Machine Learning

Birhanu Hailu (PhD)

Birhanu Hailu Belay

- Education:
  - **Feb 2021: Ph.D., Technische Universität Kaiserslautern, Germany.**
    - Computer science (Intelligent systems): Deep learning
  - **Oct 2015: M.Sc., Bahir Dar Institute of Technology, Ethiopia.**
    - Computer Science (Image processing)
  - **Jul 2010: B.Sc., Wollo University, Ethiopia.**
    - Information Communication Technology

Work experience

- Sept 2010- Aug 2011: **Arba Minch University**, Ethiopia (GA-I)
- Sept 2011- Dec 2012: **Deberetabor University**, Ethiopia.(GA-II)
- Jan 2013- Sept 2017: **Bahir Dar Instituite of Technology**, (Asst. Lecture)
- Oct 2017-July 2020: **German Research Center for Artificial Intelligence,** DFKI-Kaiserslautern, Germany (Visiting researcher)
- Since Feb 2021---: Computing Faculty, **Bahir Dar Institute of Technology**

Publication

- 2 Journal paper
- 9 conference paper
- Google and IEEE awards for 2 of my papers ($)
- link

- Link: https://www.researchgate.net/profile/Birhanu-Belay-2

Courses and thesis topics <>

- Machine learning/deep learning
- Document image analysis and Recogntion
- Artificial Neural network
- Image processing/computer vision
- ~~Natural language processing (?)~~

Contact address:

Email: birhanu.hailub@gmail.com

Course outline for Machine Learning

- **Course description:** This course is about learning through **experience** to extract statistical structure from data, for making **decisions** and **predictions**, as well as for visualization.

- **Course objective:** The goal machine learning is to **build computer systems** that learn from **experience** and that are capable to **adapt to their environments**.

- Upon completing this course, you should be able to:
  - Explain and differentiate different classical/modern machine learning techniques;
  - Identify potential application areas where machine learning techniques can be useful
  - Implement the solution, and evaluate the results.

Course outline for Machine Learning

Has 7 chapters:

1. **Overview of Machine Learning**
    - Introduction to Artificial Intelligence and machine learning
    - Basic concepts on probability and probability distributions
    - Machine Learning vs Statistical learning
    - Applications of Machine learning
    - Challenges in Machine learning
    - Basic Mathematics for Machine learning
2. **Supervised Learning**
    - Classification and regression
    - Basic steps of classification
    - Generalization, overfitting and underfitting
    - Supervised learning algorithms (DT,ANN,KNN, SVM, Bayesian, etc)
    - Parametric vs non-parametric learning

Course outline for Machine Learning

3. Artificial Neural Networks

- Neurons and biological motivation
- Perceptron
- Basics of Deep Learning algorithms (CNNs, RNNs)
- Recent topics in deep Learning (based on recently published paper)

4. Data preprocessing and representation

- Data preprocessing, Measures of data quality
- Data representation, Feature selection and encoding
- Handling missing data

5. Evaluation of Learning Algorithms

- Model selection, cross-validation
- Bootstrapping, error measures, Confusion matrix
- Precision, Specificity, Mean absolute percentage error, Root mean square error, Recall , Accuracy
- Maximum Likelihood Estimation (MLE), Interval estimation, hypothesis testing,

Course outline for Machine Learning

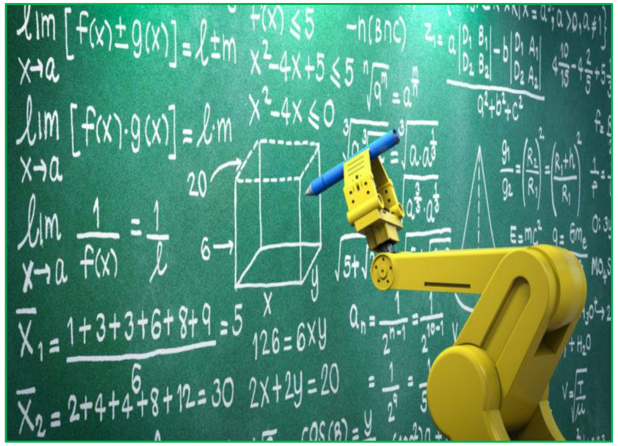
6. Unsupervised learning

- Overview of unsupervised learning
- Clustering algorithms (Partition based clustering, hierarchical clustering, density and model based)
- Based clustering, mixture of Gaussians and EM algorithm)
- Evaluation of cluster quality
- Dimensionality reduction techniques (PCA, LDA, FA,)

7. Other topics

- Semi supervised learning (S3VMs,)
- Reinforcement learning
- Bandit problems and online learning
- Monte-Carlo Methods and Temporal Difference Learning (Q-learning)
- Dynamic programming, Ensemble learning

