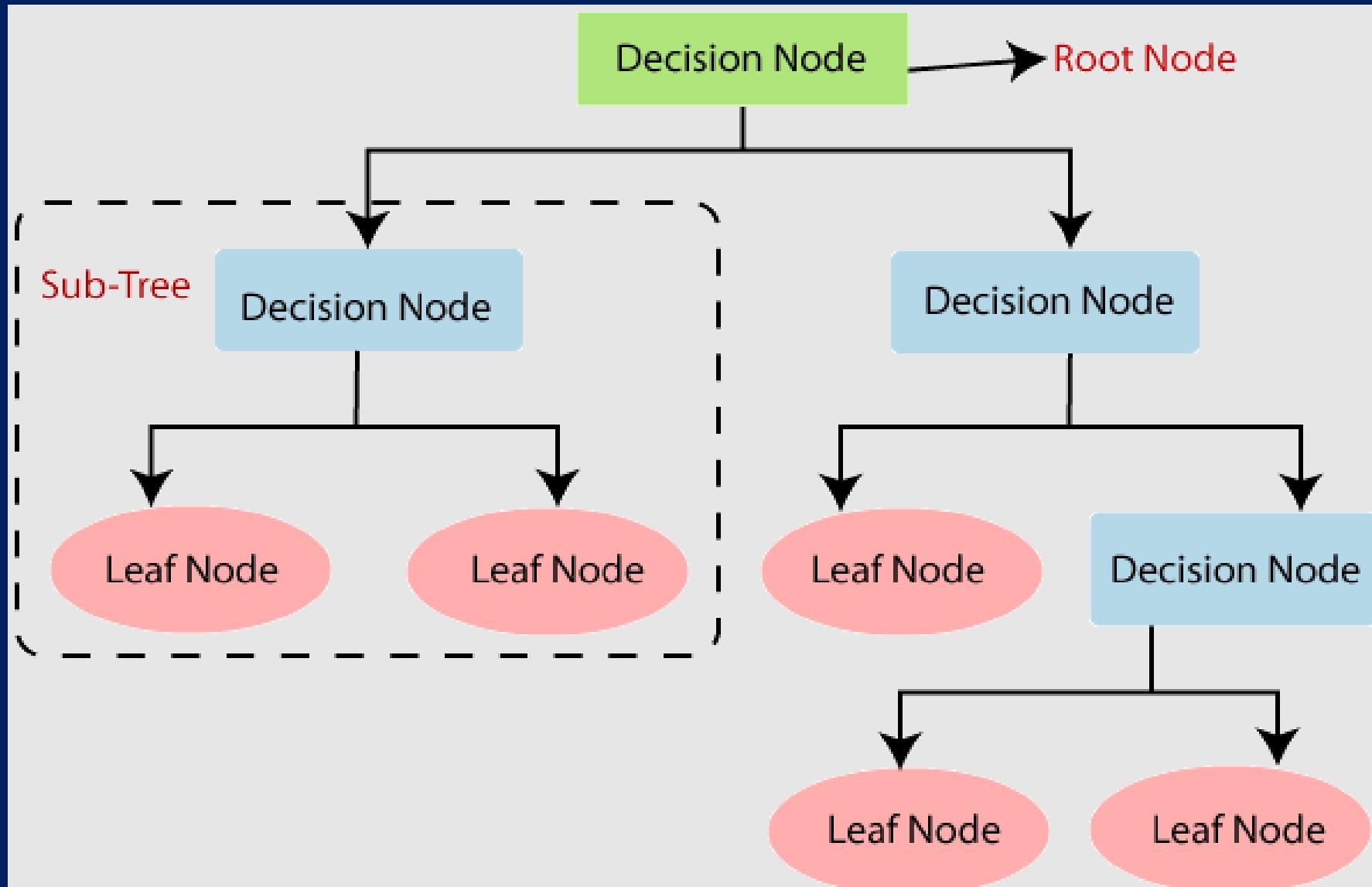
Boosting

- The term “boosting” is used to describe a family of algorithms which are able to convert weak models to strong models.
- Boosting incrementally builds an ensemble by training each model with the same dataset but where the weights of instances are adjusted according to the error of the last prediction.
- Adaboost is a widely known algorithm which is a boosting method.

Learning Decision Trees

- Decision Tree is a Supervised learning technique
- It is a tree-structured classifier, where
 - internal nodes represent the features of a dataset,
 - branches represent the decision rules and
 - each leaf node represents the outcome.
- In a Decision tree, there are two nodes, which are the
 - Decision Node and
 - Leaf Node.
- The decisions or the test are performed on the basis of features of the given dataset.
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
- In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.

Note: A decision tree can contain categorical data (YES/NO) as well as numeric data.



Why use Decision Trees?

- Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.
- The logic behind the decision tree can be easily understood because it shows a tree-like structure.

Decision trees Terminologies

- **Root Node:** Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.
- **Leaf Node:** Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.
- **Splitting:** Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.
- **Branch/Sub Tree:** A tree formed by splitting the tree.
- **Pruning:** Pruning is the process of removing the unwanted branches from the tree.
- **Parent/Child node:** The root node of the tree is called the parent node, and other nodes are called the child nodes.

How does the Decision Tree algorithm Work?

Step-1: Begin the tree with the root node, says S, which contains the complete dataset.

Step-2: Find the best attribute in the dataset using **Attribute Selection Measure (ASM)**.

Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Generate the decision tree node, which contains the best attribute.

Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Attribute Selection Measures

1. Information Gain:

- Information gain is the measurement of changes in entropy after the segmentation of a dataset based on an attribute.
 - It calculates how much information a feature provides us about a class.
 - According to the value of information gain, we split the node and build the decision tree.
 - A decision tree algorithm always tries to maximize the value of information gain, and a node/attribute having the highest information gain is split first. It can be calculated using the below formula:
 - **Information Gain**= Entropy(S)- [(Weighted Avg) *Entropy(each feature)]
 - **Entropy**: Entropy is a metric to measure the impurity in a given attribute. It specifies randomness in data. Entropy can be calculated as:
 - $\text{Entropy}(S) = -P(\text{yes})\log_2 P(\text{yes}) - P(\text{no}) \log_2 P(\text{no})$
- Where,**
- S= Total number of samples**
- P(yes)= probability of yes**
- P(no)= probability of no**

Attribute Selection Measures...

2. Gini Index:

- Gini index is a measure of impurity or purity used while creating a decision tree in the CART(Classification and Regression Tree) algorithm.
- An attribute with the low Gini index should be preferred as compared to the high Gini index.
- It only creates binary splits, and the CART algorithm uses the Gini index to create binary splits.
- Gini index can be calculated using the below formula:

$$\text{Gini Index} = 1 - \sum_j P_j^2$$

Advantages of the Decision Tree

- It is simple to understand as it follows the same process which a human follow while making any decision in real-life.
- It can be very useful for solving decision-related problems.
- It helps to think about all the possible outcomes for a problem.
- There is less requirement of data cleaning compared to other algorithms.

Disadvantages of the Decision Tree

- The decision tree contains lots of layers, which makes it complex.
- It may have an overfitting issue, which can be resolved using the **Random Forest algorithm**.
- For more class labels, the computational complexity of the decision tree may increase.



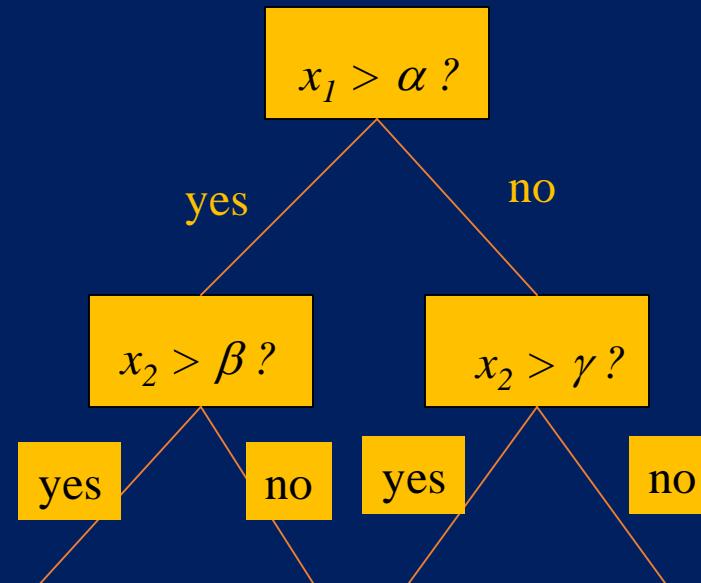
Learning Decision Trees

- let's look at a Classification Problem:

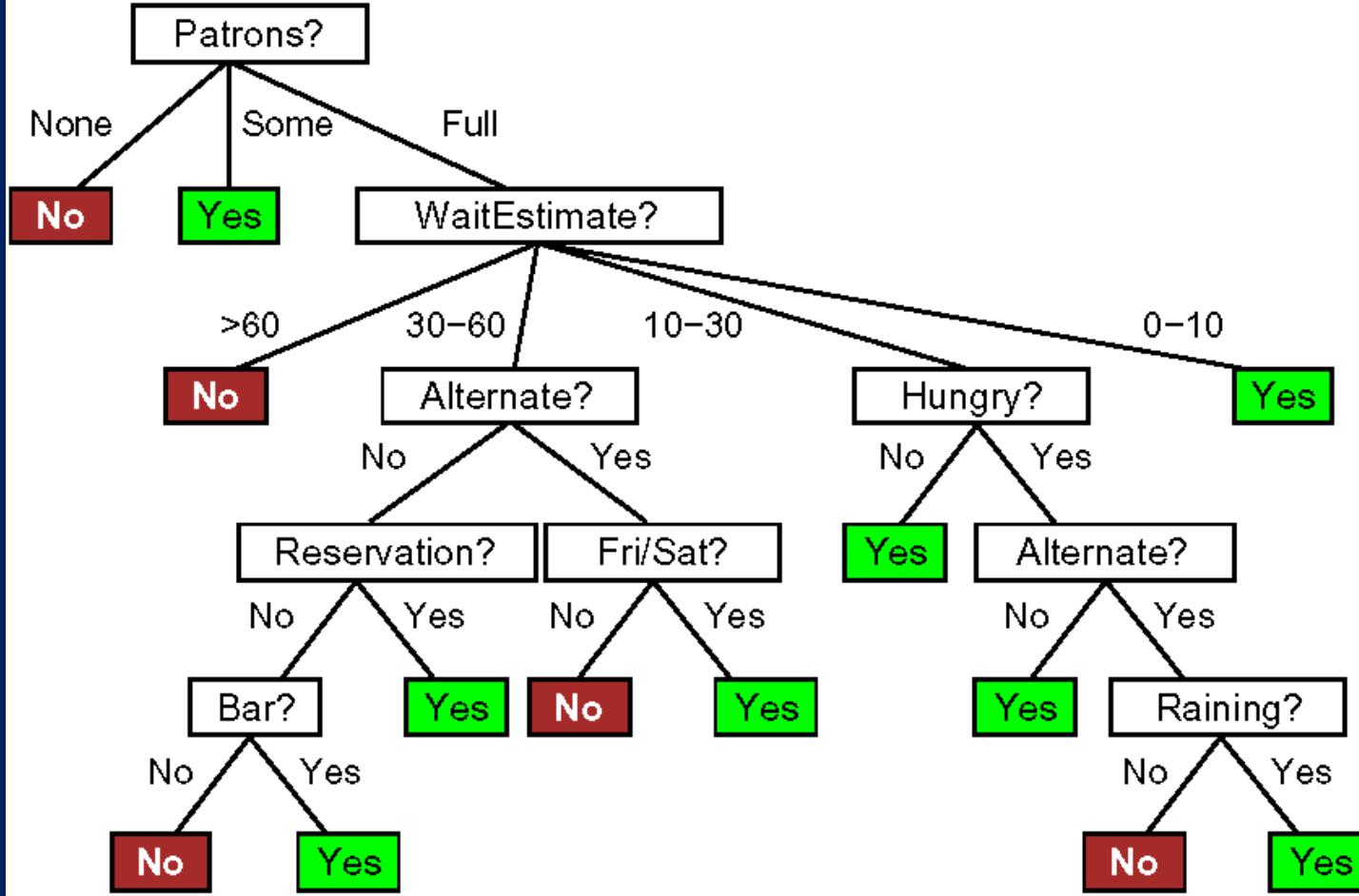
Predicting whether a certain person will choose a particular restaurant.

