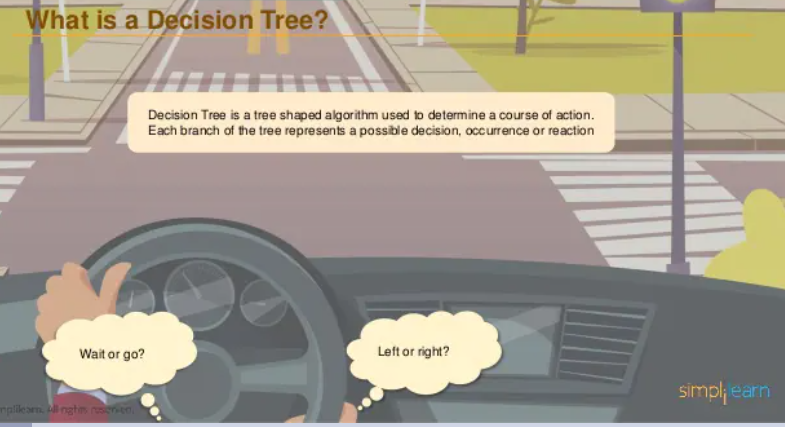
Artifical Intelligen

UNIT 2

Data mining:

fundamentals – data reduction -

Decision Tree:



1. Decision Trees are a type of Supervised Machine Learning .

2. decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions.

3. Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks.

4.The tree can be explained by two entities, namely decision nodes and leaves.

5. The leaves are the decisions or the final outcomes.

6. The decision nodes are where the data is split.

7. The decision rules are generally in form of if-then-else statements. The deeper the tree, the more complex the rules and fitter the model.

///////////example



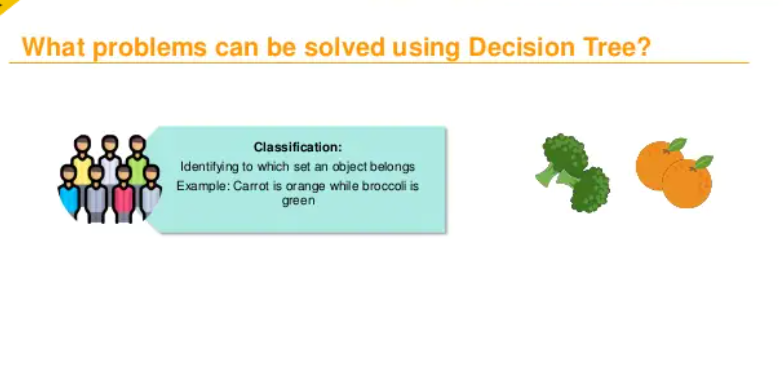
An example of a decision tree can be explained using above binary tree. Let’s say you want to predict whether a person is fit given their information like age, eating habit, and physical activity, etc. The 25 decision nodes here are questions like ‘What’s the age?’, ‘Does he exercise?’, and ‘Does he eat a lot of pizzas’? And the leaves, which are outcomes like either ‘fit’, or ‘unfit’. In this case this was a binary classification problem (a yes no type problem). ///////

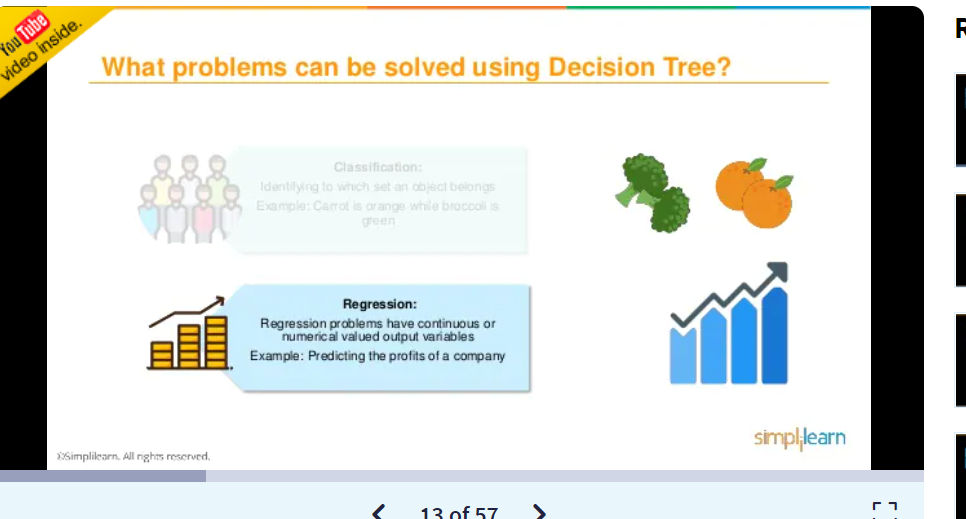
Example 2 :

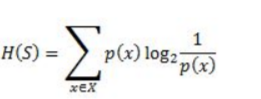
Vegitable shop

There are two main types of Decision Trees:

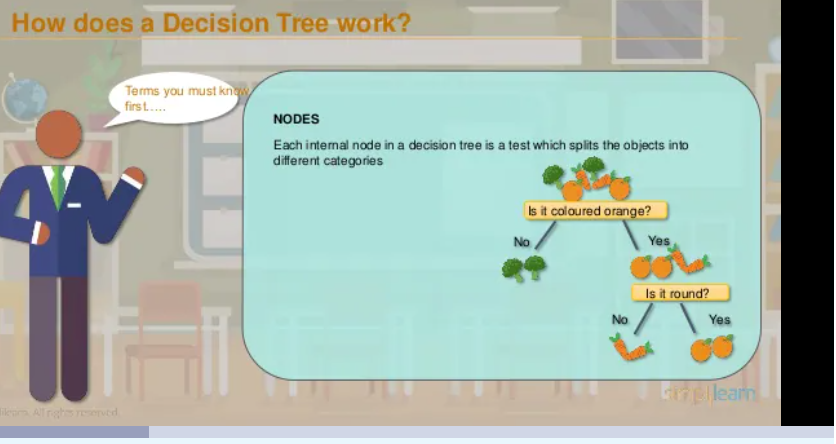
Classification





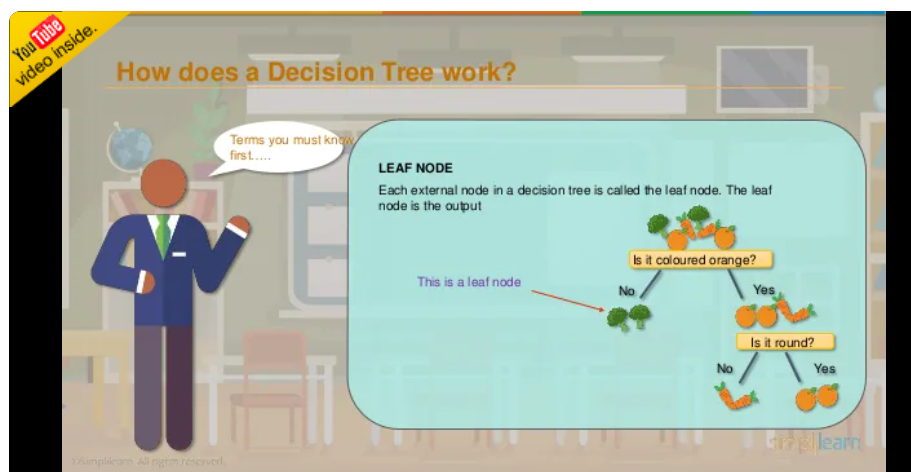


Decision work:

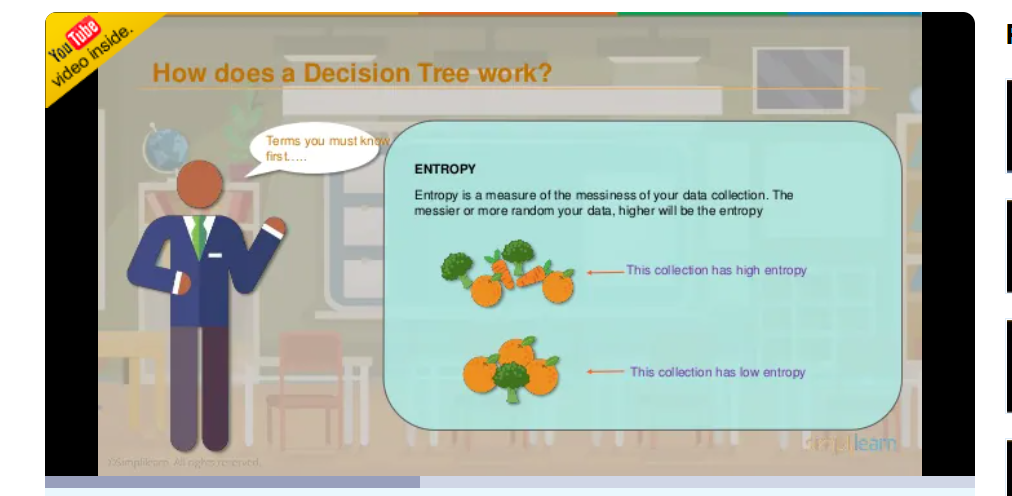




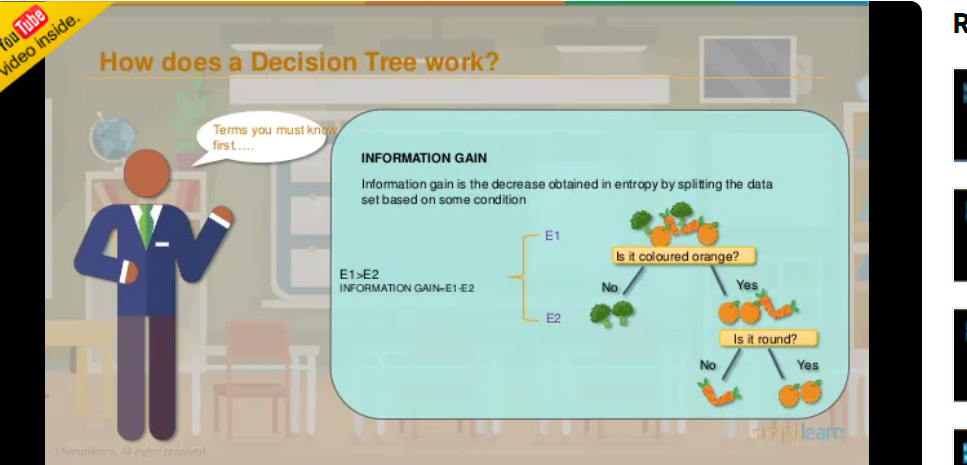
ROOT NODE The node at the top of the decision tree is called the Root node



LEAF NODE Each external node in a decision tree is called the leaf node. .



ENTROPY Entropy is a measure of the messiness of your data collection. The messier or more random your data, higher will be the entropy.



INFORMATION GAIN Information gain is the decrease obtained in entropy by splitting the data set based on some condition Is it coloured orange? Is it round? No Yes YesNo E1 E2 E1>E2 INFORMATION GAIN=E1-E2.

Decision tree algorithms:

Association rules:

An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database. It determines the set of items that occurs together in the dataset. Association rule makes marketing strategy more effective. Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item. A typical example of Association rule is Market Basket Analysis.

clustering: k-means:

\*\*importat\*\*

k-means clustering algorithm One of the most used clustering algorithm is k-means. It allows to group the data according to the existing similarities among them in k clusters, given as input to the algorithm.

A cluster refers to a collection of data points aggregated together because of certain similarities.

K-means algorithm identifies *k* number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

# How the K-means algorithm works

To process the learning data, the K-means algorithm in data mining starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids

Fuzzy c-means(Logic)

Fuzzy Logic is a form of Knowledge representation sutabil for notation that cannot be defined precisely but which depend upon their contex

Need for Fuzzy Logic:

1. It based on intution and judgement
2. There is no need for a mathematical model
3. It provides a smooth transition between members and non-members
4. It is relativly simple fast and adaptive(compare to calasification approch)
5. It can be used to implement desing objects difficult to express mathematically in linguistic or descriptive rules.

