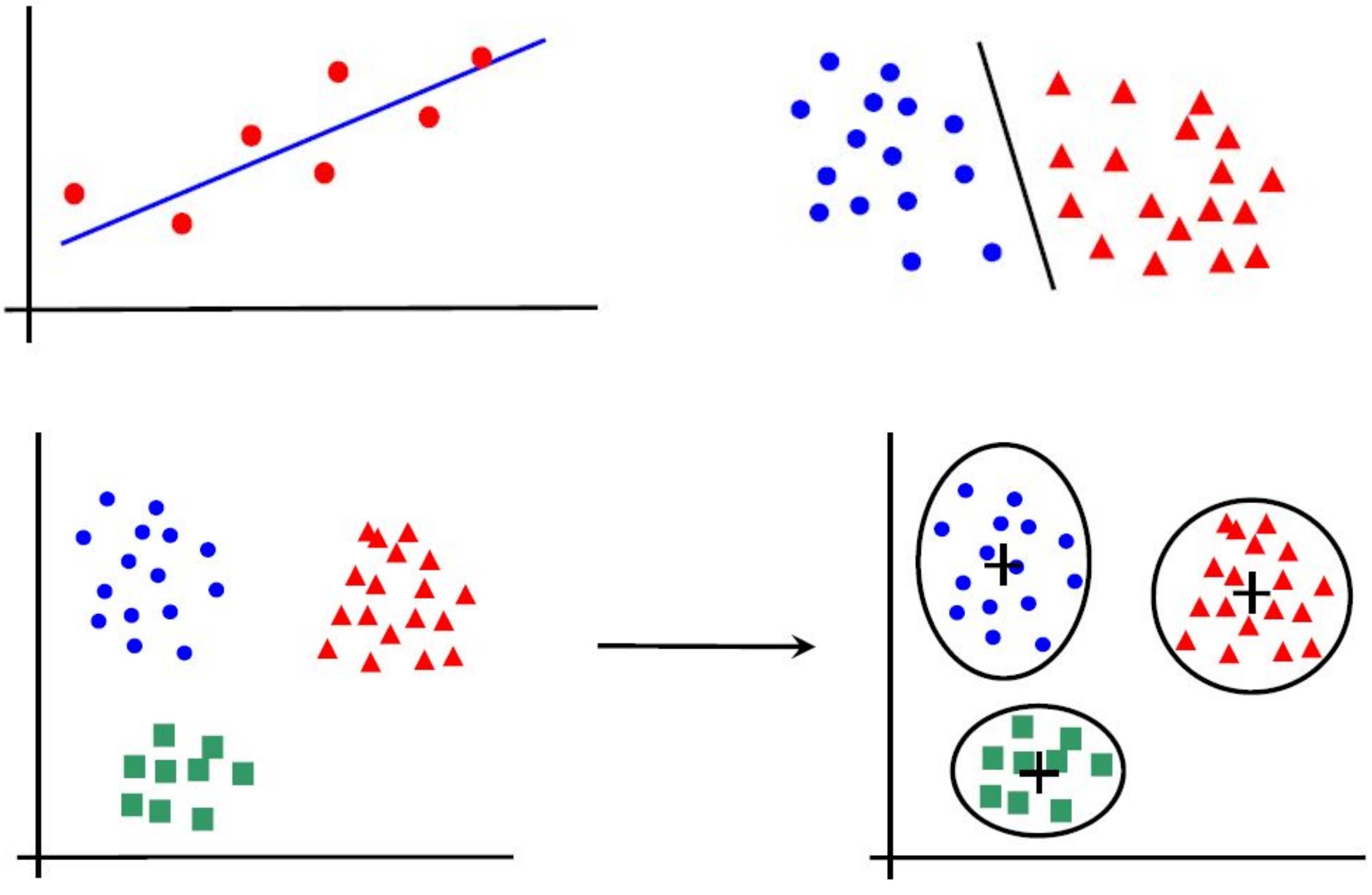
Machine Learning
Introduction

Indian Institute of Information Technology
Sri City, Chittoor

Welcome to Machine Learning Class



Today's Agenda

- Course plan
  - Pre-requisite
  - Topics
  - Textbooks/References
  - Evaluation components
  - Honor code
- Introduction to machine learning
  - What is ML?
  - When do we use ML?
  - Applications
  - Relation with AI and DL
  - Relation with other fields
  - Different machine learning paradigms

Pre-requisite

- Probability
  - Distribution, random variable, expectation, conditional probability, variance, density
- Linear algebra
  - Linear transformation, Rank, Positive definite matrix
  - Eigenvalue & Eigenvector, Diagonalization
- Calculus (Optimization – Minima, maxima)
- Basic programming
  - Python (First Priority)
  - Matlab/C/C++ (Second Priority)

Topics

- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
  - Clustering
- Semi-supervised Learning
- Reinforcement Learning

Textbooks/References

1. "Pattern Classification" by R. O. Duda, P. E. Hart and D. G. Stork.
2. "An Introduction to Statistical Learning" by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani.
3. "Pattern Recognition and Machine Learning" by Christopher M. Bishop.
4. "Introduction to Machine Learning" by Ethem Alpaydin.
5. "Pattern Recognition: An Algorithmic Approach" by M. Narasimha Murty, V. Susheela Devi.
6. "Machine learning" by Tom Mitchell.

Evaluation Components

- Mid-Exam(s): 20%
- End-Exam: 30%
- Assignments: 20%
- Scheduled Quiz: 10%
- Project: 20%

This is tentative and will be finalized in Class committee meeting.

Honor Code

Do’s

- Write down the code independently
- Submit the assignment within the deadline
- Read the books/references for detail description of the topics

Don’ts

- copy, refer to, or look at any official or unofficial previous years’ solutions in preparing the answers

Introduction to ML

- What is ML?
- Terminologies used in ML
- When do we use ML?
- Applications
- Relation with AI and DL
- Relation with other fields
- Different machine learning paradigms

What is ML?

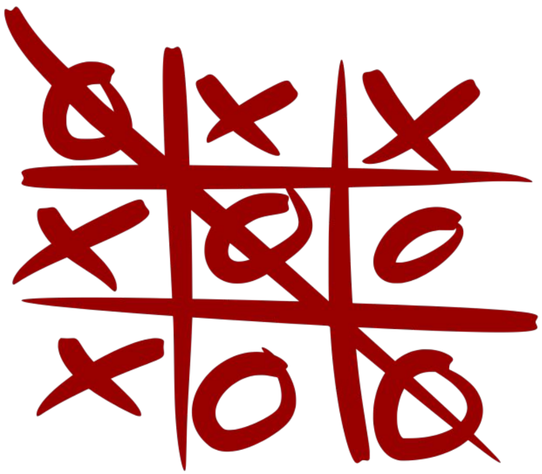
- Machine learning (ML) is the study of computer algorithms that improve automatically through experience.

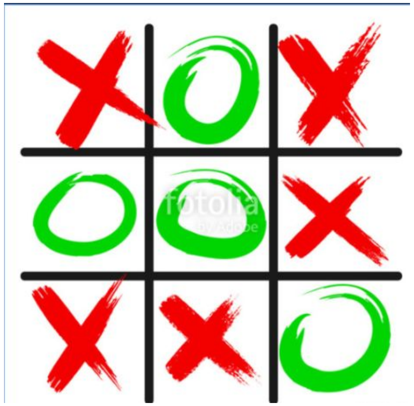
- Machine-learning algorithms use statistics to find patterns in massive amounts of data.

- Traditionally, software engineering combined human created rules with data to create answers to a problem. Instead, machine learning uses data and answers to discover the rules behind a problem – **F. Chollet, Deep Learning with Python**