Method: Decision trees

- “Divide and conquer”: Split data into smaller and smaller subsets.
- Splits usually on a single variable



The wait@restaurant decision tree



This is our true function.
 Can we learn this tree from examples?

Inductive learning of decision tree

- **Simplest:** Construct a decision tree with one leaf for every example = memory based learning.

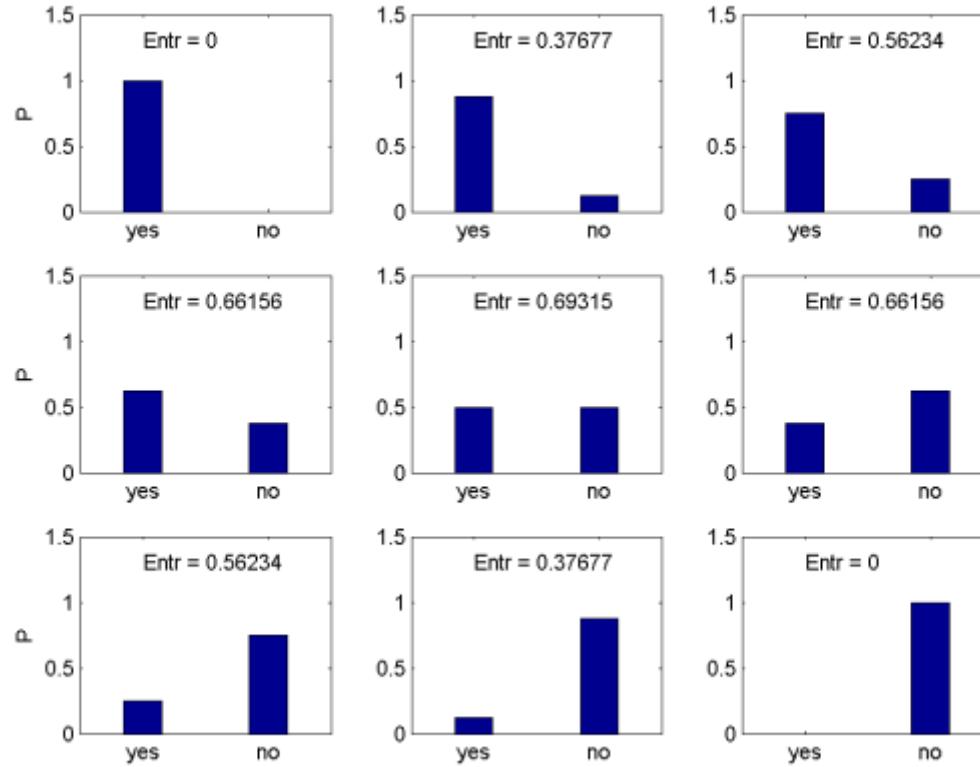
Not very good generalization.

- **Advanced:** Split on each variable so that the purity of each split increases (i.e. either only yes or only no)
- Purity measured, e.g, with entropy



General form:

$$\text{Entropy} = -\sum_i P(v_i) \ln [P(v_i)]$$



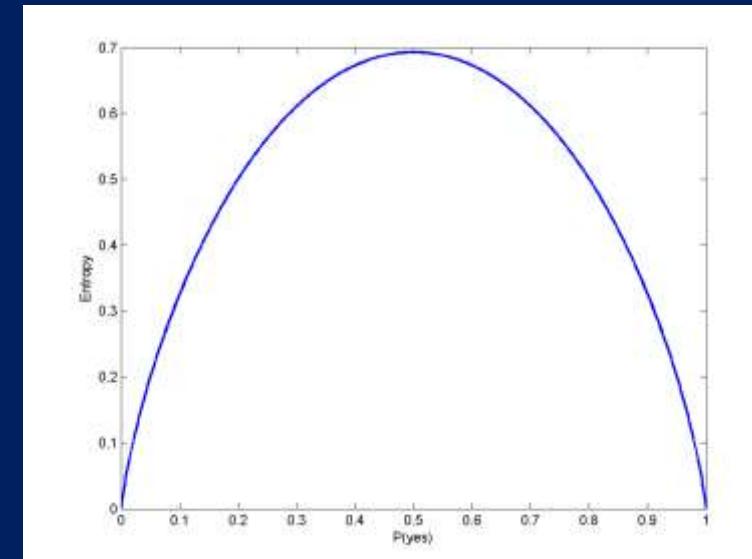
Entropy is a measure of "order" in a system.

The second law of thermodynamics:
Elements in a closed system tend
to seek their most probable distribution;
in a closed system entropy always increases

The entropy is maximal when all possibilities are equally likely.

The goal of the decision tree is to decrease the entropy in each node.

Entropy is zero in a pure "yes" node (or pure "no" node).



Decision tree learning algorithm

- Create pure nodes whenever possible
- If pure nodes are not possible, choose the split that leads to the largest decrease in entropy.

Decision tree learning example

10 attributes:

1. **Alternate:** Is there a suitable alternative restaurant nearby? {yes,no}
2. **Bar:** Is there a bar to wait in? {yes,no}
3. **Fri/Sat:** Is it Friday or Saturday? {yes,no}
4. **Hungry:** Are you hungry? {yes,no}
5. **Patrons:** How many are seated in the restaurant? {none, some, full}
6. **Price:** Price level {\$,\$\$,\$\$\$}
7. **Raining:** Is it raining? {yes,no}
8. **Reservation:** Did you make a reservation? {yes,no}
9. **Type:** Type of food {French,Italian,Thai,Burger}
10. **Wait:** {0-10 min, 10-30 min, 30-60 min, >60 min}

Decision tree learning example

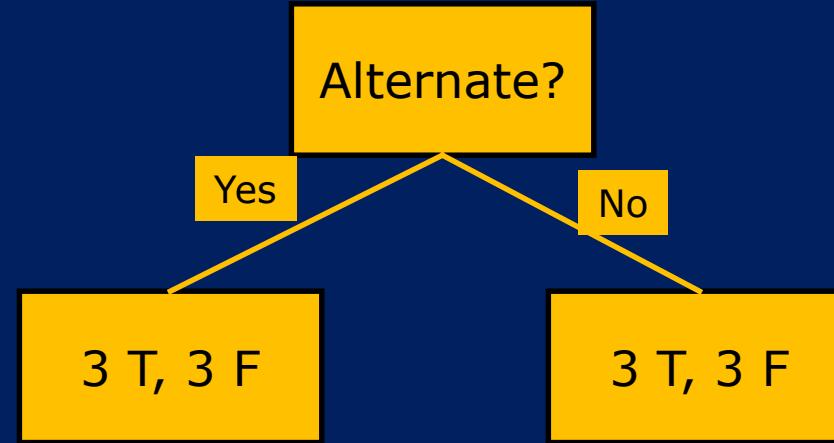
Example	Attributes											Target <i>WillWait</i>
	<i>Alt</i>	<i>Bar</i>	<i>Fri</i>	<i>Hun</i>	<i>Pat</i>	<i>Price</i>	<i>Rain</i>	<i>Res</i>	<i>Type</i>	<i>Est</i>		
X_1	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>Some</i>	<i>\$\$\$</i>	<i>F</i>	<i>T</i>	<i>French</i>	<i>0–10</i>	<i>T</i>	
X_2	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>Full</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Thai</i>	<i>30–60</i>	<i>F</i>	
X_3	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>Some</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Burger</i>	<i>0–10</i>	<i>T</i>	
X_4	<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>Full</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Thai</i>	<i>10–30</i>	<i>T</i>	
X_5	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>Full</i>	<i>\$\$\$</i>	<i>F</i>	<i>T</i>	<i>French</i>	<i>>60</i>	<i>F</i>	
X_6	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>Some</i>	<i>\$\$</i>	<i>T</i>	<i>T</i>	<i>Italian</i>	<i>0–10</i>	<i>T</i>	
X_7	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>None</i>	<i>\$</i>	<i>T</i>	<i>F</i>	<i>Burger</i>	<i>0–10</i>	<i>F</i>	
X_8	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>Some</i>	<i>\$\$</i>	<i>T</i>	<i>T</i>	<i>Thai</i>	<i>0–10</i>	<i>T</i>	
X_9	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>Full</i>	<i>\$</i>	<i>T</i>	<i>F</i>	<i>Burger</i>	<i>>60</i>	<i>F</i>	
X_{10}	<i>T</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>Full</i>	<i>\$\$\$</i>	<i>F</i>	<i>T</i>	<i>Italian</i>	<i>10–30</i>	<i>F</i>	
X_{11}	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>None</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Thai</i>	<i>0–10</i>	<i>F</i>	
X_{12}	<i>T</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>Full</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Burger</i>	<i>30–60</i>	<i>T</i>	