Hierarchical:

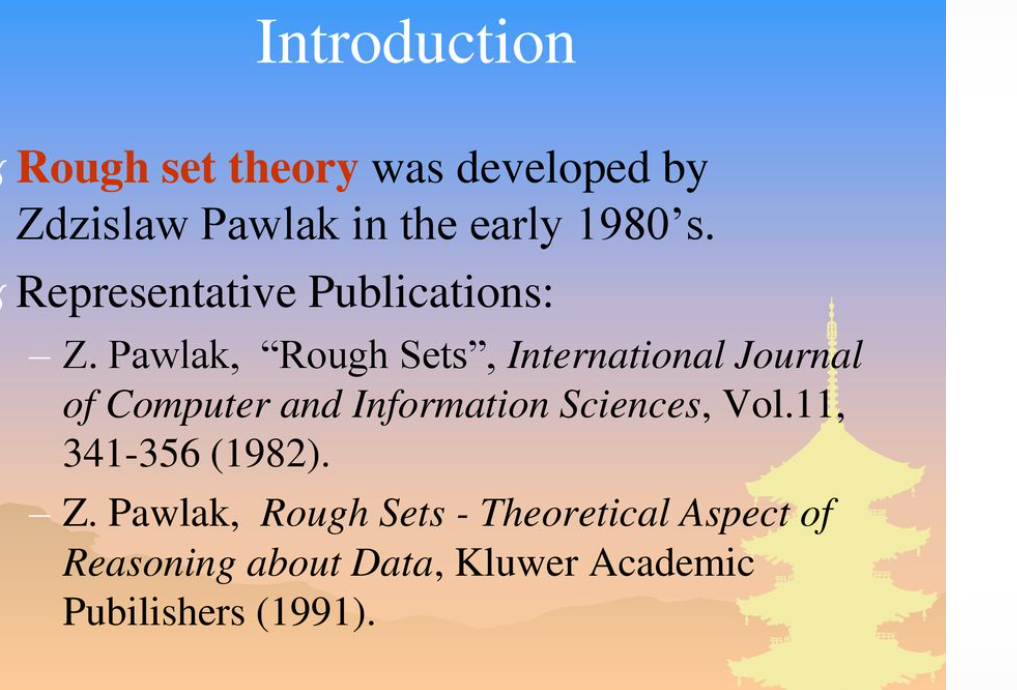
1. Hierarchical clustering is another unsupervised machine learning algorithm
2. which is used to group the unlabeled datasets into a cluster and also known as hierarchical cluster analysis or HCA.W
3. We develop the hierarchy of clusters in the form of a tree, and this tree-shaped structure is known as the dendrogram.
4. The hierarchical clustering technique has two approaches:
5. Agglomerative: Agglomerative is a bottom-up approach, in which the algorithm starts with taking all data points as single clusters and merging them until one cluster is left.
6. Divisive: Divisive algorithm is the reverse of the agglomerative algorithm as it is a top-down approach.

Sometimes the results of K-means clustering and hierarchical clustering may look similar, but they both differ depending on how they work. As there is no requirement to predetermine the number of clusters as we did in the K-Means algorithm.

Probabilistic clustering AI:

Ye wala nahi mila mujhe ans

Rough set theory:



Rough set theory was developed by Zdzislaw Pawlak in the early 1980’s.

Introduction (2)

The main goal of the rough set analysis is induction of approximations of concepts.

Rough sets constitutes a sound basis for KDD.

It offers mathematical tools to discover patterns hidden in data.

It can be used for feature selection, feature extraction, data reduction, decision rule generation, and pattern extraction (templates, association rules) etc.

identifies partial or total dependencies in data, eliminates redundant data, gives approach to null values, missing data, dynamic data and others.

Information/Decision Systems (Tables)IndiscernibilitySet ApproximationReducts and CoreRough MembershipDependency of Attributes

Basic concepts of the rough sets theory  
Consider : a data set containing the results of threemeasurements performed for 10 objects. The resultscan be organized in a matrix10x3.MeasurementsObjects

What is rough set theory:

Rough Set Theory (RST), proposed by Z Pawlak, is a new mathematical approach to vagueness and uncertainty.

Tools based on RST are found to be useful in addressing data mining tasks such as classification, clustering and rule mining.

In RST all computations are performed directly on the supplied data and works by making use of the granularity structure of the data

- Rough set theory:

definition – rule induction

Rule induction using Rough Set Theory — An application in agriculture.

By applying the proposed algorithm, a set of significant rules are generated. These rules are expected to be helpful to the farmers of the state to design their farming plans, which will enable them to improve their coconut production.

**feature selection - rough sets in data mining:**

Feature selection aims to determine a minimal feature subset from a problem

domain while retaining a suitably high accuracy in representing the original features.

Rough set theory (RST) has been used as such a tool with much success. RST enables the discovery of data dependencies and the reduction of the number of attributes contained in a dataset

using the data alone, requiring no additional information.

The main aim of feature selection (FS) is to determine a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features.

In many real world problems FS is a must due to the abundance of noisy,

irrelevant or misleading features.

For instance, by removing these factors, learning from data techniques can benefit greatly.

A detailed review of feature selection techniques devised for classification tasks can be found in (Dash & Liu, 1997).

Unit 3

Machine Learning:

Probability basics

The EM Algorithm :

The expectation-maximization algorithm is an approach for performing maximum likelihood estimation in the presence of latent variables. It does this by first estimating the values for the latent variables, then optimizing the model, then repeating these two steps until convergence. It is an effective and general approach and is most commonly used for density estimation with missing data, such as clustering algorithms like the Gaussian Mixture Model.

What is EM algorithm used for?

The EM algorithm is used to **find (local) maximum likelihood parameters of a statistical model** in cases where the equations cannot be solved directly. Typically these models involve latent variables in addition to unknown parameters and known data observations.

