

## **Disclaimer**

- This presentation is purely for academic purpose and does not carry any commercial value.
- All images and photos used in this presentation are property of respective image holder(s) and due credit is provided to them. Images are used only for indicative purpose and does not carry any other meaning.
- All information and data in this slide are collected from open domain.

MANISH GODSE, Ph.D.(IIT Bombay)

Welcome



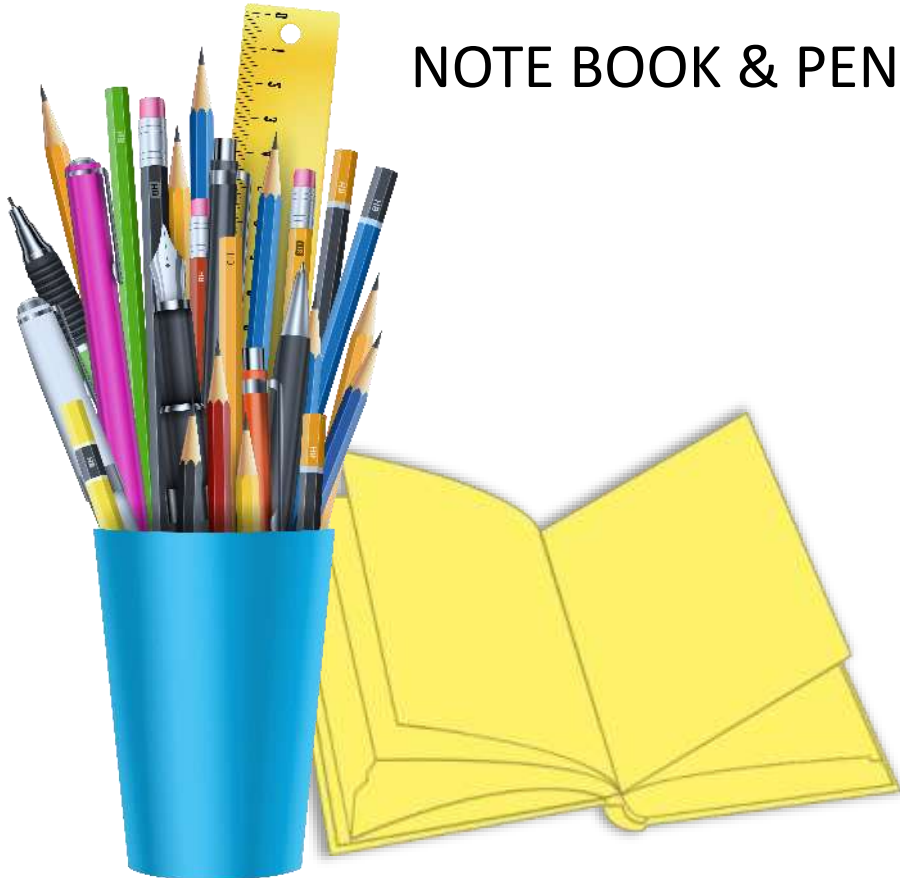
# Request & Instructions

---



# PLEASE OPEN

NOTE BOOK & PEN



CALCULATOR



LAPTOP OR DESKTOP,  
IF YOU HAVE.





# PLEASE FOLLOW THIS

# SILENCE



MUTE  
MIC



# RAISE HAND



NO  
CHAT



SILENT  
MODE



# INTRODUCTION TO MACHINE LEARNING



# BOOKS & REFERENCES

- XXX
- <https://machinelearningmastery.com/model-based-outlier-detection-and-removal-in-python/>

# Table of Contents

1. What is Learning?
2. Machine Learning
3. When to use Machine Learning?
4. Machine Learning Types





# WHAT IS LEARNING?

---

1



# WHAT IS LEARNING?

- The activity or process of gaining knowledge or skill by studying, practicing, being taught, or experiencing something.
- The activity of someone who learns.
- Knowledge or skill gained from learning.

---

*(Merriam Webster Dictionary)*

The acquisition of knowledge or skills through study, experience, or being taught.

---

*(Oxford Dictionary)*

The activity of obtaining knowledge.

---

*(Cambridge Dictionary)*

# Brain

**Humans  
have  
capability  
to learn.**



Can we write a  
**computer program**,  
which will work like a  
**human being**?



