Applied Machine Learning

BITS Pilani
Pilani Campus

Dr. Harikrishnan N B
Computer Science and Information Systems

innovate achieve lead

BITS Pilani

Pilani Campus

SE ZG568 / SS ZG568, Applied Machine Learning Lecture No. 1

Quest for Learning from Data

Image Courtesy: https://earthsky.org/human-world/johannes-kepler-birthday-dec-27-1571/
https://www.britannica.com/biography/Tycho-Brahe-Danish-astronomer

Quest for Learning from Data

Kepler (1571-1630)

Tycho Brahe (1546-1601)

Image Courtesy: https://earthsky.org/human-world/johannes-kepler-birthday-dec-27-1571/
https://www.britannica.com/biography/Tycho-Brahe-Danish-astronomer

Quest for Learning from Data

Kepler (1571-1630)

Analysis from the data:

Heliocentric Model (Copernicus's theory)

Orbits of the planets are elliptical

(Note: That was a time when telescopes were not invented, calculus, regression methods were not established.)

Image Courtesy: https://earthsky.org/human-world/johannes-kepler-birthday-dec-27-1571/
https://www.britannica.com/biography/Tycho-Brahe-Danish-astronomer

Quest for Learning from Data

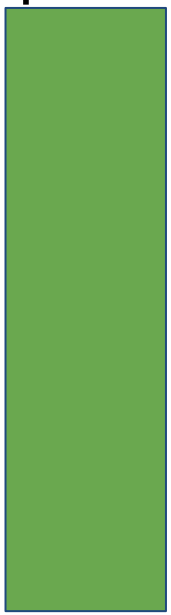
What Kepler did was from the astronomical observations (data)

1. Found patterns and validated the heliocentric model
2. Said something new about elliptical orbit of planets.

Data Driven Insights

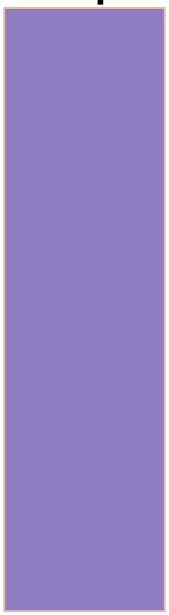
Four Pillars of Scientific Research

Empirical

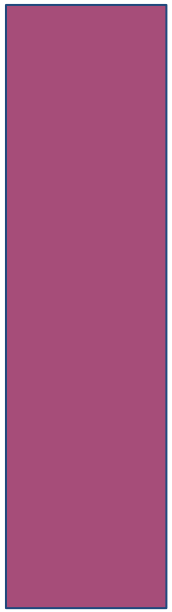


Theoretical

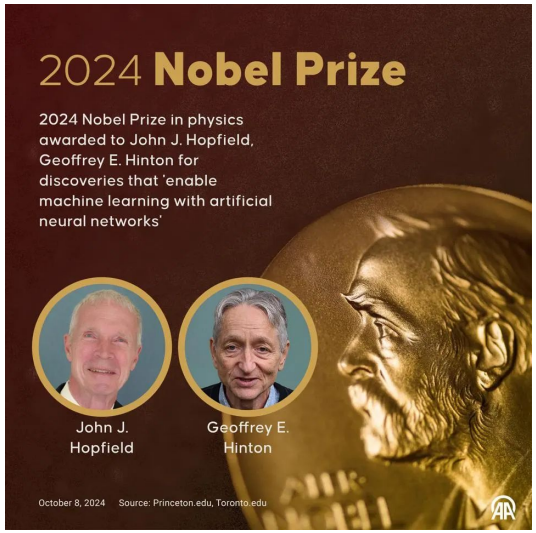
Computational

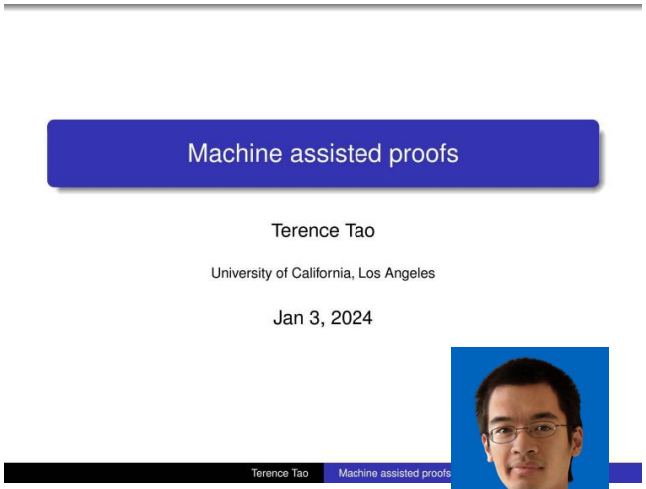


Data Driven



Machine Learning: Buzzword of the 21st century





[Video]

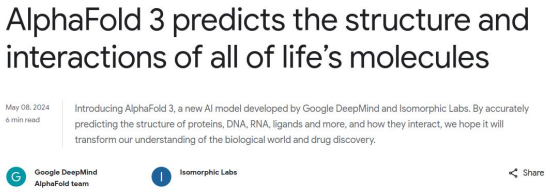
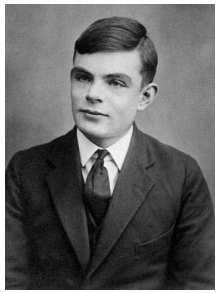


Image Reference:
https://www.forbes.com/sites/petercohan/2024/10/08/neural-networks-co-inventor-hinton-wins-physics-nobel-fears-ai/

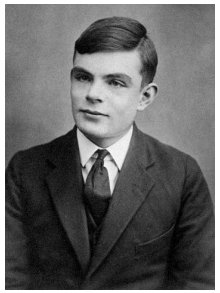
Roots of Artificial Intelligence

1950: Can Machines Think?*

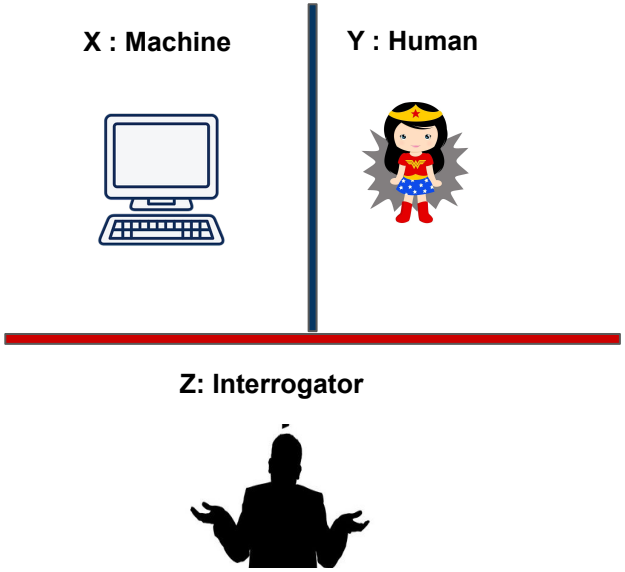


Roots of Artificial Intelligence

1950: Can Machines Think?*



Alan M. Turing



Samuel's 1952 checkers player**

*__Ref:__ Turing, Alan Mathison. "Computing Machinery and Intelligence." *Mind* 59.236 (1950): 433-460.
**__Ref__: Samuel's 1952 Checker Player: http://www.incompleteideas.net/book/ebook/node109.html

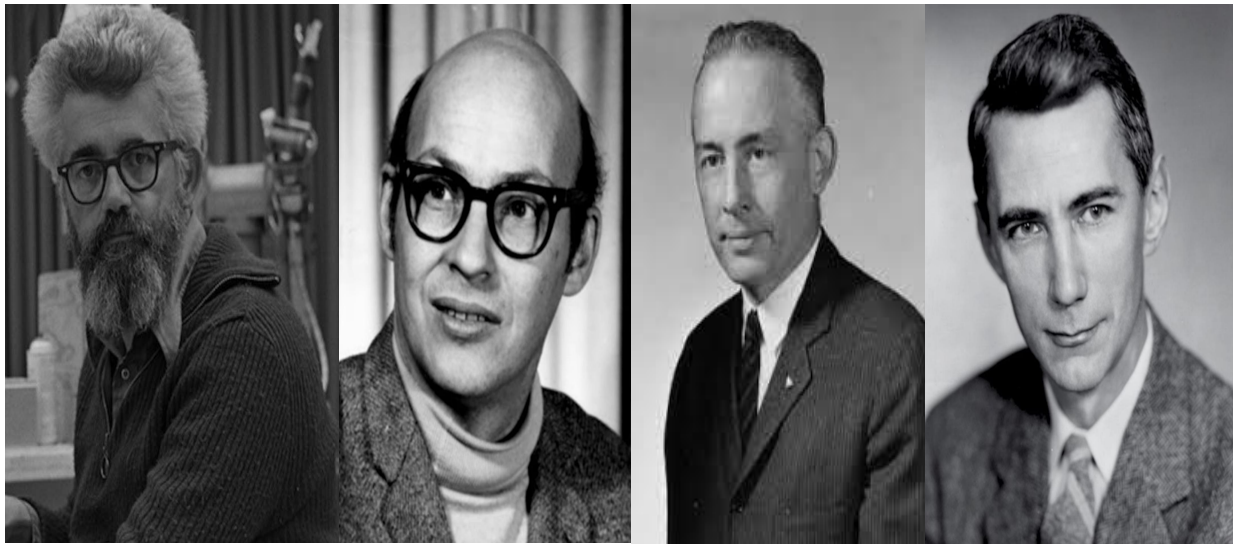
Image Source:
https://engineering.stanford.edu/news/stanfords-john-mccarthy-seminal-figure-artificial-intelligence-dead-84
https://www.semanticscholar.org/paper/Marvin-Minsky-the-father-of-AI-Verwijnen/ef31300364bb686178d7013f8d53ae91da26bac9
https://modha.org/2006/09/nathaniel-rochester-iii-1919-2001/
https://aeon.co/essays/how-a-polymath-transformed-our-understanding-of-information

When was the term Artificial Intelligence Coined?

Who coined the term Artificial Intelligence?

Birth Of Artificial Intelligence

1956: Dartmouth Summer Research Project on Artificial Intelligence**



John McCarthy | Marvin Minsky | Nathaniel Rochester | Claude E. Shannon

The proposal for this project was initiated by **August 31, 1955** authored by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. The Dartmouth Summer Research Project happened in 1956.