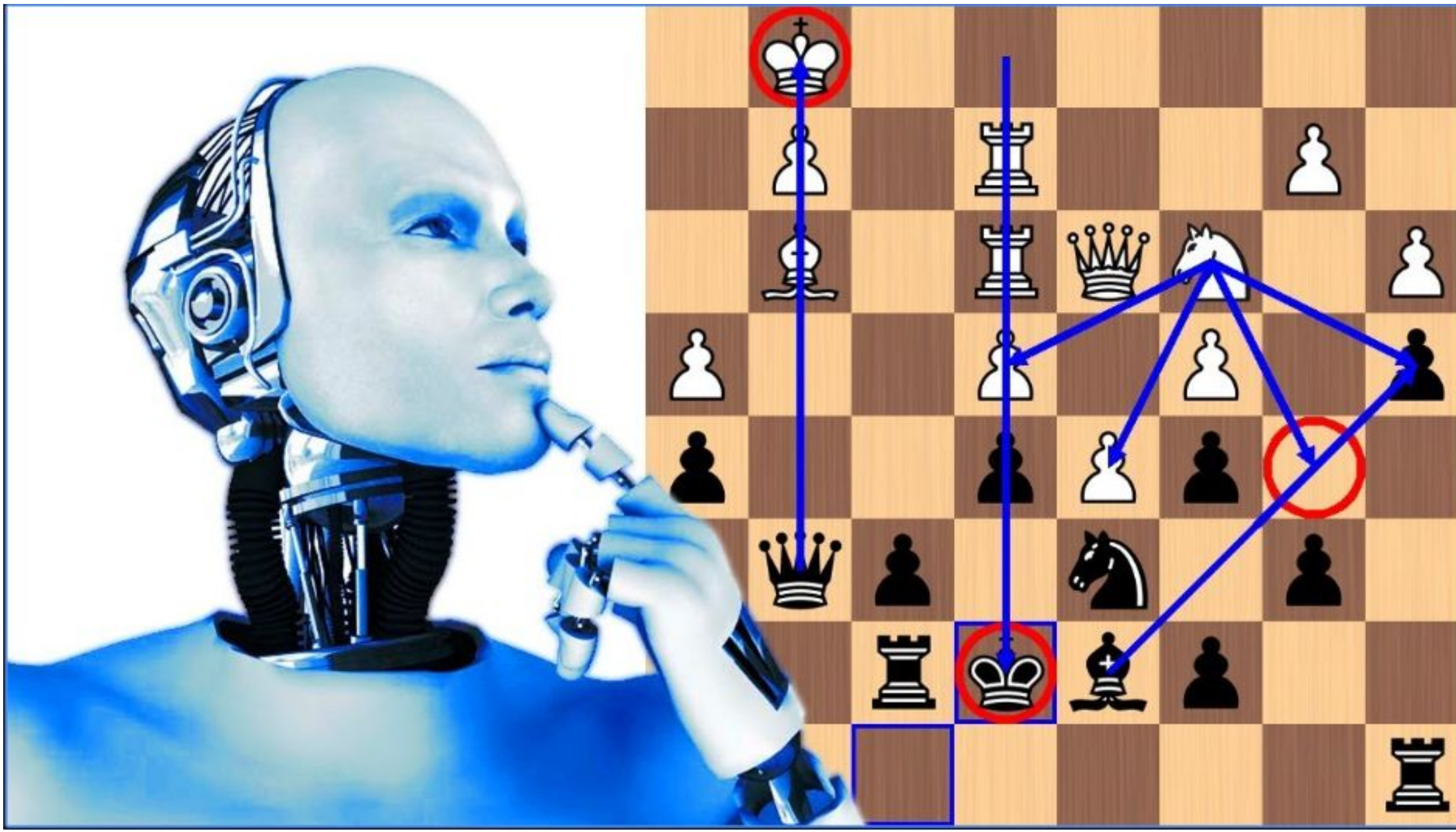
Early AI – Rule based

- Tic-tac-toe

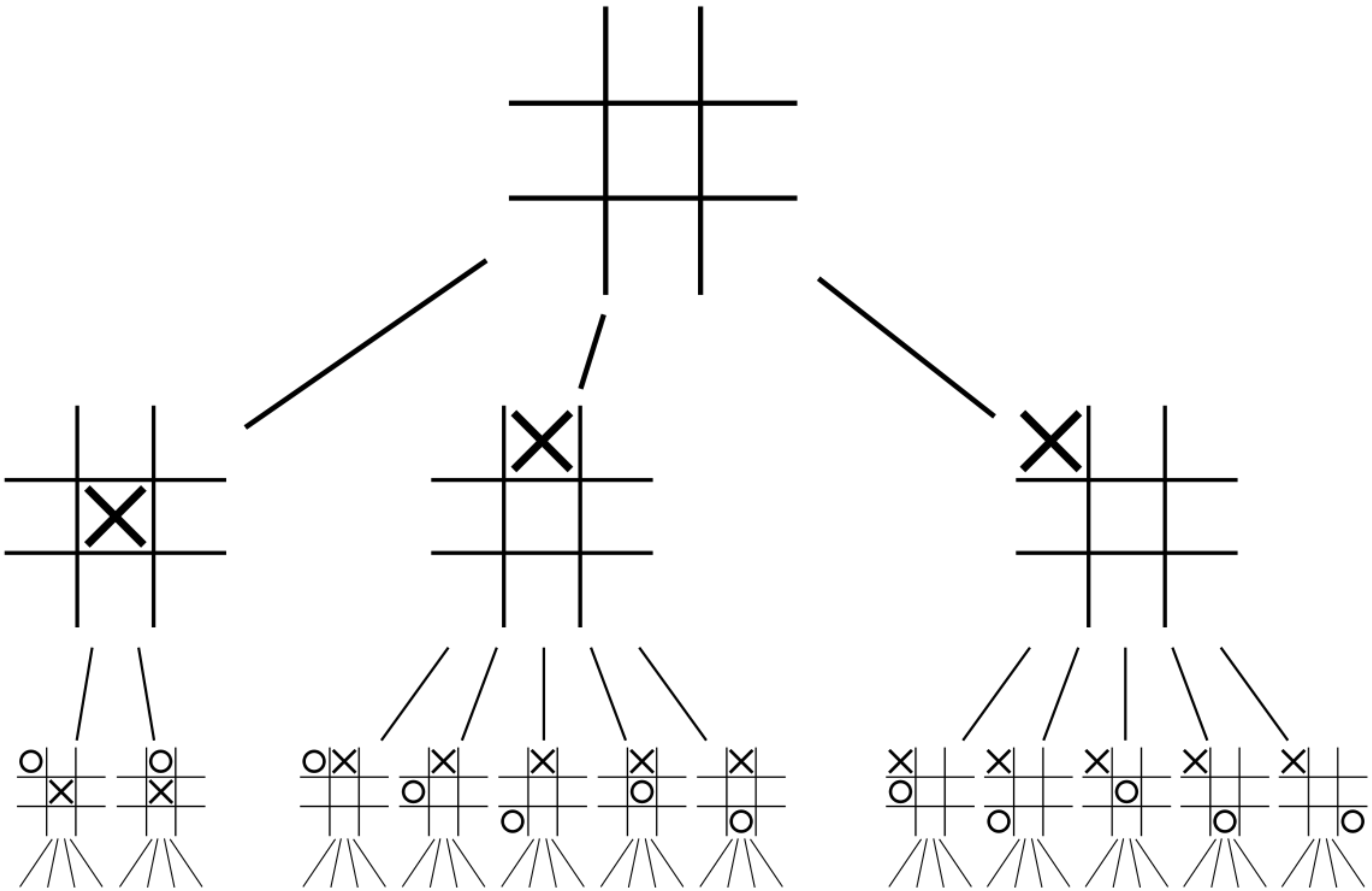




Chess --



Tic-tac-toe – how it is done?