Mode of course Delivery



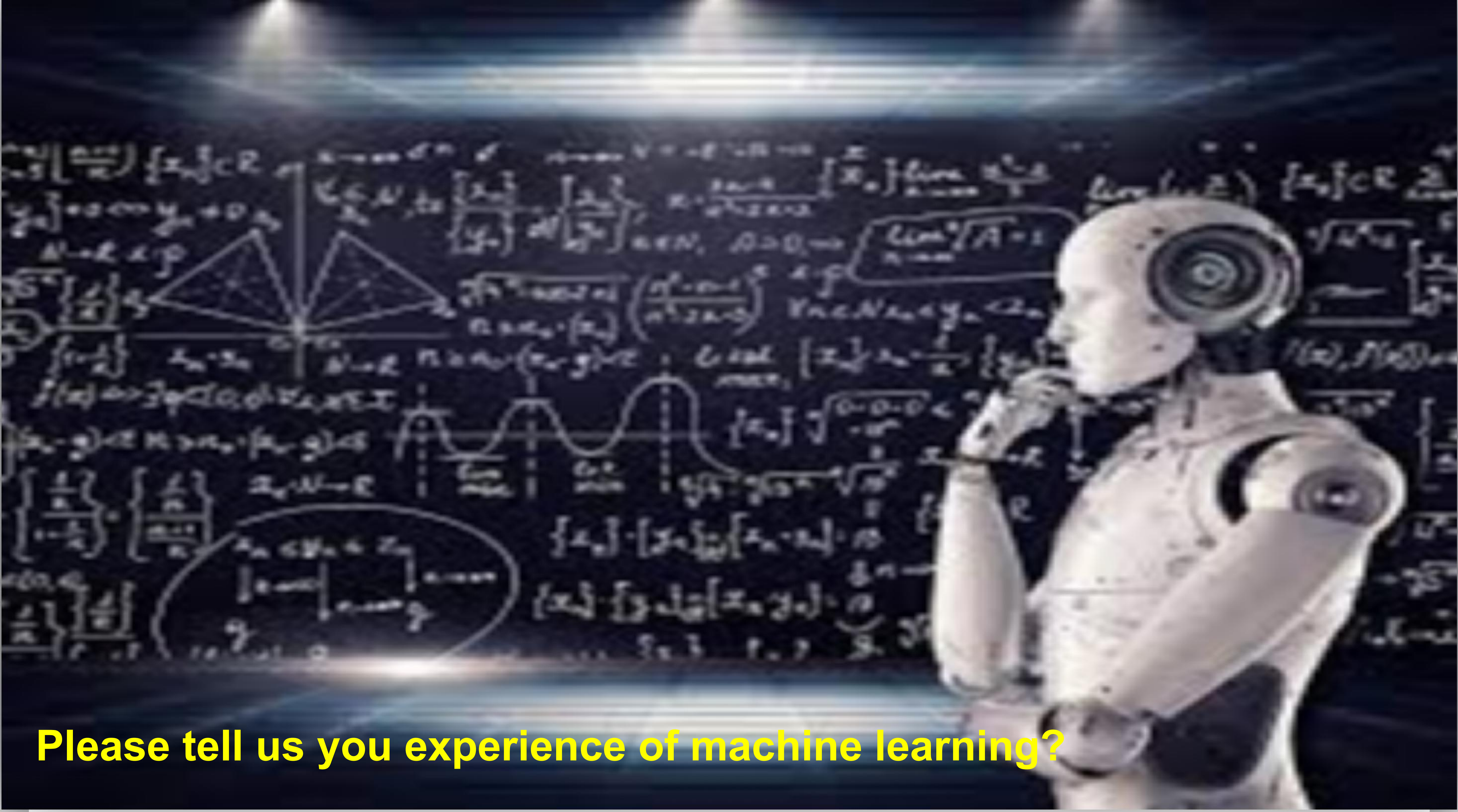
Lecture

Lab (please install these tools)

Project/Exam/Grading

References:

- Ethem Alpaydin, **Introduction to Machine Learning,** The MIT Press, 2nd Edn., 2010
- Bengio, Y., LeCun, Y., and Hinton, G. **Deep Learning**. Nature 521: 436-44, 2015.
- Marc Toussaint **Maths for Intelligent Systems,** 2017.


Please tell us you experience of machine learning?

ML-related search results

Geoffrey Hinton | Andrew Ng | Yann LeCun | Michael I. Jordan | Fei-Fei Li | Yoshua Bengio

Top ML researcher

Rank	Conference (Full Name)	Short Name
1	IEEE Conference on Computer Vision and Pattern Recognition	CVPR
2	Annual Conference on Neural Information Processing Systems	NeurIPS
3	International Conference on Computer Vision	ICCV
4	International Conference on Machine Learning	ICML
6	AAAI Conference on Artificial Intelligence	AAAI
8	Annual Meeting of the Association for Computational Linguistics	ACL

Top conference of AI/ML

Google

top cited paper in 2020

All · News · Images · Videos · Maps · More · Settings · Tools

About 222,000,000 results (0.60 seconds)

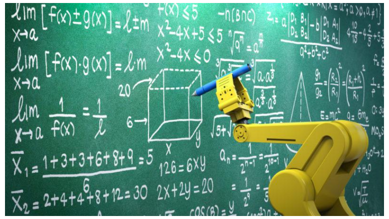
The Most cited paper in 2020

The most highly-cited paper of all, "**Deep Residual Learning for Image Recognition**", published in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, was written by a team from Microsoft in 2016. It has made a huge leap from 25,256 citations in 2019 to 49,301 citations in 2020. Jul 13, 2020

Python | Julia | Java | Scala | TypeScript

Top programming language for ML

Applications of Machine learning









Surveillance and security system

Predication (weather, medical, agricultural yield)

Spam Detection

Natural language processing

Image segmentation

Multimedia event detection

Speech Recognition

MT

sentiment classification

Image captioning

Software Engineering

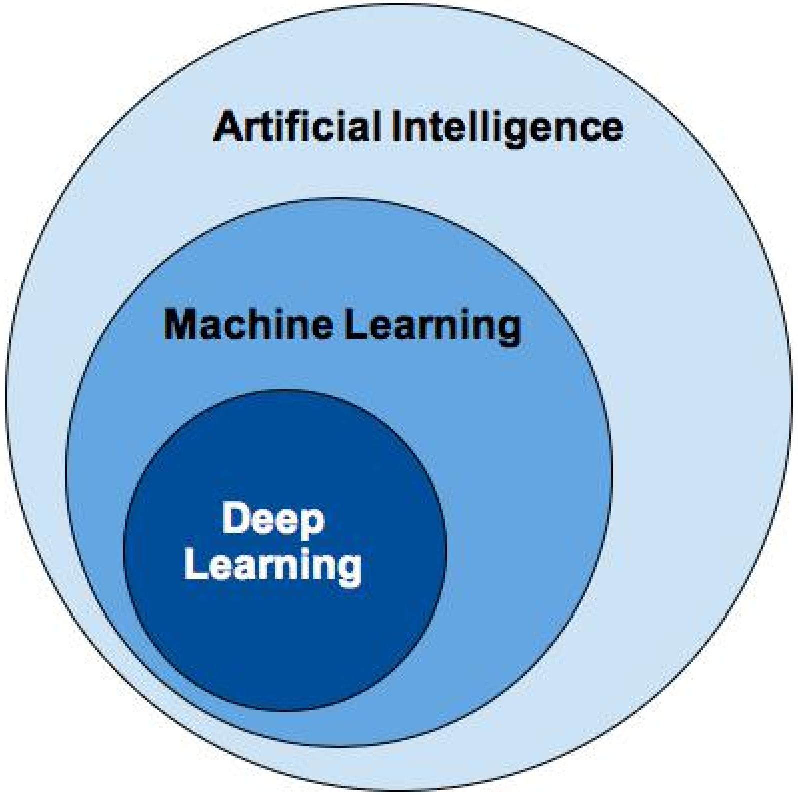
Cancer detection/classification

Character recognition

Face recognition

Object detection and recognition

AI vs ML vs DL



- **AI :** broader concept (science+Engineering) to create intelligent machines that can simulate human thinking capability.
- **ML :** subset of AI that allows machines to learn from data without being programmed explicitly.
- **DL:** subset of ML, that uses the neural networks to analyze different factors with a structure that is similar to the human neural system.

What is machine learning ?

- It means that *ML is able to perform a specified task without being directly told how to do it.*
- *Example:*
  - Distinguish between spam and valid email messages.

    Given a set of manually labeled good and bad email examples, an algorithm can automatically learn a set of rules that distinguish them.
  - Language Identification (Amharic, Ge'ez, Tigrigna, Afar, etc) *(How?)*
- Arthur Samuel (1959) defined machine learning as "*a sub-field of computer science* that gives **computers** the ability to learn *without being explicitly programmed.*"

What is machine learning ?

- A widely accepted formal definition by Tom Mitchell (1997, professor of Carnegie Mellon University):
  - 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 the tasks ***T***, as measured by ***P***, improves with the experiences.