Ref: McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4), 12-12.

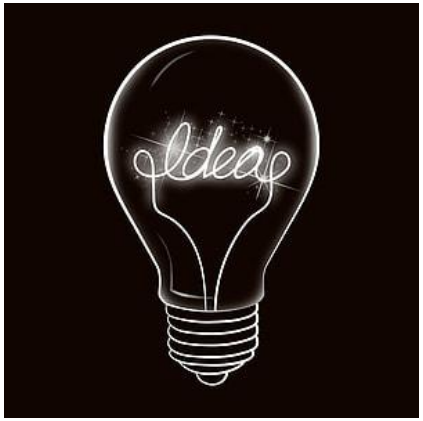
What was the goal of the Dartmouth Summer Research Project on AI?

The study is to proceed on the basis of the **conjecture** that **every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.**

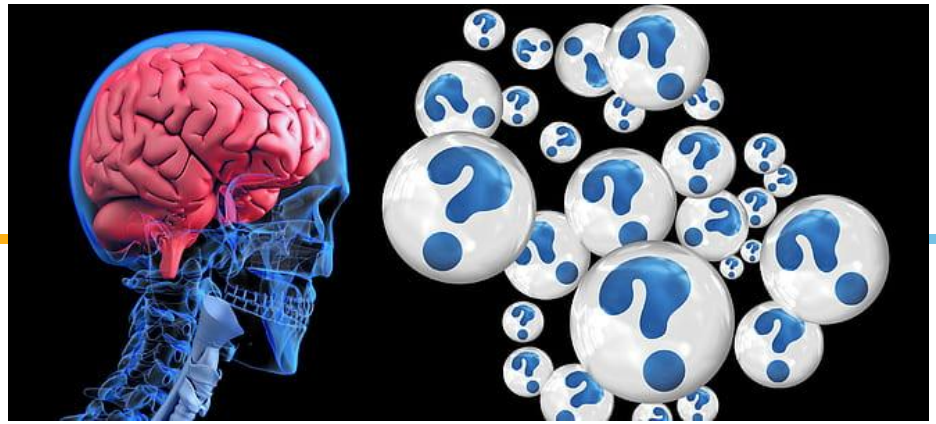
An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.

**Ref: McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4), 12-12.

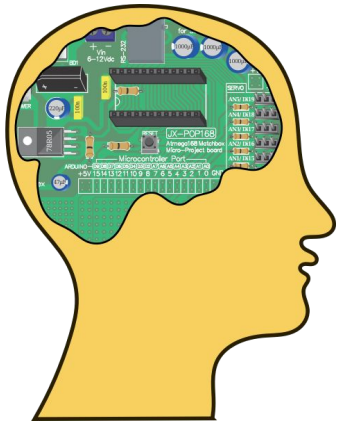
Ideas Proposed in the Conference

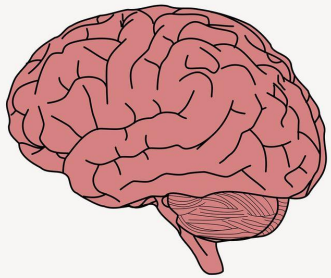


1. Natural Language Processing
2. Neural Networks
3. **Machine Learning**
4. Abstract concepts and Reasoning
5. Creativity and Randomness

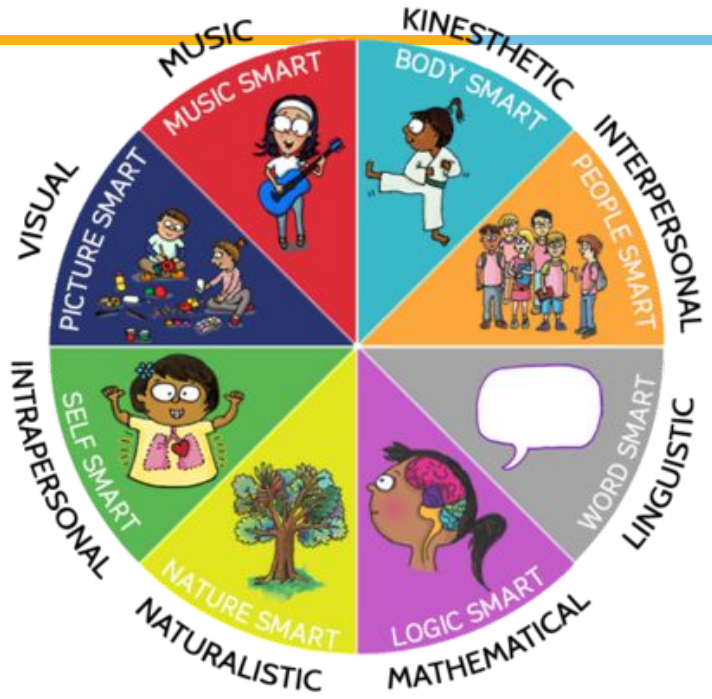


Soon… Clashes of Ideas.. Struggle in Defining Intelligence





complexity involved in understanding intelligence



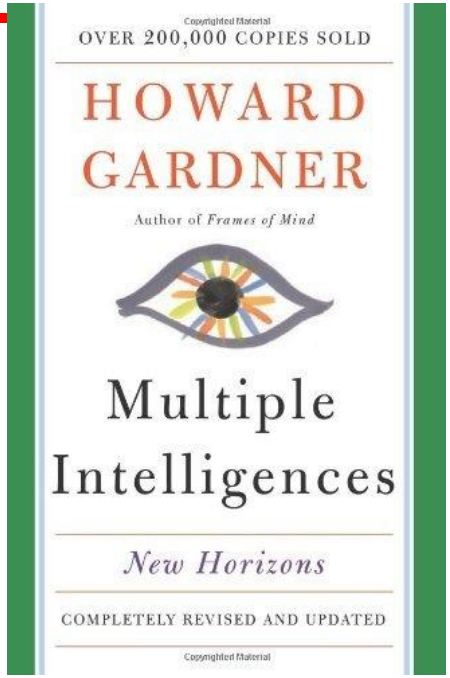


Image Source:
https://projectrangeet.com/howard-gardners-theory-of-multiple-intelligences-and-why-educators-should-be-using-the-lockdown-to-childrens-advantage/

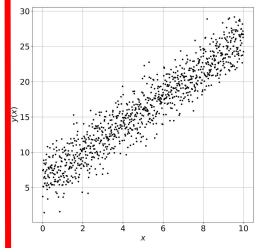
AI is a field that includes a broad set of approaches, with the goal of creating machines with intelligence.

Artificial Intelligence: An Anarchy of Methods

Symbolic AI

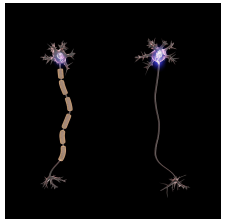
Logic
Fuzzy Logic
Decision Trees
(ID3)

Learning from Data



- Random Forest
- k-Nearest Neighbours
- AdaBoost
- Support Vector Machine
- Naive Bayes, etc.