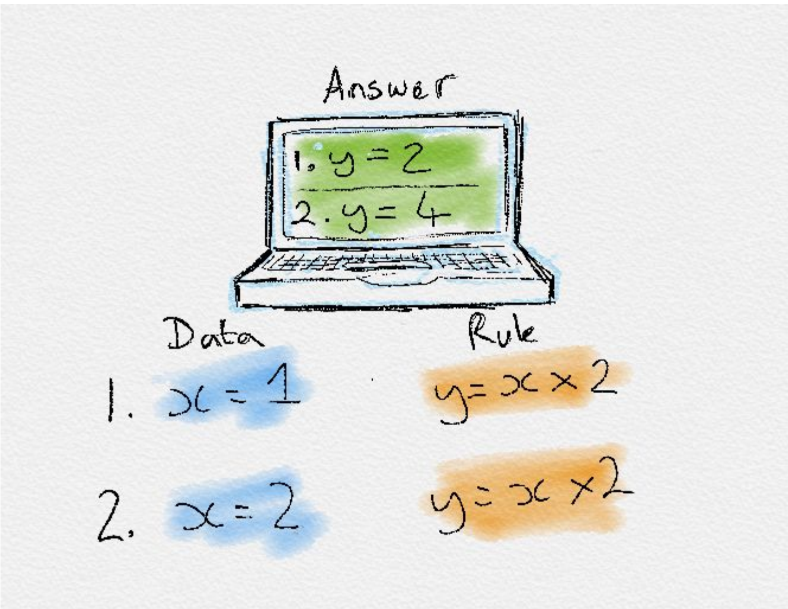


Deductive Vs Inductive

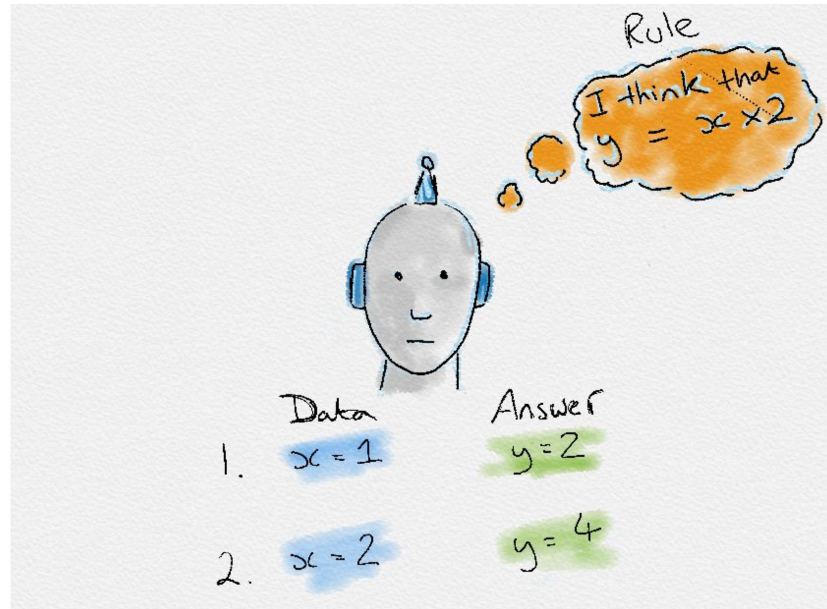
- Giving rules can work for game playing like problems.
- But, you cannot give all rules for a self-driving car.
  - This attempt failed.
- So, inductive means, we will not give explicit rules, but give examples saying which is good which is bad, so on.
  - Machine will learn rules from the data.

Deductive Vs Inductive

Deductive Reasoning	Inductive Learning



Traditional Programming



Machine Learning

Credits: Machine Learning | An Introduction | by Gavin Edwards | Towards Data Science

Machine Learning

- Learning from data

ALGORITHMS

DATA

Terminologies used in ML

- ML systems learn how to make inference from the input data samples to produce useful predictions on un-seen (test) data.
- Input data:
  - labelled examples: A labelled example includes feature(s) and the label. {features, label}: (x, y)For e.g.:

Features:	Label
Normal RBC, Normal HgB	Healthy
Low RBC, Low HgB	Anaemic

  - unlabelled examples: An unlabelled example contains features but not the label. {features, ?}: (x, ?)
    - For e.g.: (Normal RBC, Low HgB)

Features: House type	Price
4BHK,	40,000
4BHK,	15,000
2BHK,	25,000
2BHK,	8,000

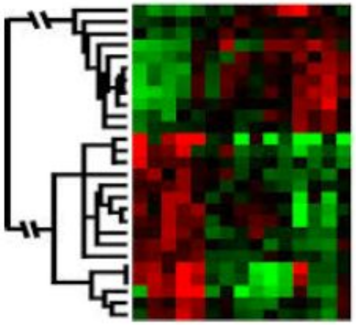
Terminologies used in ML

- Machine Learning Model:
  - A ML model defines the relationship between the features and label.
    - For e.g.: An anaemia diagnostic model might associate certain features strongly with “anaemic“ or “healthy”, and predict the labels based on the association rules it inferred.
  - Two Phases of ML model development
    - **Training** means creating or **learning** the model.
    - **Testing/Inference** means applying the trained model to unlabelled examples in order to infer the label.

When do we use ML?

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:

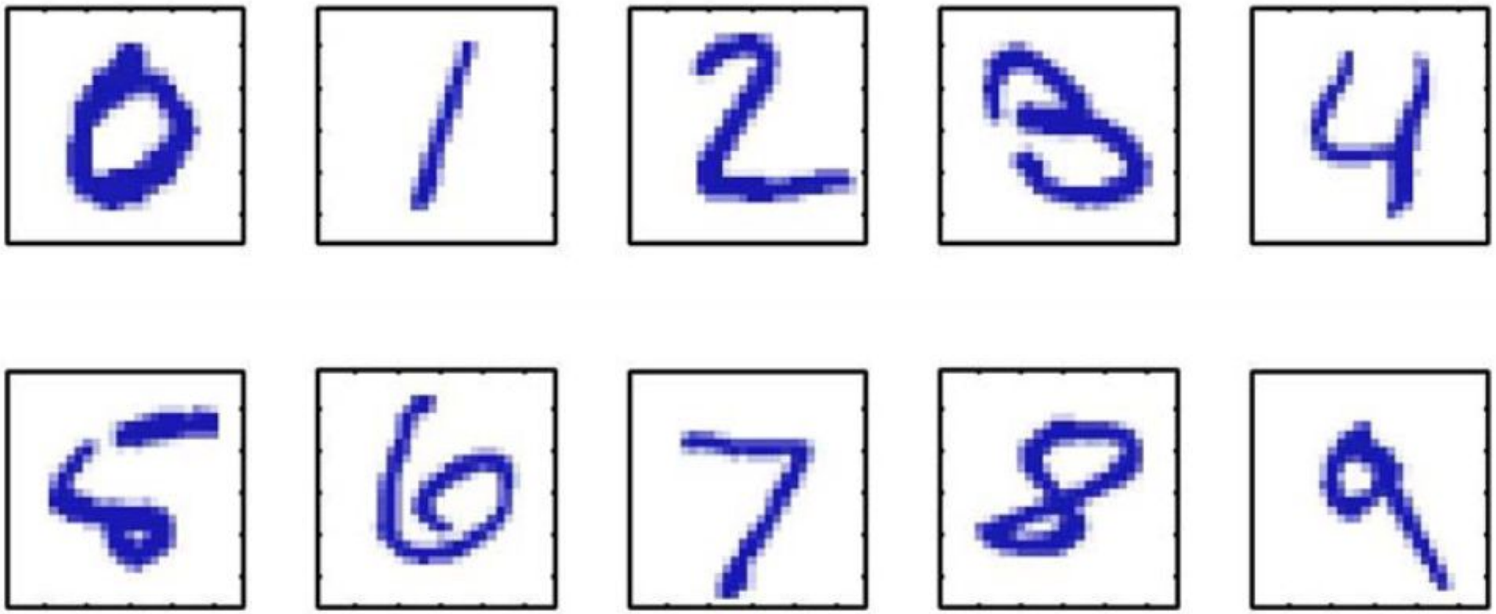
- There is no need to "learn" to calculate payroll

Slide adapted from **E Alpaydin**

Applications

- Hand-written digit recognition
- Speech recognition
- Face detection
- Object classification
- Email spam detection
- Computational biology
- Autonomous cars
- Computer-aided diagnosis