# MACHINE LEARNING

---

# 2

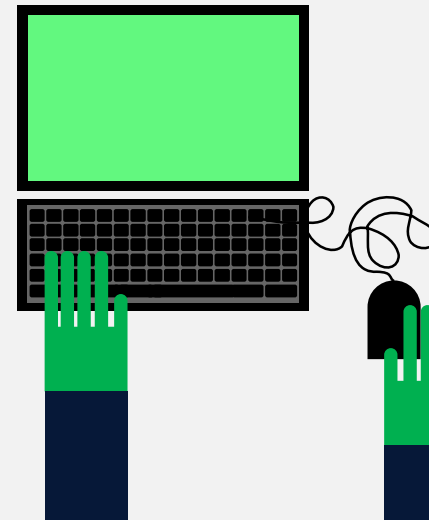


# WHAT IS LEARNING ? AND MACHINE LEARNING?



## LEARNING

The process of getting an understanding of something by studying it or by experience.



## MACHINE LEARNING

Devise learning algorithms that do the learning automatically without human intervention or assistance.

# WHAT IS MACHINE LEARNING?

Machine learning is a method of programming computers where, instead of designing the algorithm to explicitly perform a given task, the machine is programmed to learn from a set of examples.

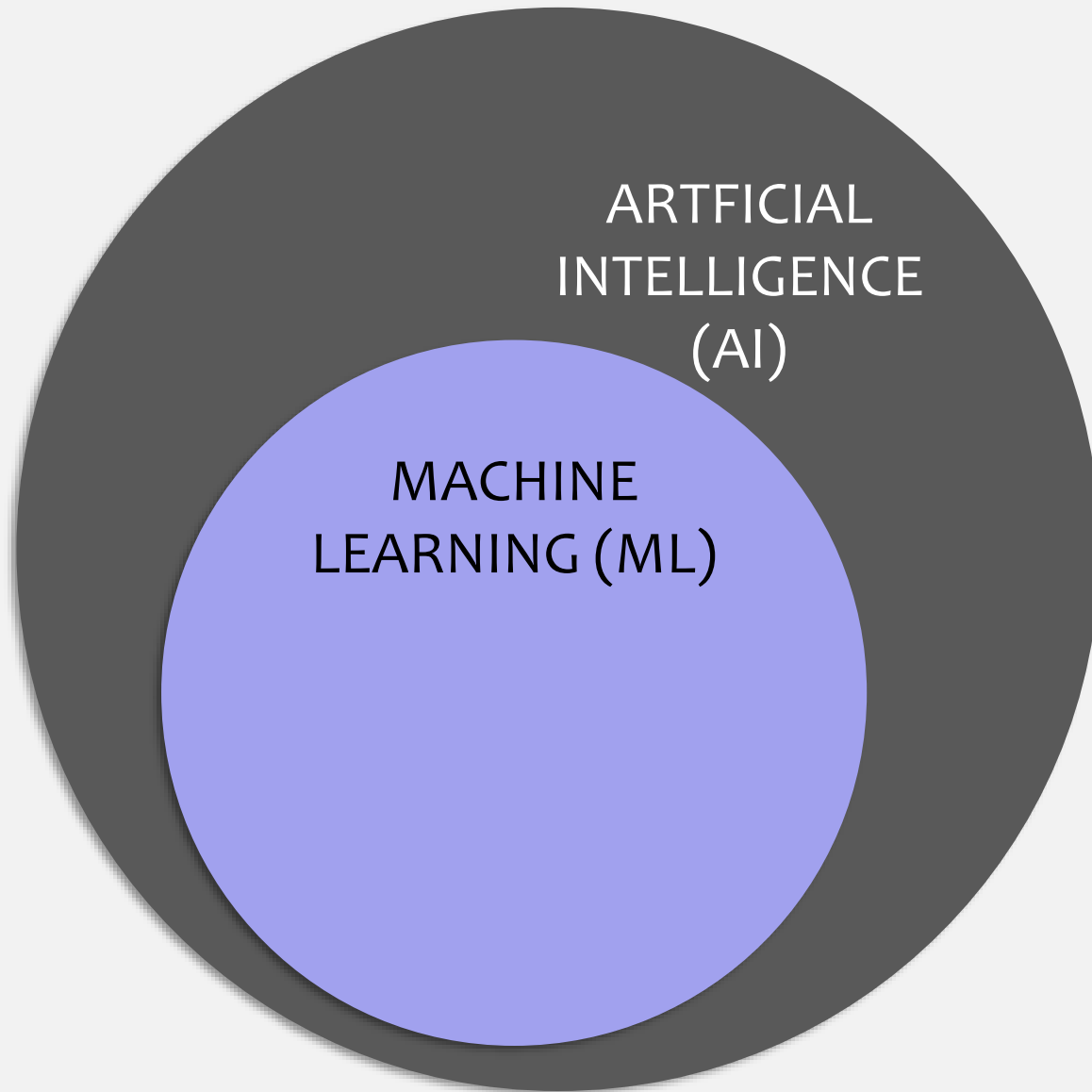
*Reference:* Introduction to Artificial Intelligence, CS 271, Fall 2007, Professor Padhraic Smyth, UCIrvine

---

The emphasis of machine learning is on **automatic** methods. In other words, the goal is to devise learning algorithms that do the learning automatically without human intervention or assistance.

