APL 745
Deep Learning in Mechanics

Instructor: Sitikantha Roy

Important Information

- **Instructor:** Sitikantha Roy
  - Email: sroy@am.iitd.ac.in
  - Office: Block IV, 243
- **TAs:** Abhilash Awasthi, Shaswata Ghosh, Sanchita Mallya, Rajnish Kumar, Hemanthika **.**
  - Email: saswath.ghosh3333@gmail.com
  - Office hours: Online timing will be announced.
- **Web:** Teams channel
- **Mailing list**: 2102-apl745@courses.iitd.ac.in

Evaluation

- Lab assignments/Projects (50%)
    - Implementing machine learning algorithms.
    - Applying them to real datasets.
  - (e.g soft robotics, biomec, InSillco medicine, optimization, comp vision, etc)
  - Exam (50%)

  - Lab: TBD
  - Class: Monday and Thursday, 5 to 6.20 pm
  - Discussion session?

Source Materials (DL)

- Bishop Christopher, M., 2006. Pattern recognition and machine learning. *Information science and statistics, New York: Springer*.
- Chollet, F., 2021. *Deep learning with Python*. Simon and Schuster.
- PyTorch documentation (https://pytorch.org/docs/stable/index.html) (visited on 23.12.2021)
- CS229, Machine Learning, Andrew NG class note.
- Tom M Mitchel, Machine Learning, McGraw-Hill, 1997
- Dive into Deep Learning, by Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J Simola (Release 0.17.5).
- Deep Learning by Ian Good Fellow, Yoshua Beningo, and Aaron Couurville: The MIT press, Cambridge, MIT.
- Deep Learning in Computational Mechanics: *An Introductory Course* by Stefan K., Davide D' Angella, Mortiz Jokeit, Leon Herrmann.
- Current literature .

History of DL revolution

Historical Notes

1943 McCulloch and Pitts proposed the McCulloch-Pitts neuron model

1949 Hebb published his book *The Organization of Behaviour*, in which the Hebbian learning rule was introduced

1958 Rosenblatt introduced the simple single layer networks called Perceptrons

1969 Minsky and Papert's book Perceptrons demonstrated the limitation of single layer perceptrons

1980 Grossberg introduced his Adaptive Resonance Theory (ART)

1982 Hopfield published a series of papers on Hopfield networks

1982 Kohonen developed the Self-Organizing Feature Maps

1986 Back-propagation learning algorithm for multi-layer perceptrons was re-discovered, and the whole field took off again

1990s ART-variant networks were developed

1990s Radial Basis Functions were developed

2000s Support Vector Machines were developed

A Few Quotes

- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet” (Tony Tether, Director, DARPA)
- Machine learning is the hot new thing” (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, Former CTO, Sun)
- “Machine learning is today’s discontinuity” (Jerry Yang, Founder, Yahoo)
- “Machine learning today is one of the hottest aspects of computer science” (Steve Ballmer, CEO, Microsoft)

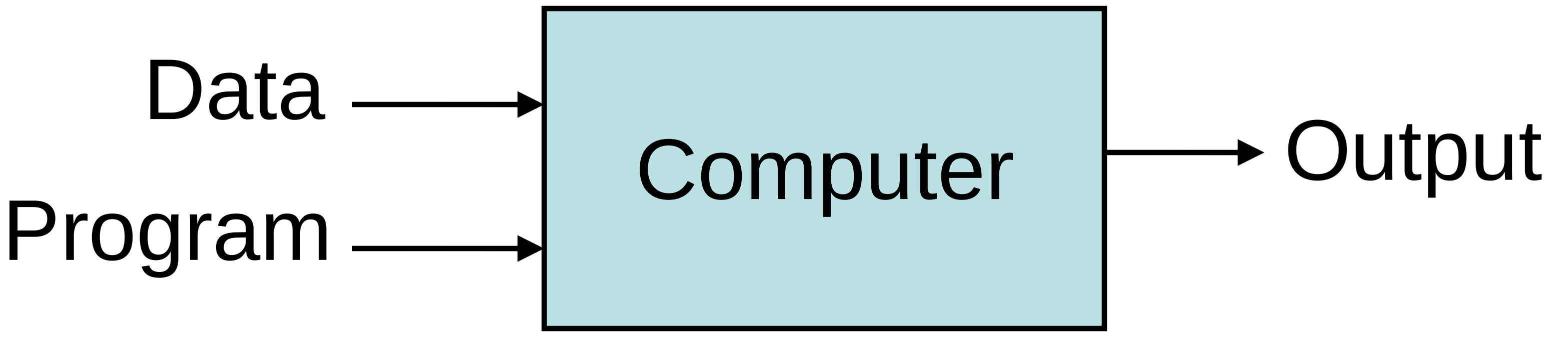
So What Is Machine Learning?