Hand-written Digit Recognition



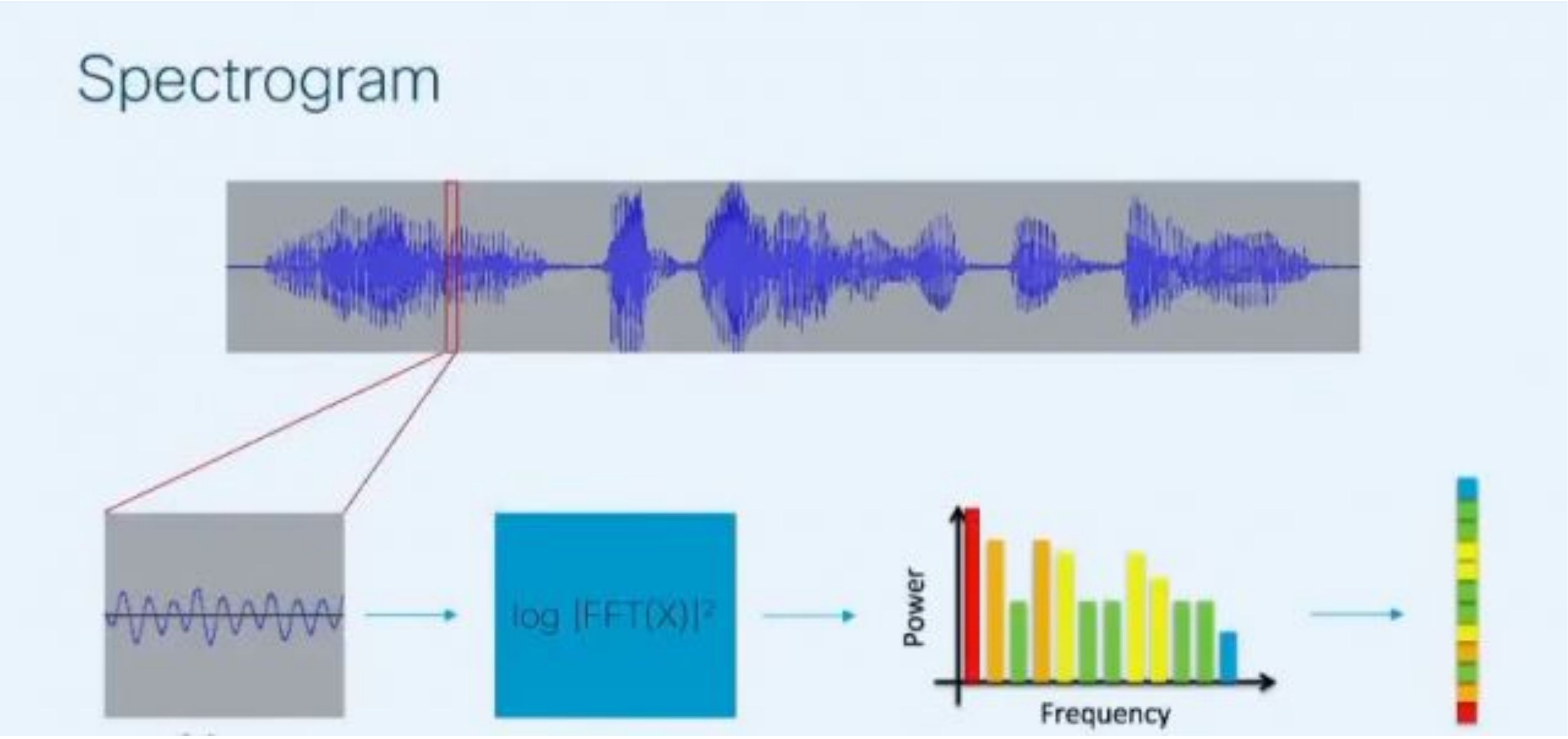
Images are 28 x 28 pixels

Represent input image as a vector $\mathbf{x} \in \mathbb{R}^{784}$

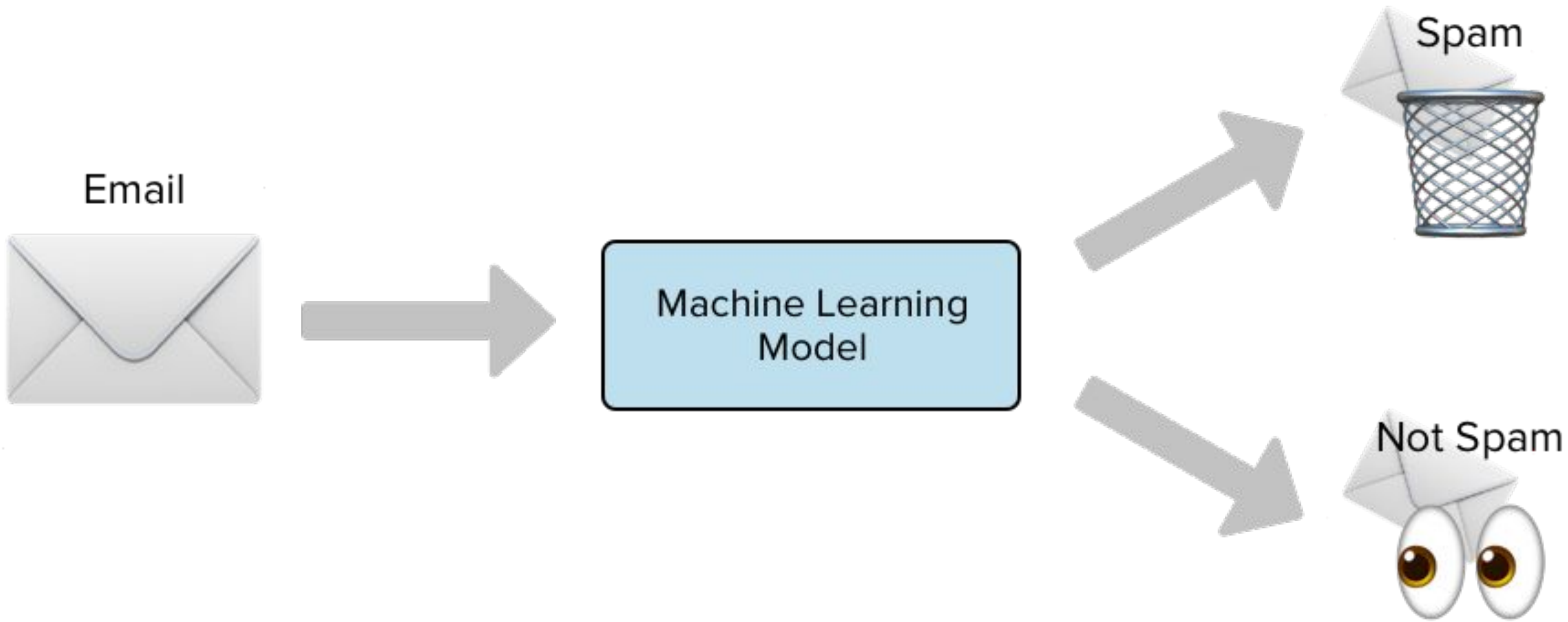
Learn a classifier $f(\mathbf{x})$ such that,