*Reference:* [www.cs.princeton.edu/courses/archive/spr08/cos511/scribe\\_notes/0204.pdf](http://www.cs.princeton.edu/courses/archive/spr08/cos511/scribe_notes/0204.pdf)

# What is Machine Learning?



In 1959, Arthur Samuel defined machine learning as a -

*"Field of study that gives computers the ability to learn without being explicitly programmed"*



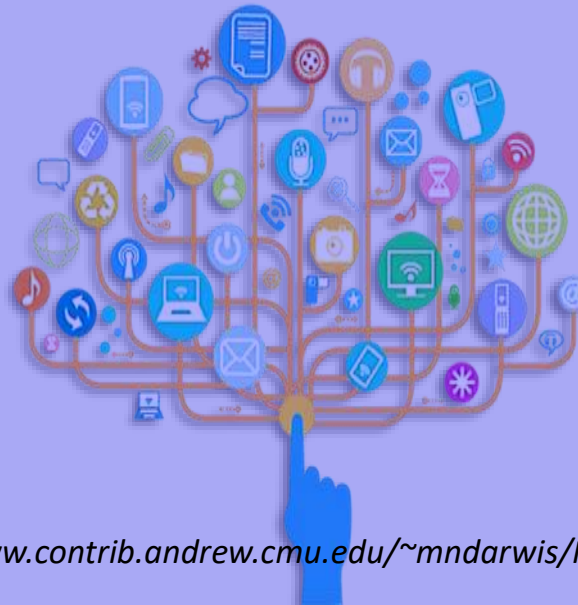
# ADVANTAGES AND DISADVANTAGES

## ADVANTAGE

- Faster than human beings
- Teaches itself from the data
- Discover features
- Discover patterns

## DISADVANTAGE

- Not suitable for every use case
- Requires lots of data
- Quality of data impacts outcome
- Need of complex computing and technology



# WHEN TO USE MACHINE LEARNING?

---

# 3



# DATA AND AUTOMATION - HEART OF MACHINE LEARNING

Machine Learning focuses on developing programs that teach computers to change when exposed to new data and to grow. ML is suitable when you are getting new data regularly.

**ADOPT TO  
NEW DATA**

**AUTOMATION**

Its goal is to understand and follow the methods by using algorithms to do that task automatically without any human assistance. ML is suitable when automation is required.

# WHEN TO USE MACHINE LEARNING?

Large  
Data Size

Repetitive Task

Manual  
Limitation Of Data  
Processing &  
Analysis

Automated  
Model  
Update &  
Data Pipeline

High Demand  
Of  
Computing  
Power

Limited  
Knowledge &  
Experience  
In Given  
Domain



# MACHINE LEARNING TYPES

---

# 4



# MACHINE LEARNING TYPES

## **SUPERVISED LEARNING**

When training data has examples of the input vectors along with their corresponding target vectors, it is known as supervised learning.

---

## **UNSUPERVISED LEARNING**

When training data consists of a set of input vectors without any corresponding target values, it is known as unsupervised learning.