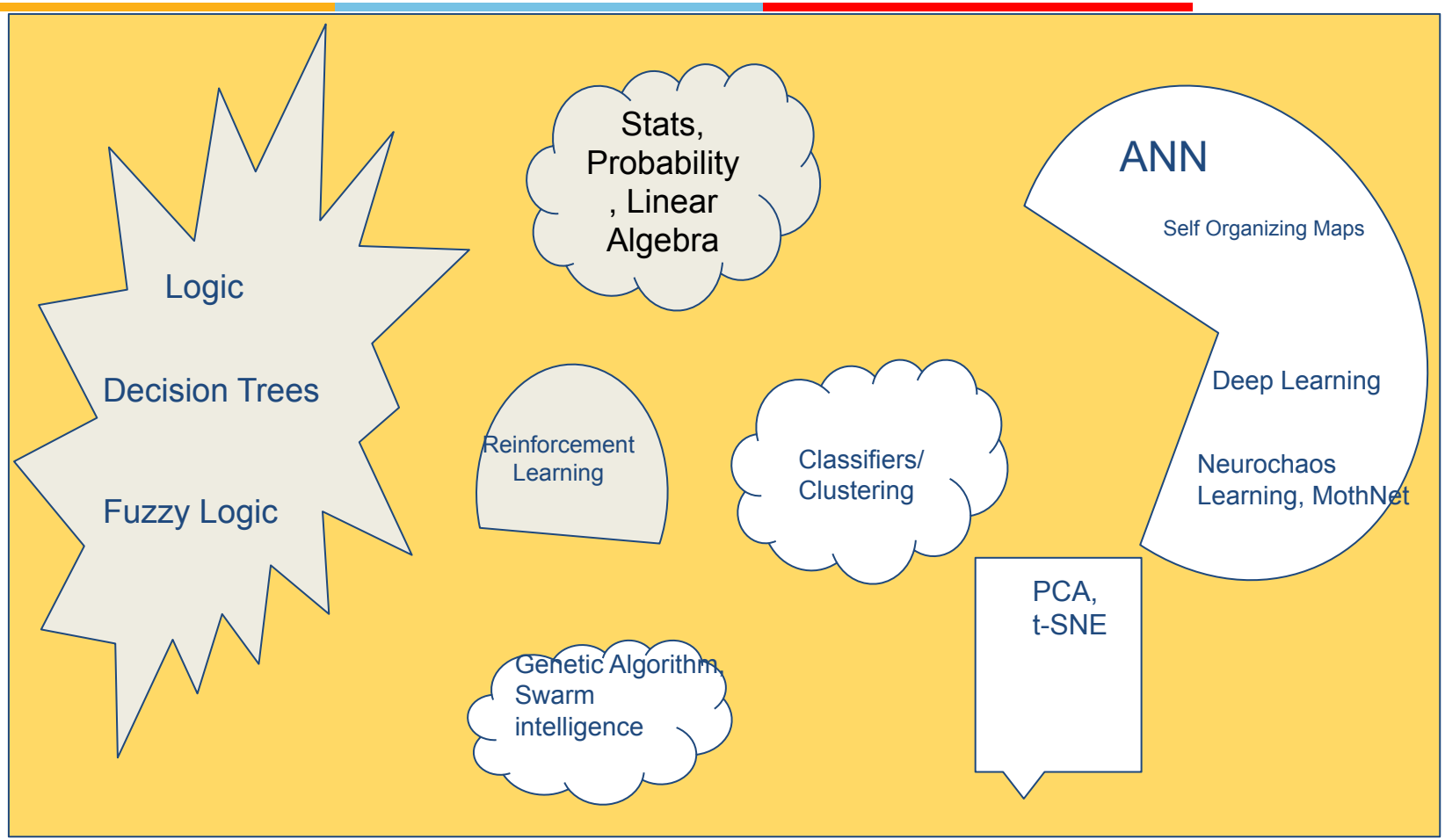
Sub-symbolic AI



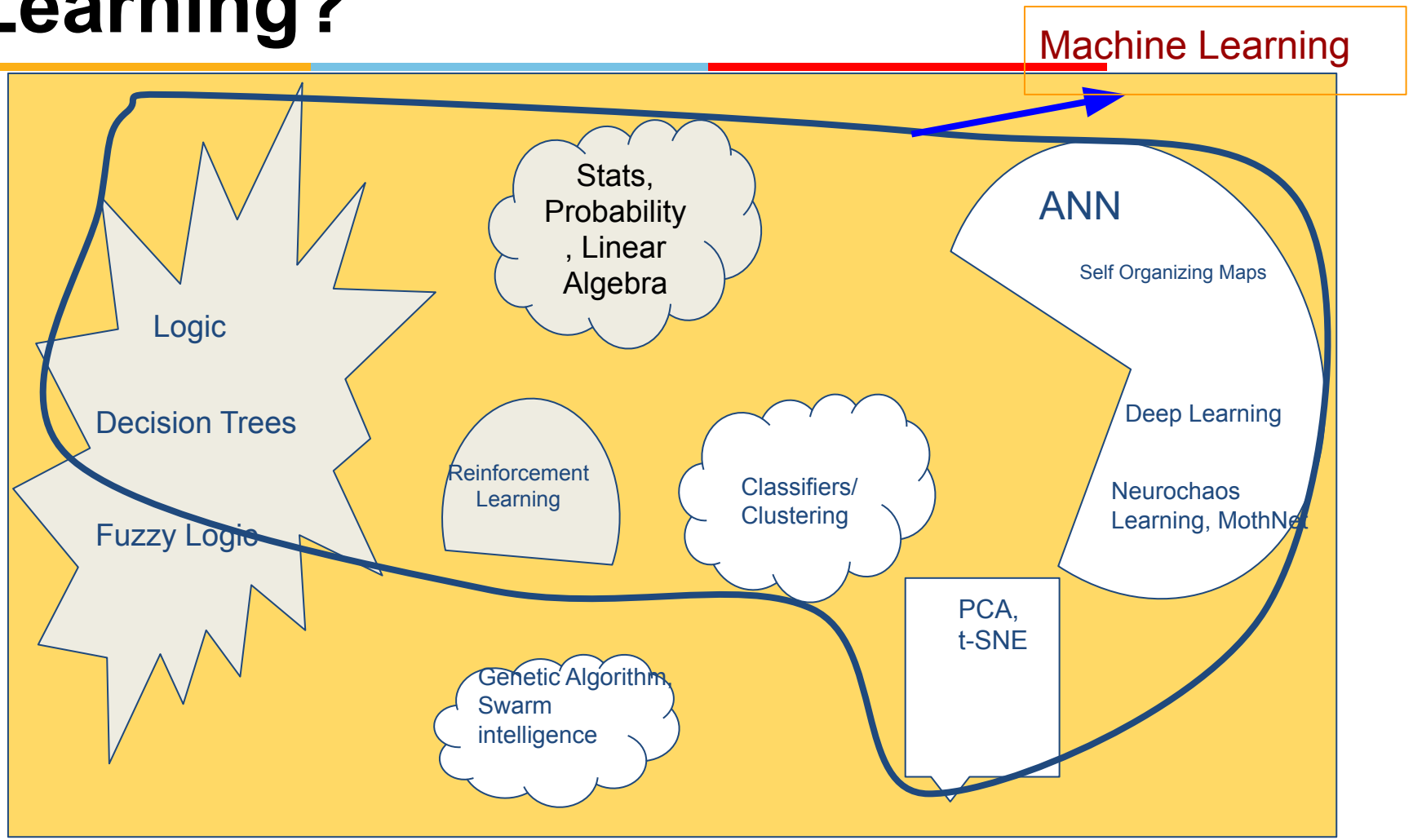
- Deep Neural Networks
- Recurrent Neural Network
- Convolutional Neural Networks
- Long Short Term Memory,
- MothNet,
- Neurochaos Learning

- **Ref:** Lehman, Joel, Jeff Clune, and Sebastian Risi. "**An anarchy of methods: Current trends in how intelligence is abstracted in ai.**" *IEEE Intelligent Systems* 29.6 (2014): 56-62.
- **Ref:** James, Gareth, et al. "**An introduction to statistical learning**." Vol. 112. New York: springer, 2013.

AI? Machine Learning? Deep Learning?

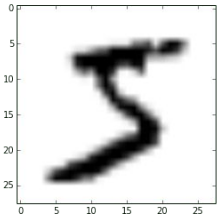


AI? Machine Learning? Deep Learning?

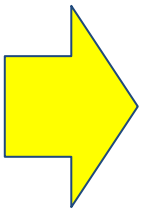


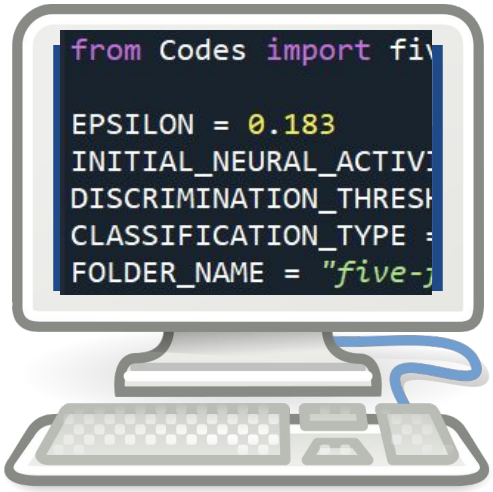
Traditional Computer Science vs. Machine Learning

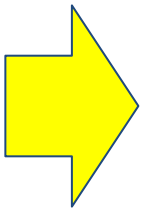
Traditional Computer Science



DATA

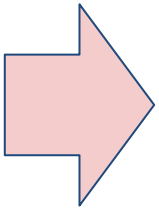




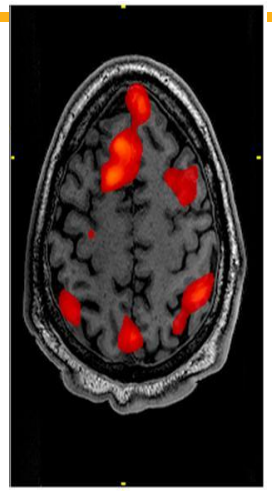


OUTPUT **5**

PROGRAM



Real World Problem

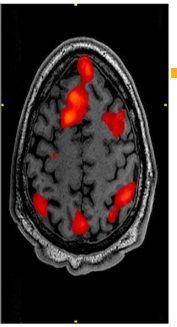


DATA

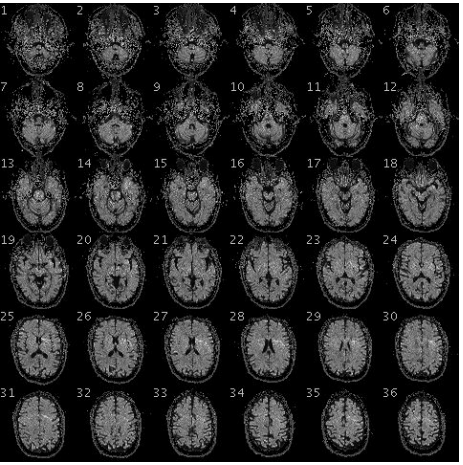
OUTPUT

PROGRAM

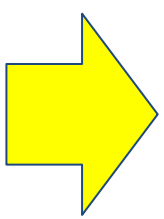
The Machine Learning Way

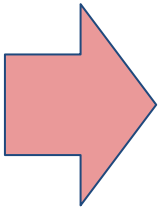


?