- In short
  - ***A set of computer programs that automatically learn from past experiences (examples or training corpus)***

- Example: According to this definition, we can reformulate the email problem as the task of identifying spam messages (***task T)*** using the data of previously labeled email messages (***experience E)*** through a machine learning algorithm with the goal of improving the future email spam labeling (***performance measure P)***

What is machine learning ?

ML aims to select, explore and extract useful knowledge from complex, often non-linear data, building a computational model capable of describing unknown patterns or correlations, and in turn, solve challenging problems.

This learning process is often carried out through repeated exposure to the defined problem (training dataset), allowing the model to achieve self-optimization and continuously enhance its ability to solve new, previously unseen problems (test dataset).

Probability Theory in Machine Learning

- **Probability** is key concept is dealing with uncertainty
  - Arises due to finite size of data sets and noise on measurements

- **Probability Theory**
  - Framework for quantification and manipulation of uncertainty
  - One of the central foundations of machine learning

Basic Concepts on Probability

- **Probability**. The word *probability* is actually undefined, but the *probability of an event* can be explained as the proportion of times, under identical circumstances, that the event can be expected **to occur** from the **known population**.
  - It is the event's long-run frequency of occurrence.
  - For example, the probability of getting a head on a coin toss = 0.5

- **Probability vs. Statistics**.
  - **Probability**, deals with predicting the likelihood of future events.
  - **Statistics**, you draw inferences about the population from the sample or analysis of the frequency of past events

Probabilistic vs Statistical Reasoning

- Suppose I know exactly the proportions of car makes in Bahir Dar. Then I can find the probability that the first car I see in the street is a Suzuki. This is probabilistic reasoning as I know the population and predict the sample.

- Now suppose that I do not know the proportions of car makes in Bahir Dar, but would like to estimate them. I observe a random sample of cars in the street and then I have an estimate of the proportions of the population. This is statistical reasoning

Key Terms

- **Classical** (or **theoretical**) **probability** is used when each outcome in a sample space is equally likely to occur. The classical probability for event *E* is given by,

$$P(E) \quad \frac{\text{Number of outcomes in event}}{\text{Total number of outcomes in sample space}}$$

Example: Find the probability of rolling a 4 on a fair die.

Answer: There are 6 possible outcomes when rolling a die: 1, 2, 3, 4, 5, and 6. The only favorable outcome is rolling a 4. = 1/6 (What do you think about rolling even number?)

- **Empirical** (or **statistical**) **probability** an event is an **estimate** that an event will occur based upon how often the event occurred after collecting data from an experiment in a large number of trials. This type of probability is based upon direct observations. Each observation in an experiment is called a trial/total frequency and is given by,

$$P(E) \quad \frac{\text{Frequency of Event } E}{\text{Total frequency}}$$

- Example: Next-slide

Key Terms

A survey was conducted to determine students' favorite brands of sneakers. Each student chose only one brand from the list of brands A, B, C, D, or E. What is the probability that a student's favorite sneaker was brand D?

Sneaker	A	B	C	D	E
Number	12	15	24	26	13

Answer: There were 12 + 15 + 24 + 26 + 13 = 90 "trials" in this experiment (each student's response was a trial).

26 out of the 90 students chose brand D.
The probability is :

$$\frac{\text{number choosing brand D}}{\text{total number choosing a brand}} = \frac{26}{90} = \frac{13}{45}$$

What is the probability of that the students favorite sneaker was not brand A or B?

Simple probability. P(A). The probability that an event (say, A) will occur.
Joint probability. P(A and B). P(A ∩ B). The probability of events A and B occurring together.
Conditional probability. P(A|B), read "the probability of A given B." The probability that event A will occur given event B has occurred.

Probability Distributions

- The probability distribution for a random variable X describes how the probabilities are distributed over the values of the random variable X.
- The probability distribution for a discrete random variable is described with a ***probability mass function*** (***PMF***).

For a discrete random variable X with possible values $x_1, x_2, x_3, \ldots, x_n$, a **probability mass function** $f(x_i)$ is a function such that

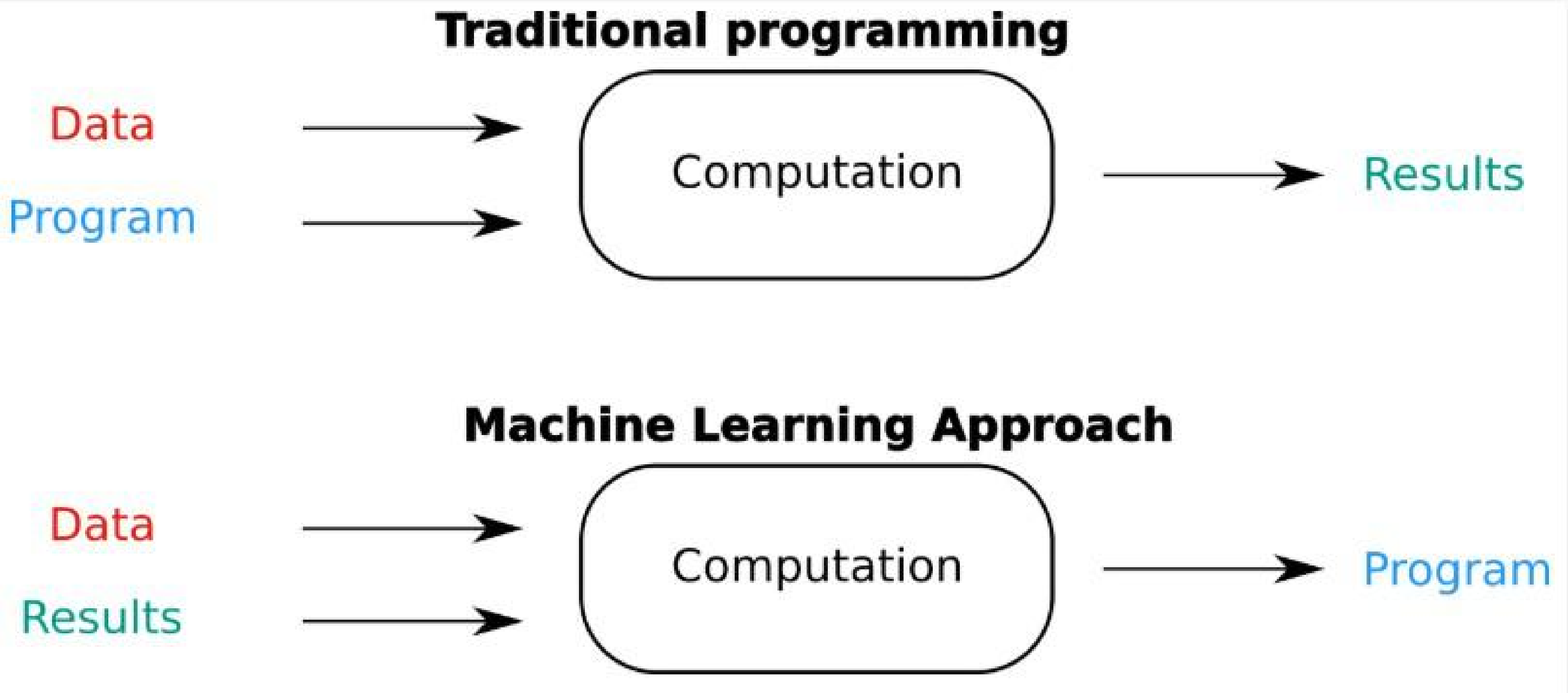
1. $f(x_i) \geq 0$
2. $\sum_{i=1}^{n} f(x_i) = 1$
3. $f(x_i) = P(X = x_i)$

- If the random variable is continuous then what could be the probability distribution?