$$f : \mathbf{x} \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

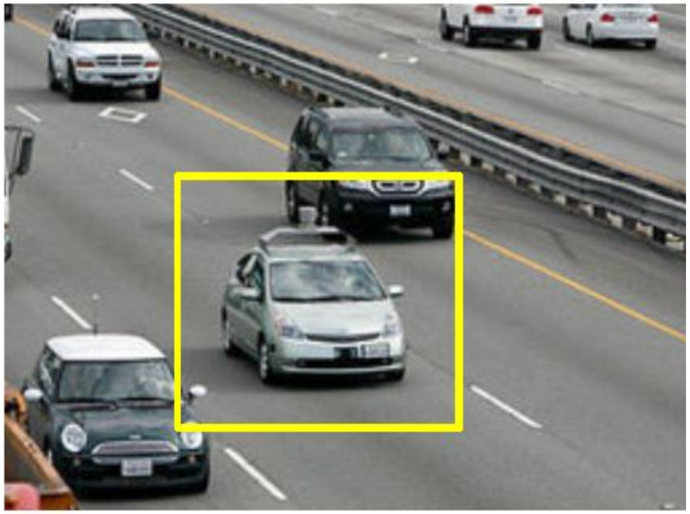
Speech Recognition



Email Spam Detection



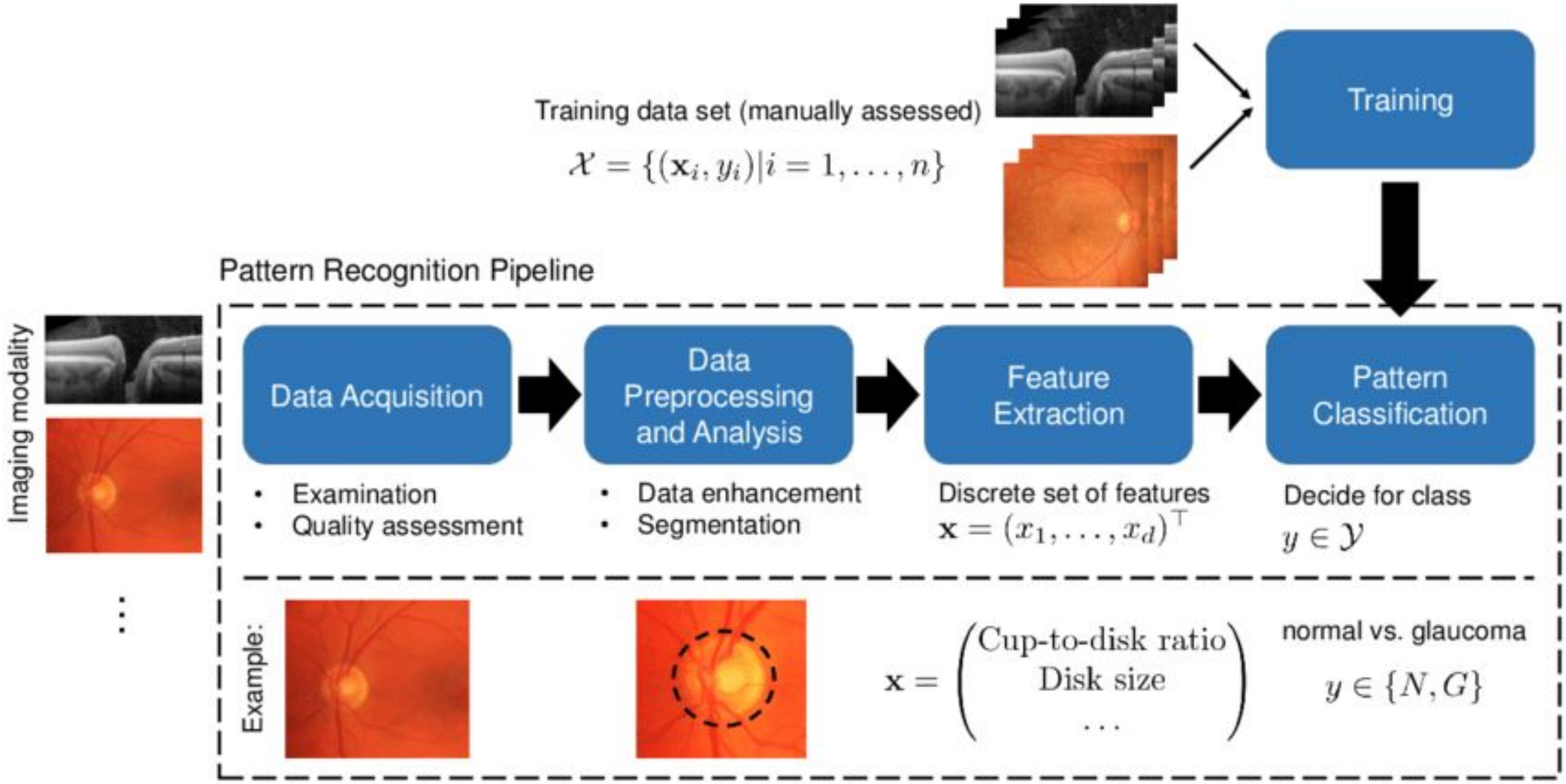
Autonomous Cars



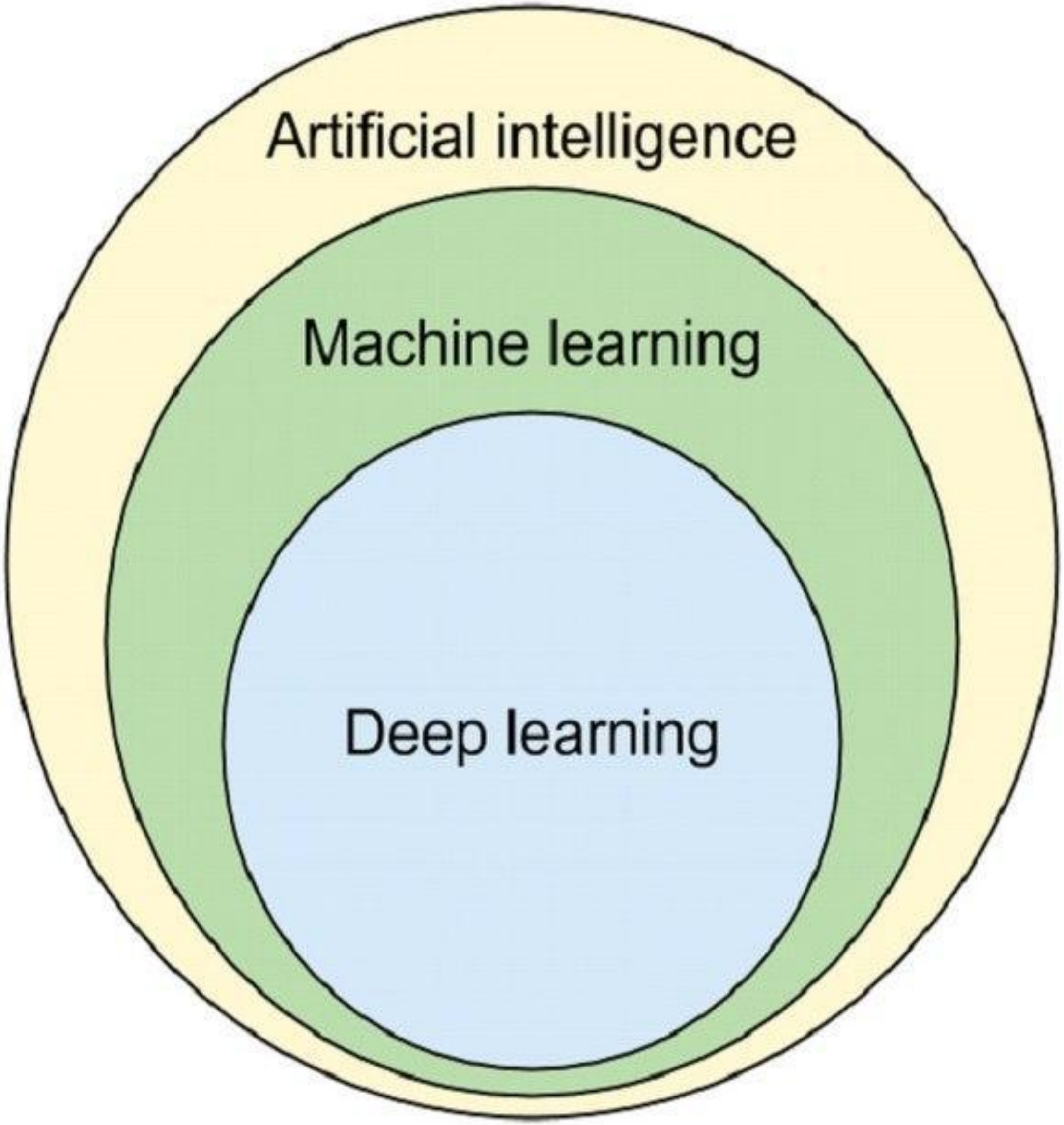
- Nevada made it legal for autonomous cars to drive on roads in June 2011
- As of 2013, four states (Nevada, Florida, California, and Michigan) have legalized autonomous cars

Penn's Autonomous Car →
(Ben Franklin Racing Team)