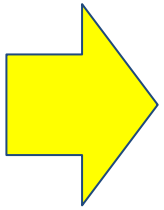
DATA



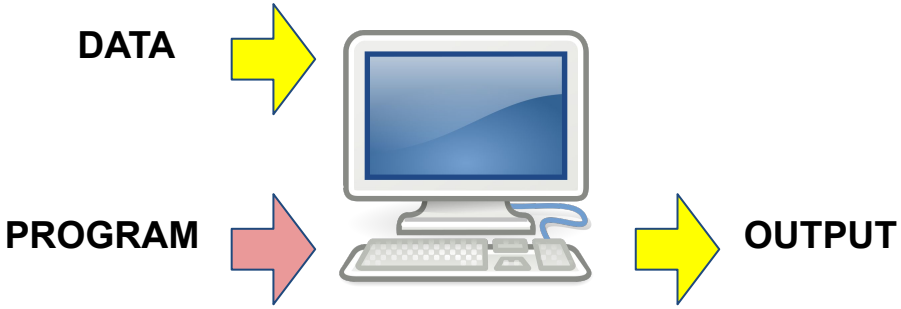


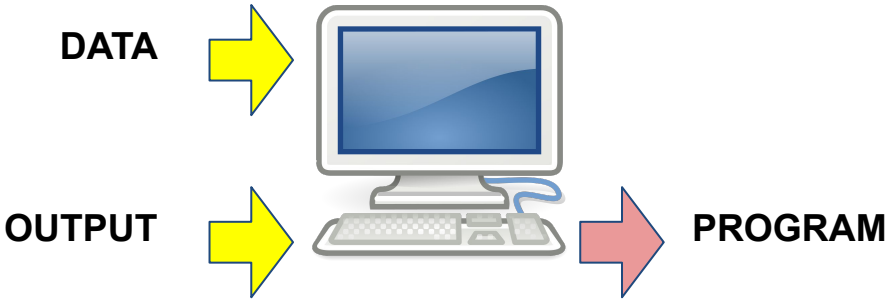
PROGRAM

OUTPUT

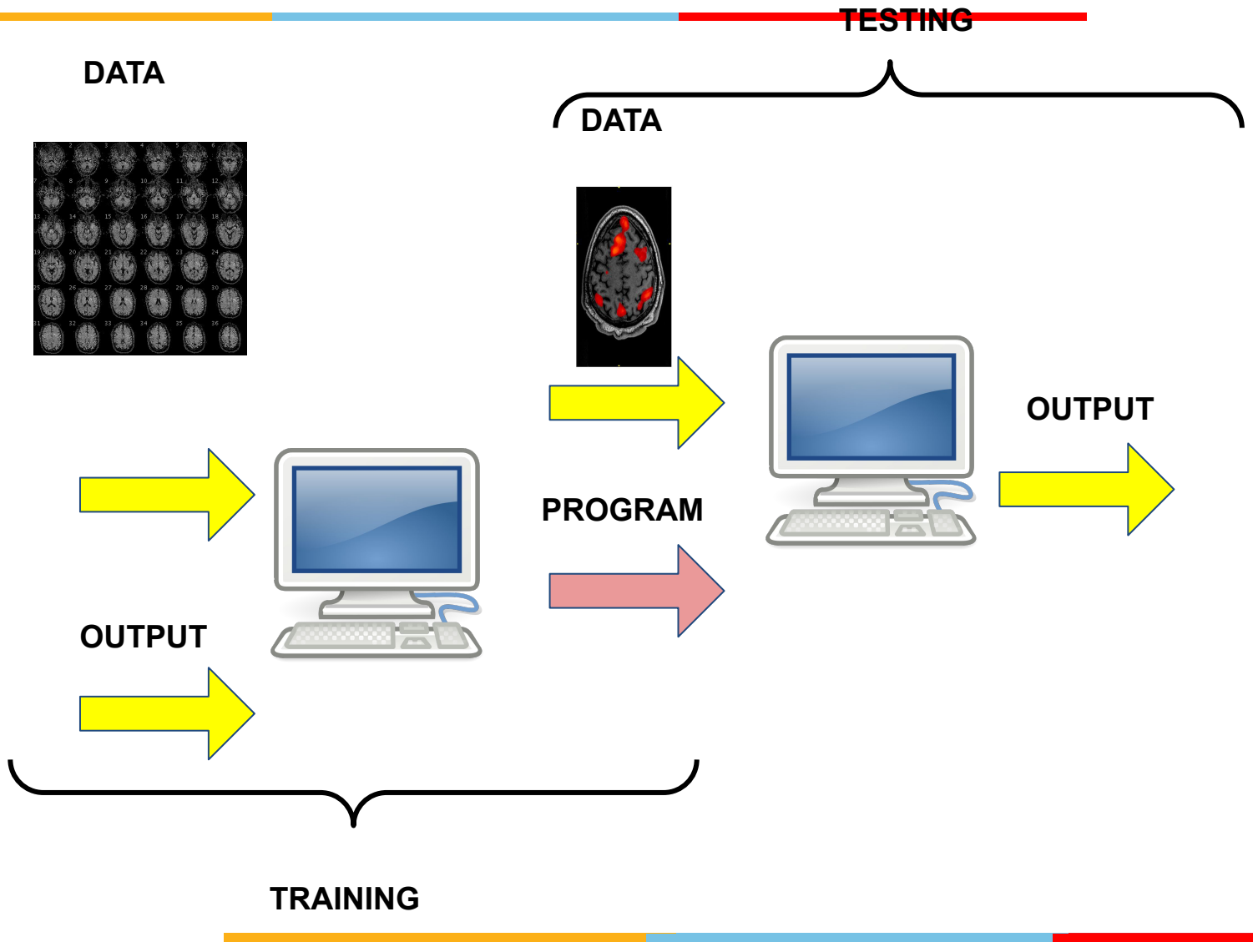


Traditional Computer Science vs. Machine Learning





Supervised Learning



Machine Learning

Study of algorithms that

- Improve their **performance** P
- At some **task** T
- With **experience** E

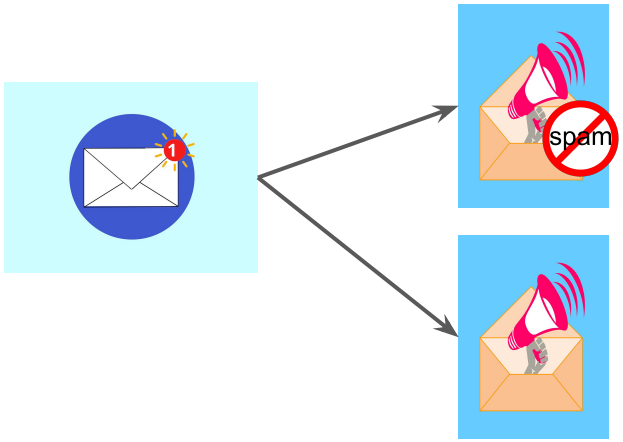
Well-defined learning task <P,T,E>

Spam Filter

Study of algorithms that

- Improve their **performance** P
- At some **task** T
- With **experience** E

Well-defined learning task <P,T,E>



Types of Learning

1. **Supervised Learning**
2. **Unsupervised Learning**
3. Semi-Supervised Learning
4. Reinforcement Learning