The need of probability distribution:

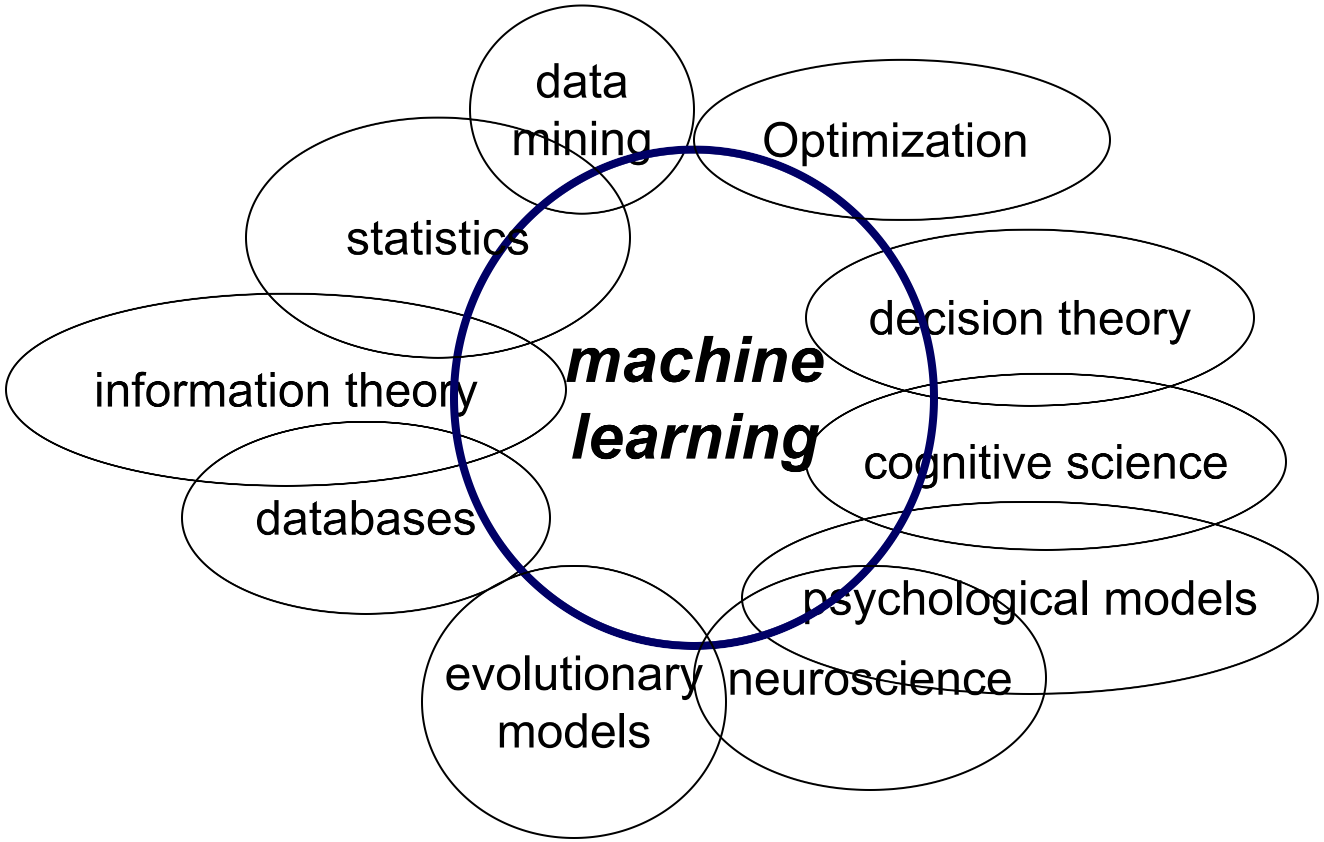
- To calculate confidence intervals for parameters
- To calculate critical regions for hypothesis tests

Traditional Programming vs ML Approach



Traditional programming VS ML approach

Related Fields



Machine learning is primarily concerned with the **accuracy** and **effectiveness** of the *computer system* in **performing complex tasks.**

