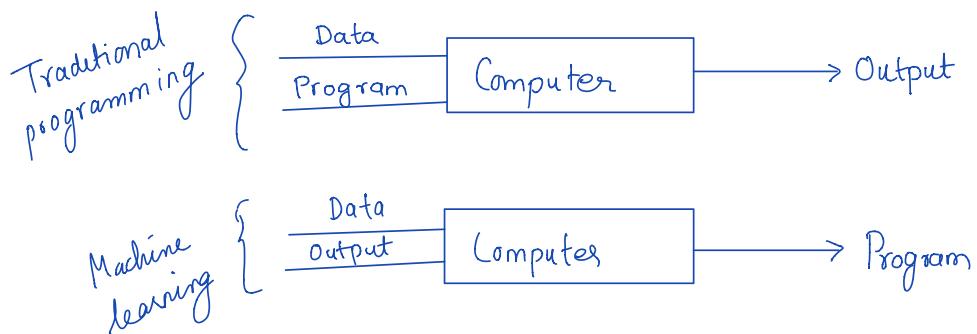


① { Machine learning }

Machine learning is a field of CS that uses statistical techniques to give computer systems the ability to "learn" with data, without being explicitly programmed.

↓
explicit programming means we write code for every scenario.



② { When machine learning is useful? }

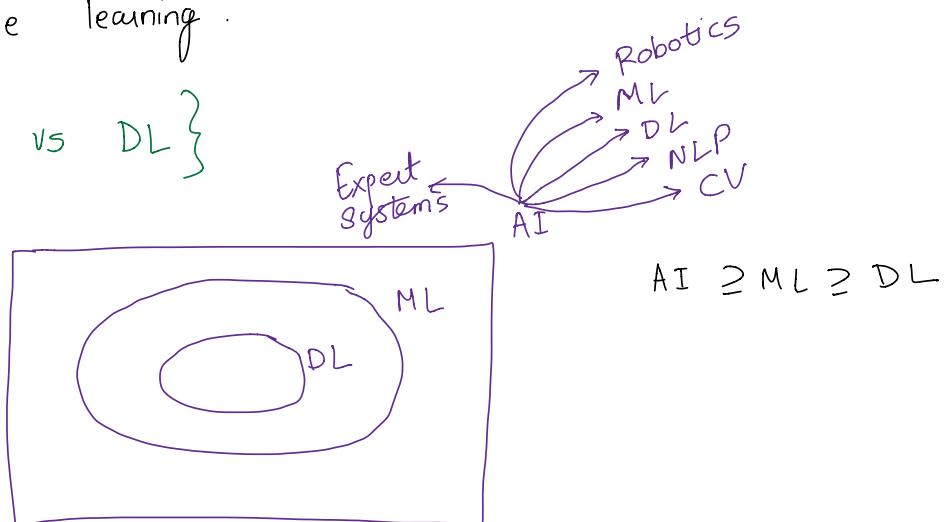
- (a) Scenarios where you cannot do explicit programming: For example, SPAM - HAM classifier. In the case of SPAM classifier, it is not possible to write each and every scenario. Also, if in future, more SPAM varieties come, then in case of traditional programming, we need to revise logic everytime the data changes.
- (b) Data mining: To find the hidden patterns in data which are not possible by plotting the graph.

③ { History of ML }

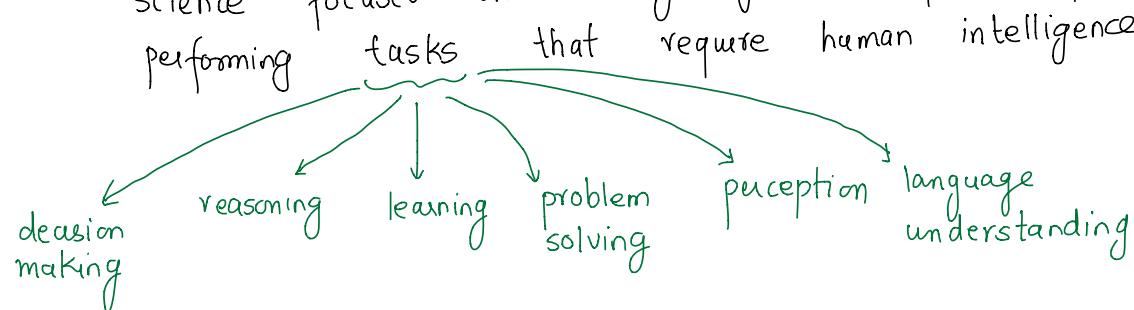
- * has history like Nawazuddin Siddiqui
- * started to gain recognition around 2010.