How does EM algorithm work?

It works **by choosing random values for the missing data points, and using those guesses to estimate a second set of data**.

Reinforcement Learning:

* What Is Machine Learning?
* Ans: Machine learning is a subset of artificial intelligence (AI) which provides machines the ability to learn automatically & improve from experience without being explicitly programmed. They look the same! Cherry Apple Orange Data Algorithm
* Types Of Machine Learning: Reinforcement Learning Supervised Learning Unsupervised Learning

Q: What Is Reinforcement Learning?

Ans: Reinforcement learning is a type of Machine Learning where an agent learns to behave in a environment by performing actions and seeing the results.

Reinforcement learning is one of three basic machine learning paradigms, alongside supervised learning and unsupervised learning.

Key areas of Interest :

* Environment
* Action
* Reward
* State



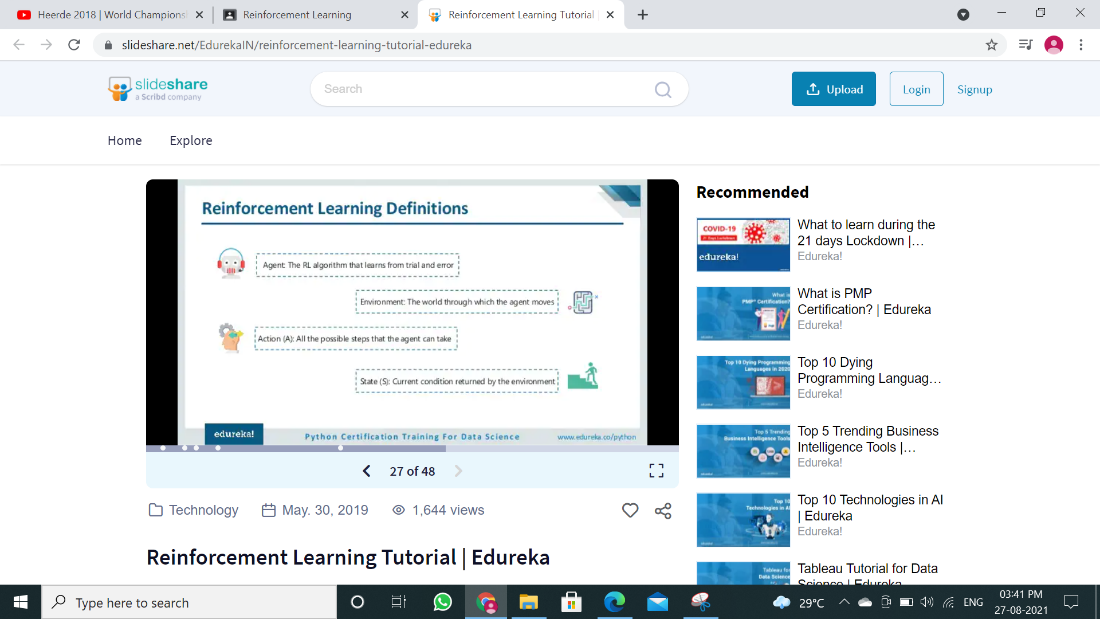
Reinforcement Learning Process :

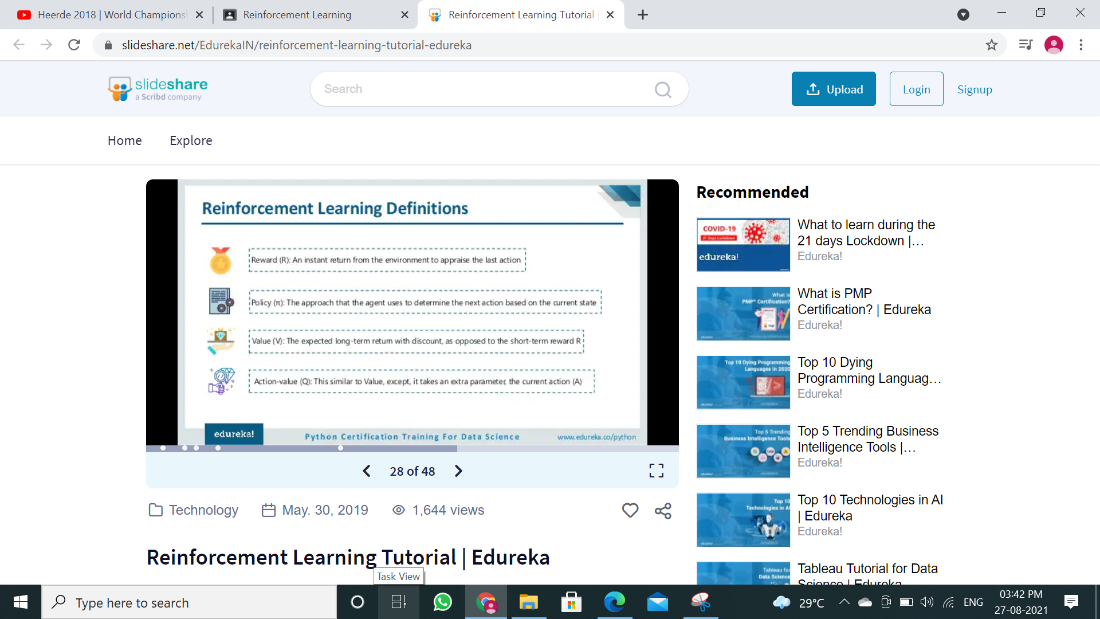
Reinforcement Learning system is comprised of two main components: • Agent • Environment

* Reinforcement Learning –
* Counter Strike Example:
* ounter Strike Example 1. The RL Agent (Player1) collects state S⁰ from the environment 2. Based on the state S⁰, the RL agent takes an action A⁰, initially the action is random 3. The environment is now in a new state S¹ 4. RL agent now gets a reward R¹ from the environment 5. The RL loop goes on until the RL agent is dead or reaches the destination

\*\*\*IMPORTANT\*\*\*

Reinforcement Learning Definitions:





* Reinforcement Learning Definitions

1. **Agent:** The RL algorithm that learns from trial and error
2. **Environment:** The world through which the agent moves
3. **Action (A)**: All the possible steps that the agent can take
4. **State (S):** Current condition returned by the environment
5. **Reward (R):** An instant return from the environment to appraise the last action
6. **Policy (π):** The approach that the agent uses to determine the next action based on the current state
7. **Value (V):** The expected long-term return with discount, as opposed to the short-term reward R
8. **Action-value (Q):** This similar to Value, except, it takes an extra parameter, the current action (A)

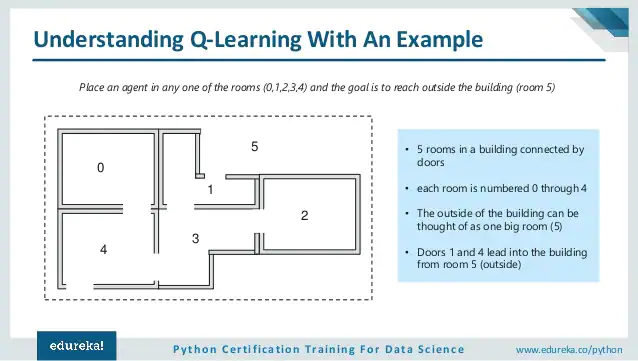
Exploration & Exploitation

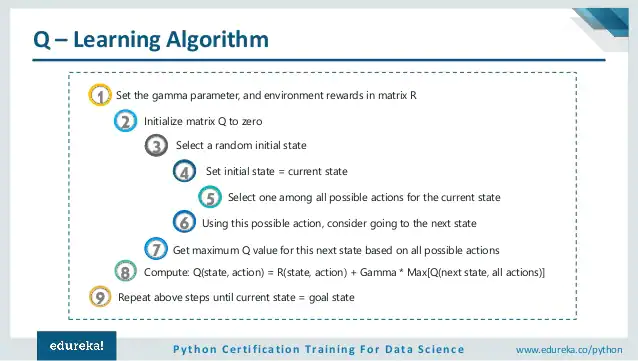
Exploration is about exploring and capturing more information about an environment Agent Opponent Reward

Exploitation is about using the already known exploited information to heighten the rewards

**Markov Decision Process**

The mathematical approach for mapping a solution in reinforcement learning is called Markov Decision Process (MDP)





**Q – Learning Algorithm:**

1. Set the gamma parameter, and environment rewards in matrix R
2. Initialize matrix Q to zero
3. Select a random initial state
4. Set initial state = current state
5. Select one among all possible actions for the current state
6. Using this possible action, consider going to the next state
7. Get maximum Q value for this next state based on all possible actions
8. Compute: Q(state, action) = R(state, action) + Gamma \* Max[Q(next state, all actions)]
9. Repeat above steps until current state = goal state

**Unit 3 Test Topic**

**1]  Supervised Learning**

Supervised learning

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

2. The goal of supervised learning is to train the model so that it can predict the output when it is given new data.

3. Supervised learning can be used for those cases where we know the input as well as corresponding outputs.

4. A supervised learning algorithm analyzes the training data and produces a function, which can be used for mapping new examples.

5. Supervised learning can be categorized in Classification and Regression problems.

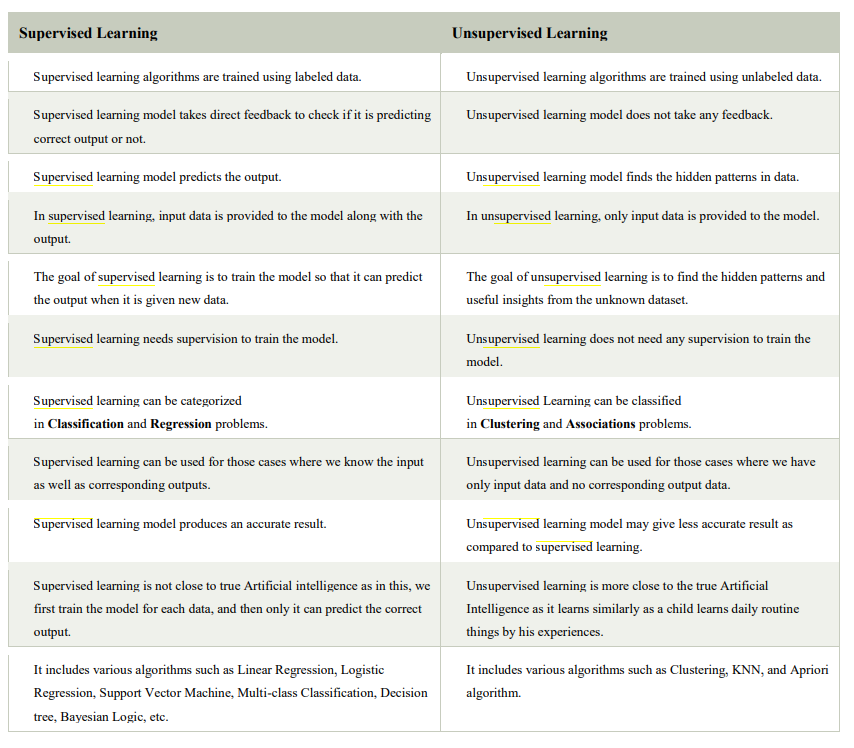
6. A wide range of supervised learning algorithms are available, each with its strengths and weaknesses.

7. A training set of examples with the correct responses is provided and based on this training set, the algorithm generalises to respond correctly to all possible inputs.