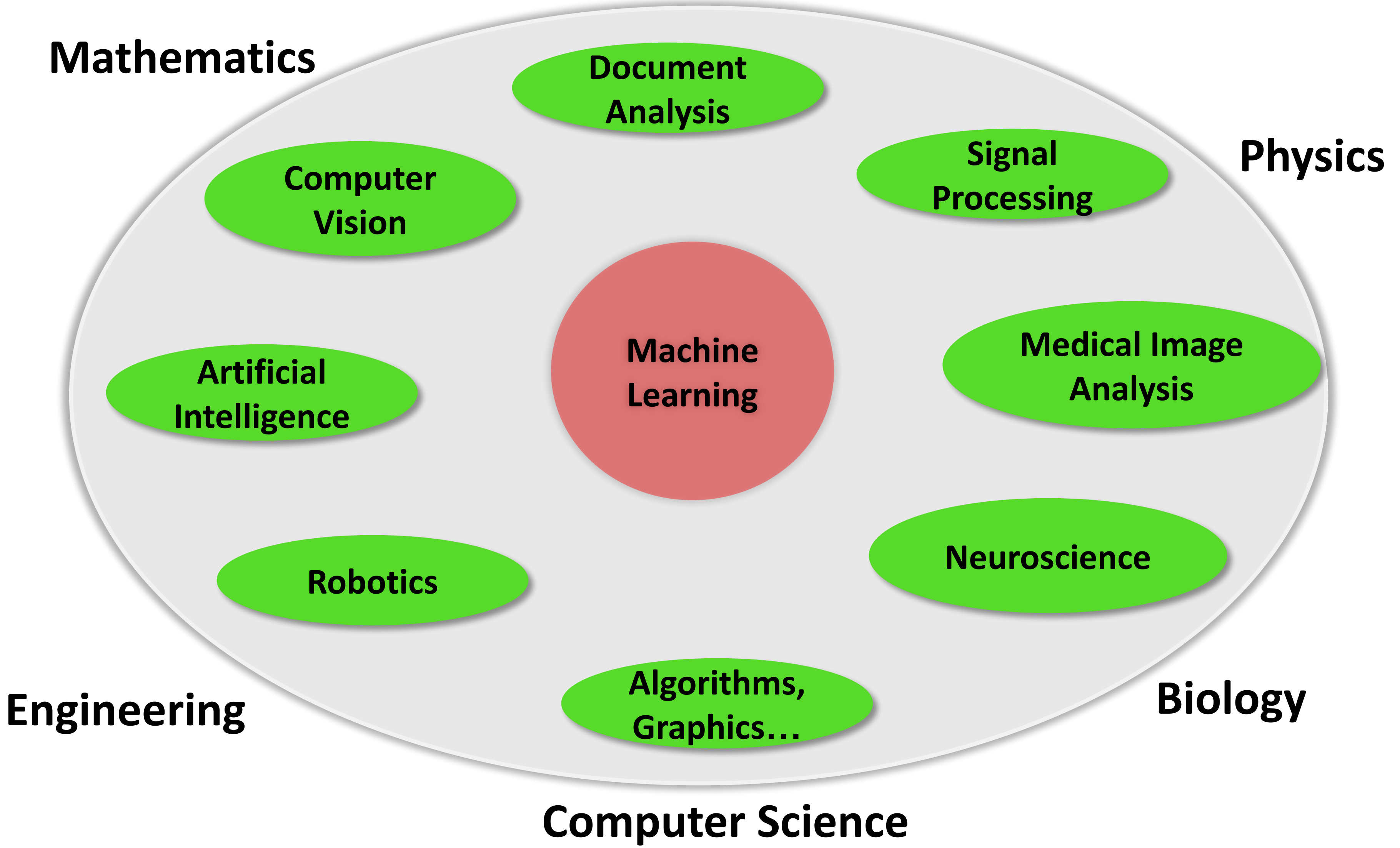
Computer-aided Diagnosis



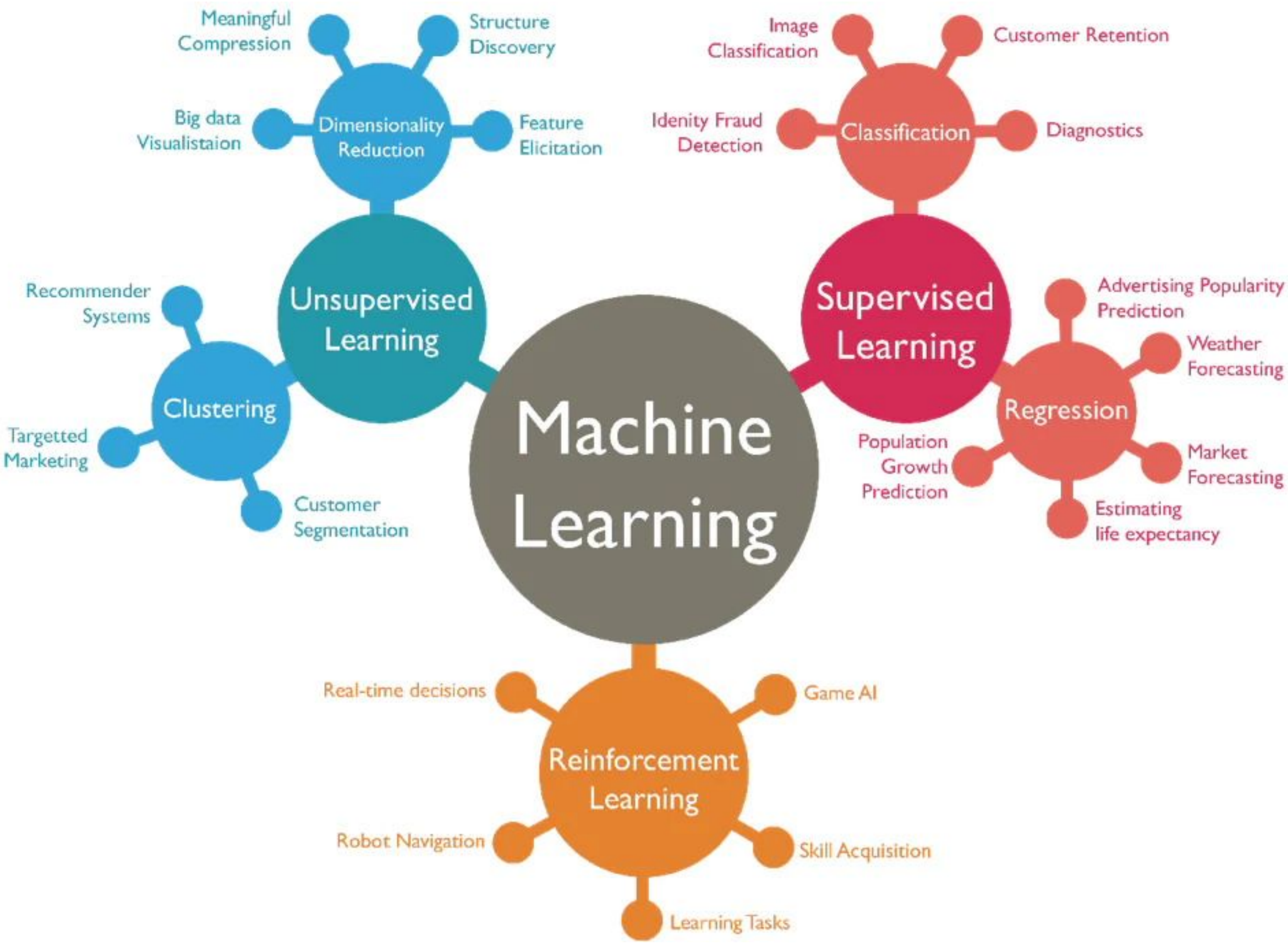
Relation with AI and DL



Relation with Other Fields



Different Machine Learning Paradigms



Credits: J. Chugh, "Types of machine learning and top 10 algorithms everyone should know," 2018

Supervised Learning

- To learn an unknown target function ***f***
- Input: a training set of labeled examples (x_j, y_j) where $y_j = f(x_j)$
- E.g., x_j is an image, $f(x_j)$ is the label "giraffe"
- Output: hypothesis *h* that is "close" to *f*, i.e., predicts well on unseen examples ("test set")
- Many possible hypothesis families for *h* – Linear models, logistic regression, neural networks, decision trees, examples (nearestneighbors) etc.

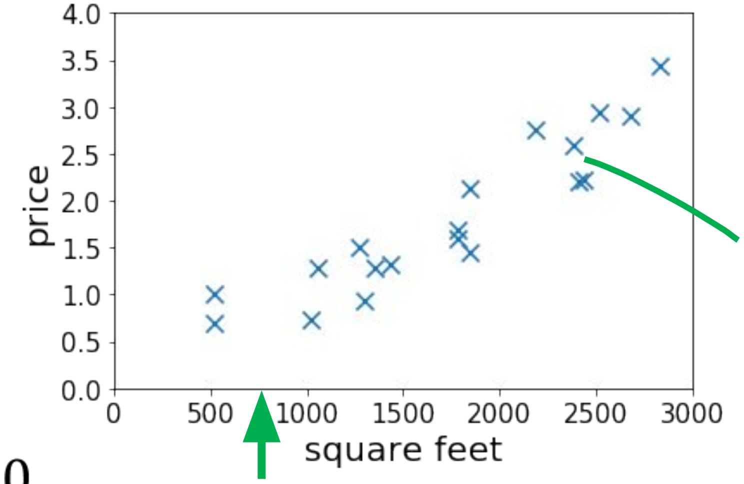
Slide Adapted from Stuart Russell

Housing Price Prediction

- Given: a dataset that contains n samples

$$\left(x^{(1)}, y^{(1)}\right), \ldots \left(x^{(n)}, y^{(n)}\right)$$

- Task: If a residence has x square feet, predict its price?



15th sample
$(x^{(15)}, y^{(15)})$

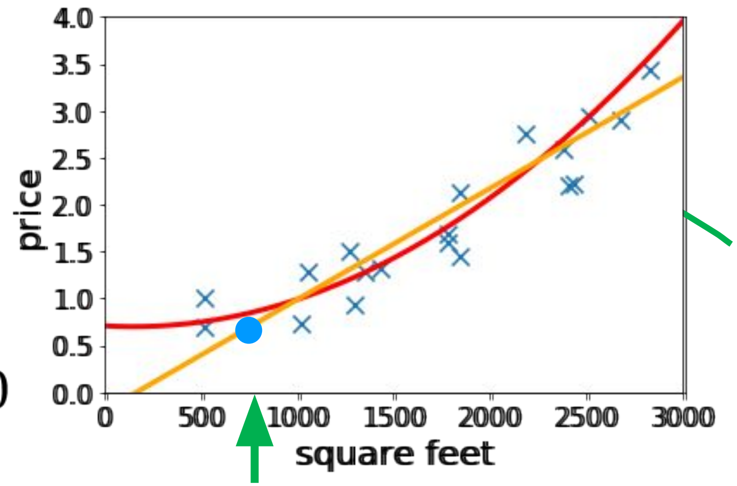
$x = 800$
$y = ?$

Housing Price Prediction

- Given: a dataset that contains n samples

$$\left(x^{(1)}, y^{(1)}\right), \ldots \left(x^{(n)}, y^{(n)}\right)$$

- Task: If a residence has x square feet, predict its price?



15th sample
$(x^{(15)}, y^{(15)})$

$x = 800$
$y = ?$

- Solution: fitting linear/quadratic functions to the dataset.