Supervised Learning

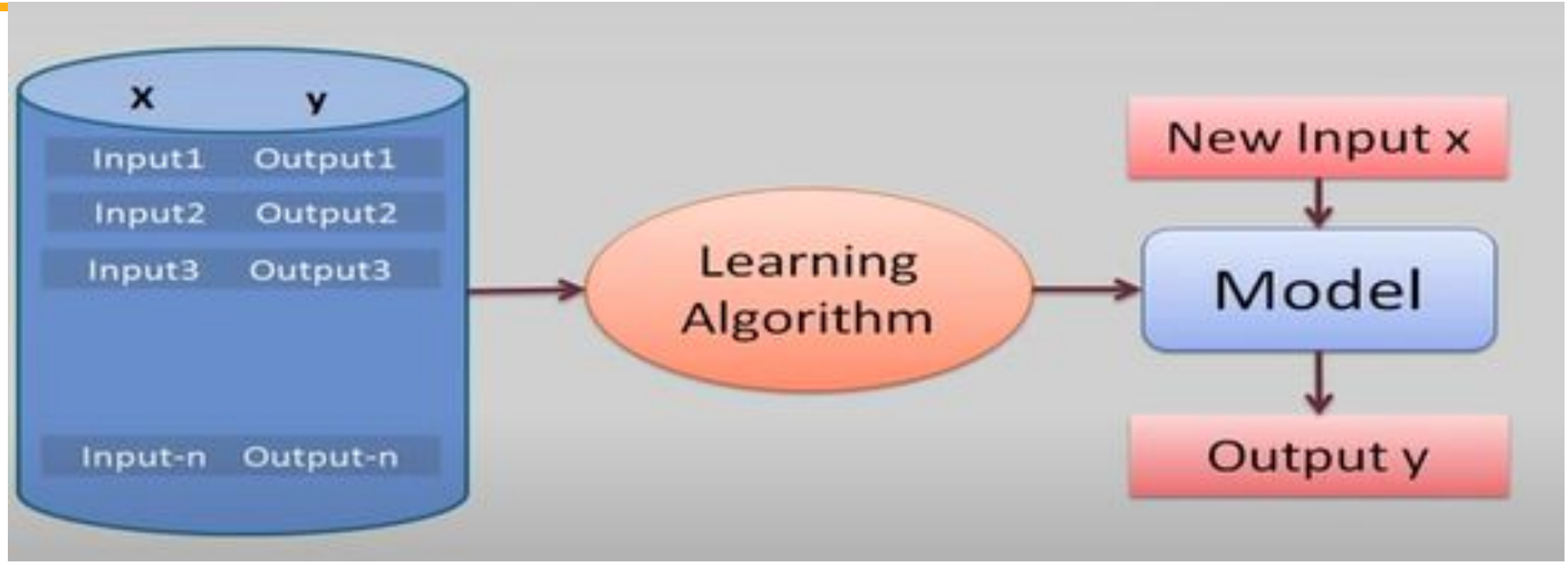


Image Source: Prof. Sudeshna Sarkar's lecture

Unsupervised Learning

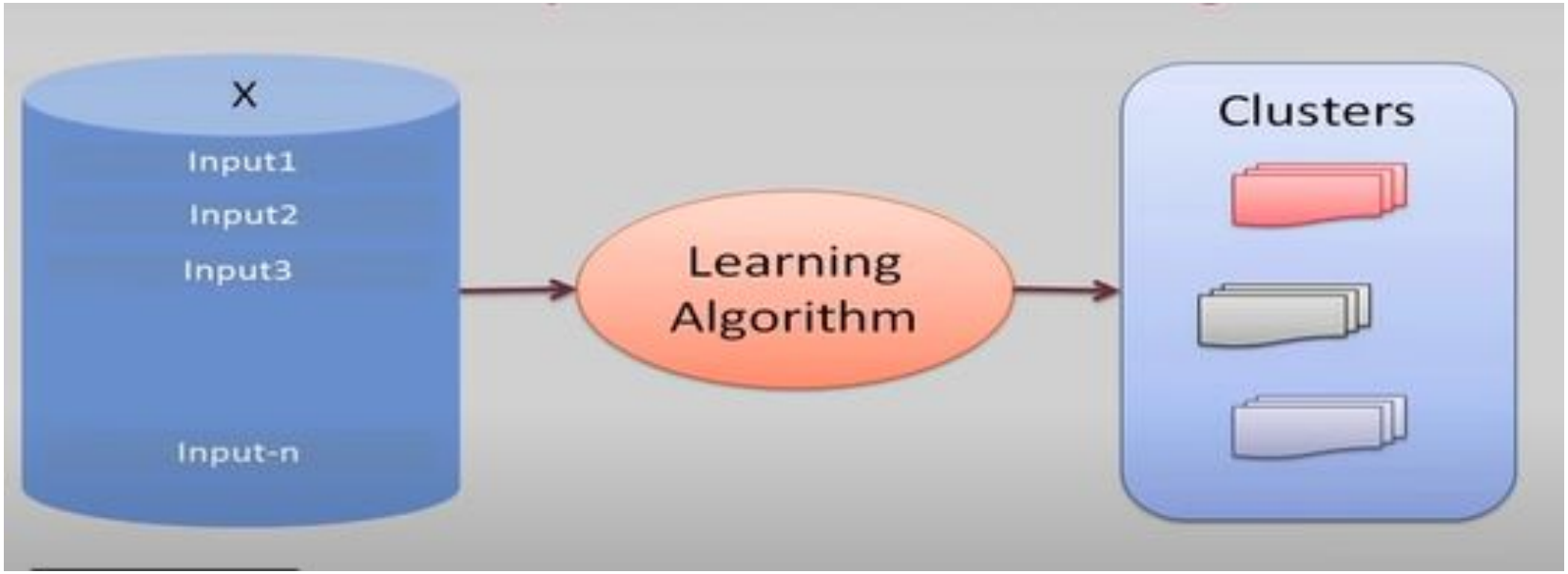


Image Source: Prof. Sudeshna Sarkar's lecture

Semi-Supervised Learning

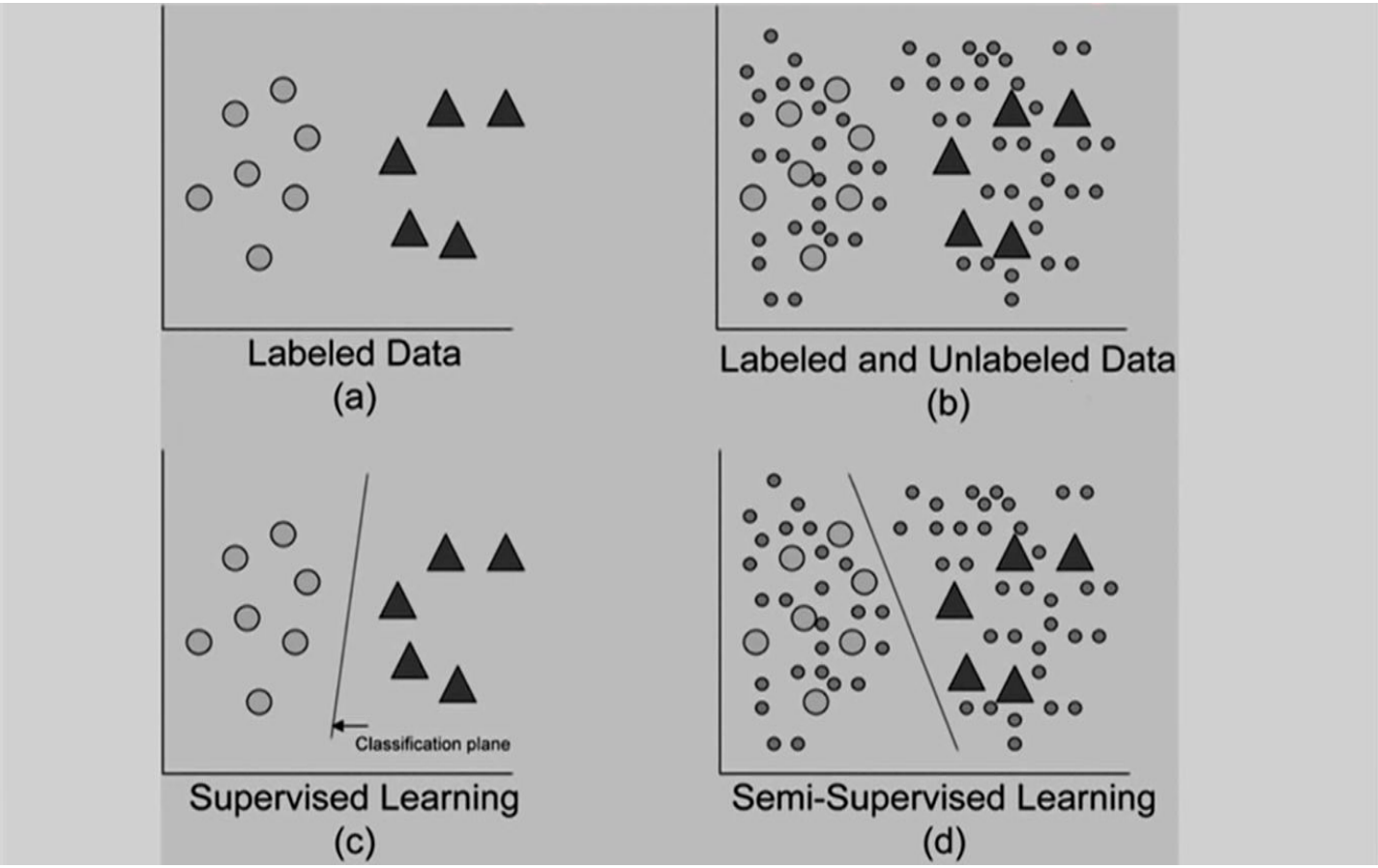
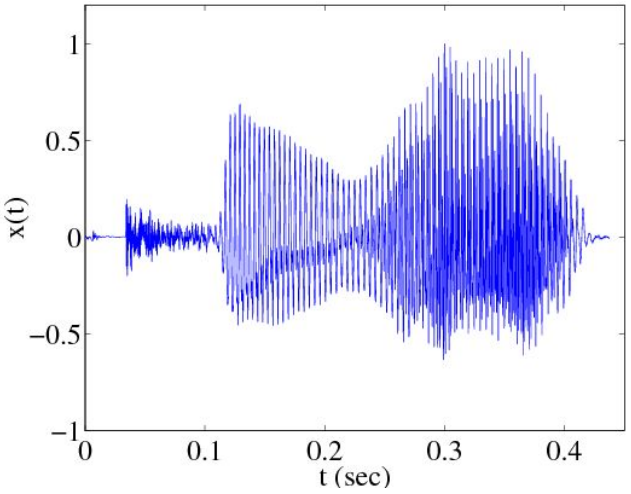
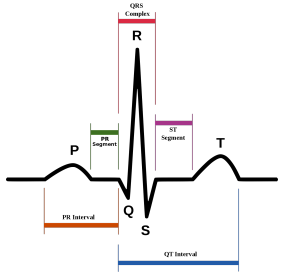


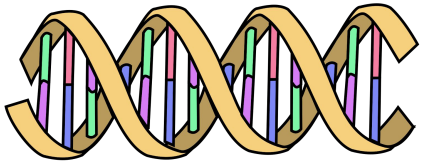
Image Source: Prof. Sudeshna Sarkar's lecture

Data is the king - AI Community

What is **Data?**







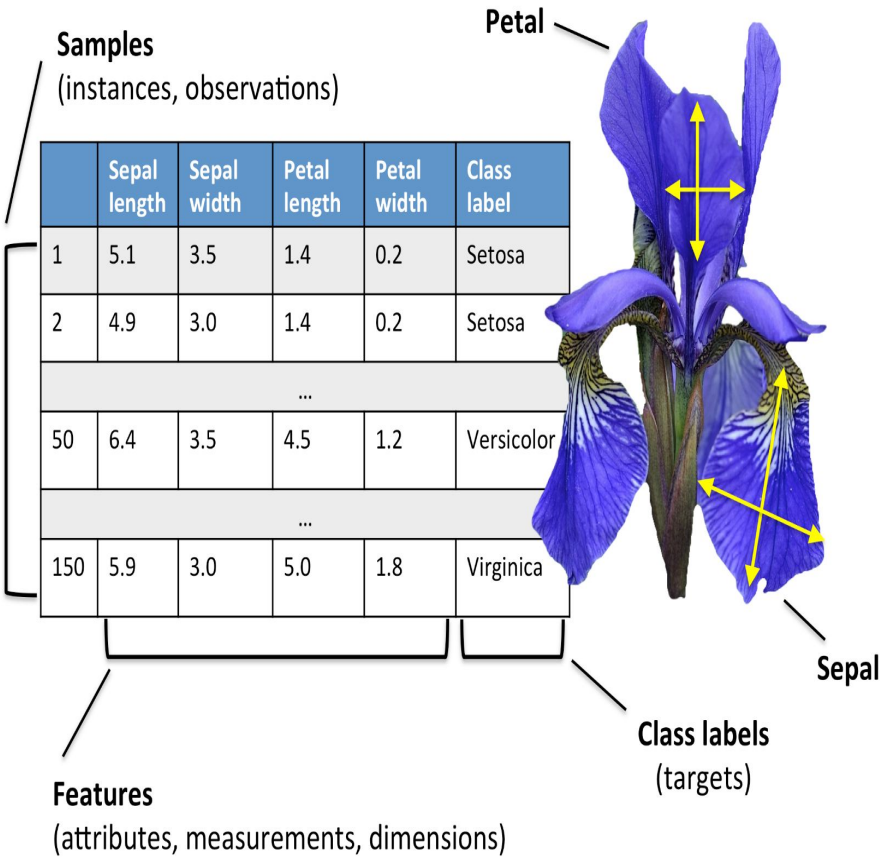
Data is everywhere!

Supervised Learning Data Format

f_1	f_2	….	f_n	Output (Y)
a_1	a_2	….	a_n	y_1
b_1	b_2	….	b_n	y_2
.	.	….	.	.
.	.	….	.	.

- $f_1, f_2, \ldots f_n$ are called features
- Each row is a data instance or a feature vector
- *Y* is the output variable.