8. In supervised learning each example in the training set is a pair consisting of an input object and an output value.

9. It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc

|  |  |
| --- | --- |
| **Supervised Learning** | **Unsupervised Learning** |
| Supervised learning algorithms are trained using labeled data | Unsupervised learning algorithms are trained using unlabeled data. |
| Supervised learning model takes direct feedback to check if it is predicting correct output or not. | Unsupervised learning model does not take any feedback. |
| Supervised learning model predicts the output. | Unsupervised learning model finds the hidden patterns in data. |
| In supervised learning, input data is provided to the model along with the output. | In unsupervised learning, only input data is provided to the model. |
| The goal of supervised learning is to train the model so that it can predict the output when it is given new data | The goal of unsupervised learning is to find the hidden patterns and useful insights from the unknown dataset. |
| Supervised learning needs supervision to train the model | Unsupervised learning does not need any supervision to train the model. |
| Supervised learning can be categorized in Classification and Regression problems. | Unsupervised Learning can be classified in Clustering and Associations problems. |
| Supervised learning can be used for those cases where we know the input as well as corresponding outputs. | Unsupervised learning can be used for those cases where we have only input data and no corresponding output data |
| Supervised learning model produces an accurate result. | Unsupervised learning model may give less accurate result as compared to supervised learning. |
| Supervised learning is not close to true Artificial intelligence as in this, we first train the model for each data, and then only it can predict the correct output | Unsupervised learning is more close to the true Artificial Intelligence as it learns similarly as a child learns daily routine things by his experiences |
| It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc | It includes various algorithms such as Clustering, KNN, and Apriori algorithm |



**2] Learning Decision Trees**

1. Decision Trees are a type of Supervised Machine Learning

2. This is explain what the input is and what the corresponding output is in the training data.

3. The data is continuously split according to a certain parameter.

4. The tree can be explained by two entities, namely decision nodes and leaves.

5. The leaves are the decisions or the final outcomes And the decision nodes are where the data is split.

ID3 Algorithm:

ID3 Algorithm will perform following tasks recursively

1. Create root node for the tree

2. If all examples are positive, return leaf node positive

3. Else if all examples are negative, return leaf node negative

4. Select the attribute which has maximum value of IG(S, x)

5. For each attribute, calculate the entropy with respect to the attribute „x‟ denoted by H(S, x)

6. Calculate the entropy of current state H(S)

7. Remove the attribute that offers highest IG from the set of attributes 8. Repeat until we run out of all attributes, or the decision tree has all leaf nodes.

**the highest information gain is given by the attribute Outlook.**

**3] Regression and Classification with Linear Models**

Classification Trees:

1. A classification tree is an algorithm where the target variable is fixed or categorical.

2. The algorithm is then used to identify the “class” within which a target variable would most likely fall.

An example of a classification-type problem would be determining who will or will not subscribe to a digital platform; or who will or will not graduate from high school.

These are examples of simple binary classifications where the categorical dependent variable can assume only one of two, mutually exclusive values.

In other cases, you might have to predict among a number of different variables.

For instance, you may have to predict which type of smartphone a consumer may decide to purchase.

In such cases, there are multiple values for the categorical dependent variable. Here’s what a classic classification tree looks like.

**Linear Regression in Machine Learning Linear:**

1. The linear regression model provides a sloped straight line representing the relationship between the variables.

2. Regression is one of the easiest and most popular Machine Learning algorithms.

3. It is a statistical method that is used for predictive analysis.

4. linear regression show linear relationship and finds how the value of the dependent variable is changing according to the value of the independent variable.

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

6. It makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

Some popular applications of linear regression are:

Analyzing trends and sales estimates

Salary forecasting

Real estate prediction

Arriving at ETAs in traffic.

ssss

Regression Trees:

A regression tree refers to an algorithm where the target variable is and the algorithm is used to predict it’s value.

As an example of a regression type problem, you may want to predict the selling prices of a residential house, which is a continuous dependent variable.

This will depend on both continuous factors like square footage as well as categorical factors like the style of home, area in which the property is located and so on.

**Linear Regression:**

1. Linear regression is a statistical regression method which is used for predictive analysis.
2. It is one of the very simple and easy algorithms which works on regression and shows the relationship between the continuous variables.
3. It is used for solving the regression problem in machine learning.
4. Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), hence called linear regression.
5. If there is only one input variable (x), then such linear regression is called simple linear regression. And if there is more than one input variable, then such linear regression is called multiple linear regression.
6. The relationship between variables in the linear regression model can be explained using the below image. Here we are predicting the salary of an employee on the basis of the year of experience.

**4] Artificial Neural Networks**

1. Artificial neural network (ANN) learning algorithm usually called "neural network" (NN)

2. It is a learning algorithm that is vaguely inspired by biological neural networks.

3. They are usually used to model complex relationships between inputs and outputs.

4. To find patterns in data or to capture the statistical structure in an unknown joint probability distribution between observed variables.5. Computations are structured in terms of an interconnected group of artificial neurons, processing information using a connectionist approach to computation.

6. Modern neural networks are non-linear statistical data modeling tools.

**5] Nonparametric Models**

1. Nonparametric machine learning algorithms try to make assumptions about the data given the patterns observed from similar instances.
2. Nonparametric Models are statistical models that do not often conform to a normal distribution, as they rely upon continuous data, rather than discrete values.
3. This algorithms are **those which do not make specific assumptions about the type of the mapping function**.
4. They are prepared to choose any functional form from the training data, by not making assumptions.
5. Nonparametric algorithms are often slower and require large amounts of data, they are rather flexible as they minimize the assumptions they make about the data.
6. Nonparametric statistics often deal with ordinal numbers, or data that does not have a value as fixed as a discrete number.
7. The only assumption it makes about the data set is that the training patterns that are the most similar are most likely to have a similar result. While

Example:

a popular nonparametric machine learning algorithm is the K-Nearest Neighbor algorithm it is looks at similar training patterns for new instances

**6] Support Vector Machines**

1. Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms.
2. It is used for Classification as well as Regression problems.
3. The SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.
4. This best decision boundary is called a hyperplane.
5. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.
6. SVM algorithm can be used for Face detection, image classification, text categorization, etc.

**Types of SVM :**

Linear SVM:

Linear SVM is used for linearly separable data. (which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier).

Non-linear SVM:

Non-Linear SVM is used for non-linearly separated data. ( which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.)

Example: SVM example that we have used in the KNN classifier.

