UVA CS 4774: Machine Learning

Lecture 1: Introduction

Dr. Yanjun Qi

University of Virginia
Department of Computer Science

Roadmap

- Course Logistics
- History and Now
- A Rough Plan of Course Content

ATT:

- video-recording of my lectures are available
- Each session will include multiple small modules --- each about 20mins (each into a different recorded video)

Welcome

- Course Website: (now please open your Collab)
- Discussion via Slack Channel
- Announcement via email (see collab overview)
- Communication with instructor team via: email (see collab overview)

We focus on learning fundamental principles, algorithm design and deep learning methods and applications.

Objective

- To help students get able to build simple machine learning tools
  - (not just a tool user!!!)

- Key Results:
  - Able to build a few simple machine learning methods from scratch
  - Able to understand a few complex machine learning methods at the source code level

Course Staff

- Instructor: Prof. Yanjun Qi
  - QI: /ch ee/
  - You can call me “professor”, “professor Qi”;
  - I have been teaching Graduate-level and Under-Level Machine Learning course for years!
  - My research is about machine learning

- TA and Office Hour information @ CourseWeb

Course Material

- Text books for this class is:
  - NONE
  - Multiple good reference books are shared via CourseWeb
- My slides – **if it is not mentioned in my slides, it is not an official topic of the course**
- Your UVA Collab for Assignments and Project
- Google Forms for Quizzes

Course Background Needed

- **Background Needed**
  - Calculus, Basic linear algebra,
  - Basic probability and Basic Algorithm
  - Statistics is recommended.
  - Python is required for all programming assignments

Assignments

- Assignments (75%, with five assignments)

- See policy in CourseWeb

Quiz

- Class quizzes (25%): Each takes 10 mins via google form;
  - We will have a total of 12 quizzes
  - Your top 10 scored will be counted into 25%
  - Within Our Zoom Synchronous sessions
  - Will be close book (please follow honor code!)

- See policy in CourseWeb

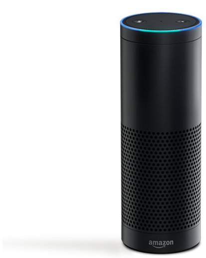

Thank You

Roadmap

- Course Logistics
- History and Now
- A Rough Plan of Course Content

Artificial Intelligence Today

Watson (IBM)



Echo
(Amazon)

SIRI
(Apple)

Boston Dynamics

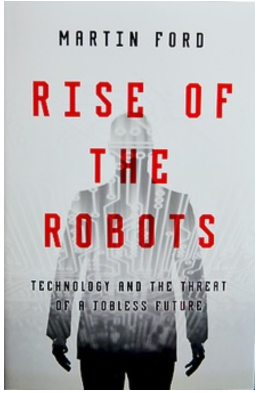
DeepMind (Google)

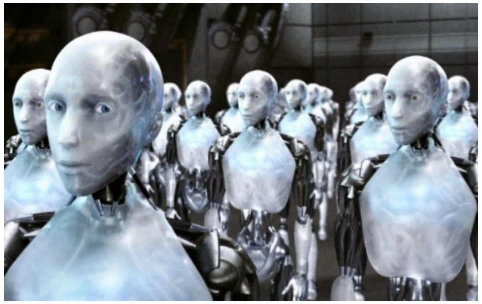
Impact: Good and Bad?

Economic, cultural, social, health… endless disruption

Labor - McKinsey >50% of jobs automated

Martin Ford, Rise of the Robots





Elon Musk, artificial intelligence... existential threat

Image credits: sibaraki@acm.org

Artificial intelligence (AI)

The study of computer systems that attempt to model and apply the intelligence of the human mind.

What defines "intelligence"?

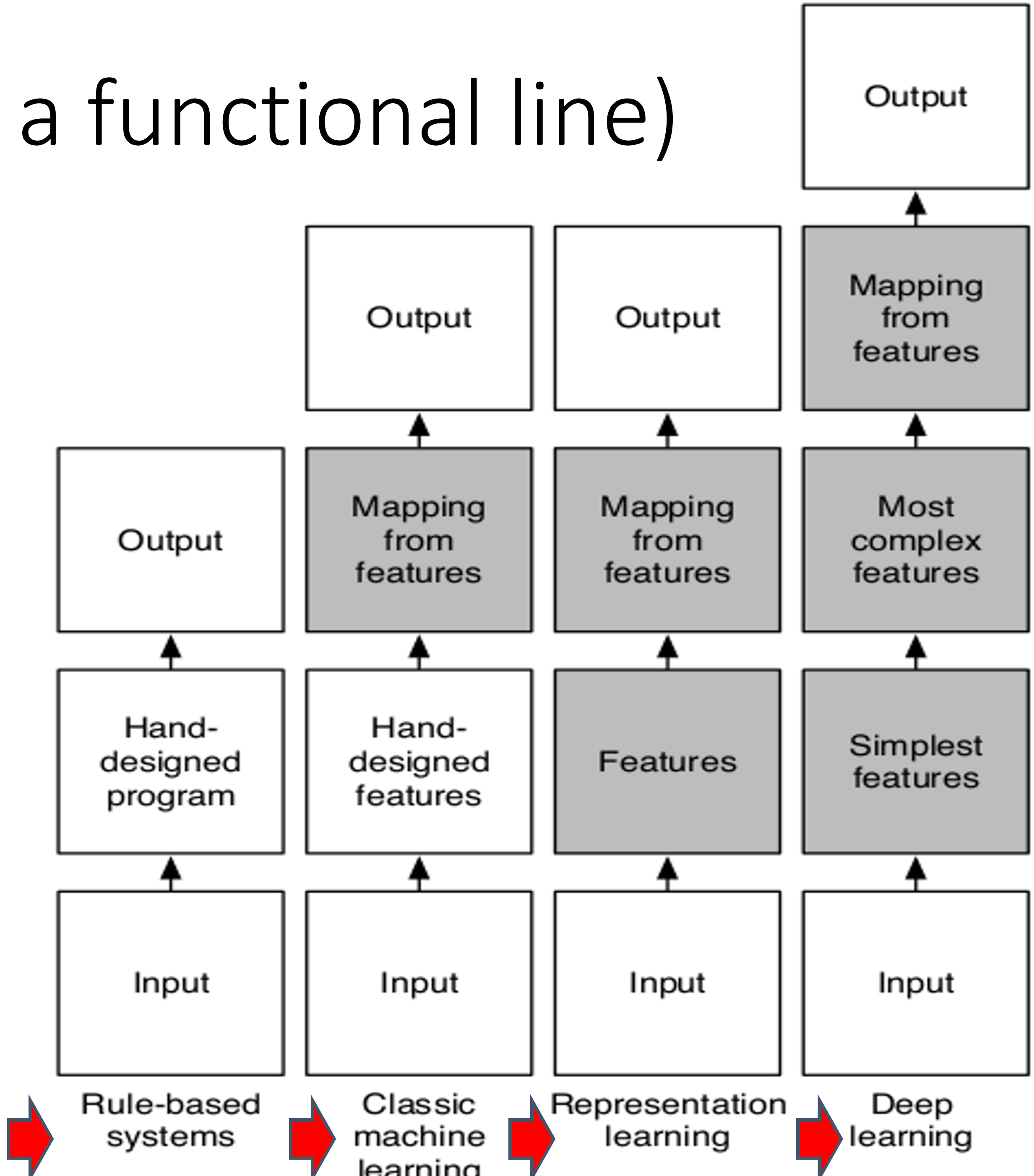
Why is it that we assume humans are intelligent?