Example Dataset:



	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

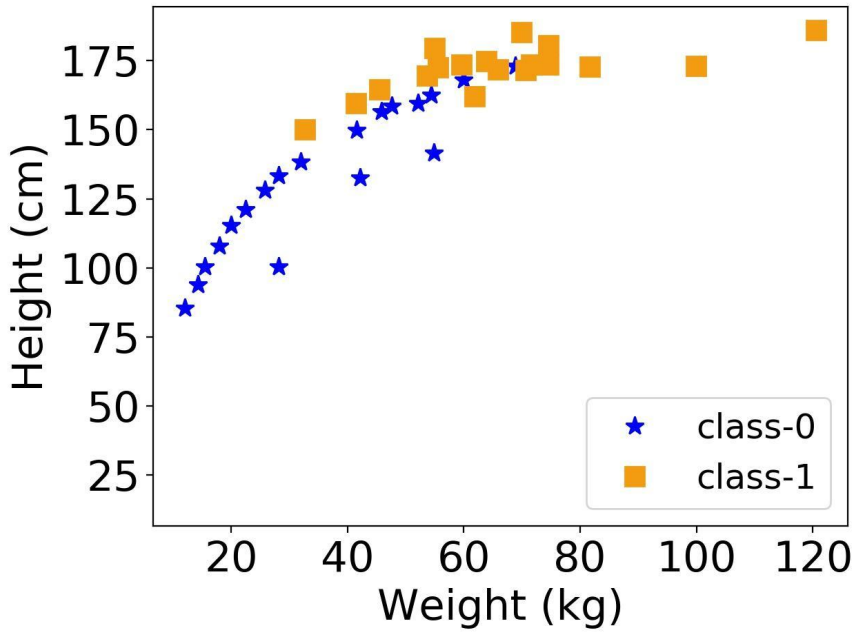
A	B	C	D	E
R&D Spend	Administration	Marketing Spend	State	Profit
165349.2	136897.8	471784.1	New York	192261.83
162597.7	151377.59	443898.53	California	191792.06
153441.51	101145.55	407934.54	Florida	191050.39
144372.41	118671.85	383199.62	New York	182901.99
142107.34	91391.77	366168.42	Florida	166187.94
131876.9	99814.71	362861.36	New York	156991.12
134615.46	147198.87	127716.82	California	156122.51
130298.13	145530.06	323876.68	Florida	155752.6
120542.52	148718.95	311613.29	New York	152211.77
123334.88	108679.17	304981.62	California	149759.96
101913.08	110594.11	229160.95	Florida	146121.95
100671.96	91790.61	249744.55	California	144259.4

Image Source: https://setscholars.net/applied-machine-learning-coding-turicreate-in-python-iris-dataset/

Understanding Classification Via Example

A	B	C
Weight (kg)	Height (cm)	Class
12	85.5	0
14.2	94	0
15.4	100.3	0
17.9	107.9	0
199	115.5	0
22.4	121.1	0
25.8	128.2	0
28.1	133.3	0
28.1	100.4	0
31.9	138.4	0
74.6	173.6	1
71.6	173.6	1
59.6	173.6	1
55.6	172.6	1
74.6	180.5	1
70.6	171.6	1
53.6	169.6	1
45.5	164.6	1
41.44	159.65	1
32.68	149.99	1
54.93	179.62	1

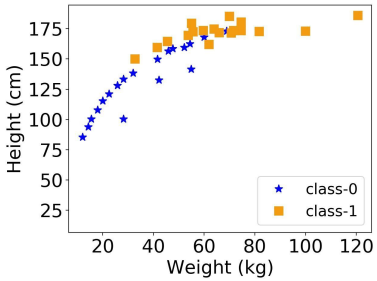
Class-0	below 15 years
Class-1	above 15 years

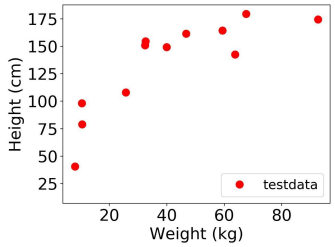


Supervised Learning

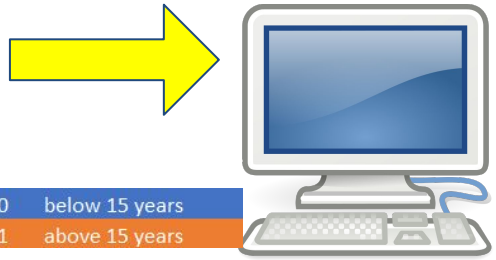
TESTING

TESTDATA





TRAINDATA



OUTPUT

Class-0	below 15 years
Class-1	above 15 years

PROGRAM

OUTPUT

TRAINING

Learning - Function Approximation

Problem Setting

- Set of possible instances *X.*
- Unknown target function *f: X—> Y*
- Set of function hypotheses *H = {h| h: X—> Y}*

Input

Training Examples of unknown target function *f.*

Output

Hypothesis $h \in H$ that best approximates target function *f.*

Probabilistic Interpretation

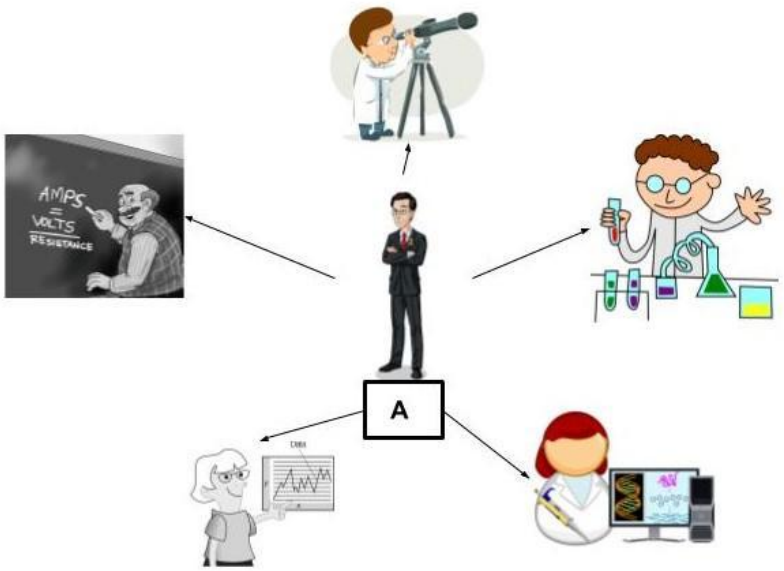
Learning as compression

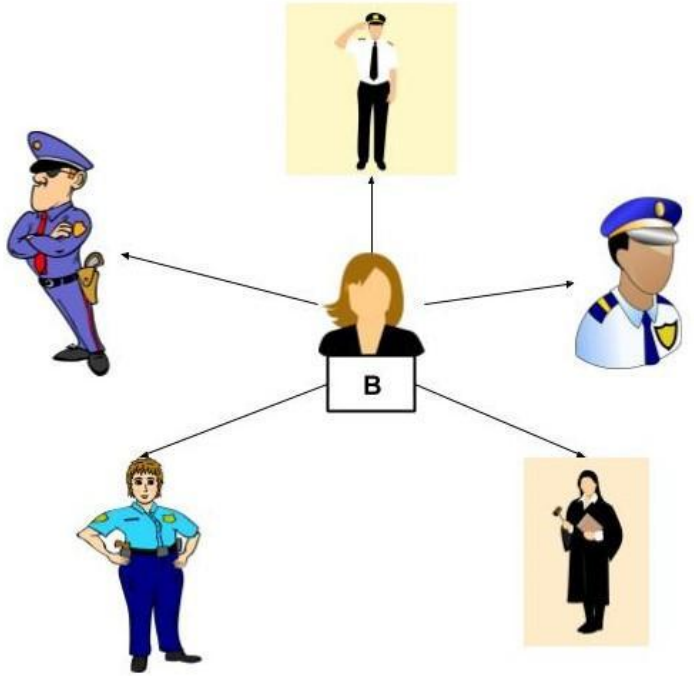
Learning

Understand how to do Machine Learning in a Principled Way

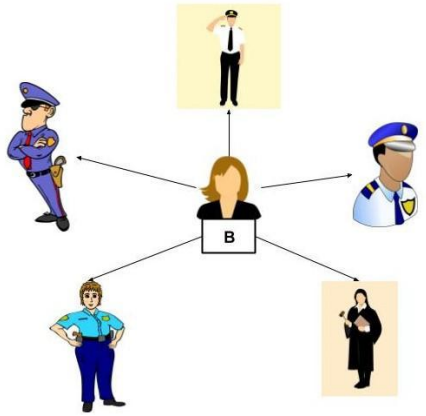
Let us play a game!

Goal: Identify the profession of two individuals A and B.
Rule: You only know the professions of five people with whom A and B spend their maximum time.

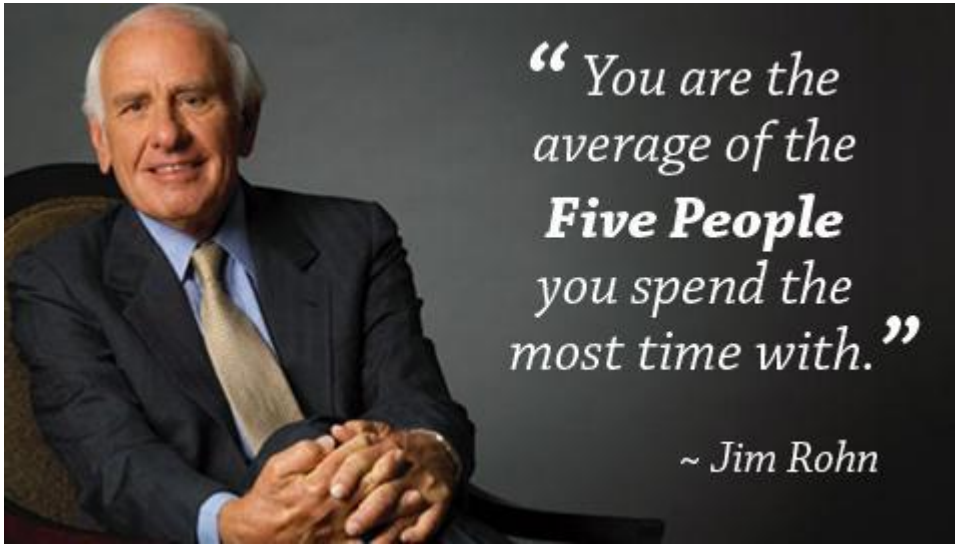


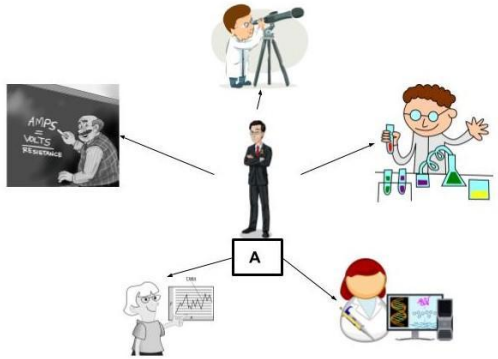


Let us come up with an algorithm to identify their professions.