T = True,
F = False



6 True,
 6 False

Decision tree learning example

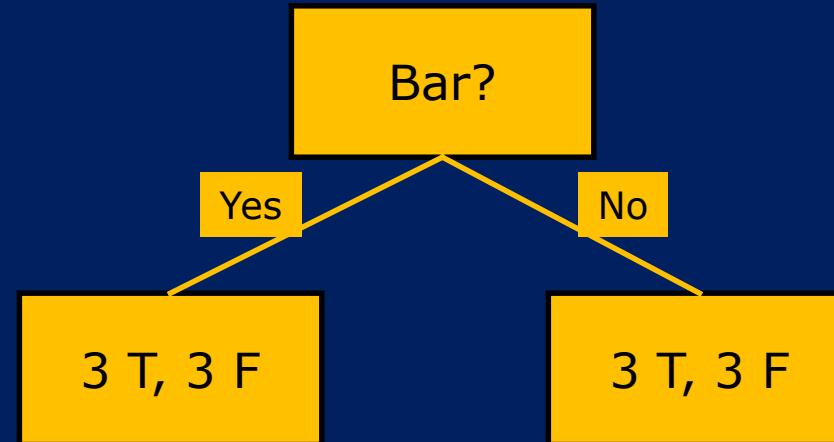


Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T



Entropy decrease = 0.30 – 0.30 = 0

Decision tree learning example

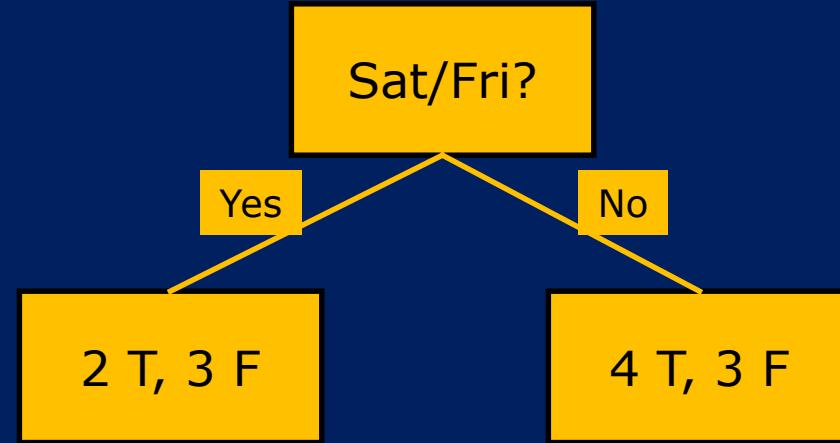


Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0–10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30–60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0–10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10–30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0–10	T
X ₇	F	T	F	F	None	\$	T	F	Burger	0–10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0–10	T
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10–30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0–10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30–60	T



Entropy decrease = 0.30 – 0.30 = 0

Decision tree learning example

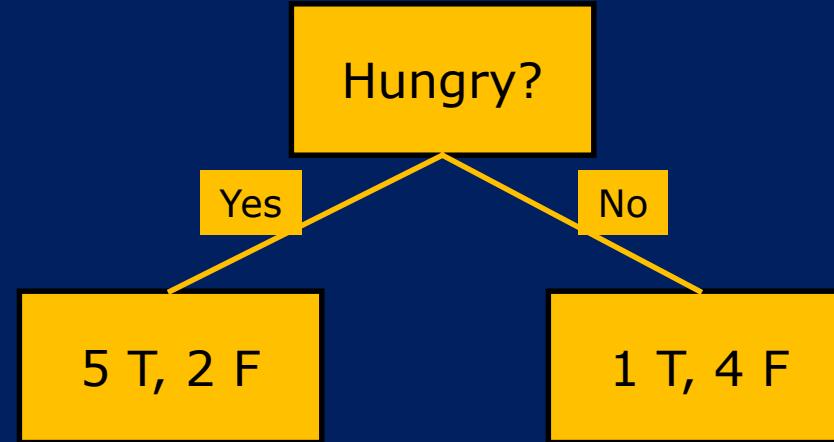


Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T



Entropy decrease = $0.30 - 0.29 = 0.01$

Decision tree learning example

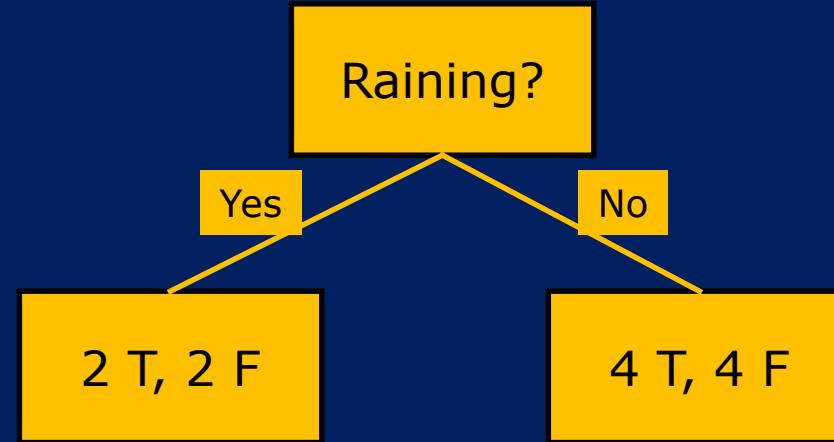


Example	Attributes											Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est		
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T	
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F	
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T	
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T	
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F	
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T	
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F	
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T	
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F	
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F	
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F	
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T	



Entropy decrease = $0.30 - 0.24 = 0.06$

Decision tree learning example

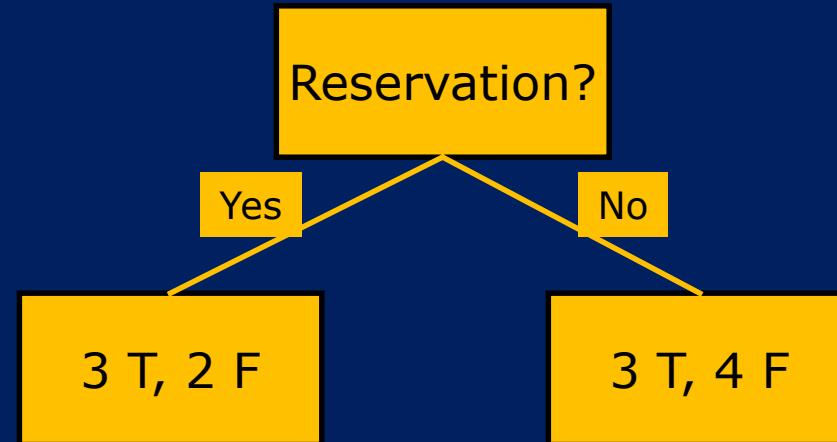


Example	Attributes									Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60