High-dimensional Features

- $x \in \mathbb{R}^d$ for large d
- E.g.,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ \vdots \\ \vdots \\ x_d \end{bmatrix} \begin{array}{l} \text{--- living size} \\ \text{--- lot size} \\ \text{--- \# floors} \\ \text{--- condition} \\ \text{--- zip code} \\ \qquad\vdots \\ \\ \end{array} \longrightarrow \quad y \text{ --- price}$$

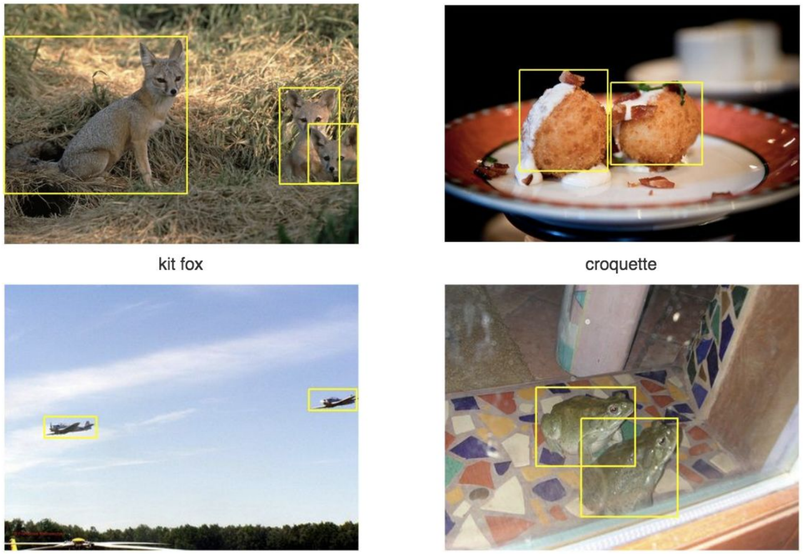
Supervised Learning in CV

- Image Classification
  - x = raw pixels of the image, y = the main object

ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

Supervised Learning in CV

- Object localization and detection
  - x = raw pixels of the image, y = the bounding boxes

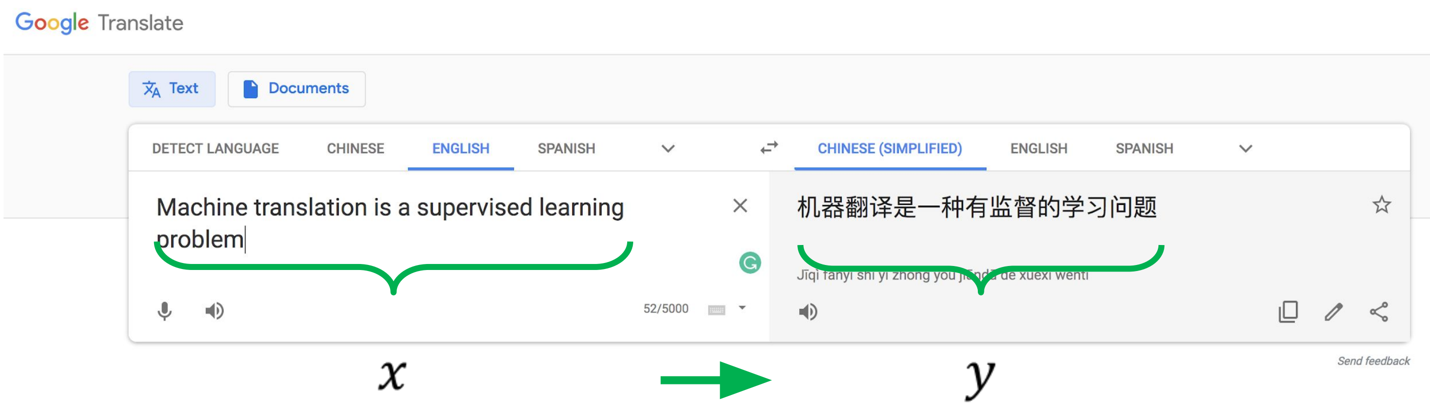


kit fox, croquette, airplane, frog

ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

Supervised Learning in Natural Language Processing

- Machine translation



- Note: This course only covers the basic and fundamental techniques of supervised learning (which are not enough for solving hard vision or NLP problems.)

Supervised Learning

Advantage

- Well established field.
  - has good quantified methods (often with theoretical bounds)

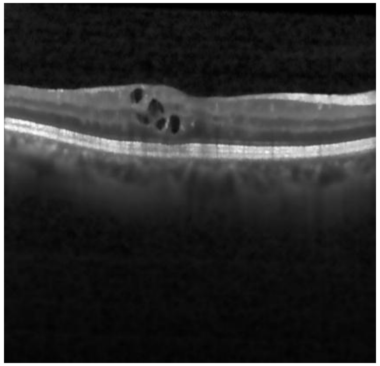
- You can easily test and debug your learning machine.
  - Since the labelled data is available you can easily inspect its output and find out what errors it's making on what type of input data.

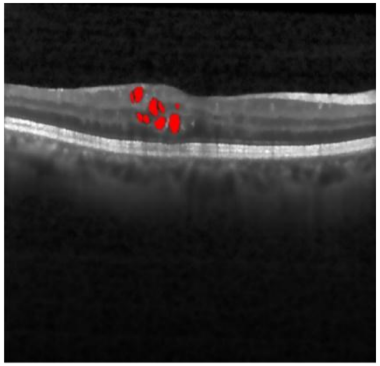
Supervised Learning

Disadvantage

- Collecting and labelling data is expensive and time-consuming.

  Example: Speech Recognition, Medical Image Analysis, etc.

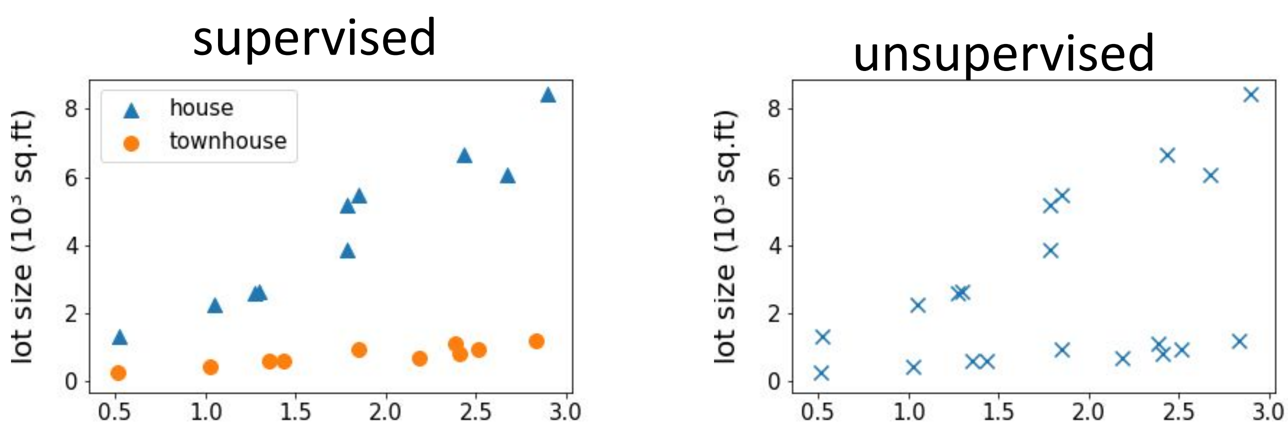

Original Retinal Scan


Intra-retinal cysts annotated scan

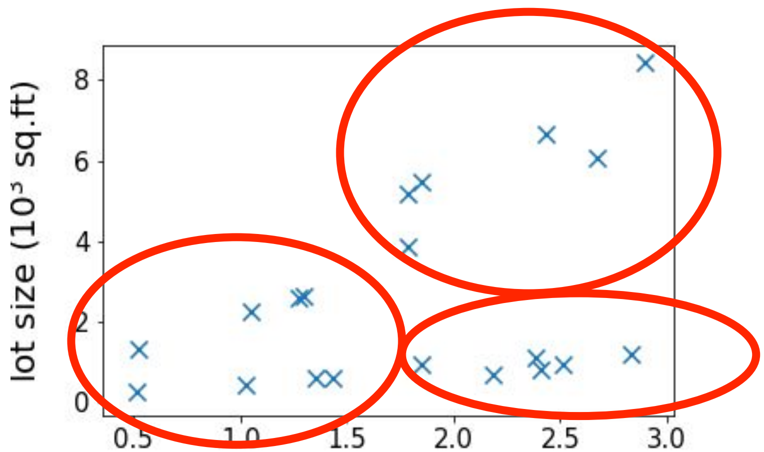
- Errors in your training data might confuse your algorithm and lower its accuracy. Garbage-in -> Garbage-out