Statistics vs. Machine Learning

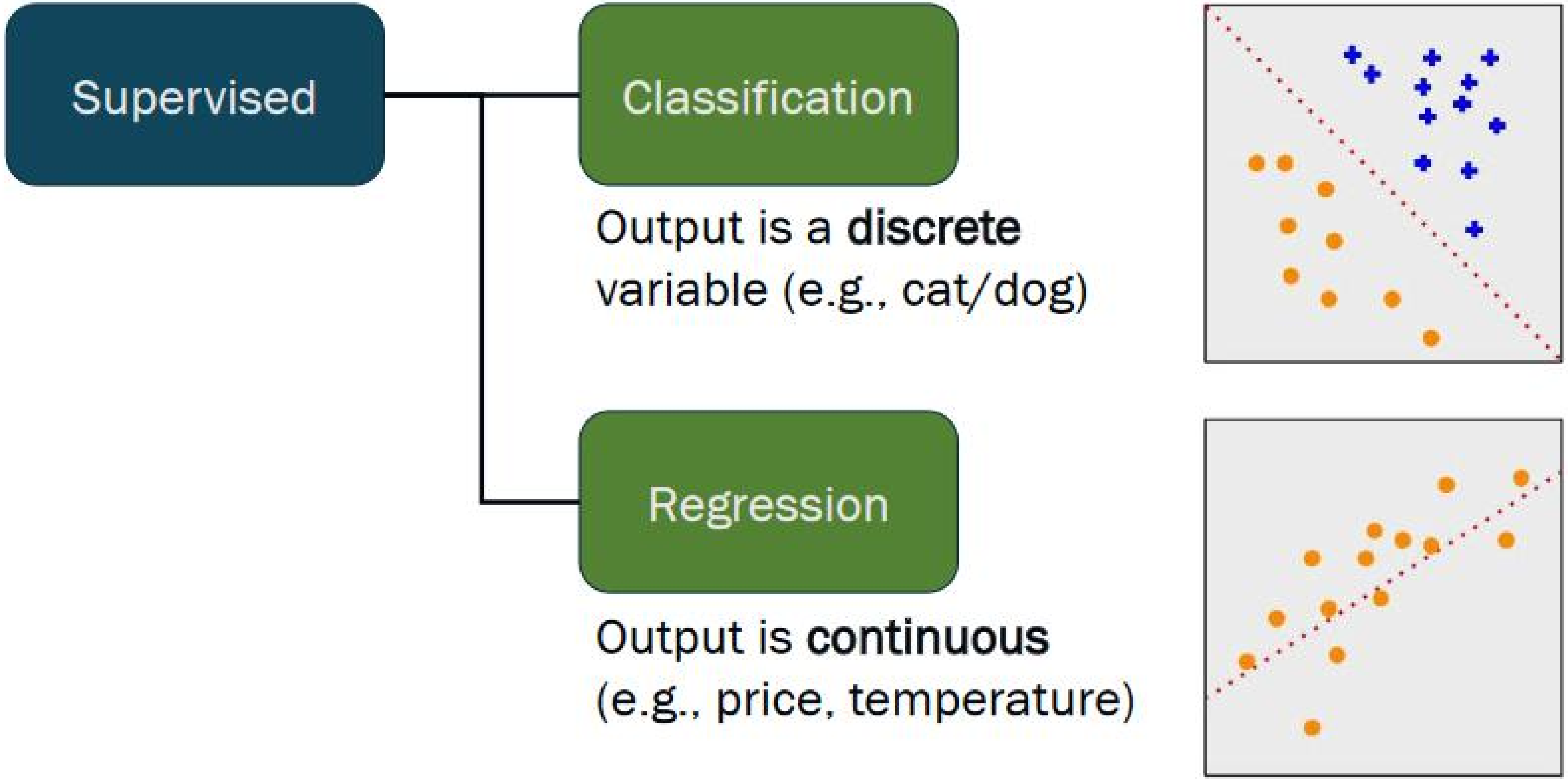
Statistics	Machine Learning
Inference	Prediction
Small data sets/low-dimensional data	Large data sets/high-dimensional data
Specific assumptions and hypotheses	Large flexibility and free from a priori assumptions/hypothesis free
Computation of the P values to accept or reject a null hypothesis	ROC curve, cross-validation, etc.
Fitting a parsimonious model to produce an easy to understand and interpretable results	Considers complex non-linear patterns, a sophisticated model that is not easy to understand or interpret.

- **Inference**: drawing conclusions about something in the text using the text evidence, your own background knowledge and common sense. Inferences are made about what happened in the past or what is currently happening.
- **Prediction**: using the text evidence ,background knowledge and common sense to make a guess to what will happen in the future.

Classes of Machine Learning problem

- **Supervised Learning**
  - Learn to predict output when given an input vector
  - Training data includes desired outputs
- **Unsupervised Learning**
  - The aim is to uncover the underlying structures (classes or clusters) in the data
  - Training data does not include desired outputs. This is the new frontier of machine learning because most big datasets do not come with labels.
- **Semi-supervised Learning**
  - Desired outputs or classes are available for only a part of the training data.
  - This approach is useful when it is impractical or too expensive to access or measure the target variable for all participants
- **Reinforcement Learning**
  - Learning method that interacts with its environment by **producing actions** and **discovers** errors or rewards.
  - On the basis of trial and error, to discover what actions maximize reward and minimize the penalty.