Entropy decrease = 0.30 – 0.30 = 0

Decision tree learning example

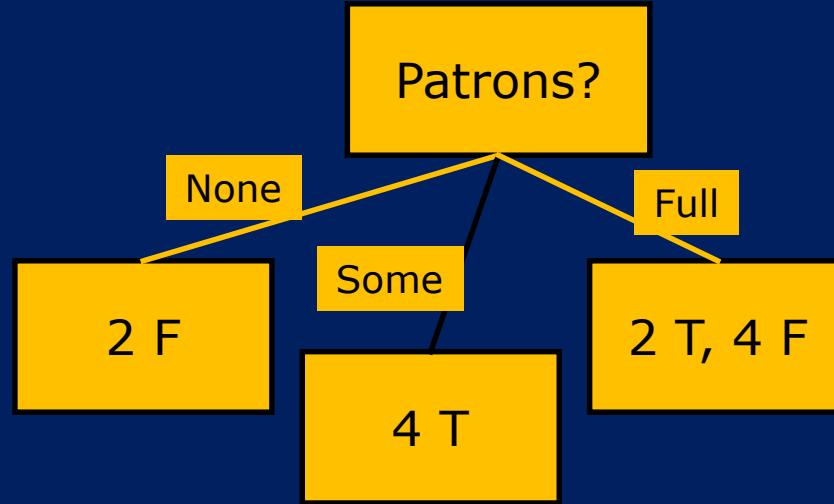


Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T

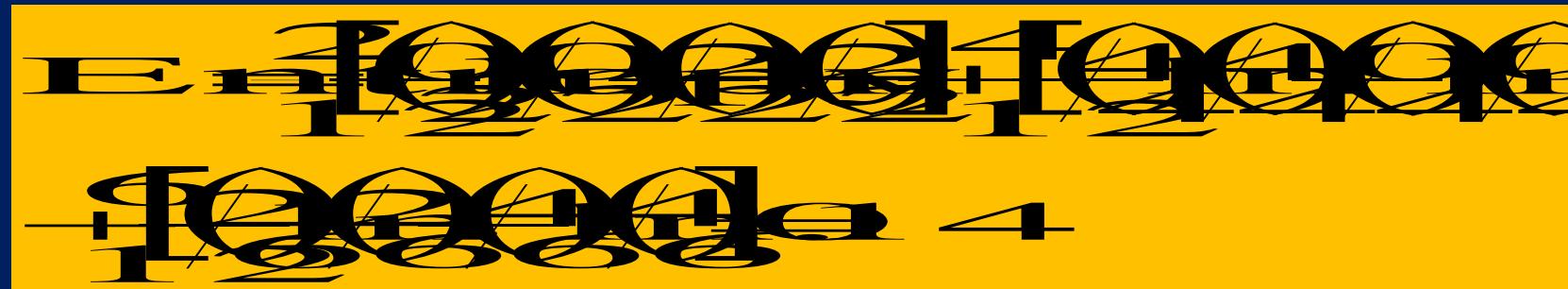


Entropy decrease = 0.30 - 0.29 = 0.01

Decision tree learning example

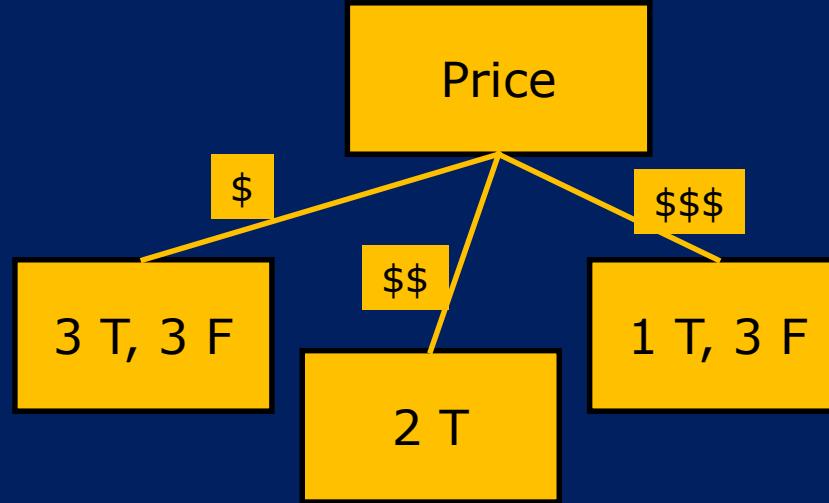


Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T

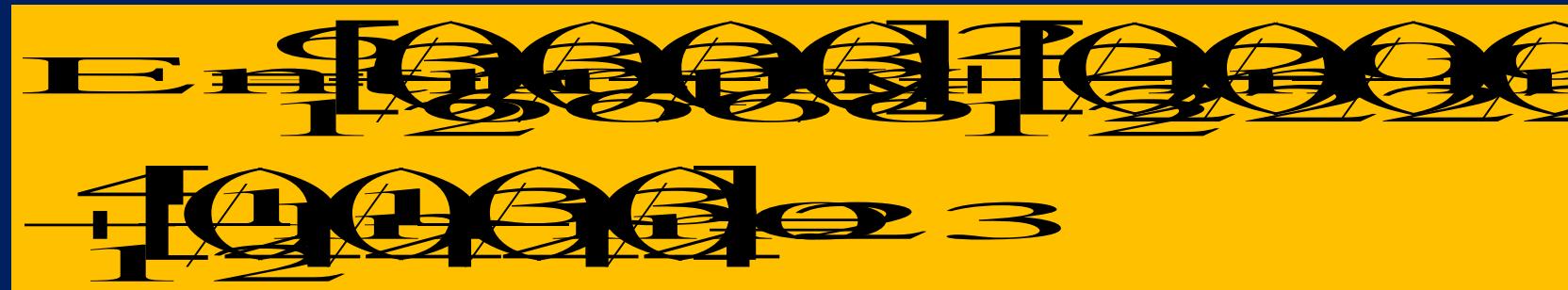


Entropy decrease = $0.30 - 0.14 = 0.16$

Decision tree learning example

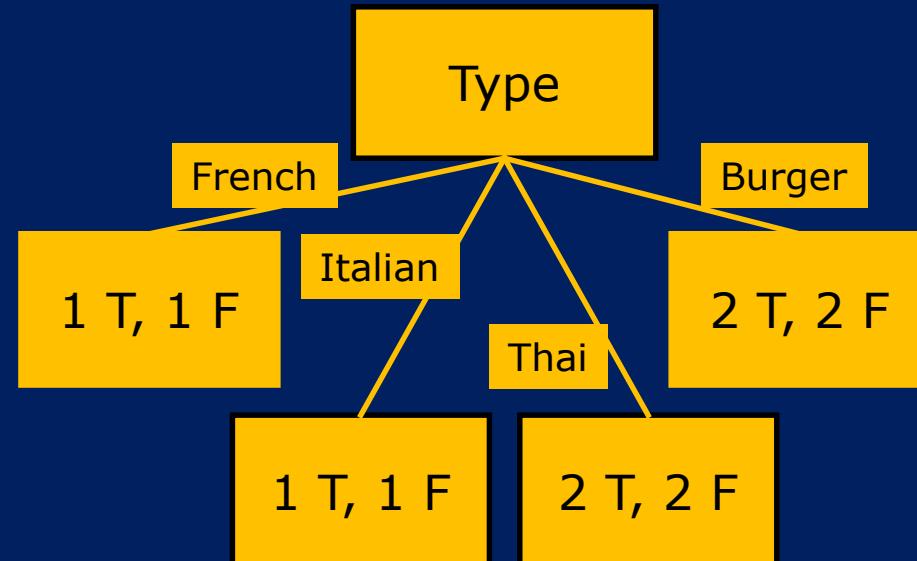


Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T



Entropy decrease = 0.30 - 0.23 = 0.07

Decision tree learning example

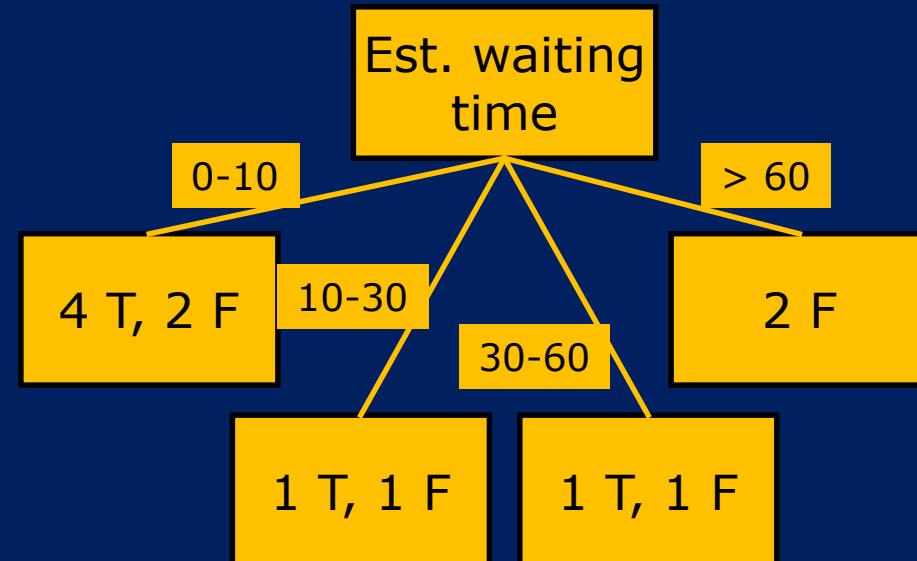


Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T



Entropy decrease = 0.30 - 0.30 = 0

Decision tree learning example

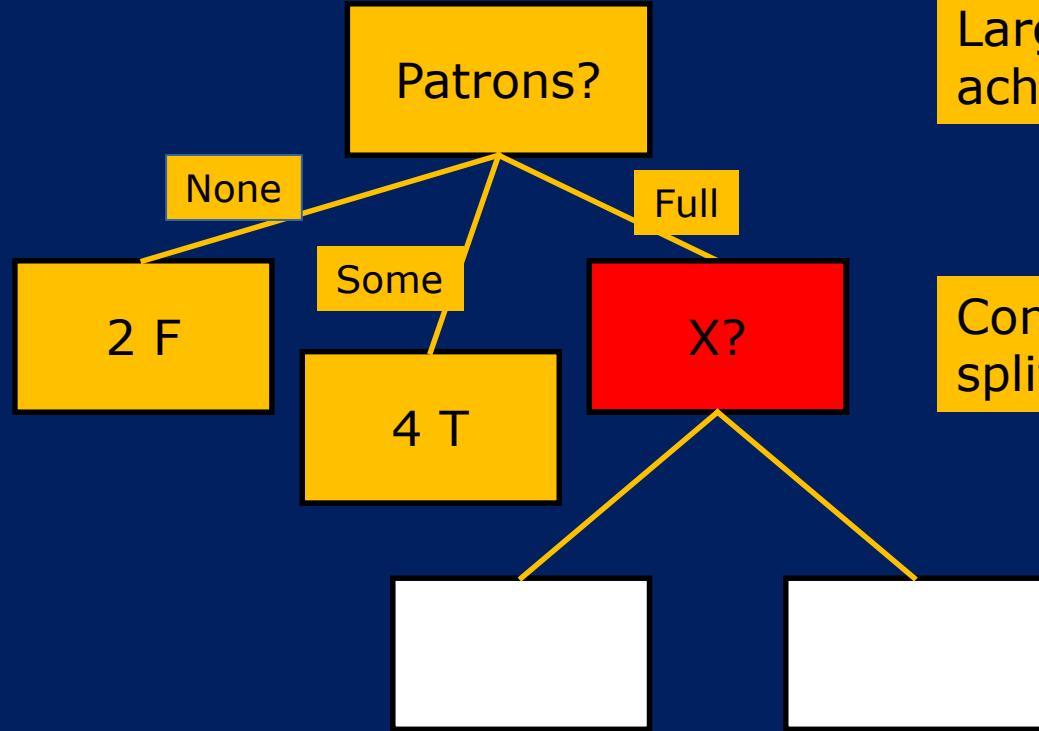


Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T



Entropy decrease = $0.30 - 0.24 = 0.06$

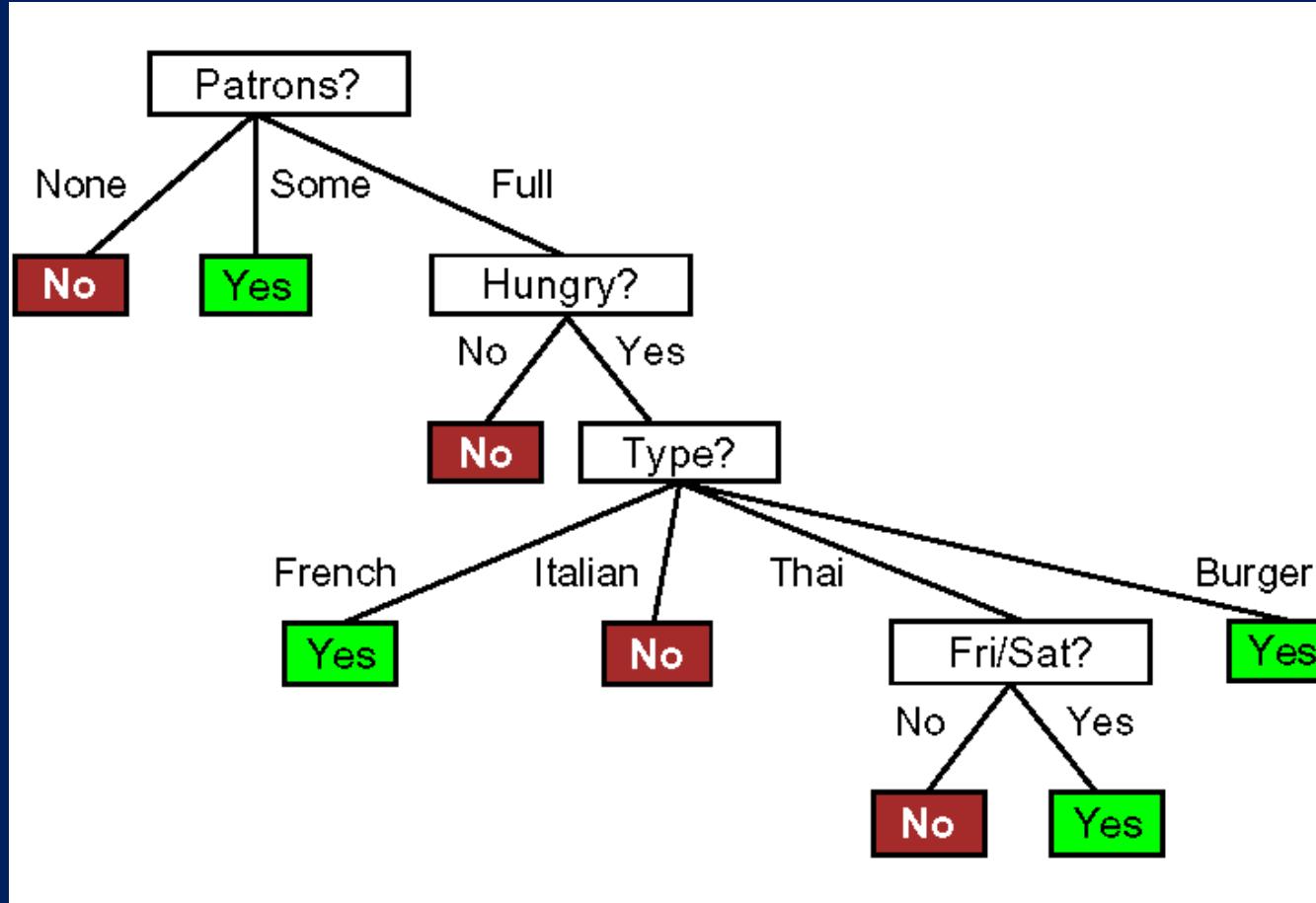
Decision tree learning example



Largest entropy decrease (0.16) achieved by splitting on Patrons.

