

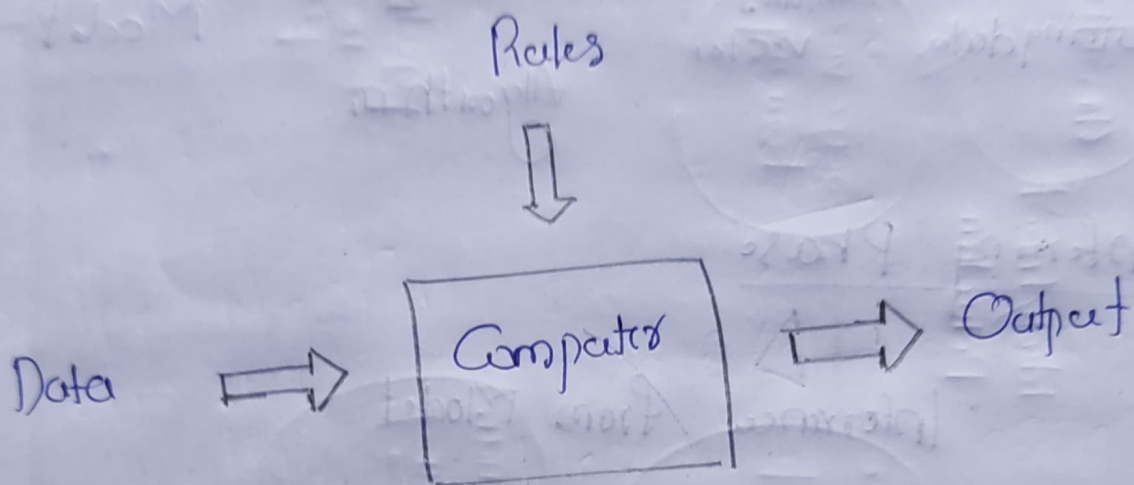
8/01/2022

How Learning Works.

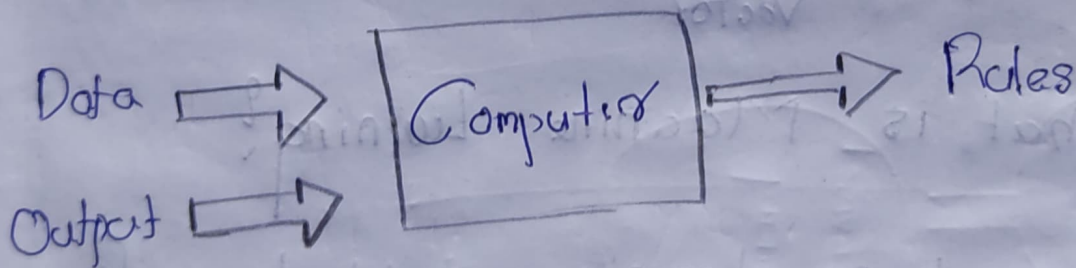
- The way the machine learns is similar to the human being.
- Humans learn from experience.
- The more we know, the more easily we can predict. (By analogy, when we face an unknown situation, the likelihood of success is lower than the known situation).
- Machines are trained the same.
- To make an accurate prediction, the machine sees an example. When we give the machine

a similar example. It can figure out the outcome. (However, like a human, if its feed a previously unseen example, the machine has difficulties to predict).