There is a cat that also has some features of dogs if we want a model that identify it is a cat or dog this type of model create using the SVM algorithm. First we train our model to images of cats and dogs so it can learn about different features of cats and dogs and after that we test it. So as support vector creates a decision boundary between cat and dog and choose extreme cases (support vectors), it will see the extreme case of cat and dog. On the basis of the support vectors, it will classify it as a cat.

//////////Extra//////

**machine learning algorithms can be classified into three types.**

• Supervised learning

• Unsupervised learning

• Reinforcement learnin

**Unsupervised learning:**

1. Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses.

2. The statistical approach to unsupervised learning is known as density estimation.

3. In Unsupervised learning algorithms a classification or categorization is not included in the observations.

4. Unsupervised learning refers to the use of artificial intelligence ([AI](https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence)) algorithms to identify patterns in data sets.

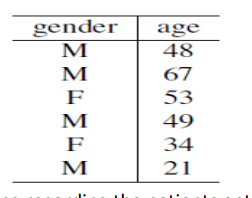
5.  In Unsupervised learning AI system will group unsorted information according to similarities and differences even though there are no categories provided.

6. unsupervised learning can be more unpredictable than a supervised learning model.

Example: The most common unsupervised learning method is cluster analysis, which is used for exploratory data analysis to find hidden patterns 14 or grouping in data.

given to the learner are unlabeled, the accuracy of the structure that is output by the algorithm cannot be evaluated.

Example Consider the following data regarding patients entering a clinic.



The data consists of the gender and age of the patients. Based on this data, can we infer anything regarding the patients entering the clinic?

**Reinforcement learning:**

1. It is a machine learning training method based on rewarding desired behaviors and/or punishing undesired ones
2. It is give only feedback to the program's actions in a dynamic environment such as driving a vehicle or playing a game against an opponent.
3. The algorithm gets told when the answer is wrong, but does not get told how to correct it.
4. Reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error.
5. It has to explore and try out different possibilities until it works out how to get the right answer.
6. Reinforcement learning is sometime called learning with a critic because of this monitor that scores the answer, but does not suggest improvements.
7. This is somewhere between supervised and unsupervised learning.
8. Reinforcement learning is the problem of getting an agent to act in the world so as to maximize its rewards.
9. learner does not told what actions to take in most forms of machine learning, but instead must discover which actions yield the most reward by trying them.
10. Actions may affect not only the immediate reward but also the next situations and, through that, all subsequent rewards.

Ex: Train computers to do many tasks, such as playing backgammon or chess, scheduling jobs, and controlling robot limbs.

Example Consider teaching a dog a new trick: we cannot tell it what to do, but we can reward/punish it if it does the right/wrong thing. It has to find out what it did that made it get the reward/punishment.

Reinforcement learning is different from supervised learning.

Supervised learning is learning from examples provided by a knowledgeable expert.

Entropy **:**

Entropy, also called as Shannon

Entropy is denoted by H(S) for a finite set S, is the measure of the amount of uncertainty or randomness in data.

It tells the predictability of a certain event.

Example, consider a coin toss whose probability of heads is 0.5 and probability of tails is 0.5. Here the entropy is the highest possible, since there’s no way of determining what the outcome might be. **this event has no randomness hence it’s entropy is zero.**

Information Gain:

Information gain is also called as Kullback-Leibler divergence denoted by IG(S, A) for a set S is the effective change in entropy after deciding on a particular attribute A.

It measures the relative change in entropy with respect to the independent variables

Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).

**Unit 4**

**1] Introduction to Deep Learning Models**

**Deep Lerning**

ANS: Deep learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network.

3) examples of deep learning:

**Virtual assistants, vision for driverless cars, money laundering, face recognition and many more**.

**Deep Learning Models:**

1) Deep learning models are trained by using large sets of labeled data and neural network architectures that learn features directly from the data without the need for manual feature extraction.

2) One of the most popular types of deep neural networks is known as convolutional neural networks (CNN or ConvNet).

3) It's essentially a neural network with many layers.

4) These models can be enormous in size often with more than 50 million parameters. 5) The algorithm is not new, but because we now have bigger data with more computing power.

6) In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound.

7) Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance.

#####EXTRA#####

* Companies are working on solutions for Deep Learning acceleration:

1. NVIDIA NVIDIA created a entire plaftorm stack dedicated to work with Deep Learning, called DIGITS. Their GPUs are widely used in Deep Learning.
2. AMAZON Amazon AWS also create EC2 instances with NVIDIA GPUs (with 4GB of memory and 1536 CUDA cores). Lots of AMIs with Deep Learning software ecosystem already installed.
3. MICROSOFT Microsoft announced that it will offer NVIDIA GPUs on its Azure cloud platform.

**2] Autoencoder**

1. autoencoders belong to the neural network family.
2. It is used to learn efficient codings of unlabeled data (unsupervised learning).
3. The autoencoder learns a representation (encoding) for a set of data, typically for dimensionality reduction, by training the network to ignore insignificant data (“noise”).
4. autoencoder is a sort of unsupervised neural network that is used for dimensionality reduction and feature extraction.
5. In Autoencoders (AE) the input is the same as the output.
6. AE work compressing the input into a latent-space representation, and then reconstructing the output from this representation.
7. Autoencoders can be employed for creating generative models, using dimensionality reduction, and detecting anomalies.
8. autoencoder is a neural network model that can be used to learn a compressed representation of raw data.
9. Autoencoder components are encoder, code and decoder.
10. The encoder compresses the input and produces the code, the decoder then reconstructs the input only using this code.

**3] Convolutional Neural Networks**

1) CNN is a feed forward neural network that is generally used to analyze visual images by processing data with grid like topology.

2) A CNN is also known as a “ConvNet” Convolution operation forms the basis of any Convolution Neural Network In CNN

3) CNN is a supervised deep learning method.