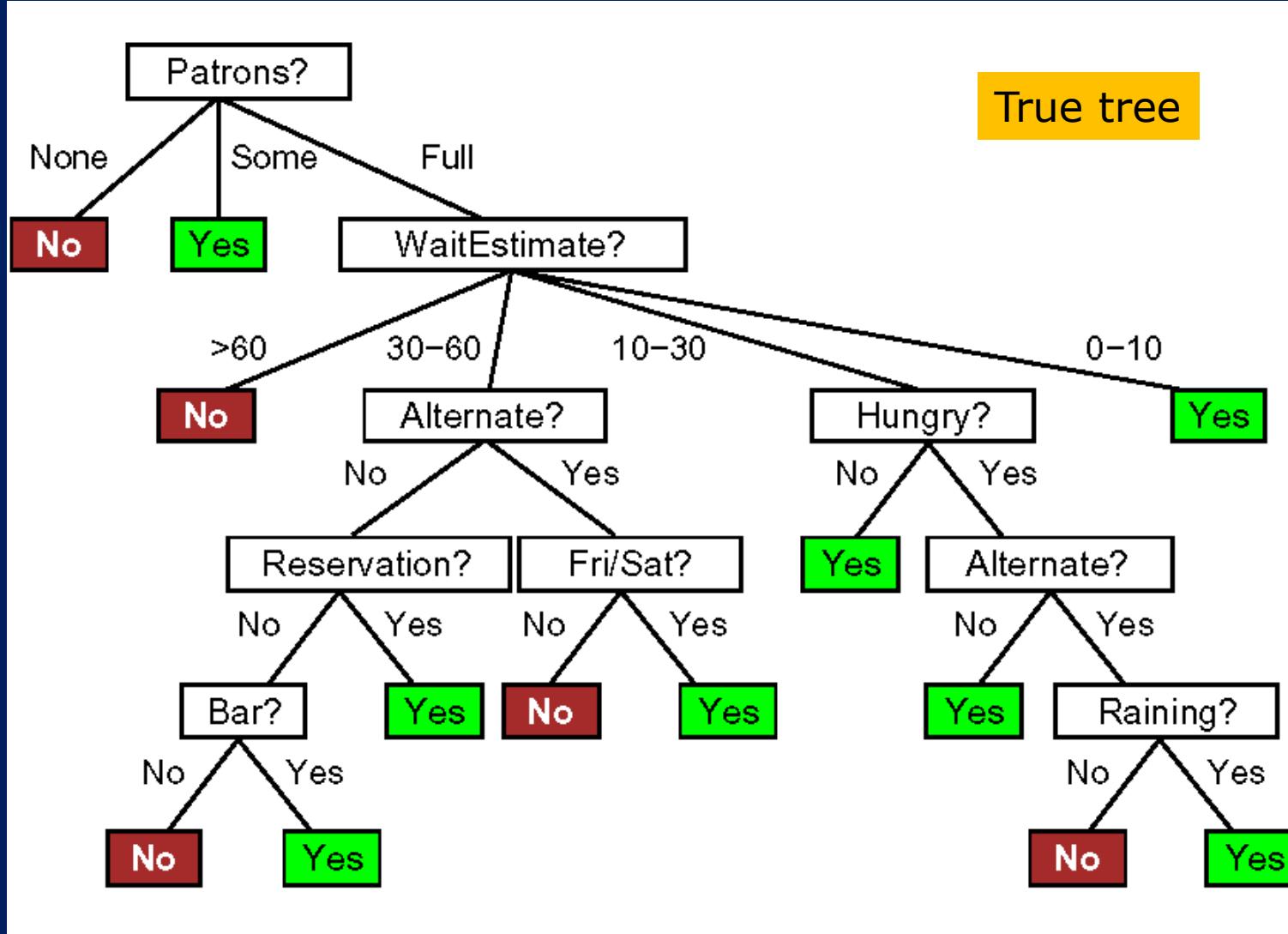
Continue like this, making new splits, always purifying nodes.

Decision tree learning example

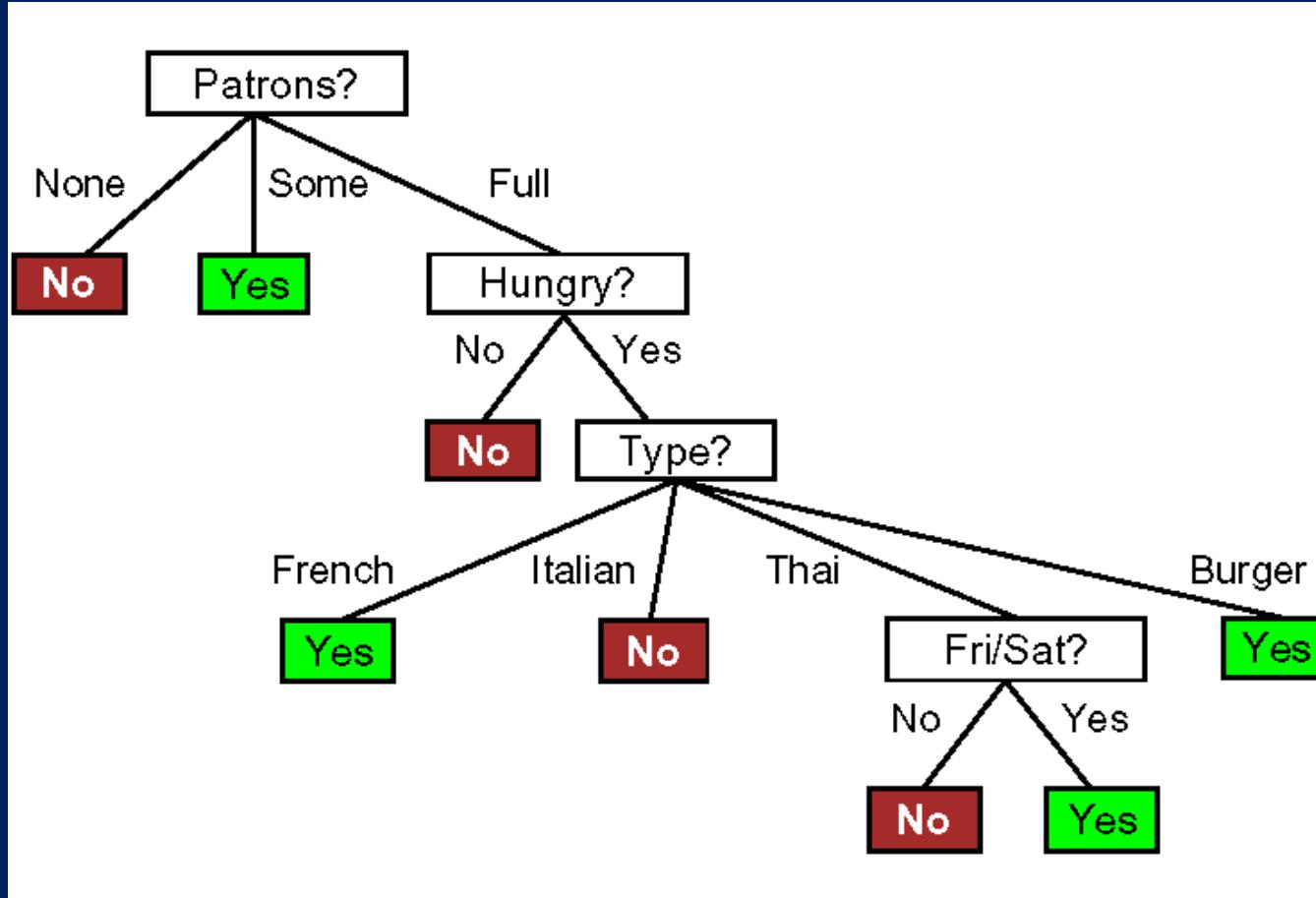


Induced tree (from examples)

Decision tree learning example



Decision tree learning example



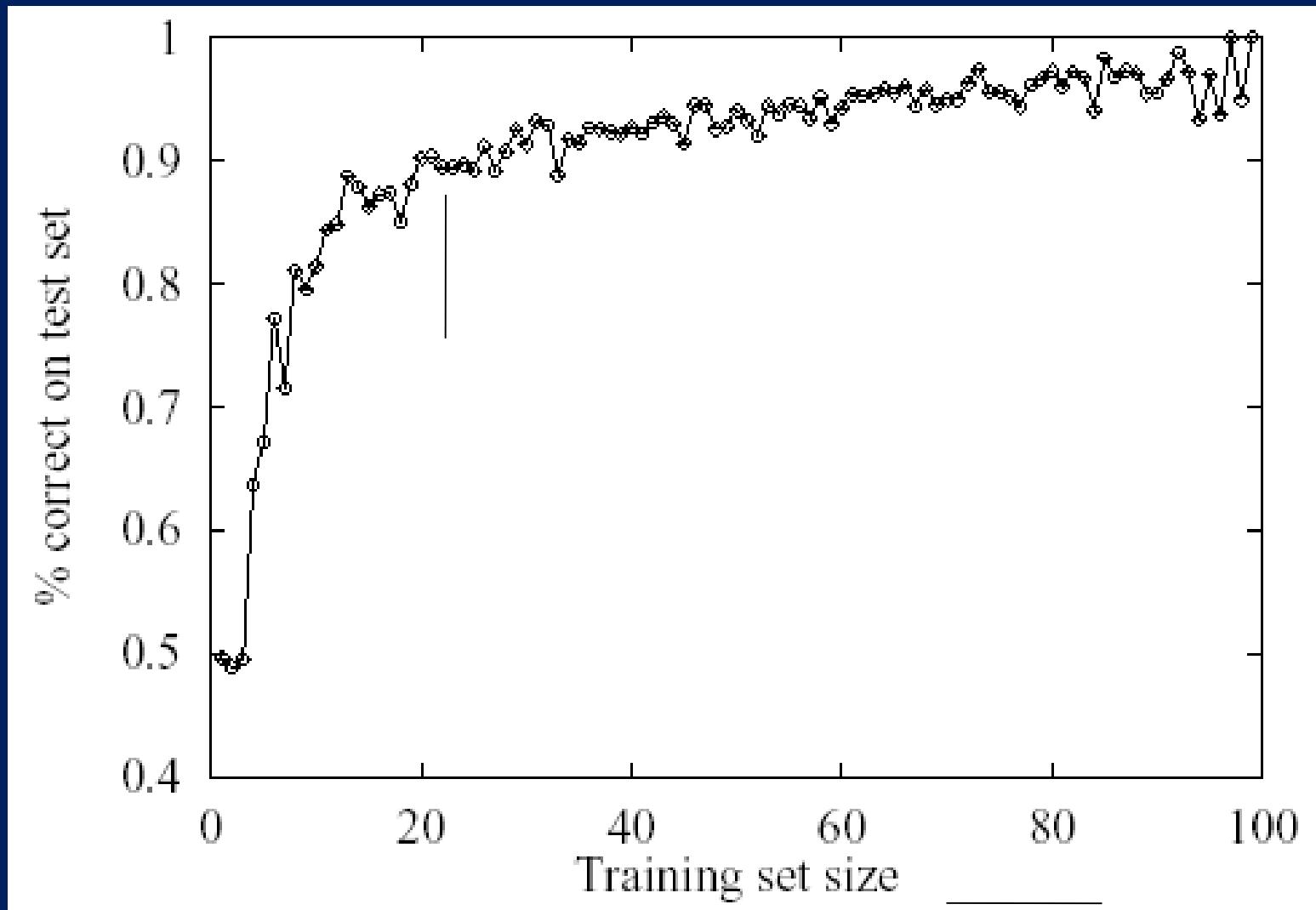
Induced tree (from examples)

Cannot make it more complex than what the data supports.