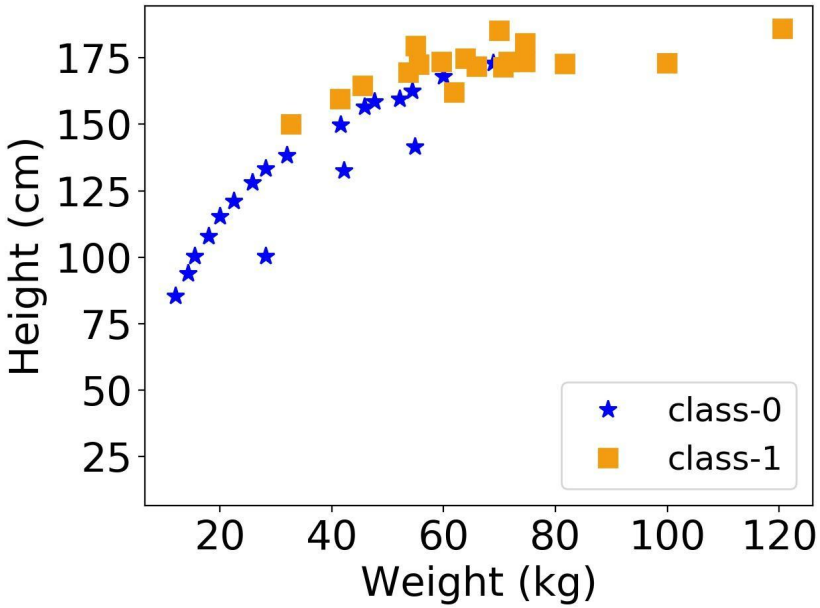
Model



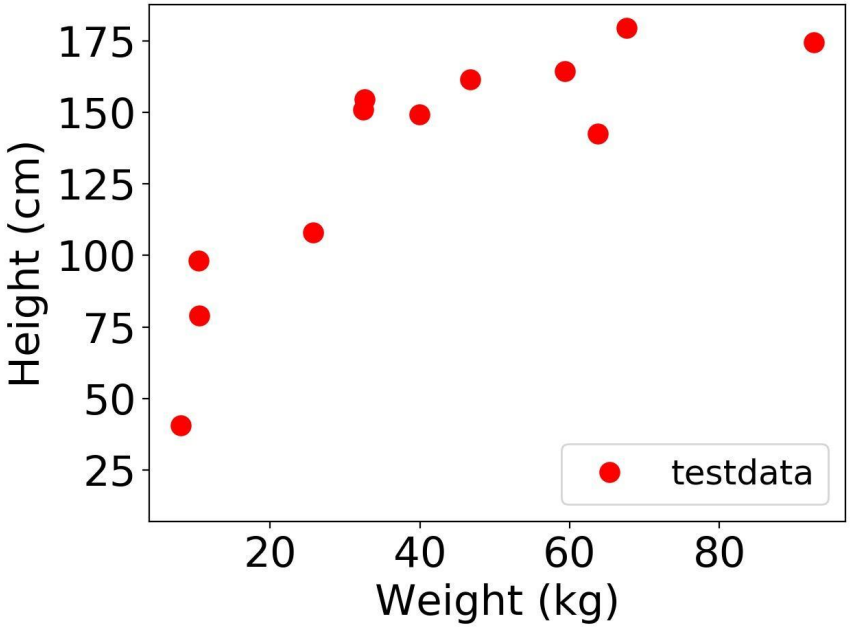


K-Nearest Neighbour Classifier

TRAINDATA



TESTDATA



$$dist(x, z) = \left(\sum_{j=1}^{r} |x_j - z_j|^p \right)^{\frac{1}{p}}$$

Mathematical Description

Let the training data $x_i \in R^2$ and its corresponding labels $y_i \in \{0, 1\}$ where $i \in \{0, 1, \ldots, N\}$

$$D = \{(x_0, y_0), (x_1, y_1), \ldots, (x_N, y_N)\}$$

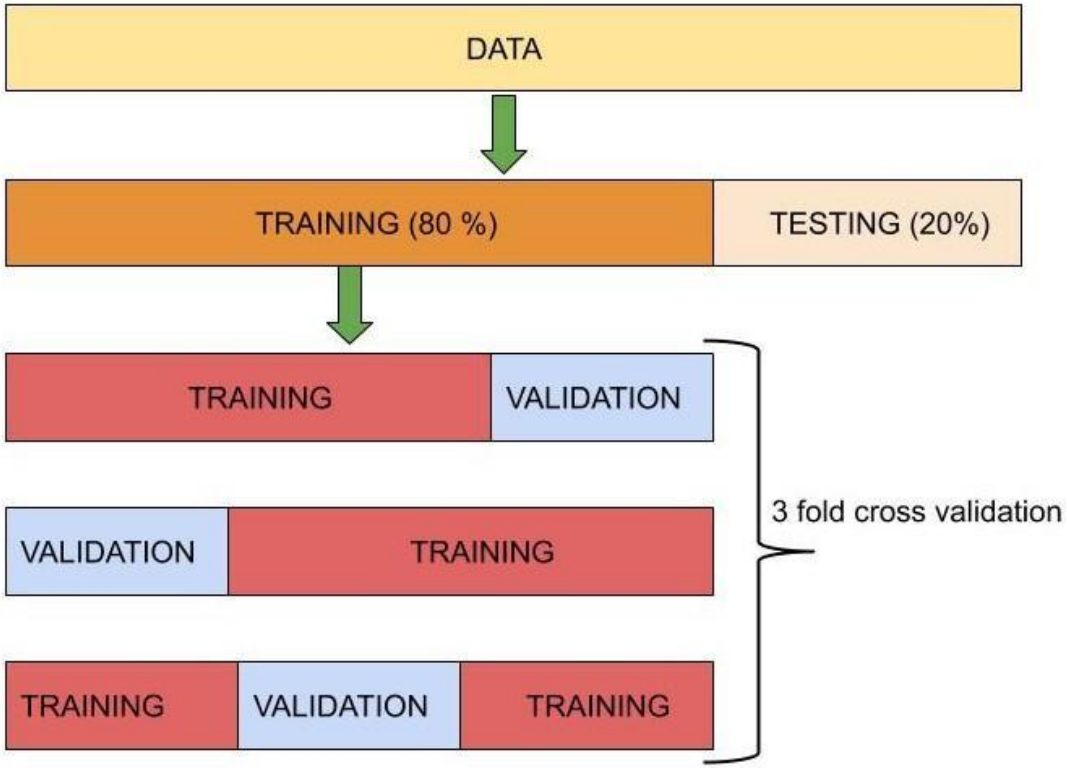
Let the set of k nearest neighbours of z be S_x where $S_x \subseteq D$ and $|S_x| = k$ (no. of elements in S_x is k).

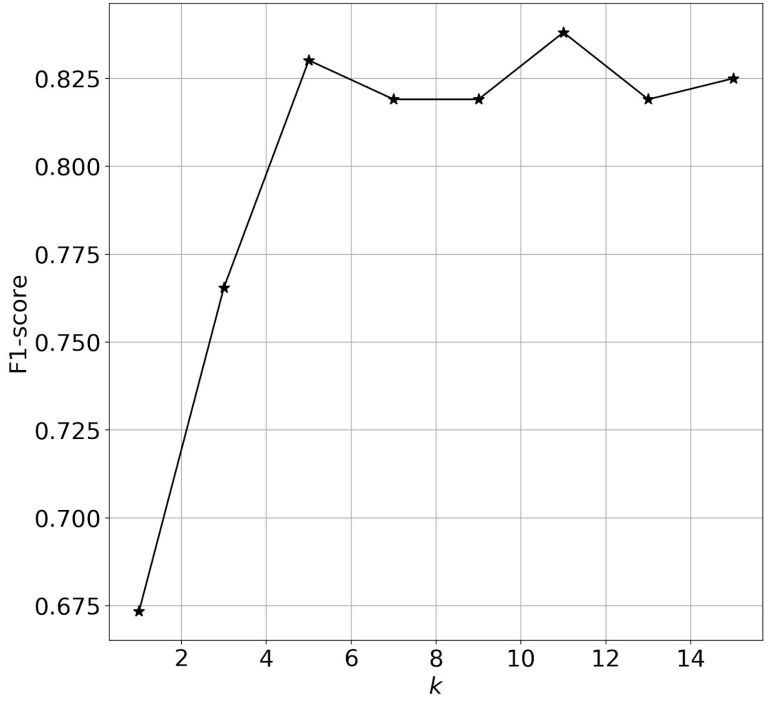
Now $\forall (x', y') \in D \backslash S_x$
(for all data instances in D but not in S_x) the following is true.

$$dist(z, x') \geq \max_{x'', y'' \in S_x} dist(z, x'')$$

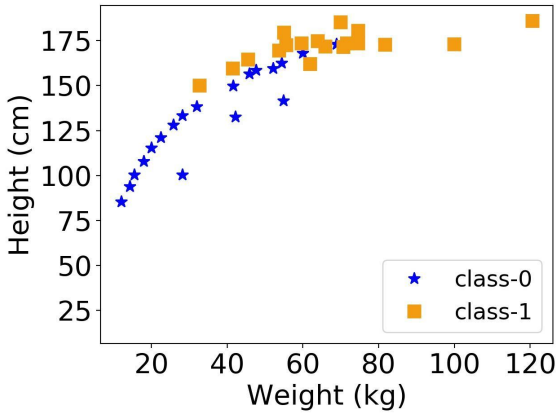
$$dist(x, z) = \left(\sum_{j=1}^{r} |x_j - z_j|^p \right)^{\frac{1}{p}}$$

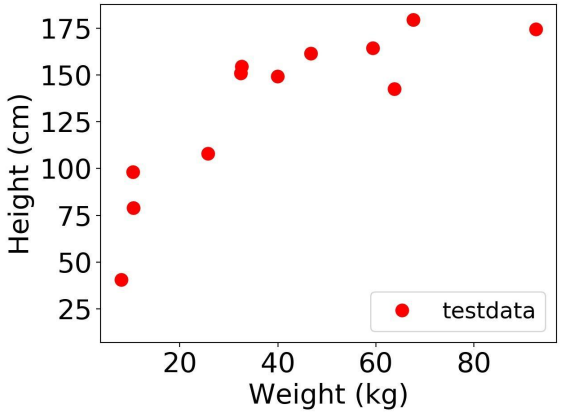
How to choose K? Principled way of doing ML