- Automating automation
- Getting computers to program themselves
- Writing software is the bottleneck
- Let the data do the work instead!

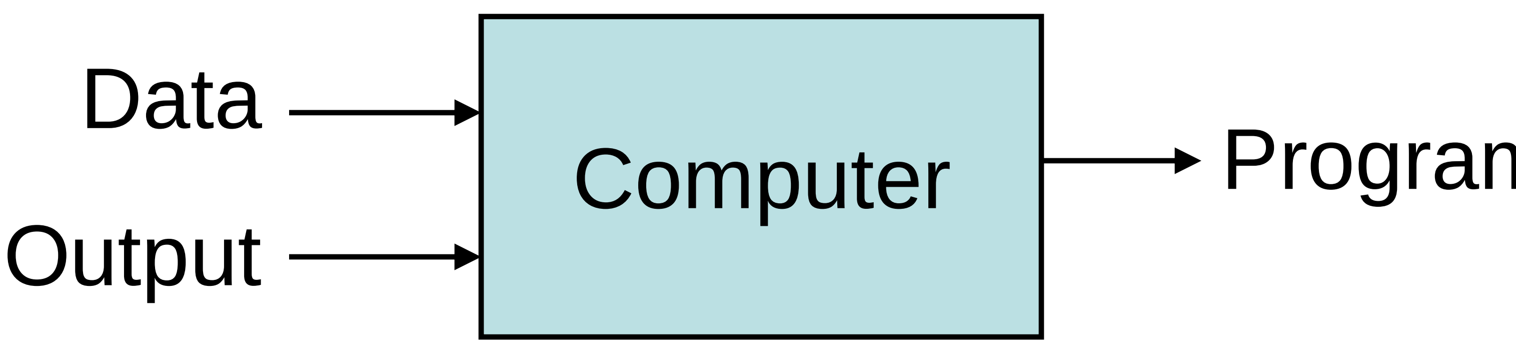
Ref: Dapino

- Hollywood movies (on AI)
  - Trancendence (Jonny Depp)
  - Her (Joaquin Phoenix)
  - Imitation Game (Based on Enigma machine)

Traditional Programming



Machine Learning



What is ML

- ML system are trained rather than explicitly programmed.
- It is presented with many examples relevant to a task and find statistical significance in these examples (DATA)
- The examples (training data) eventually allows the system to produce rules for automating the task.

What is ML?

- A formal definition

by Mitchell states that “a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E” [Mit97]. Taking image recognition as an example, the task T is to classify previously unseen images, the performance measure P corresponds to the amount of correctly classified images, and the experience E includes all images that have been used to train the algorithm.

Taking image recognition as an example,

<u>Task T</u>: Classify previously unseen images

<u>Performance measure P</u>: Corresponds to the amount of correctly classified images.

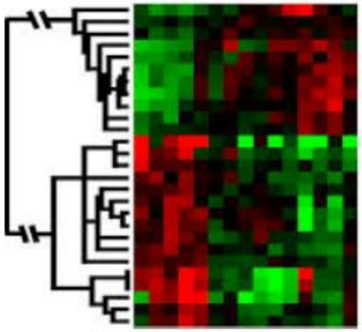
<u>The experience E</u>: Includes all the images that have been used to train the algorithm.

When Do We Use Machine Learning?