The main problem with the machine learning till 2010 was that we did not have sufficient data. Around 2010, we started creating more and more data. This was the boom of machine learning.

④ { AI vs ML vs DL }



AI :- Artificial Intelligence is a broad field of computer science focused on creating systems capable of performing tasks that require human intelligence.



ML : A subset of AI that involves the use of algorithms and statistical models to enable computers learn from and make predictions or decisions based on data.

It has wide range of techniques like regression, classification, clustering and reinforcement learning.

DL :- A specialized subset of ML that uses neural networks with many layers (hence "deep") to model and understand complex patterns in data.

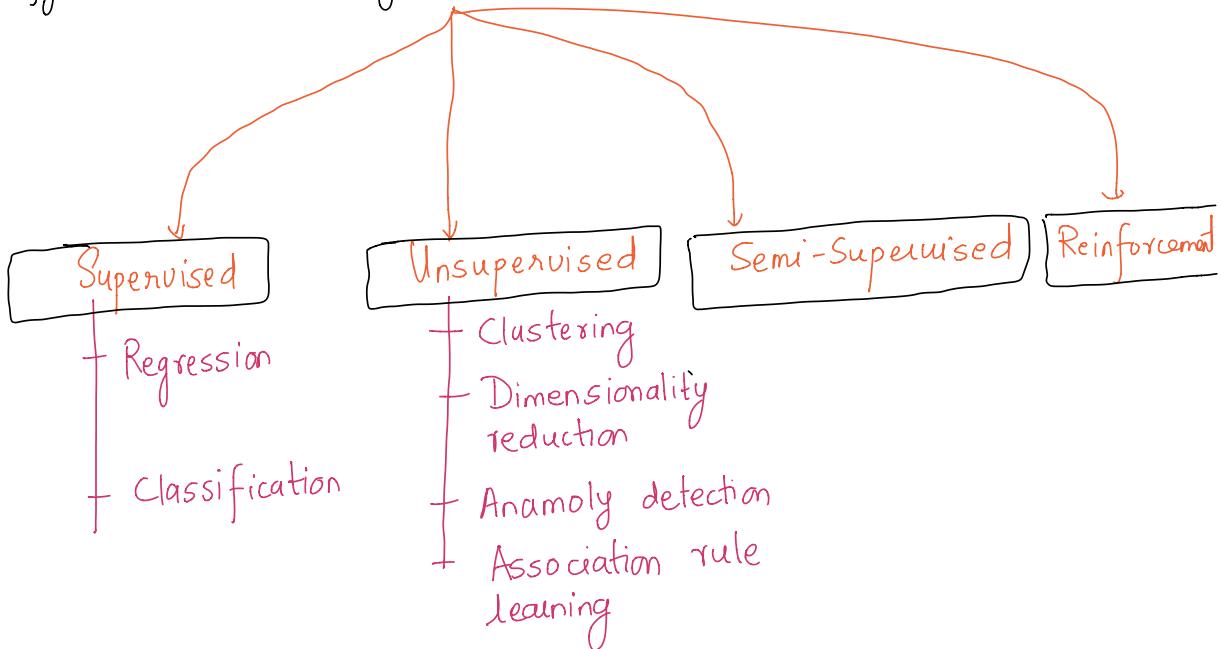
This is particularly useful when we have large amounts of data and complex patterns, such as image and speech recognition.

⑤ { ML vs DL }

Machine learning	Deep learning
<ul style="list-style-type: none">• works with smaller datasets and can achieve good performance with appropriate feature engineering.• Techniques are supervised learning, unsupervised learning and reinforcement learning.• Includes a variety of algorithms such as decision trees, SVMs, k-NN, LR, MLR, Logistic regression etc.	<ul style="list-style-type: none">• Requires large amount of data to perform well because deep neural networks have many parameters• Uses backpropagation for training deep neural networks.• Includes neural network architectures such as Convolutional Neural Networks (CNNs) for image data, Recurrent Neural Network (RNNs) for sequential data and Transformers for NLP.