How do we know it is correct?

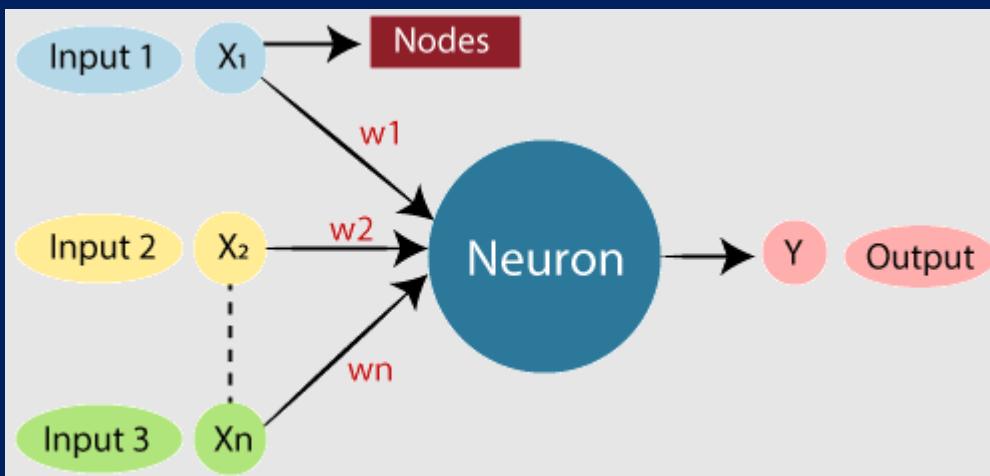
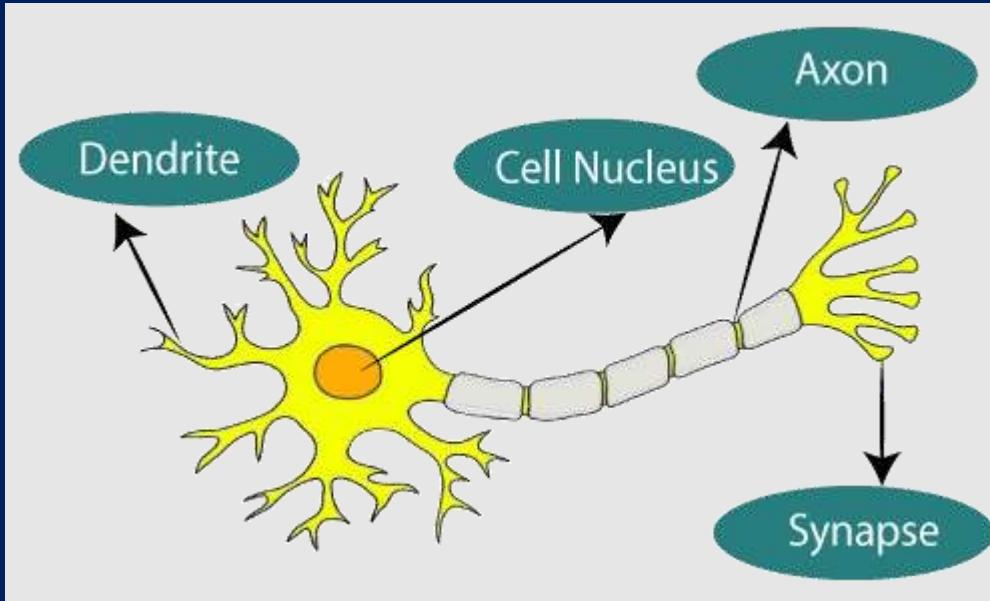
- How do we know that $h \approx f$?
(Hume's Problem of Induction)
 - Try h on a new **test set** of examples
(cross validation)
- ...and assume the "principle of uniformity", i.e. the result we get on this test data should be indicative of results on future data. Causality is constant.



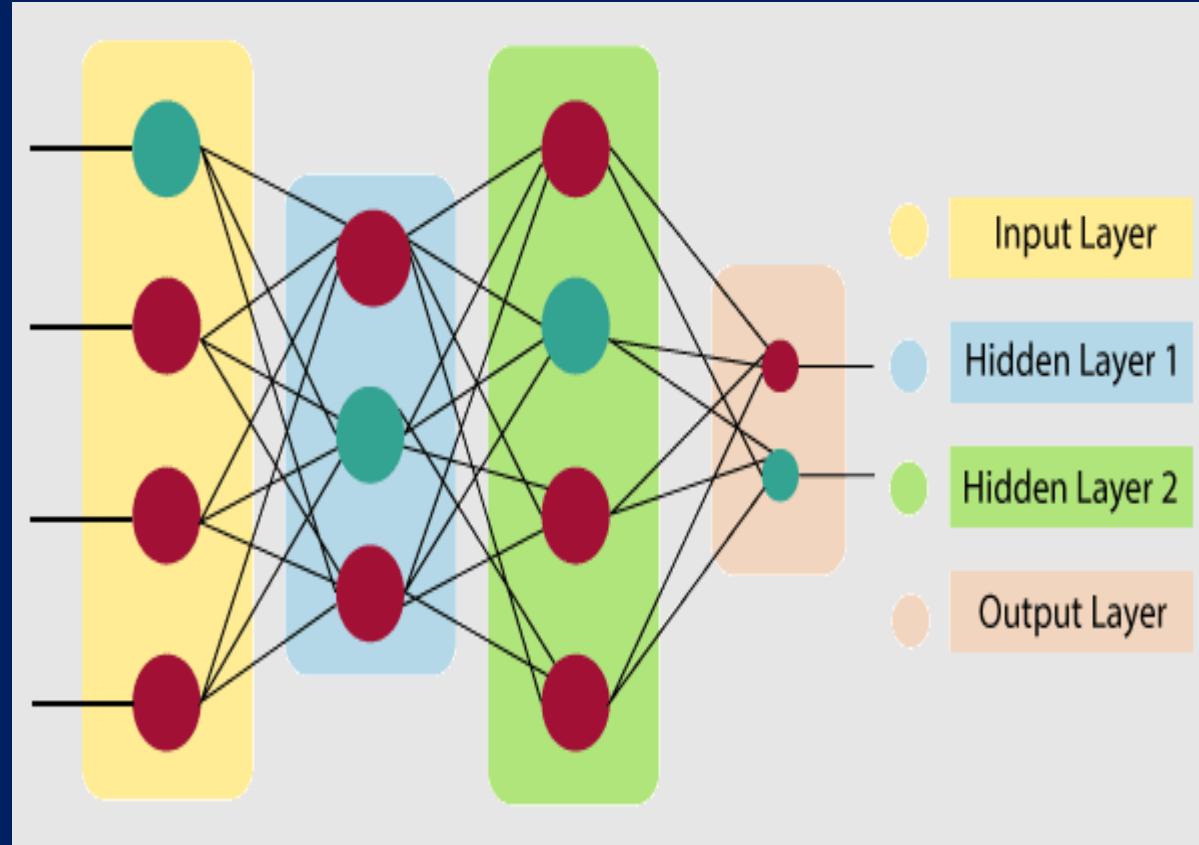
Learning curve for the decision tree algorithm on 100 randomly generated examples in the restaurant domain.
The graph summarizes 20 trials.

Artificial Neural Networks

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



The architecture of an artificial neural network:



Advantages & Disadvantages of ANN

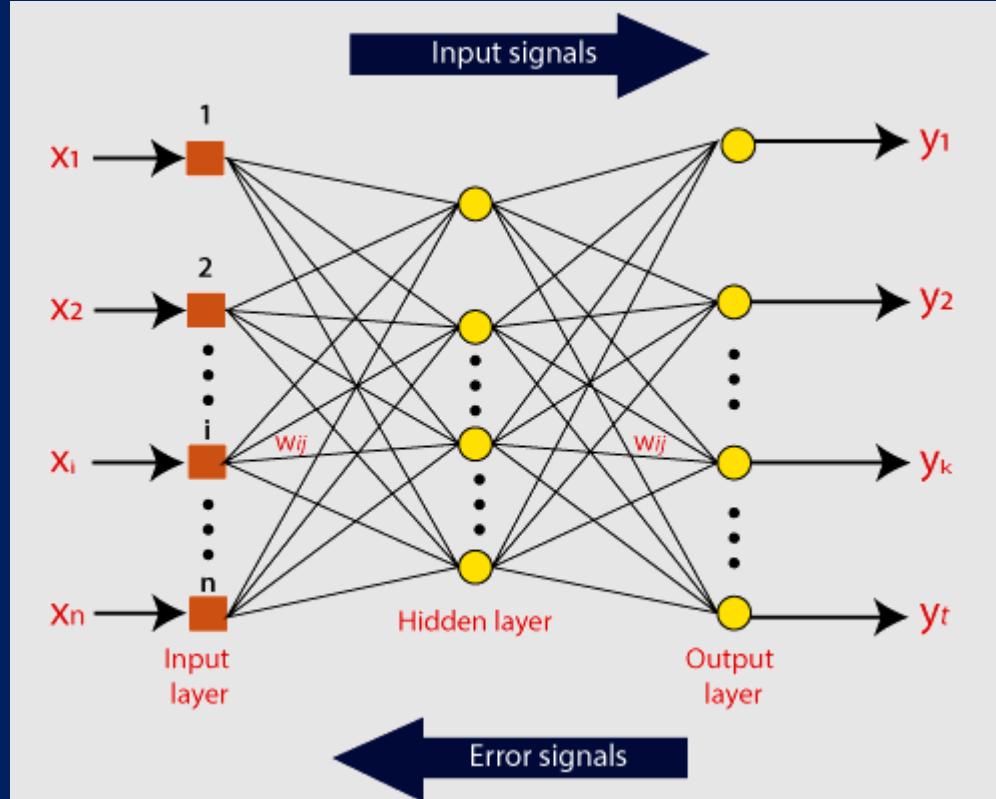
Advantages

- Parallel processing capability
- Storing data on the entire network
- Capability to work with incomplete knowledge
- Having a memory distribution
- Having fault tolerance

Disadvantages

- Assurance of proper network structure
- Unrecognized behavior of the network
- Hardware dependence
- Difficulty of showing the issue to the network
- The duration of the network is unknown

How do artificial neural networks work?