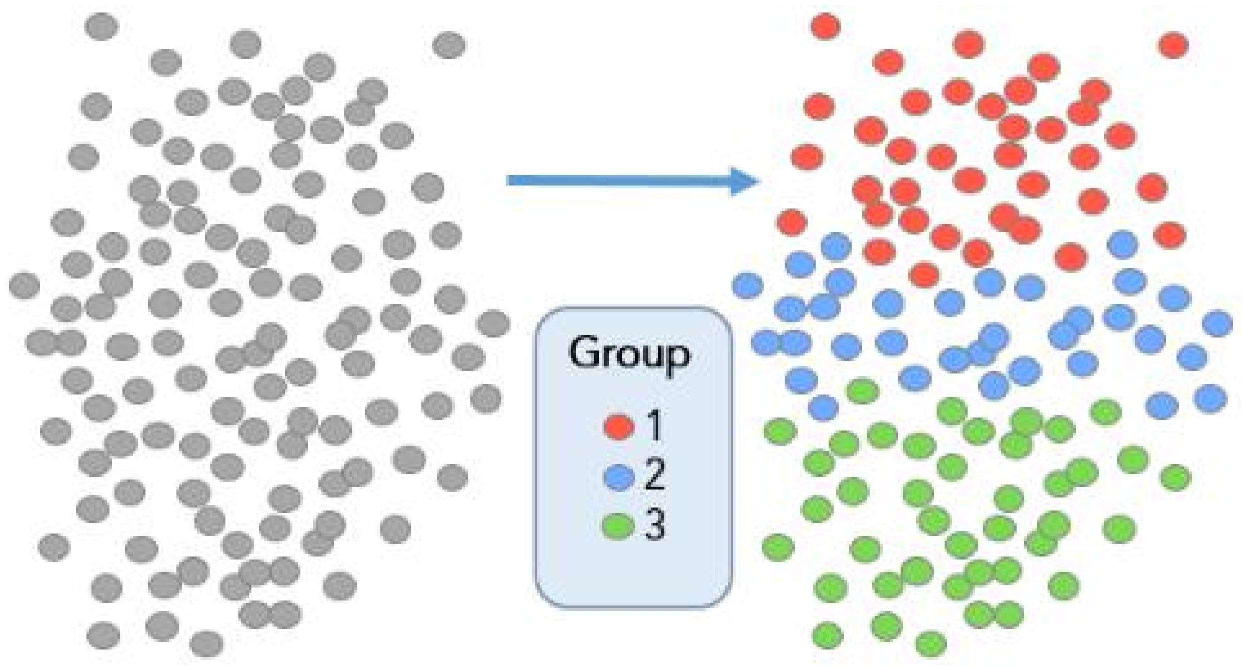
Classes of machine learning Problem



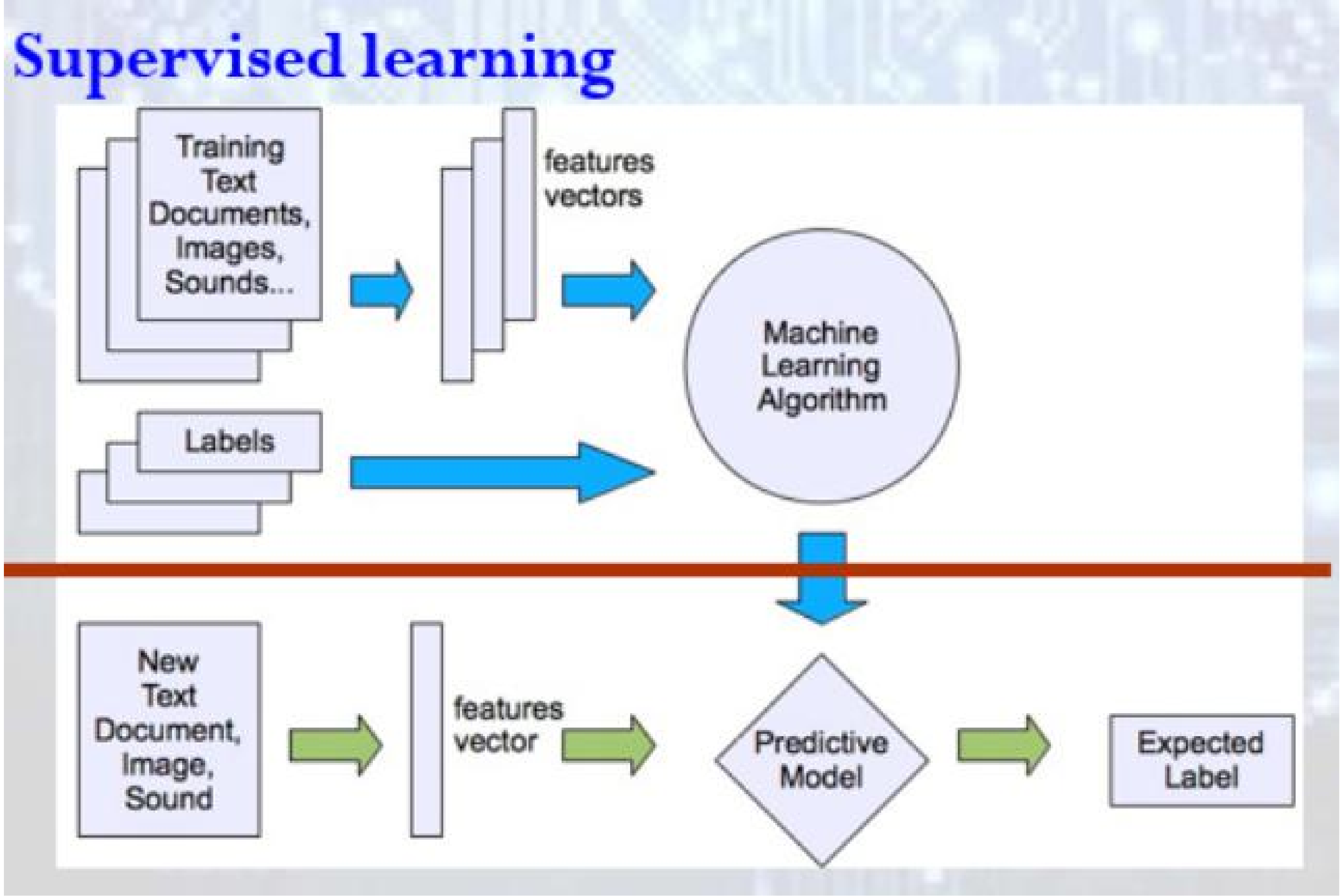
Classes of machine learning Problem

Unsupervised

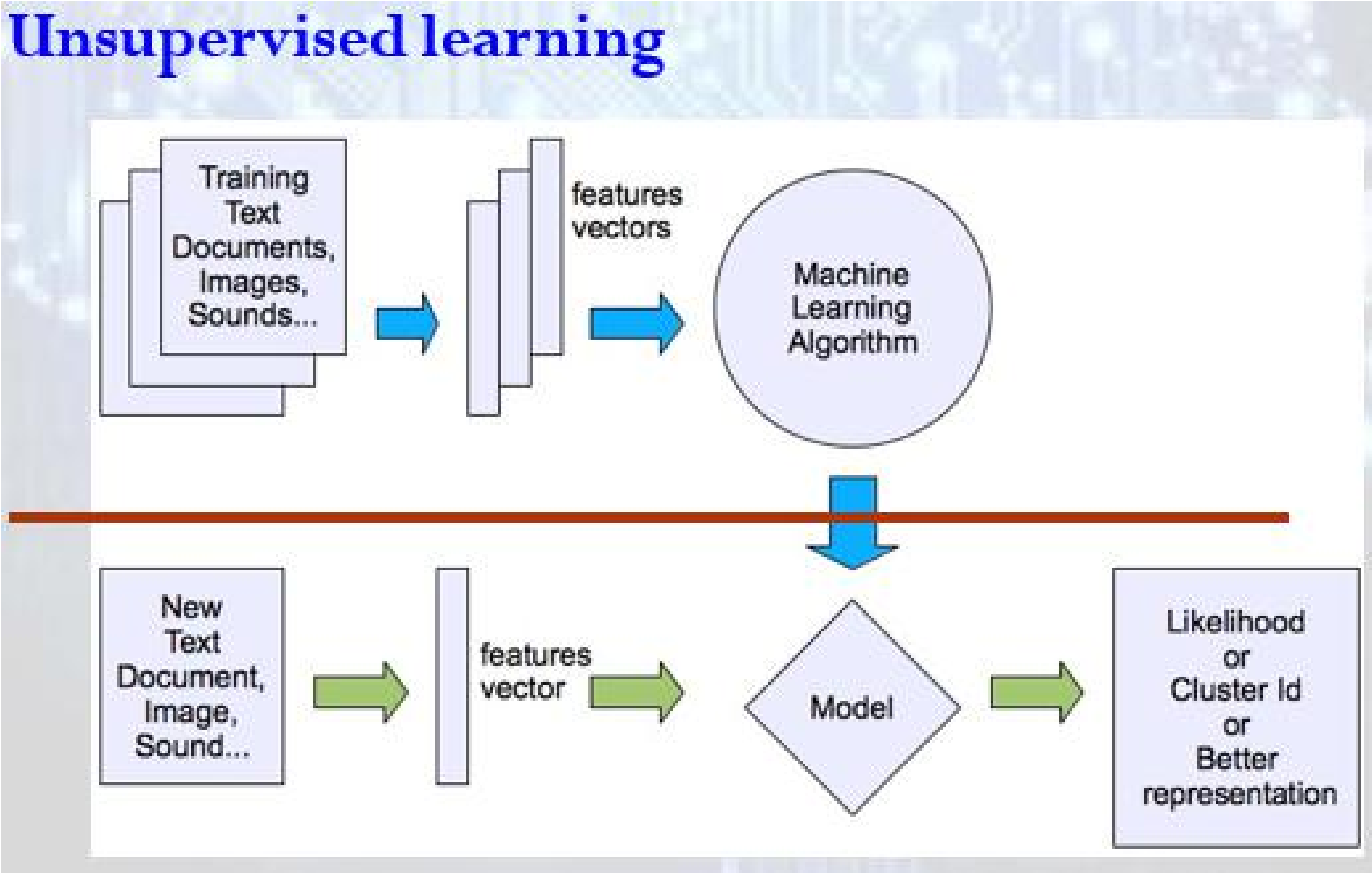
- Clustering
- Dimensionality reduction



Machine learning structure



Machine learning structure



The Learning Problem

- Given <x,f(x)> pairs, infer f

x	f(x)
1	1
2	4
3	9
4	16
5	?