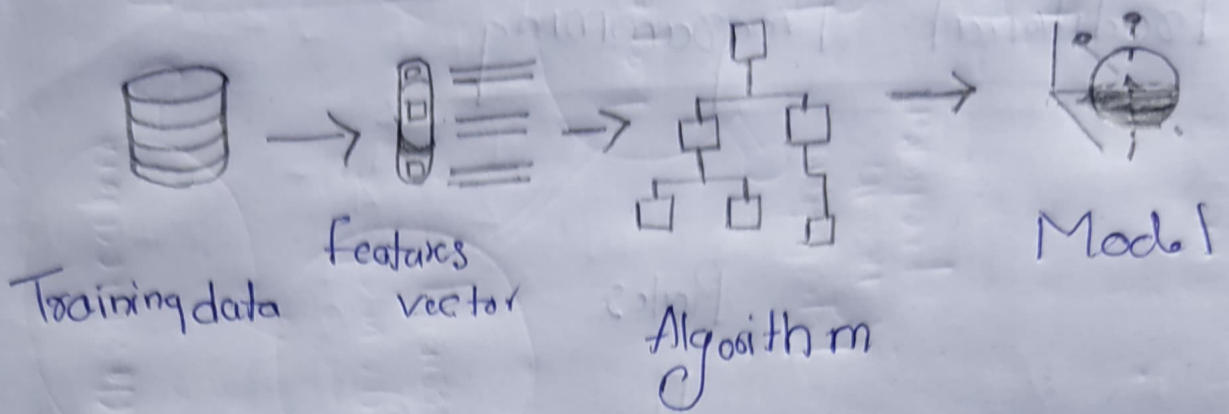
Traditional Programming



Machine Learning

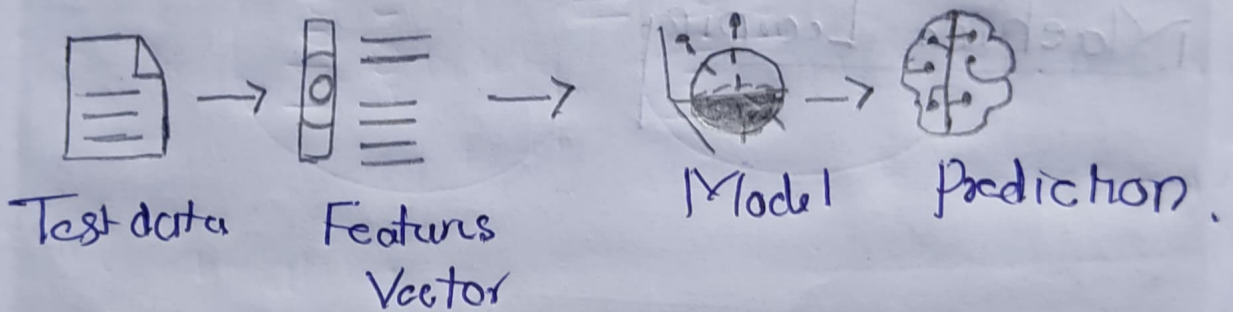


Learning Phase



Infering Phase

Inference from Model



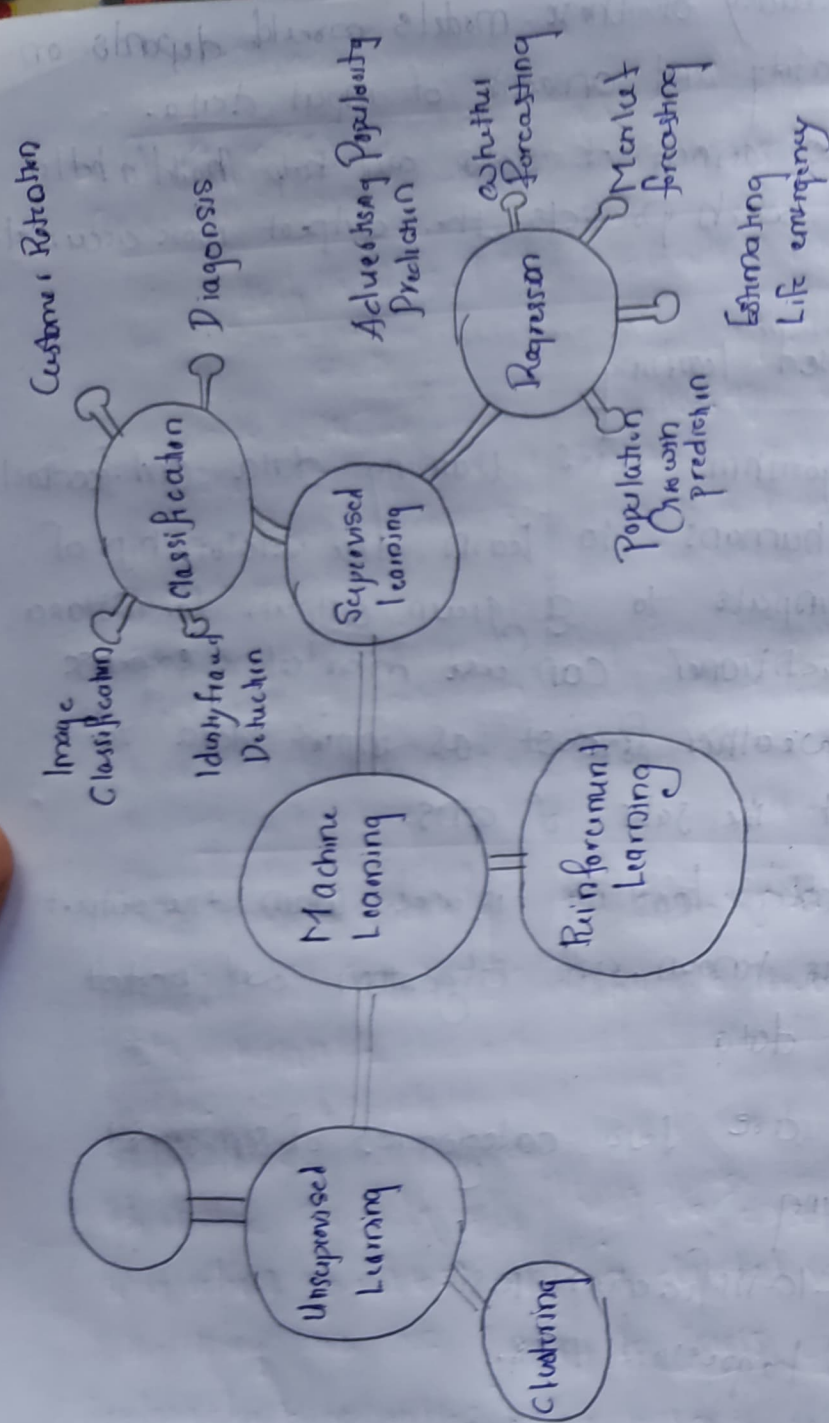
What is Machine Learning?

- > A machine learning model learns from the historical data fed to it.
- > Builds prediction algorithms to predict the output for the new set of data.

- > The accuracy of these models would depend on the quality and amount of input data.
- > A large amount of data will help build a better model which predicts the output more accurately.

Supervised learning

- An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output. For instance a practitioner can use marketing expense and weather forecast as input data to predict the sales of cars.
- Supervised learning is used when the output data is known. The algorithm will predict new data.
- There are two categories of supervised learning
 - i) Classification tasks
 - ii) Regression task.



• Classification

Imagine you want to predict the gender of a customer for a commercial. You will start gathering data on the height, weight, job, salary, purchasing basket etc. From your customer database. You know the gender of each of your customers, it can only be male or female. The objective of the classifier will be to assign a probability of being a male or a female (i.e. label) based on the information (i.e. features you have collected). When the model learned how to recognize male or female, you can use new data to make a prediction. For instance, you just got new information from an unknown customer, and you want to know if it is male or female. If the classifier predicts male = 70%, it means the algorithm is sure at 70% that this customer is male and 30% it is a female.

• The label can be of two or more classes.

Regression

- A regression problem has a real number (a number with a decimal point) as its output.