Types of Artificial Neural Network

1. Feedback ANN

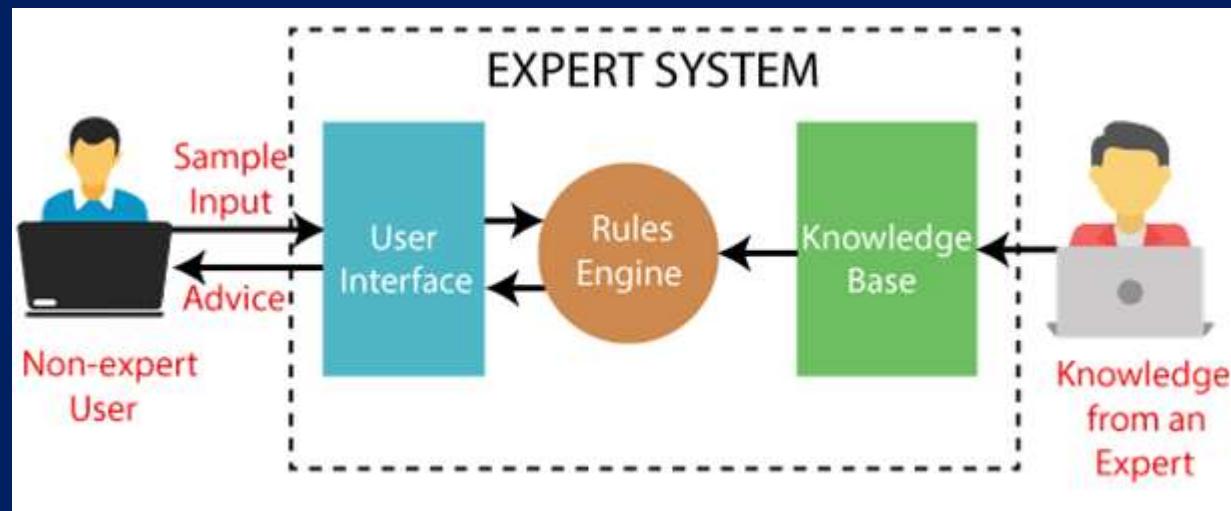
- In this type of ANN, the output returns into the network to accomplish the best-evolved results internally.
- The feedback networks feed information back into itself and are well suited to solve optimization issues.
- The Internal system error corrections utilize feedback ANNs.

2. Feed-Forward ANN

- A feed-forward network is a basic neural network comprising of an input layer, an output layer, and at least one layer of a neuron.
- Through assessment of its output by reviewing its input, the intensity of the network can be noticed based on group behavior of the associated neurons, and the output is decided.
- The primary advantage of this network is that it figures out how to evaluate and recognize input patterns.

What is an Expert System?

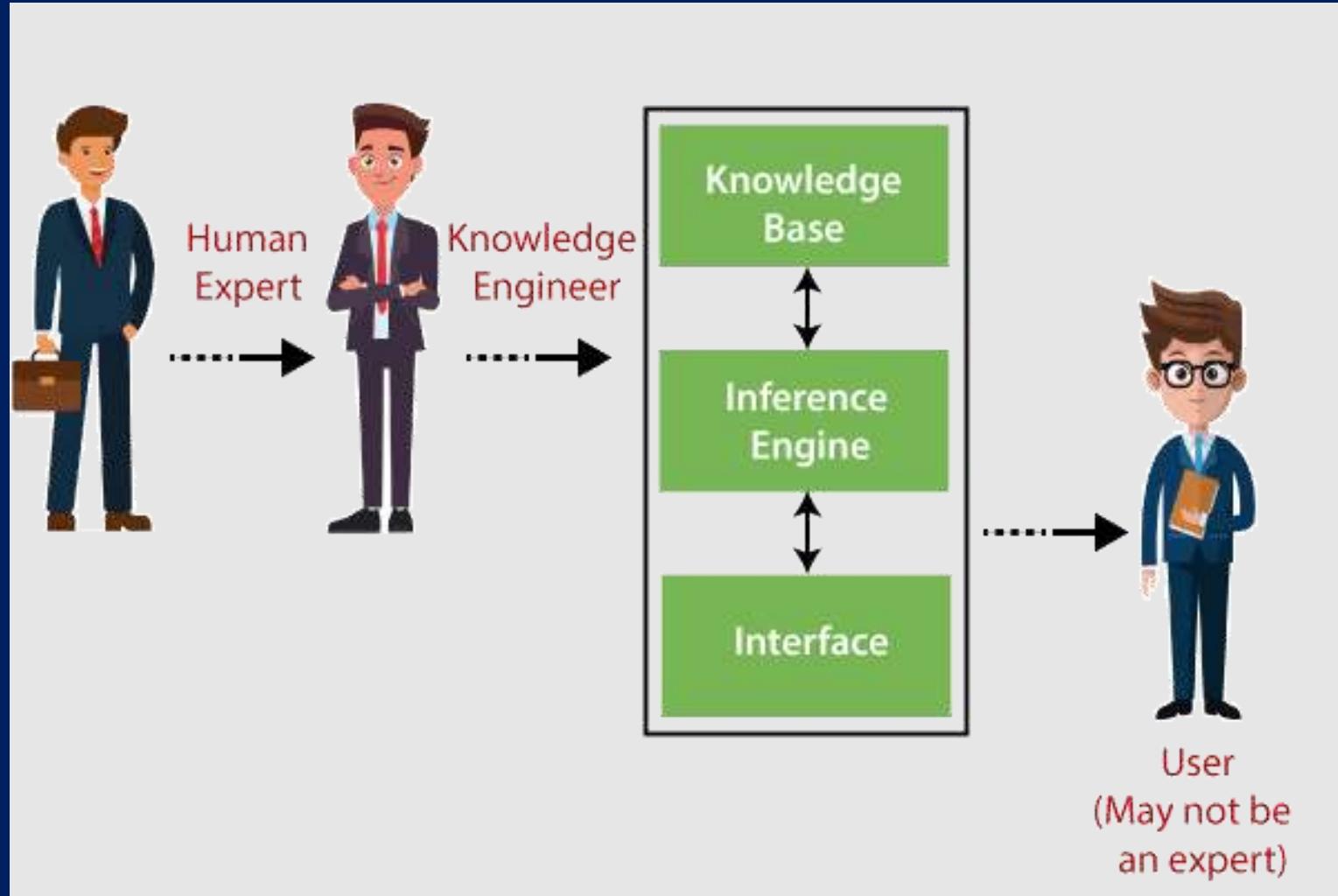
- An expert system is a computer program that is designed to solve complex problems and to provide decision-making ability like a human expert.
- It performs this by extracting knowledge from its knowledge base using the reasoning and inference rules according to the user queries.



Characteristics of Expert System

- **High Performance:** The expert system provides high performance for solving any type of complex problem of a specific domain with high efficiency and accuracy.
- **Understandable:** It responds in a way that can be easily understandable by the user. It can take input in human language and provides the output in the same way.
- **Reliable:** It is much reliable for generating an efficient and accurate output.
- **Highly responsive:** ES provides the result for any complex query within a very short period of time.

Components of Expert System



Capabilities of the Expert System

- **Advising:** It is capable of advising the human being for the query of any domain from the particular ES.
- **Provide decision-making capabilities:** It provides the capability of decision making in any domain, such as for making any financial decision, decisions in medical science, etc.
- **Demonstrate a device:** It is capable of demonstrating any new products such as its features, specifications, how to use that product, etc.
- **Problem-solving:** It has problem-solving capabilities.
- **Explaining a problem:** It is also capable of providing a detailed description of an input problem.
- **Interpreting the input:** It is capable of interpreting the input given by the user.
- **Predicting results:** It can be used for the prediction of a result.
- **Diagnosis:** An ES designed for the medical field is capable of diagnosing a disease without using multiple components as it already contains various inbuilt medical tools.

Why Expert System?

Why Expert System



No emotion

High Efficiency

Expertise in a domain

No Memory limitation

Regular updates improve the performance

High Security

Considers all facts

Advantages of Expert System

- These systems are highly reproducible.
- They can be used for risky places where the human presence is not safe.
- Error possibilities are less if the KB contains correct knowledge.
- The performance of these systems remains steady as it is not affected by emotions, tension, or fatigue.
- They provide a very high speed to respond to a particular query.

Limitations of Expert System

- The response of the expert system may get wrong if the knowledge base contains the wrong information.
- Like a human being, it cannot produce a creative output for different scenarios.
- Its maintenance and development costs are very high.
- Knowledge acquisition for designing is much difficult.
- For each domain, we require a specific ES, which is one of the big limitations.
- It cannot learn from itself and hence requires manual updates

Applications of Expert System

- **In designing and manufacturing domain**

It can be broadly used for designing and manufacturing physical devices such as camera lenses and automobiles.

- **In the knowledge domain**

These systems are primarily used for publishing the relevant knowledge to the users. The two popular ES used for this domain is an advisor and a tax advisor.

- **In the finance domain**

In the finance industries, it is used to detect any type of possible fraud, suspicious activity, and advise bankers that if they should provide loans for business or not.

- **In the diagnosis and troubleshooting of devices**

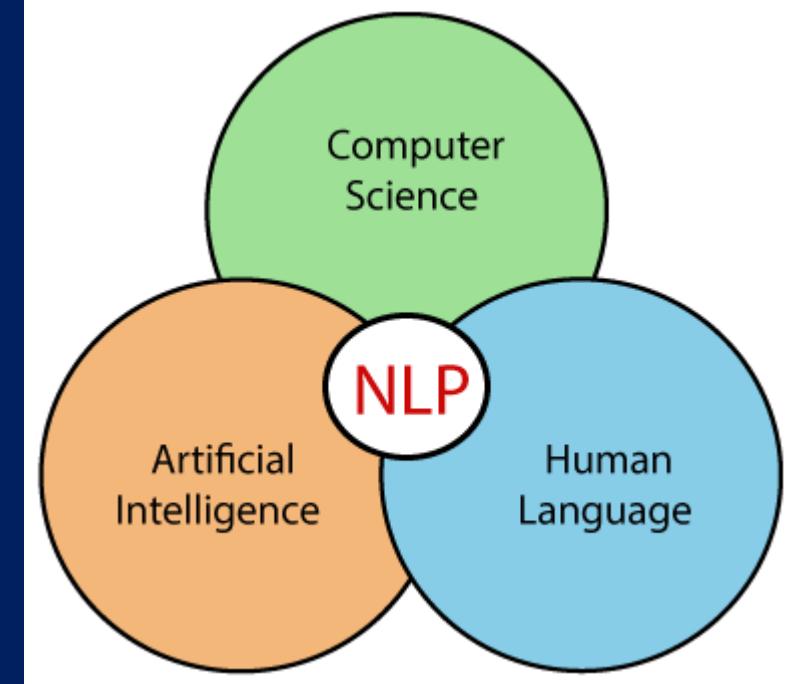
In medical diagnosis, the ES system is used, and it was the first area where these systems were used.

- **Planning and Scheduling**

The expert systems can also be used for planning and scheduling some particular tasks for achieving the goal of that task.

Natural Language Processing

- NLP stands for **Natural Language Processing**, which is a part of **Computer Science**, **Human language**, and **Artificial Intelligence**.
- It is the technology that is used by machines to understand, analyse, manipulate, and interpret human's languages.
- It helps developers to organize knowledge for performing tasks such as
 - **Translation**
 - **Automatic summarization**
 - **Named Entity Recognition (NER)**
 - **Speech recognition**
 - **Relationship extraction** and
 - **Topic segmentation**.



Components of NLP

There are the following two components of NLP –

1. Natural Language Understanding (NLU)

- Natural Language Understanding (NLU) helps the machine to understand and analyse human language by extracting the metadata from content such as concepts, entities, keywords, emotion, relations, and semantic roles.
- NLU mainly used in Business applications to understand the customer's problem in both spoken and written language.

NLU involves the following tasks -

- It is used to map the given input into useful representation.
- It is used to analyze different aspects of the language.