---

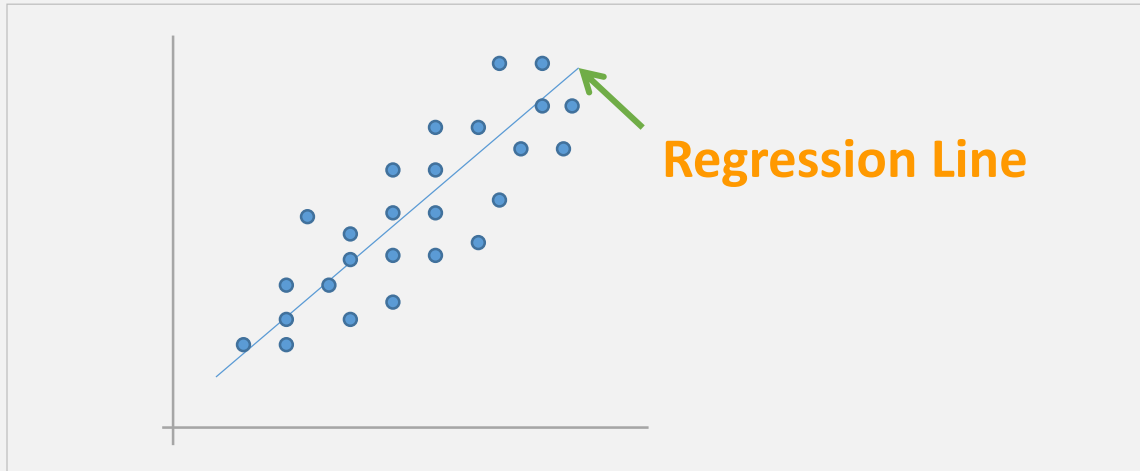
## **SEMI-SUPERVISED LEARNING**

Semi-supervised learning uses both labeled and unlabeled data to perform an otherwise supervised learning or unsupervised learning task.

# SUPERVISED LEARNING

- The goal of supervised learning is to **predict** an outcome based on a set of variables.
- In supervised learning, the outcome variable **supervises** or guides the process of learning and to predict future predictions.
- A **prediction** task requires the estimation of the value for a **continuous** outcome (e.g., sales revenue). Linear regression is one of the supervised learning approaches from classical statistics. It has data of a quantitative outcome (the dependent y-variable) and one or more corresponding variables (the independent x-variables) are used to model an equation for estimating y values.
- A **classification** task requires the identification of the value for a **categorical** outcome (e.g., loan default or no loan default).

# SUPERVISED LEARNING



For each observation of the predictor measurement(s)  $X_i$ ,  $i = 1, \dots, n$  there is an associated response measurement  $Y_i$ .

We wish to fit a model that relates the response to the predictors, with the aim of accurately predicting the response for future observations (prediction) or better understanding the relationship between the response and the predictors (inference).

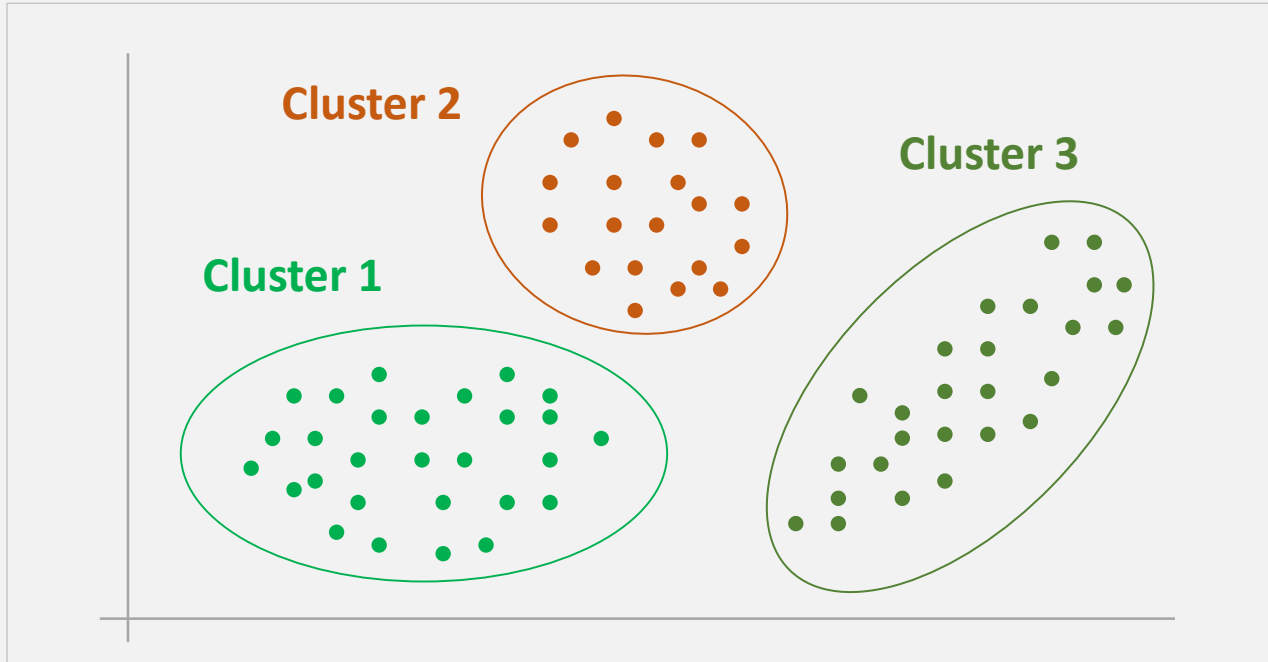
- Training data includes both the **input** and the desired **result**.
- **Data** is used to prepare **model** and model is trained till desired level of **accuracy** is achieved. Then trained model is used for intended purpose.
- E.g. Regression, Classification
- Classify email as spam/not-spam or predict stock price.