Height (inches)	Weight (Pounds)
65.38	112.99
71.52	136.49

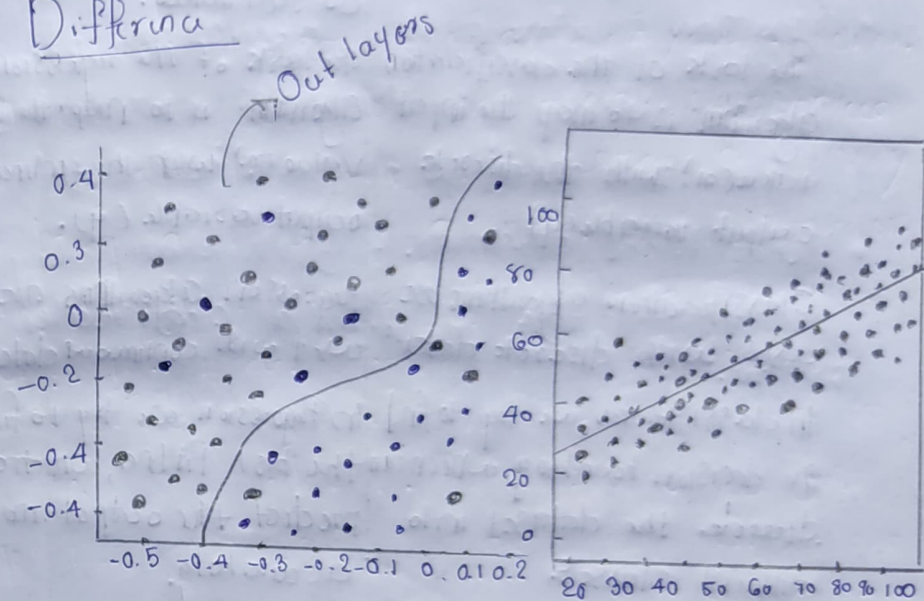
We have to identify an appropriate mapping function of input data (height) to the output data (weight).

Regression

- We have an independent variable (or set of independent variables) and a dependent variable (the thing we are trying to guess our independent variables). For instance, we could say height is the independent variable and weight is the dependent variable.

Also, each row is called an example, observation or data point, while each column (not including the label/dependent variable) is often called a predictor, dimension, independent variable, or a feature.

Difference



Classification

Regression

Difference Summary

Classification	Regression
The output variable must be a discrete value.	The output variable must be of continuous nature or real value.
The task of the classification algorithm is to map the input value (x) with the discrete output variable (y).	The task of the regression algorithm is to map the input value (x) with the continuous output variable (y).
Classification algorithms are used with discrete data.	Regression algorithms are used with continuous data.
In classification, we try to find the decision boundary, which can divide the dataset into different classes.	In Regression, we try to find the best fit line, which can predict the output more accurately.
Classification Algorithms can be used to solve classification problems such as identification of spam emails, speech recognition, identification of cancer cells, etc.	Regression algorithms can be used to solve the regression problems such as weather predictions, house price predictions, etc.

Supervised Learning Methods:

Algorithm Name	Description
Linear Regression	Finds a way to correlate each feature to the output to help predict future values.
Logistic Regression	Extension of linear regression that's used for classification tasks. The output variable is binary (e.g. only black or white) rather than continuous (e.g. an infinite list of potential colors).
Decision Tree	Highly interpretable classification or regression model that splits data - feature values into branches at decision nodes (e.g. if a feature is a color, each possible color becomes a new branch) until a final decision output is made.
Naive Bayes	The Bayesian method is a classification method that makes use of the Bayesian theorem. The theorem updates the prior knowledge of an event with the independent probability of each feature that can affect the event.

Unsupervised Learning Methods

- K-means clustering Puts data into some groups (K) that each contains data with similar characteristics (as determined by the model, not in advance by humans).
- Gaussian Mixture model A generalization of K-means clustering that provides more flexibility in the size and shape of groups (clusters).
- Hierarchical clustering: - Splits clusters along a hierarchical tree to form a classification system.