ML is used when:

- Human expertise does not exist (navigating on Mars)
- Humans can't explain their expertise (speech recognition)
- Models must be customized (personalized medicine)
- Models are based on huge amounts of data (genomics)



Learning isn't always useful:

- There is no need to "learn" to calculate payroll

Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- **Robotics**
- Information extraction
- Social networks
- Debugging
- **Mechanics**

Related Disciplines

- Artificial Intelligence
- Data Mining
- Probability and Statistics
- Information theory
- Numerical optimization
- Computational complexity theory
- Control theory (adaptive)
- Psychology (developmental, cognitive)
- Neurobiology
- Linguistics
- Philosophy
- Computational Neuroscience

ML in a Nutshell

- Tens of thousands of machine learning algorithms
- Hundreds new every year
- Every machine learning algorithm has three components:
  - **Representation (Model)**
  - **Evaluation**
  - **Optimization**

Representation

- Decision trees
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks-> Deep Learning/ DL
- Support vector machines
- Model ensembles
- Etc.

Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- Etc.

Optimization

- Combinatorial optimization
  - E.g.: Greedy search
- Convex optimization
  - E.g.: Gradient descent
- Constrained optimization
  - E.g.: Linear programming

Another way of looking

- Most DL algorithms can be subdivided into the following
  - A dataset
  - A parameterized model
  - A cost function or loss function
  - An optimization procedure

DL how it works

- Generally, the cost function/loss function defines an objective function for an optimization criterion by relating the data to the model parameters.
- Typical optimization problem
- $\theta = \cancel{f}(\theta)$ $\theta = \arg\min f(\theta)$

$$\textit{Subjected to} \quad h(\theta) = 0,\ g(\theta) \leq 0$$

- θ becomes the model parameters.

Terminologies & Data structure

- We will deal with data with multiple features/attributes, n
- Every data point $\boldsymbol{x}^i$, is a vector with i, number of features

$$\boldsymbol{x}^i = \begin{pmatrix} x_1^i \\ x_2^i \\ x_3^i \\ x_4^i \end{pmatrix}$$

features

Feature vector

one data pt

Data structure

- When we have multiple data points, each having these features---

m: # examples.
n: # of features.

$$\underline{X}_{m \times n} = \left[\begin{array}{c} \longleftarrow \underline{x}^{(1)} \longrightarrow \\ \longleftarrow \underline{x}^{(2)} \longrightarrow \\ \longleftarrow \underline{x}^{(3)} \longrightarrow \\ \vdots \\ \longleftarrow \underline{x}^{(m)} \longrightarrow \end{array}\right] \begin{array}{l} \text{ex } 1 \\ \text{ex } 2 \\ \text{ex } 3 \\ \\ \text{ex } m \end{array}$$

m data points in $\mathbb{R}^n$ [n # of features]