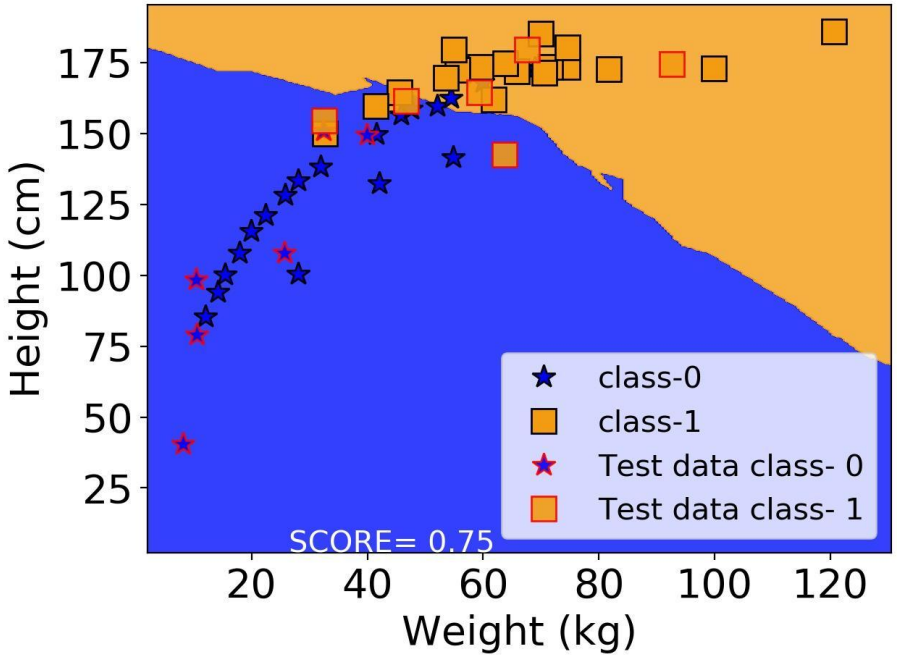




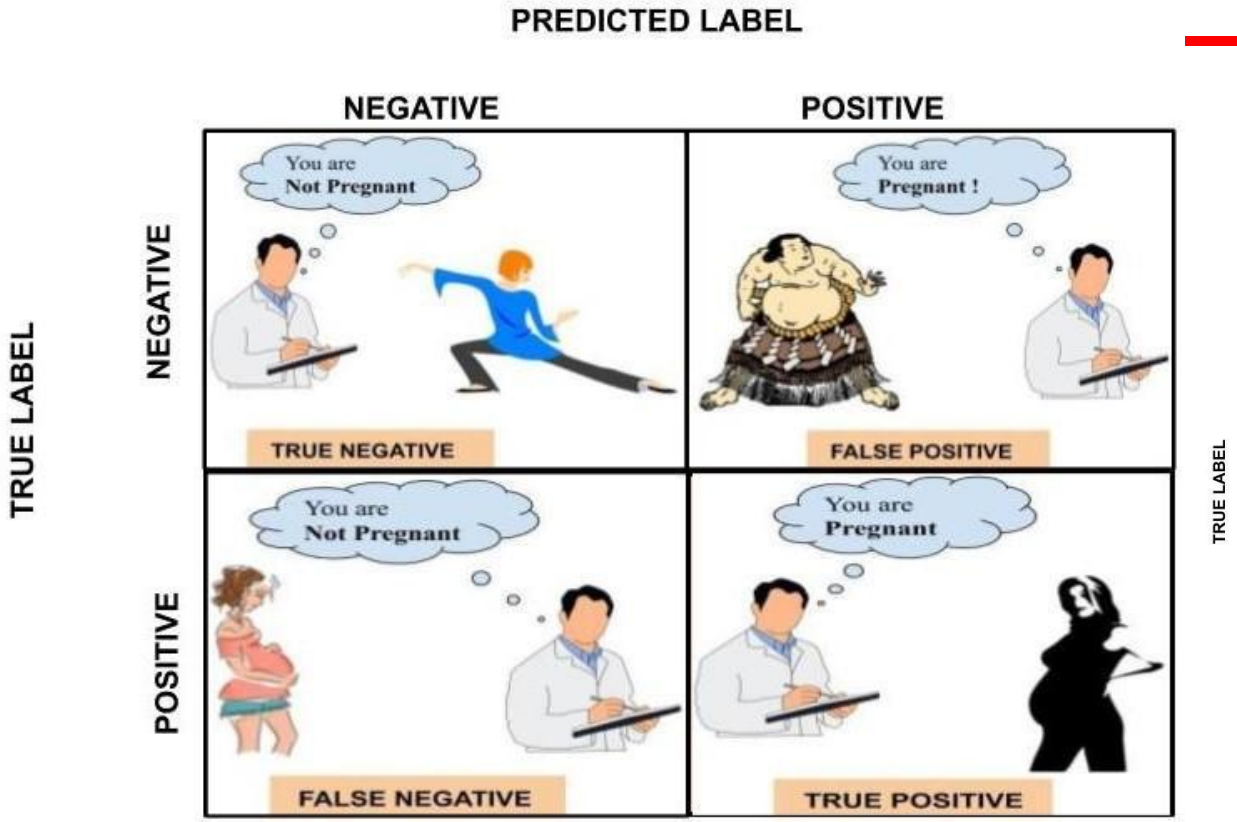
Let us see how the model performs for Test data

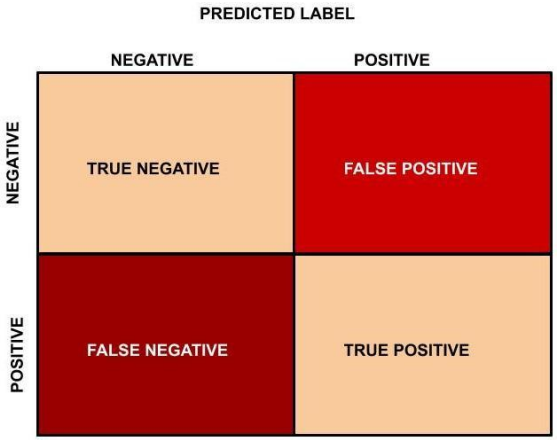




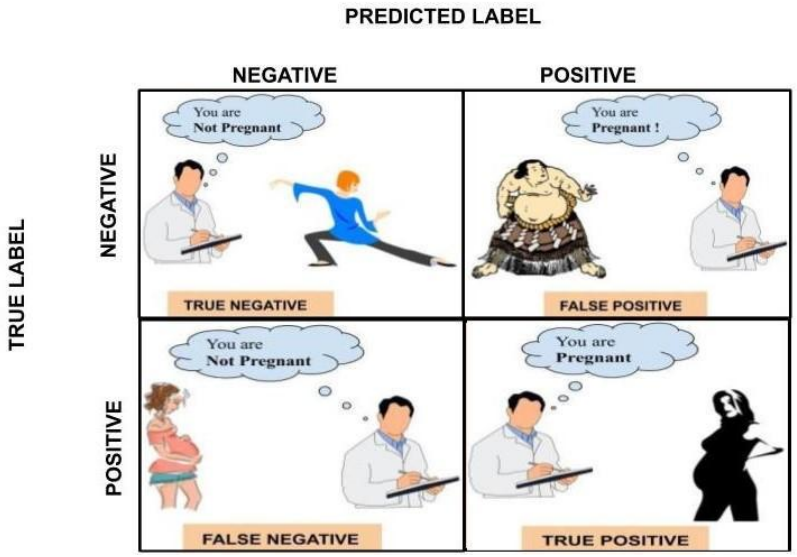


Performance Measures





Performance Measures



If we consider a binary classification problem,
C_{00} represents the count of true negative
C_{01} represents the count of false positive
C_{10} represents the count of false negative and
C_{11} represents the count of true positive.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Example

So far…

- Knowledge of the problem that you are trying to solve
- Load the data correctly
- Analyze the data
  - How many classes?
  - Number of data instances per class
  - Train-test split
- Algorithm used for solving the problem
  - What are the hyperparameters of the algorithm?
  - Crossvalidation to find the best hyperparameters.
- Retrain the training data with the tuned hyperparameters.
- Used the trained model (with best hyperparameters obtained after crossvalidation) on test data to evaluate the performance of unseen test data. This should be done only once.

Demo for KNN

Course Objective

This course covers the foundations of Machine Learning. This includes the theoretical and applications of various algorithms.

The following topics shall be covered: Decision Tree, Naive Bayes, Logistic Regression, Linear Regression (both matrix and matrix free methods), Bias -Variance Trade off, k-Nearest Neighbours, Principal Component Analysis, Singular Value Decomposition, Reservoir computing for forecasting, Learning Theory, Boosting, Kernels, Support Vector Machine. After the successful completion of the course, the student will learn (a) the principled way to do machine learning, (b) read, understand and implement Machine Learning research papers.

Course Resources

1. **Textbook (T1):** Tom M. Mitchell. *Machine learning*. Vol. 1, no. 9. New York: McGraw-hill, 1997.
2. **Textbook (T2):** Bishop, Christopher M., and Nasser M. Nasrabadi. *Pattern recognition and machine learning*. Vol. 4, no. 4. New York: springer, 2006.
3. **Textbook (T3)**: Aurelien Geron, “Hands-On Machine Learning with Scikit-Learn, Keras and Tensorflow”, O’Reilly, 2020
4. **Reference (R1):** Hastie, Trevor, Robert Tibshirani, Jerome H. Friedman, and Jerome H. Friedman. The elements of statistical learning: data mining, inference, and prediction. Vol. 2. New York: springer, 2009.
5. **Reference (R2):** Strang, Gilbert. "Introduction to linear algebra." (2020).

Evaluative Components

EC1 = Assignment (30%)

(Timeline: 10 days)

EC2 = Mid Semester [Closed Book] (30%)

(Timeline: 2 hrs), Closed Book

EC3 = Comprehensive Exam (40%)

(Timeline: 3hrs), Open Book

Honour Code

"I affirm that I have not given or received any unauthorised help on this assignment, and that this work is my own. Any part of the text or material of the assignment submission, if copied from internet sources, need to be cited and quoted, without which I will be held liable to have committed **plagiarism**. For group projects, I accept responsibility for my role in ensuring the integrity of the work submitted by the group in which I participated. I understand that any violation of the stated policies, as determined by the FIC, shall be subject to scrutiny by the FIC and is liable for punishment."