Given a finite sample, it is often impossible to guess the true function f.

Approach: Find some pattern (called a ***hypothesis***) in the training examples, and assume that the pattern will hold for future examples too.

The machine learning framework

$$y = f(\mathbf{x})$$

output prediction function Image feature

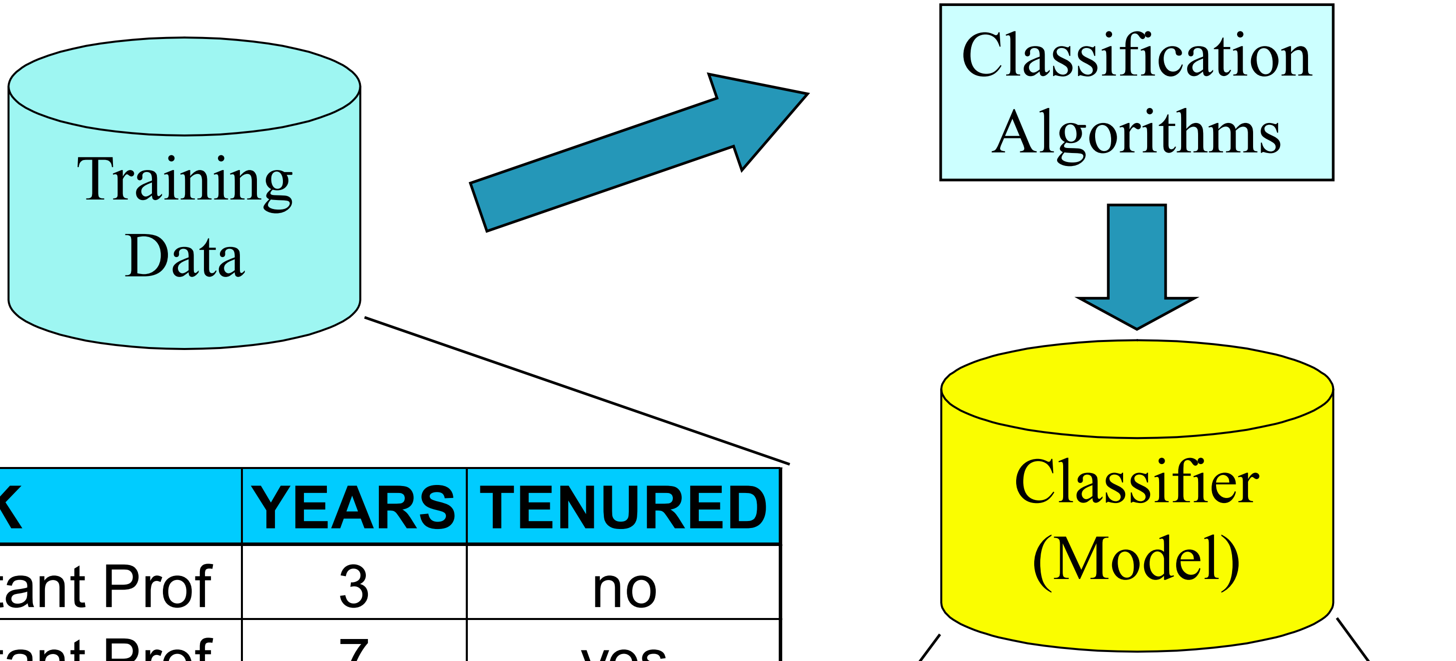
- **Training:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the prediction function **f** by minimizing the prediction error on the training set
- **Testing:** apply **f** to a never before seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

Learning—A Two-Step Process

- **Model construction:**
  - A training set is used to create the model.
  - The model is represented as classification rules, decision trees, or mathematical formula

- **Model usage:**
  - the test set is used to see how well it works for classifying future or unknown objects

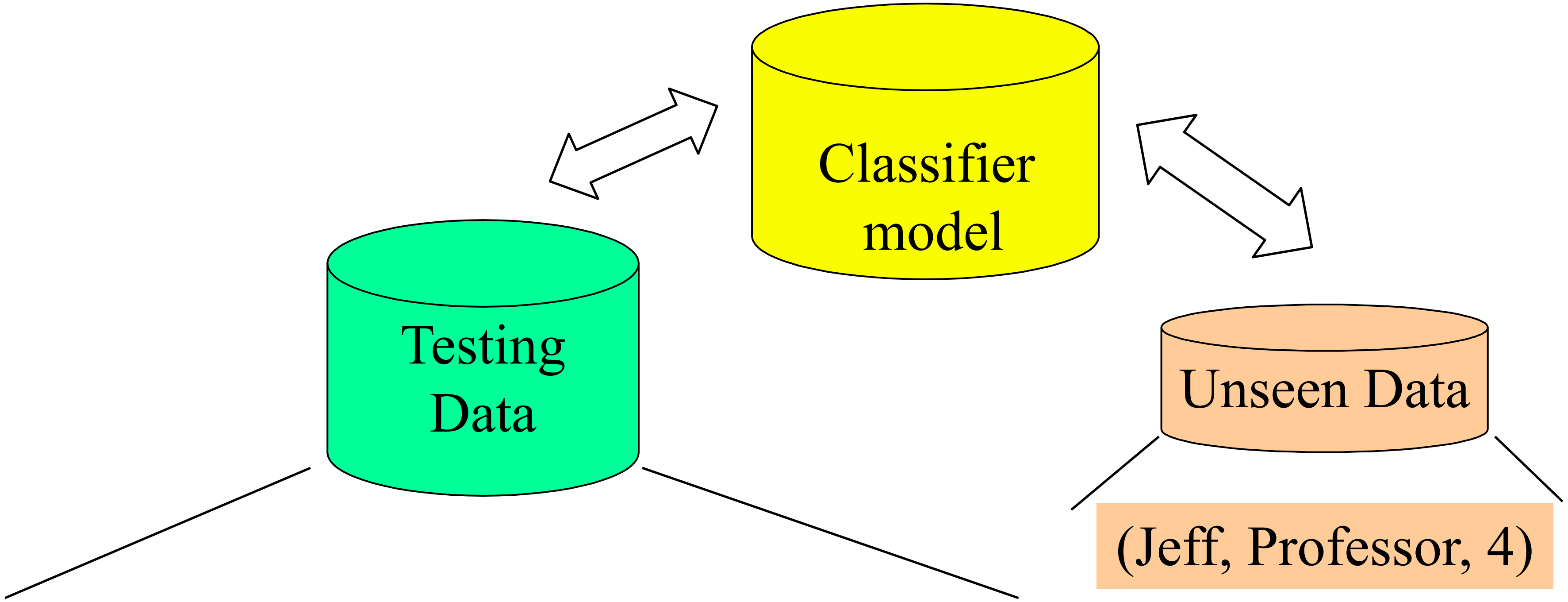
Step 1: Model Construction



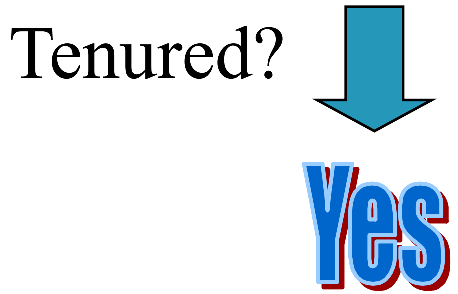
NAME	RANK	YEARS	TENURED
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no