Unsupervised Learning

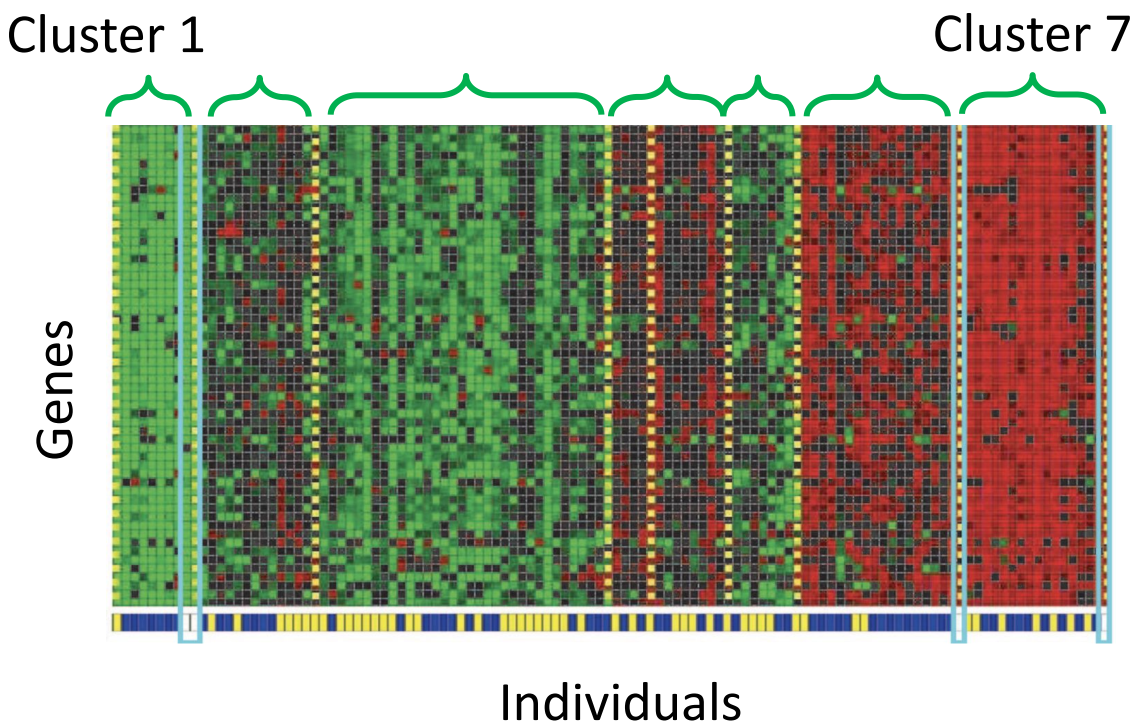
- Dataset contains no labels: $x^{(1)}, \ldots x^{(n)}$
- Goal (vaguely-posed): to find interesting structures in the data



Clustering



Clustering Genes



Identifying Regulatory Mechanisms using Individual Variation Reveals Key Role for Chromatin Modification. [Su-In Lee, Dana Pe'er, Aimee M. Dudley, George M. Church and Daphne Koller. '06]

Need for Unsupervised Learning

- Annotating large datasets is very costly and time consuming. Example: Speech Recognition, Medical Image Analysis, etc.

- There may be cases where we don't know how many/what classes is the data divided into.

  Example: Data Mining, Sentimental Analysis.

- We may want to use clustering to gain some insight into the structure of the data before designing a classifier.

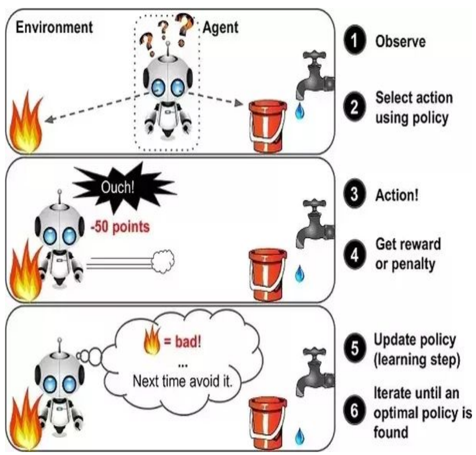
Disadvantages of Unsupervised Learning

- Unsupervised Learning is harder as compared to Supervised Learning. Since, quantifying the inference made is difficult, since the ground-truth is missing.

- How do we know if results are meaningful since it has unlabelled data?
  - External evaluation- Expert analysis.
  - Internal evaluation- Objective function.

Reinforcement Learning

A reinforcement learning algorithm, or agent, learns by interacting with its environment. The agent receives rewards by performing correctly and penalties for performing incorrectly.



Reinforcement learning example: Fire fighting agent

Need for Reinforcement Learning

- Reinforcement learning can be used to solve very complex problems that cannot be solved by conventional techniques.
- In the absence of a training dataset, it is bound to learn from its experience.
- Reinforcement learning models can outperform humans in many tasks and learning process is similar to human learning.
- DeepMind’s AlphaGo program, a reinforcement learning model, beat the world champion *Lee Sedol* at the game of *Go* in March 2016.

Disadvantages of Reinforcement Learning

- Reinforcement learning needs a lot of data and a lot of computation. It is data-hungry.
  - So for solving video games and puzzles it performs well.

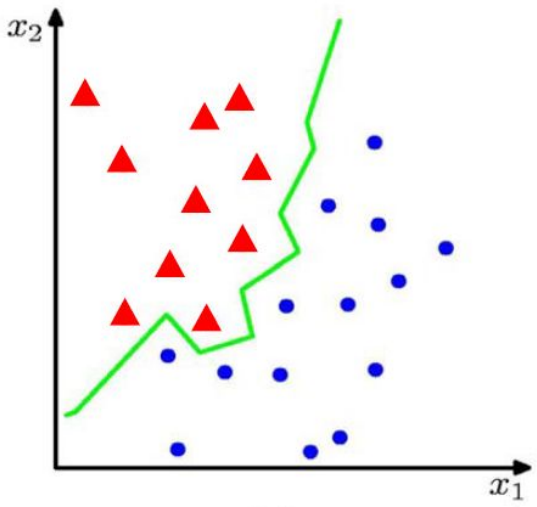
- Reinforcement learning assumes the world is Markovian, which it is not.
  - The Markovian model describes a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

What is classification problem?

- Let there are two classes of objects.
  - Class 1: Set of dog pictures
  - Class 2: Set of cat pictures

- Problem is
  - Given a picture, you should say whether it is cat or dog.
  - For a human being it is easy…, but for a machine it is a non-trivial problem.

What is classification problem?



- Suppose we are given a training set of N observations

$$(x_1, \ldots, x_N) \text{ and } (y_1, \ldots, y_N), x_i \in \mathbb{R}^d, y_i \in \{-1, 1\}$$

- Classification problem is to estimate f(x) from this data such that

$$f(x_i) = y_i$$

Classification: Supervised Learning

Training Phase

We have shown a set of dog pictures and a set of cat pictures to a child.

Classification: Supervised Learning

Testing Phase

→

→ DOG

This picture as it is
may not be in the
training set

Child has done more than just remembering

What is Learning?

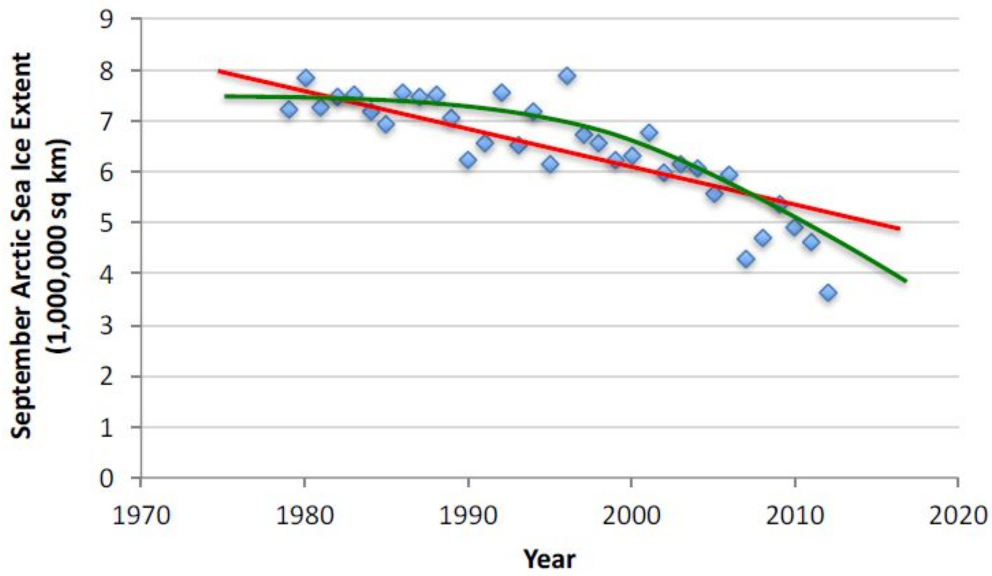
- Child has learnt what is it that is common among dogs … and, what is it that is common among cats… also, what are the distinguishing features/attributes.

- Child has learnt the pattern (regularity) behind all dogs and the pattern behind all cats.

- Child then recognized a test image as having a particular pattern that is unique to dogs.

What is Regression Problem?

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is real-valued == regression



Popular ML algorithms

Classification

- Linear Classifiers
- Support Vector Machines
- Decision Trees
- K-Nearest Neighbor
- Random Forest

Regression

- Linear Regression
- Logistic Regression
- Polynomial Regression

Resources: Journals

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- IEEE Transactions on Neural Networks
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Annals of Statistics
- Journal of the American Statistical Association

Resources: Conferences

- International Conference on Machine learning (ICML)
- European Conference on Machine Learning (ECML)
- Neural Information Processing Systems (NIPS)
- Computational Learning
- International Joint Conference on Artificial Intelligence (IJCAI)
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)
- IEEE Int. Conf. on Data Mining (ICDM)

Thank You:
Question?