⑥ { Types of machine learning }

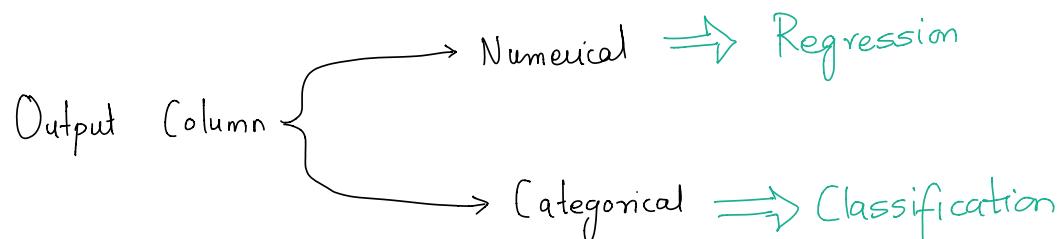
On the amount of external supervision needed, we can classify machine learning algorithms into 4 types.



⑦ { Supervised machine learning model }

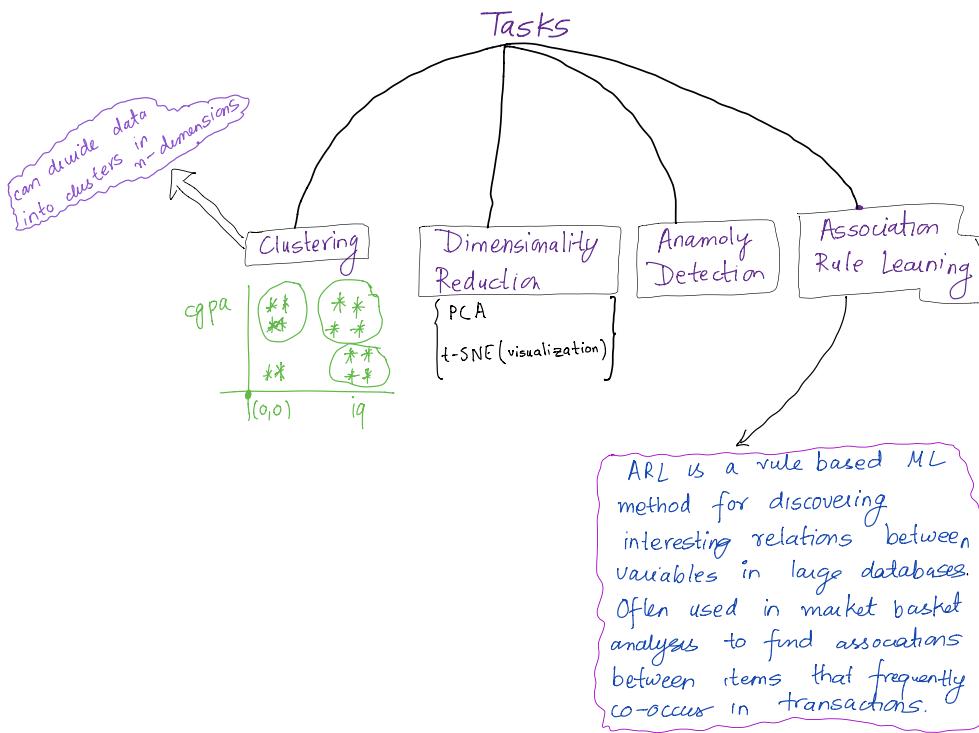
Def It is a type of machine learning where the model is trained on labeled data. The algorithm learns to map inputs to outputs based on this learning

It is basically of 2 types



⑧ { What is unsupervised machine learning? }

In unsupervised machine learning, we do not have an output column. Thus, the task of prediction does not make sense.



⑨ { What is semi-supervised learning? }

⑩ { Reinforcement learning }

Def

Reinforcement learning (RL) is a type of machine learning where an agent learns to make decisions by taking actions in an environment to maximize some notion of cumulative reward.

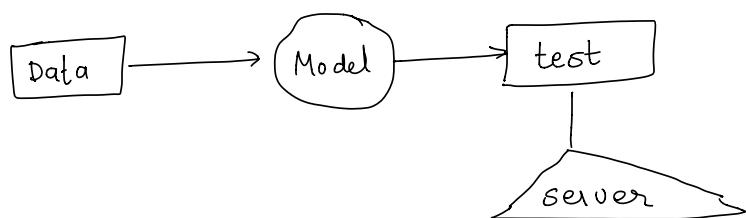
In RL, the algorithm (agent) learns from the consequences of actions through trial and error.

One such example is α go which is a RL model developed by deepmind .

⑪ { Batch Vs Online ML }

Batch machine learning , also known as offline learning , refers to the training of machine learning models using a fixed dataset that is available all at once . In this setting , the entire training dataset is used to train the model .