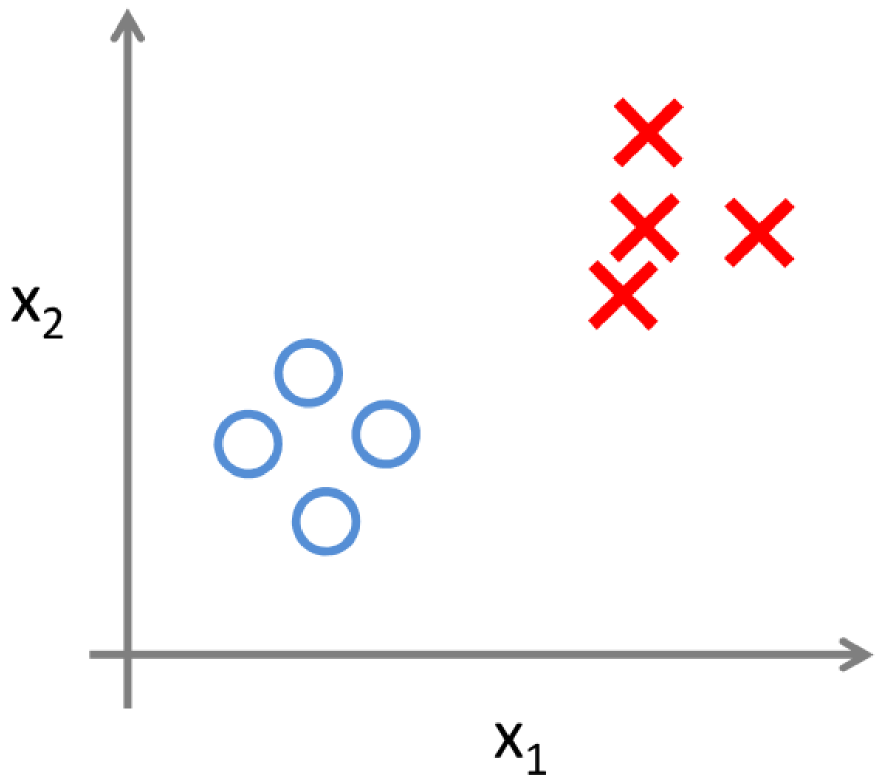
Training data/testing data

- Common practice to split the data into different subsets, namely a **training set** and **testing set**.
- Majority of data are used for training/or finding out the model parameters.
- The remaining part is used for validation purpose.

Types of Deep Learning

- **Supervised (inductive) learning**
  - Training data includes desired outputs (labelled dataset)
  - Example: Image classification; Each image has previously been tagged into a certain category.
  - The supervised learning algorithm studies the dataset and learns to classify the images into given category, by comparing its prediction with given ground truth level.
  - Examples (**Example: Is it a cat or a dog?, House prices, How's the weather today?**)