Components of NLP...

2. Natural Language Generation (NLG)

- Natural Language Generation (NLG) acts as a translator that converts the computerized data into natural language representation. It mainly involves Text planning, Sentence planning, and Text Realization.

Note: The NLU is difficult than NLG

Difference between NLU and NLG

NLU	NLG
NLU is the process of reading and interpreting language.	NLG is the process of writing or generating language.
It produces non-linguistic outputs from natural language inputs.	It produces constructing natural language outputs from non-linguistic inputs.

Phases of NLP

1. Lexical Analysis and Morphological

- This phase scans the source code as a stream of characters and converts it into meaningful lexemes. It divides the whole text into paragraphs, sentences, and words.

2. Syntactic Analysis (Parsing)

- Syntactic Analysis is used to check grammar, word arrangements, and shows the relationship among the words.

3. Semantic Analysis

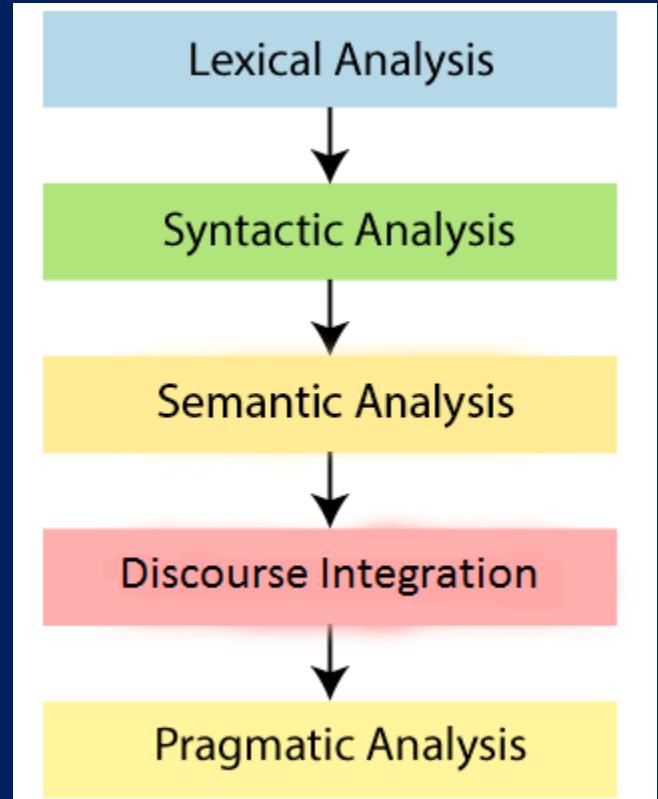
- Semantic analysis is concerned with the meaning representation. It mainly focuses on the literal meaning of words, phrases, and sentences.

4. Discourse Integration

- Discourse Integration depends upon the sentences that precede it and also invokes the meaning of the sentences that follow it.

5. Pragmatic Analysis

- Pragmatic is the fifth and last phase of NLP. It helps you to discover the intended effect by applying a set of rules that characterize cooperative dialogues.



How to build an NLP pipeline

Step1: Sentence Segmentation

Step2: Word Tokenization

Step3: Stemming

Step 4: Lemmatization

Step 5: Identifying Stop Words

Step 6: Dependency Parsing

Step 7: POS (Parts of Speech) tags

Step 8: Named Entity Recognition (NER)

Step 9: Chunking

Advantages of NLP

- NLP helps users to ask questions about any subject and get a direct response within seconds.
- NLP offers exact answers to the question means it does not offer unnecessary and unwanted information.
- NLP helps computers to communicate with humans in their languages.
- It is very time efficient.
- Most of the companies use NLP to improve the efficiency of documentation processes, accuracy of documentation, and identify the information from large databases.

Disadvantages of NLP

A list of disadvantages of NLP is given below:

- NLP may not show context.
- NLP is unpredictable
- NLP may require more keystrokes.
- NLP is unable to adapt to the new domain, and it has a limited function that's why NLP is built for a single and specific task only.

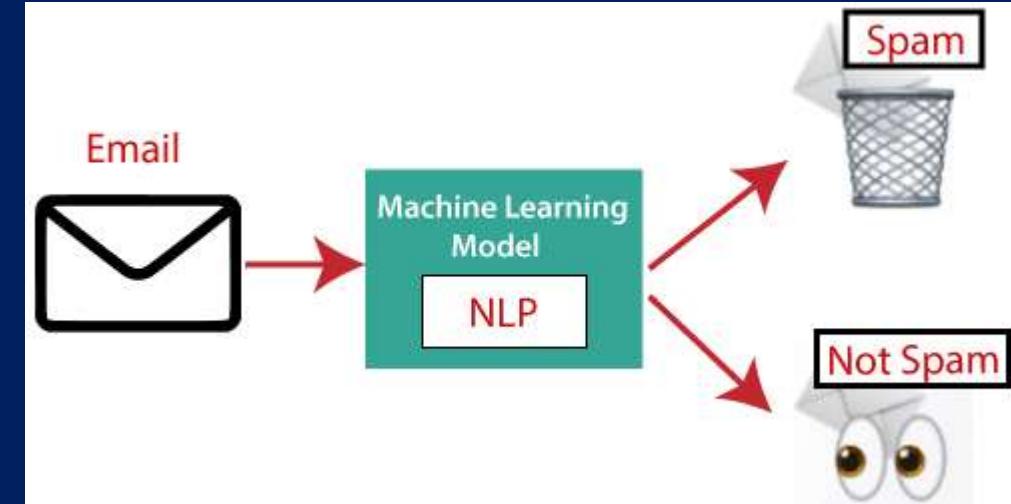
Applications of NLP

There are the following applications of NLP

1. Question Answering



2. Spam Detection

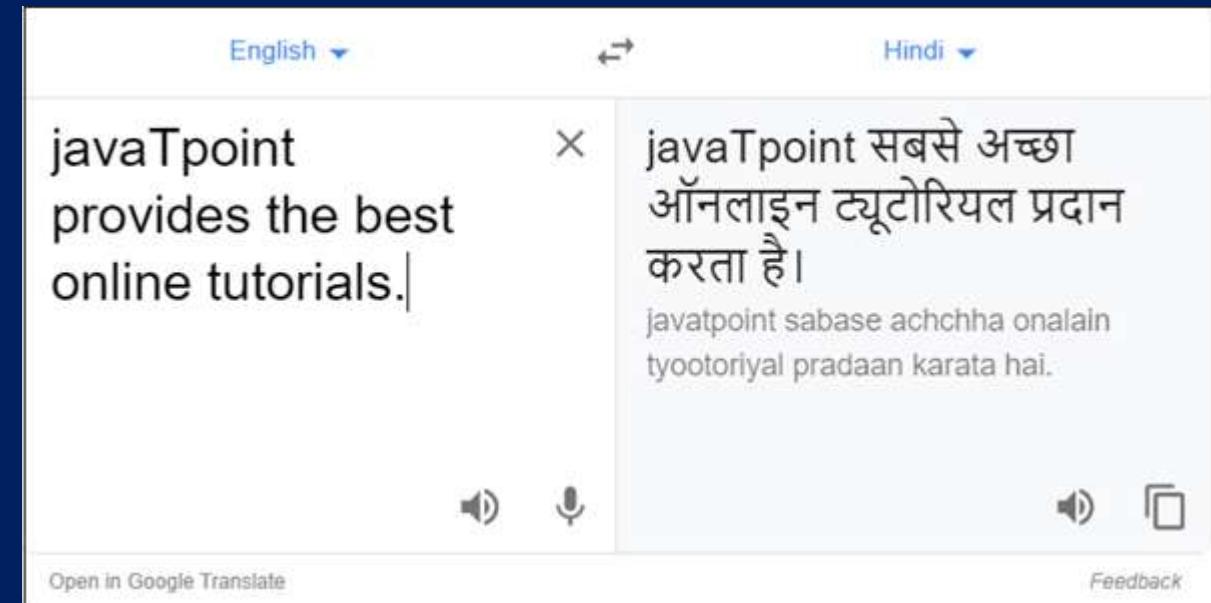


Applications of NLP...

3. Sentiment Analysis



4. Machine translation

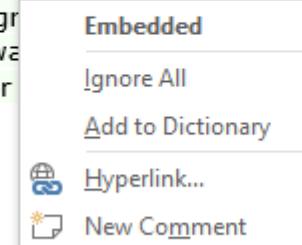


The screenshot shows a Google Translate interface. On the left, under "English", the text "javaTpoint provides the best online tutorials." is displayed. On the right, under "Hindi", the text "javaTpoint सबसे अच्छा ऑनलाइन ट्यूटोरियल प्रदान करता है।" is shown. Below the Hindi text, a smaller box contains the transliterated text "javatpoint sabase achchha onalain tyootoriyal pradaan karata hai." At the bottom of the interface, there are "Open in Google Translate" and "Feedback" buttons, along with audio and microphone icons.

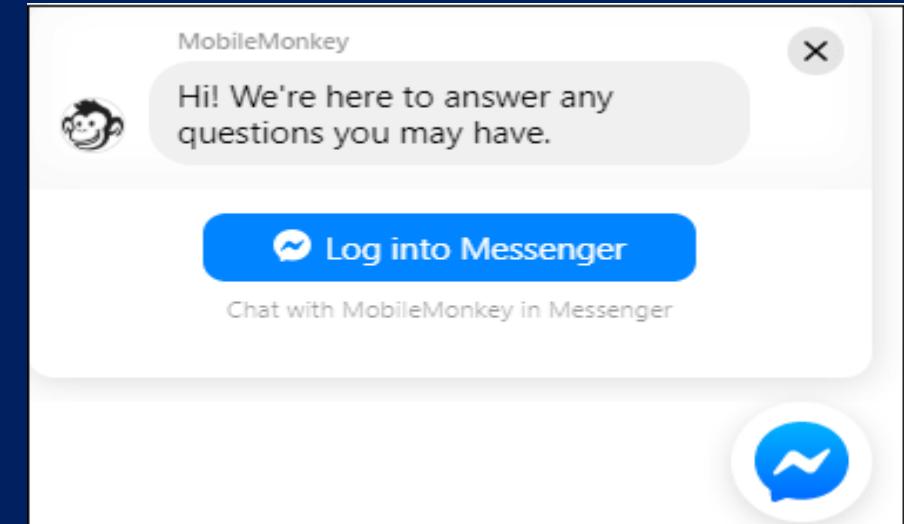
Applications of NLP...

5. Spelling Correction

JavaTpoint offers **Corporate Training, Summer Training, Online Training** and **Winter Training** on Java, Blockchain, Machine Learning, Meanstack, Artificial Intelligence, Kotlin, Cloud Computing, Angular, React, IOT, DevOps, RPA, Virtual Reality, **Embedded Systems**, Robotics, PHP, .Net, Big Data and Hadoop, Spark, Data Analytics, R Programming, Python, Oracle, Web Designing, Spring, Hibernate, Software technologies. For more details visit javatpoint.com



6. Chatbot



Applications of NLP...

7. Speech Recognition

8. Information Extraction

9. Natural Language Understanding (NLU)