IF rank = ‘professor’
OR years > 6
THEN tenured = ‘yes’

Step 2: Using the Model in Prediction



NAME	RANK	YEARS	TENURED
Tom	Assistant Prof	2	no
Merlisa	Associate Prof	7	no
George	Professor	5	yes
Joseph	Assistant Prof	7	yes



Challenges in Machine Learning

- Efficiency and scalability of machine learning algorithms
- Handling high-dimensionality
- Handling noise, incomplete and imbalanced data
- Pattern evaluation and knowledge integration
- Protection of security, integrity, and privacy in machine learning
- Data acquisition and representation issues
- Degree of interpretability for predictive power
- Deployment issues

Basic Steps in Machine Learning

1. **Data collection**

   “training data”, **mostly** with “labels” provided by a “teacher”;

2. **Data preprocesing**

   Clean data to have homogenity

3. **Feature engineering**

   Select represenatative features to improve performance

4. **Modeling**

   choose the class of models that can describe the data

5. **Estimation/Selection**

   find the model that best explains the data: simple and fits well;

6. **Validation**

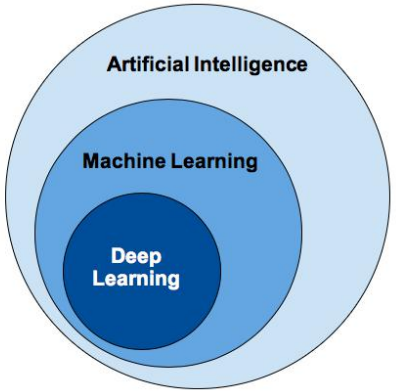
   evaluate the learned model and compare to solution found using other model classes;

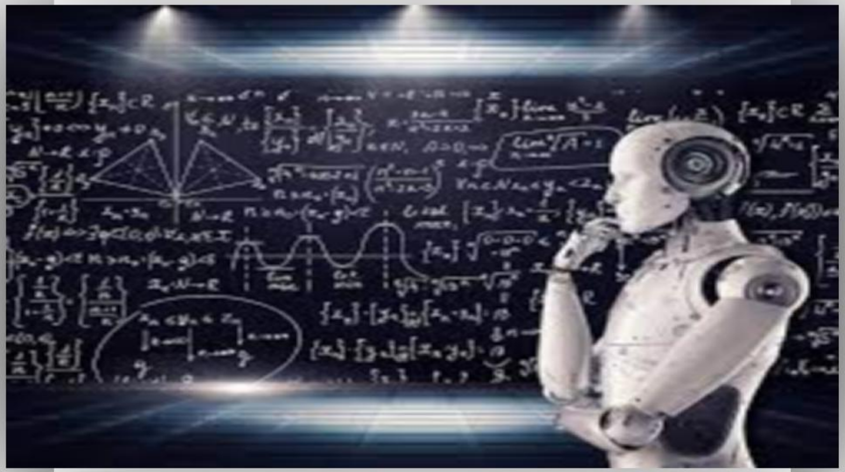
7. **Operation**

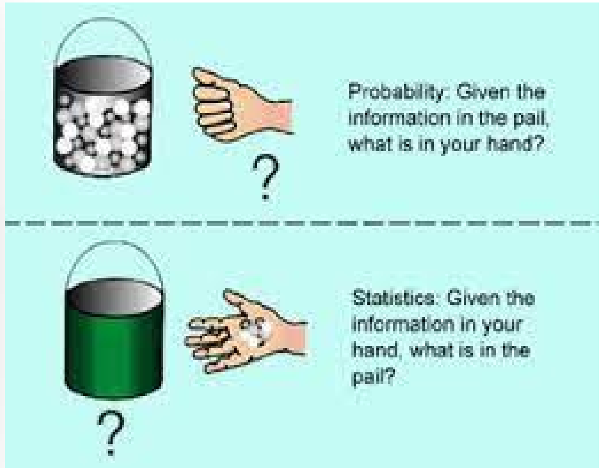
   Apply learned model to new “test” data or real world instances

Did you attend the following courses?

- Linear Algebra
- Calculus





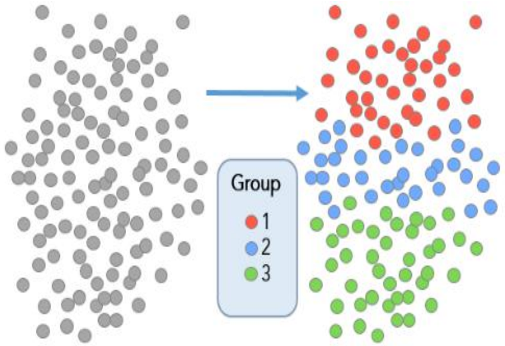


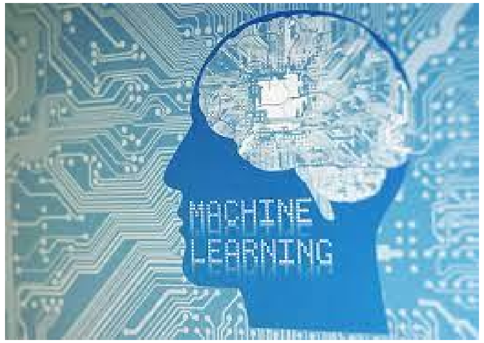
$y = f(\mathbf{x})$

Classifier
(Model)

$$P(E)\ \frac{\text{Number of outcomes in event}}{\text{Total number of outcomes in sample space}}.$$

Unsupervised





Course Project Guidelines

- Project in a group of 2 or 3.
  - Proposal/project description, June 9, 2021: <=1/2 page
  - Presentation, 2 TBD (30%)
  - Final Report, 1 TBD 6-10 pages (70%)
- Projects which have well-designed experiments, on real problems, and a thorough analysis of the results are scored higher.
- The writing style and the clarity of the written paper is an asset to score good grade [latex] -- will provide a template
- You should prepare slides for a 10 minute presentation of your project, with 5 minutes for questions
- The final Report of your project should consist **Abstract**, **Introduction**, **related work**, **Experimental results** and **analysis**, and **conclusions**.