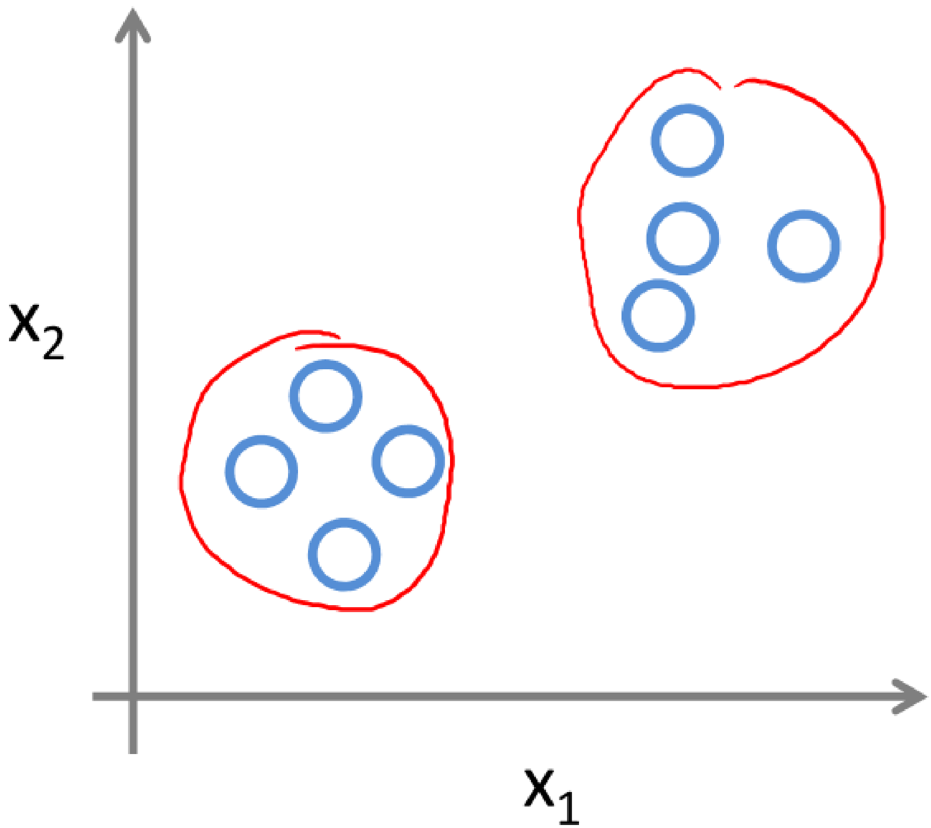
Supervised Learning

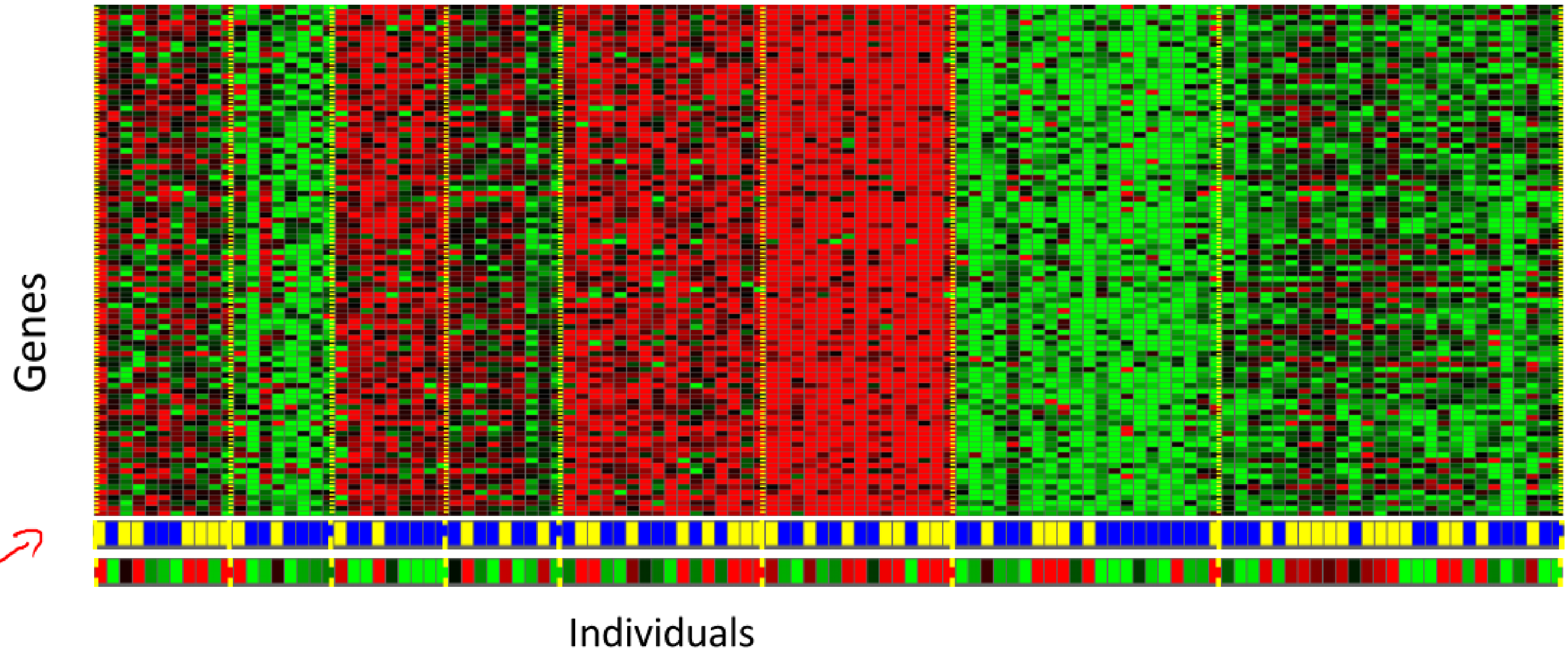


Types of Learning

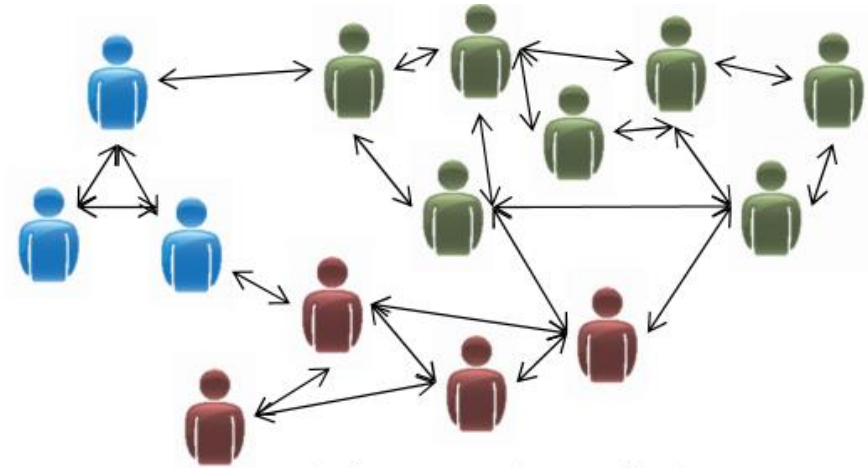
- **Unsupervised learning**
  - The goal of unsupervised learning is to find a structure, or more precisely, the probability distribution in the provided data.
  - The data is NOT labeled; hence no explicit prediction is possible.
  - USL can find inherent structure from large dataset.
  - Example: For instance, anomaly detection algorithms are used to identify fraudulent credit card transactions that differ from the usual purchasing behavior of the customer.

Unsupervised Learning



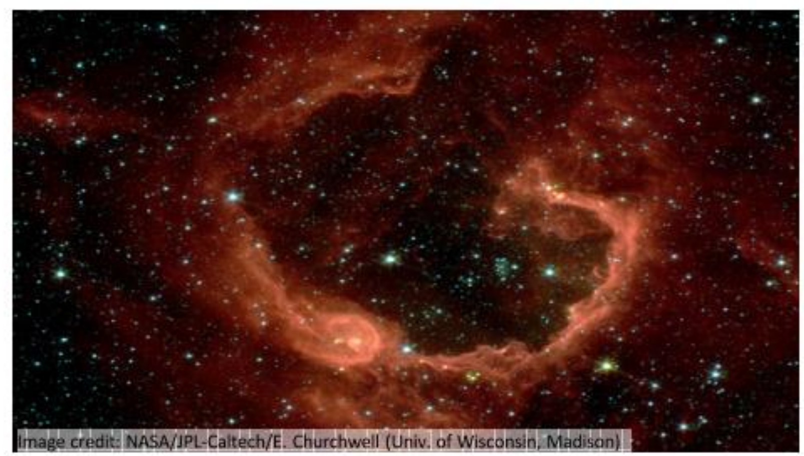

Genes
Individuals

Organize computing clusters



Social network analysis

Market segmentation



Astronomical data analysis

Type of learning

- **Semi-supervised learning**
  - Combines two preceding concepts
  - In cases where small samples of data are labeled, the semi-supervised learning helps to improve the performance of the supervised learning algorithm
  - When training DATA is sparse.

Types of Learning

- **Reinforcement Learning**
  - The basic idea of reinforcement learning is that an algorithm interacts with an environment, to learn a certain decision behavior maximizing the expected average reward.
  - It is used for problems involving sequential decision-making in order to fullfil a long term goal.

Supervised (Inductive) Learning

- **Given** examples of a function$(x^i, y^i = h_w(x^i))$
- **Predict** function $h_w(x^i)$ for new examples x^i.
  - Discrete $h_w(x^i)$: Classification
  - Continuous $h_w(x^i)$: Regression
  - $h_w(x^i)$ = Probability(x^i): Probability estimation

What We'll Cover

- **Supervised learning**
  - Regression (Single variate/multi variate)
  - Classification (Binary/multi class)
  - Neural networks
  - Fully Connected Neural Network
  - Forward propagation
  - Back propagation
  - Convolution Neural network
  - Recurrent Neural network
  - Physics Informed Neural Net (PINNS and its variants)
- **Unsupervised learning (may be)**
  - Clustering
  - Dimensionality reduction

DL in Practice

- Understanding domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learning models
- Interpreting results
- Consolidating and deploying discovered knowledge
- Loop