4) Every image is represented in the form of arrays of pixel values Real Image of the digit 8 Represented in the form of an array 0 0 0 0 1 0 1 0 0 0 0 0 0 Digit 8 represented in the form of pixels of 0’s and 1’s

5) The origin of the applications of deep learning to object recognition tasks can be traced to the convolutional neural networks (CNNs) in the early 90s.

6) The CNNbased architectures have captured intense interest in computer vision since October 2012 shortly after the ImageNet competition results were released.

CNN(or convnets) are based on the following principles:

Local receptive ﬁelds

Shared weights

Pooling (or down-sampling)

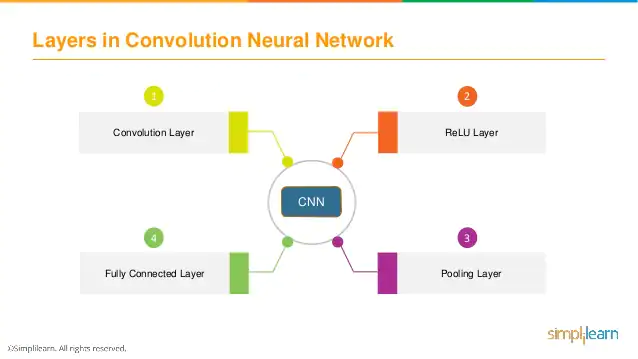
* This special neural network architecture takes advantage of the spatial structure of data.
* CNN ARCHITECTURE
* They are very similar to ordinary NNs.:
* They are made up of neurons that have learnable weights and biases.
* Each neuron receives some inputs, performs a dot product and optionally follows it with a nonlinearity.
* Their connectivity is now restricted to be local spatially.

CNNs are comprised of three types of layers:

* Convolutional layers
* Pooling layers
* Fully-connected layers

Unlike standard ANNs, the neurons within Convolutional and Poolinglayers will only connect to a small region of the layer preceding it

The neurons in a layer will only be connected to a small region of the layer before it, instead of all of the neurons in a fully-connected manner.



**4] Recurrent Neural Networks:**

1. Recurrent Neural Network(RNN) are a type of [Neural Network](https://www.geeksforgeeks.org/tag/neural-network/) where the output from previous step are fed as input to the current step.
2. RNN have a “memory” which remembers all information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks.
3. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence.
4. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence.

//////// continu///

1. The brain is a recurrent network because activation is fed back to the units that caused it.
2. recurrent networks have internal state stored in the activation levels of the units.
3. This also means that computation can be much less orderly than in feed-forward networks.
4. Recurrent networks can become unstable, or oscillate, or exhibit chaotic behavior.
5. recurrent networks can implement more complex agent designs and can model systems with state.
6. recurrent networks require some quite advanced mathematical method.
7. Hopfield networks are probably the best-understood class of recurrent networks

### **Advantages of Recurrent Neural Network**

1. RNN remembers each and every information through time.
2. It is used with convolutional layers to extend the effective pixel neighborhood.
3. It is useful for the time series prediction because of the feature to remember previous inputs as well.
4. It is called Long Short Term Memory.

**Disadvantages of Recurrent Neural Network**

1. Training an RNN is a very difficult task.
2. Gradient vanishing and exploding problems.
3. It cannot process very long sequences if using tanh or relu as an activation function.

**5] LSTM,**

1. Long Short Term Memory is a kind of recurrent neural network.
2. LSTM was designed by Hochreiter & Schmidhuber.
3. LSTM networks are an extension of recurrent neural networks (RNNs) mainly introduced to handle situations where RNNs fail.
4. It is an artificial recurrent neural network (RNN) architecture used in the field of deep learning.
5. It is used for processing, predicting and classifying on the basis of time series data.
6. LSTM can by default retain the information for long period of time.
7. LSTM has been designed that the vanishing gradient problem is almost completely removed, while the training model is left unaltered.
8. In lstm there is no need to keep a finite number of states from beforehand as required in the hidden Markov model (HMM).
9. LSTMs provide with a large range of parameters such as learning rates, and input and output biases.
10. The complexity to update each weight is reduced to O(1) with LSTMs, similar to that of Back Propagation Through Time (BPTT), which is an advantage.
11. LSTM models need to be trained with a training dataset prior to its employment in real-world applications.
12. Some Applicate are Language modelling or text generation, Image processing, Music generation , Language Translation and Speech, Handwriting Recognition and Question Answering Chatbots.

/////Extra/////

1. LSTM has a chain structure that contains four neural networks and different memory blocks called **cells**.
2. Information is retained by the cells and the memory manipulations are done by the **gates.** There are three gates –

**Forget Gate:**

**Input gate:**

**Output gate**

A**pplications of LSTM include:**

1. Robot control.
2. Time series prediction.
3. Speech recognition.
4. Rhythm learning.
5. Music composition.
6. Grammar learning.
7. Handwriting recognition.
8. Human action recognition.

**6] Network Architecture Search (NAS)**

1. Neural architecture search (NAS)[[1]](https://en.wikipedia.org/wiki/Neural_architecture_search#cite_note-survey-1)[[2]](https://en.wikipedia.org/wiki/Neural_architecture_search#cite_note-survey2-2) is a technique for automating the design of [artificial neural networks](https://en.wikipedia.org/wiki/Artificial_neural_network) (ANN), a widely used model in the field of [machine learning](https://en.wikipedia.org/wiki/Machine_learning).
2. NAS can be seen as a subfield of AutoML and has a significant overlap with hyperparameter optimization.
3. It is a research field concerned with utilizing optimization algorithms to design optimal neural network architectures.
4. it goes with the principle “Better the design, Better the performance” and NAS helps to minimize the time and cost involved in design experimentation
5. we need to provide a NAS system with a dataset and a task (classification, regression, etc), and it will give us the architecture.
6. This architecture will perform best among all other architecture for that given task when trained by the dataset provided.
7. NAS has been used to design networks that are on par or outperform hand-designed architectures.
8. Methods for NAS can be categorized according to the search space, search strategy and performance estimation strategy used.

There are many approaches concerning the architectural search spaces, optimization algorithms, as well as candidate architecture evaluation methods.