This is the conventional way of training ML model . The training is NOT done in incremental manner .



The problem with this approach is that it is static in nature. Modern world business problem evolve with time. Due to this, we train our model periodically. For this, whenever we get new data, we merge it with our old data. Now we pull down our model and train it with the new merged data. After training is done, we test it and deploy the model on server. This is often done in a period of 1 week or a month or sometimes even on 24 hrs basis.

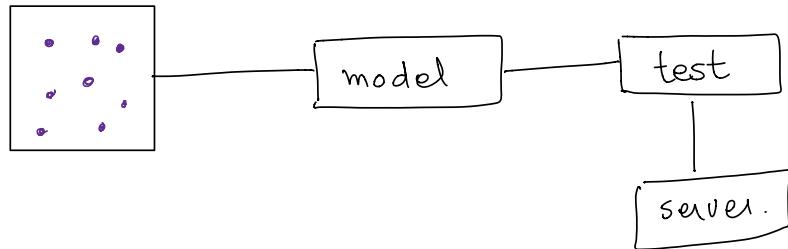
{ Disadvantages of Batch ML }

- + Lots of data
- + Hardware limitation
- + Availability \Rightarrow Sometimes, we need to update the model before its updation frequency. Thus, availability is a problem.

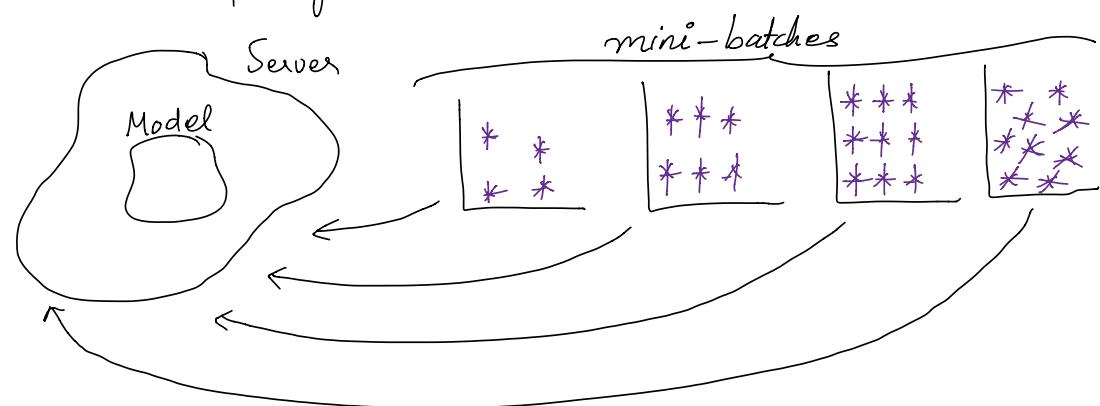
② { Online ML }

Online learning is done incrementally. We train the models with batches of data. In this technique, we train the model with mini batches of data. This is done on server.

We initially train the model, test it and then deploy it on server.



Now as the new data comes, the model is trained with this new data without pulling down the model from the server



⑬ { When to use Online ML? }

- + where there is concept drift \Rightarrow whenever the nature of problem keeps changing in that case we should use online ML .
- + cost effective
- + Faster solution

⑭ { How to implement online ML ? }

There are several libraries that support online machine learning.

These are :-

- + Scikit - learn
- + River { A library specifically designed for online ML }
- + Vowpal Wabbit { A fast and scalable online ML system }
- + Tensorflow { can be used to implement custom online learning algorithms }

⑮ { Out of core learning }

Out of core learning, also known as external memory learning, is a technique used to handle datasets that are too large to fit into a computer's main memory (RAM). Instead of loading the entire dataset into memory at once, out-of-core learning processes data in chunks, allowing machine learning algorithms to work with very large datasets efficiently.



This is quite similar to online learning.

⑯ { Instance Vs Model Based Learning }

In general, there are 2 ways of learning any thing.



In Memorization, we do not try to learn the underlying concept. We just memorize the data. However, in generalization, we try to learn the underlying concept.

The same way, machine learning models also learn. There are algorithms which just memorize the data and there are ML algorithms which try to understand the underlying pattern in the data, using mathematical equation.