# UNSUPERVISED LEARNING

- The unsupervised learning approach doesn't focus to predict an output value.
- The unsupervised learning used to detect **patterns** and **relationships** in the data.
- The unsupervised learning tasks may be **clustering** of observations and developing **association rules** between items in an observation.

# UNSUPERVISED LEARNING



- Input data is not labelled and does not have a known result.
- A model is prepared by inferring structures present in the input data.
- e.g. association mining, k-means clustering, association

Reference: <http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>

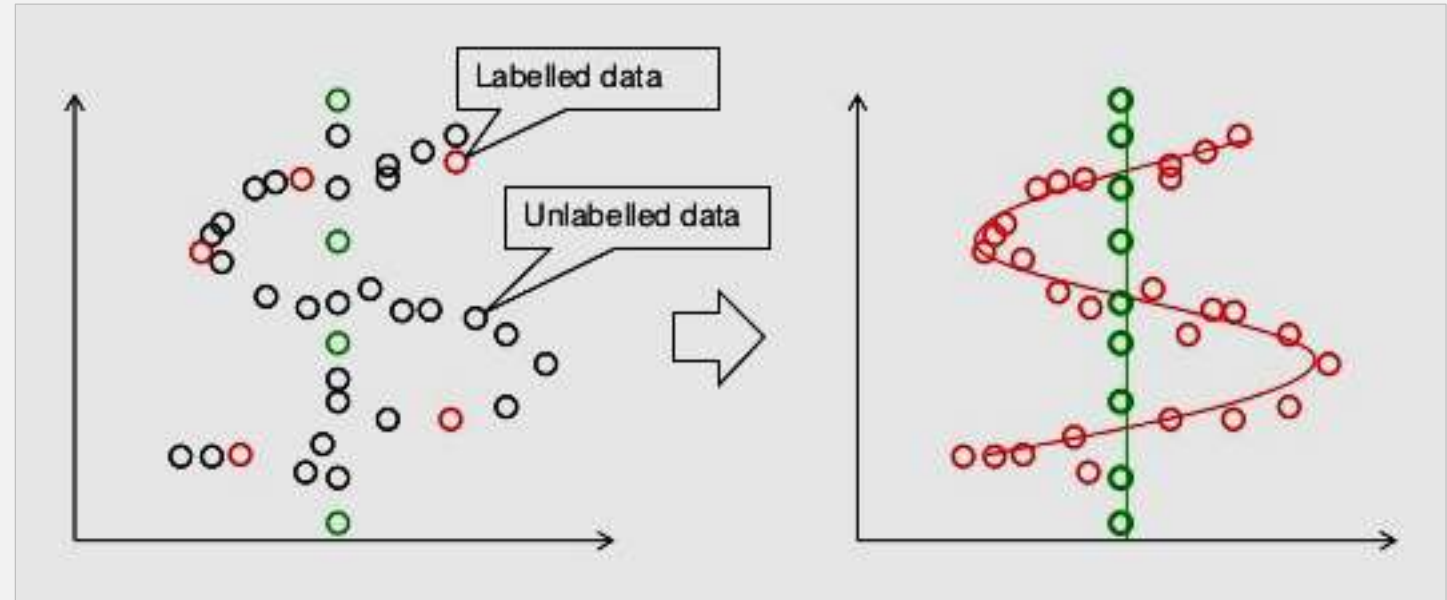
In unsupervised learning, for every observation  $i = 1, \dots, n$ , we observe a vector of measurements  $X_i$  but no associated response  $Y_i$ . It is not possible to fit a linear regression model, since there is no response variable to predict.

In this situation, we are working blind to get some output; the situation is referred to as unsupervised because we lack a response variable that can supervise our analysis.

# SEMI-SUPERVISED LEARNING

Semi-supervised learning uses both labeled and unlabeled data to improve supervised learning.

This learning is in between supervised and unsupervised learning.



Reference: <https://www.slideshare.net/alembert2000/130201-deep-learning-130131220>

- In semi-supervised learning, input data is large but only some of them are labelled.
- E.g. Photo archive where some of the images are labeled, (e.g. name, date) and the rest images are unlabeled.
- Distribution of labelled examples should be well distributed to get the better model.

# SUMMARY

	Supervised	Unsupervised	Semi-supervised
Training Data	Labeled	Unlabeled	Mix of Labeled and Unlabeled
Algorithm Purposed	Predict the output from input data	Find patterns	Mixture of supervised and unsupervised learning

# QUESTION AND ANSWERS