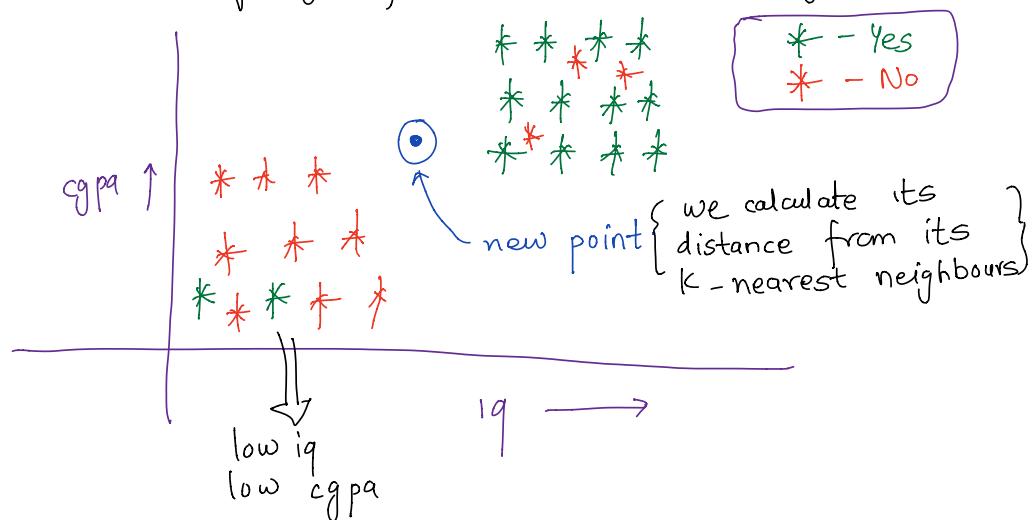
The algorithm which works on the principle of memorization is called **Instance based** ML model. Some examples are k-NN algorithm.

The algorithm which works on the principle of generalization (understanding the underlying concept) is called **model based** ML algorithm. Most of the ML algorithms are model based only.

Let us understand this with the below example dataset.

cgpa	iq	placement
...

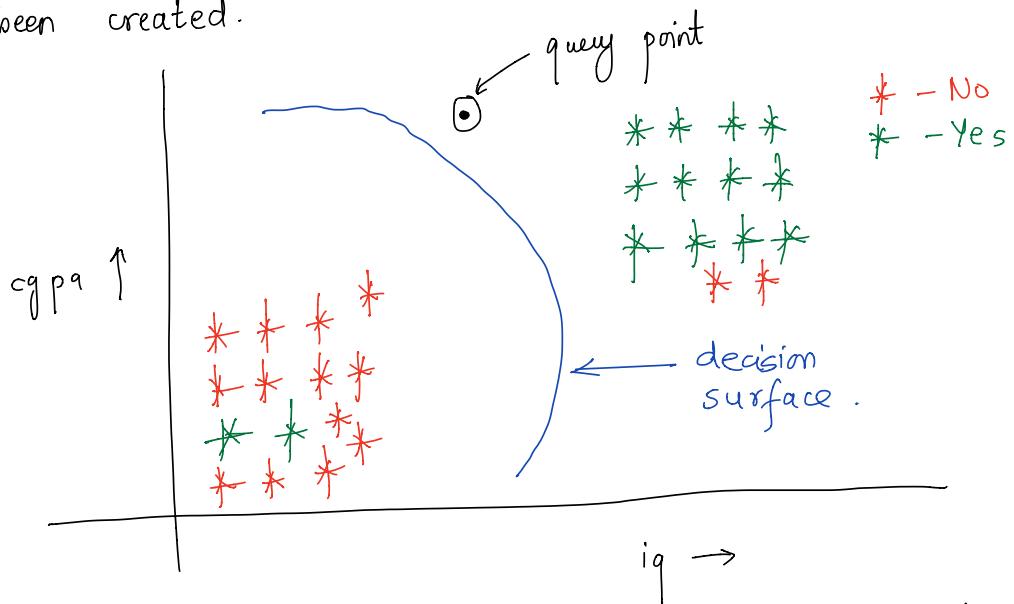
An instance based ML algorithm would just plot the data and whenever a new point comes, it would measure the distances of its k -neighbours. The point would be put in the category where majority of its k -nearest neighbours are



In these algorithms, we do not train our data. We just wait for the new query point to come. Since we use our original data during prediction for new query point we need our data even after the creation of model.

This, however, is not the case with model based ML algorithms where in we try to understand the underlying mathematical formulation. In the above case, we would have created a decision surface. Now since after the model has been trained and mathematics behind it has been understood, we do not require the original data after the model

has been created.



After the creation of decision surface, we are no longer dependent on original data. we might as well delete the original data.