Are monkeys intelligent? Dogs? Ants? Pine trees?

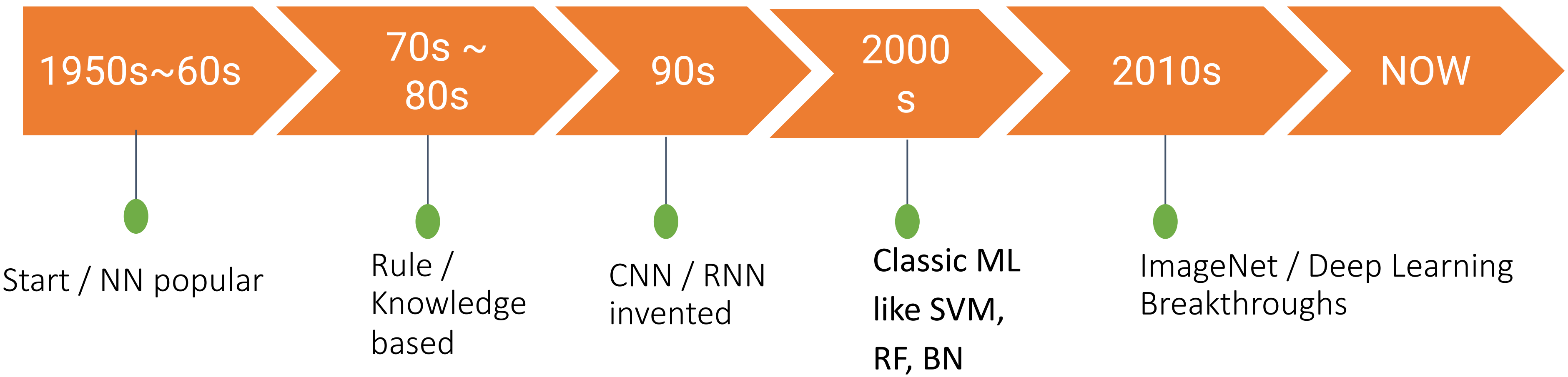
How to build more intelligent computer / machine ?

- Able to perceive the world,
  - e.g., objective recognition, speech recognition, …
- Able to understand the world,
  - e.g., machine translation, text semantic understanding
- Able to Interact with the world,
  - e.g., AlphaGo, AlphaZero, self-driving cars, …
- Able to think / reason / learn,
  - e.g., learn to program programs, learn to search deepNN architecture, …
- Able to imagine / to make analogy,
  - e.g., learn to draw with styles, ……

History (on a functional line)



History (on a time line)



Auther Samuel popularized the term "machine learning" in 1959.[4] The Samuel Checkers-playing Program was among the world's first successful self-learning programs;

Early History

- In 1950 English mathematician Alan Turing wrote a landmark paper titled "Computing Machinery and Intelligence" that asked the question: "Can machines think?"
- Further work came out of a 1956 workshop at Dartmouth sponsored by John McCarthy. In the proposal for that workshop, he coined the phrase a "study of artificial intelligence"
- Expert systems (70s, 80s)
  - A software system based the knowledge of human experts;
  - Rule-based system
  - processes rules to draw conclusions
  - Idea is to give AI systems lots of information to start with

MIT Technology Review

10 Breakthrough Technologies
2013

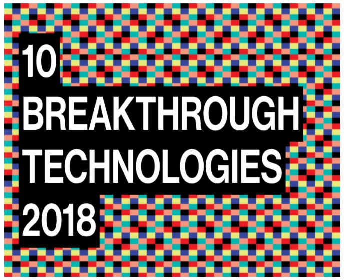
Think of the most frustrating, intractable, or simply annoying problems you can imagine. Now think about what technology is doing to fix them. That's what we did in coming up with our annual list of 10 Breakthrough Technologies. We're looking for technologies that we believe will expand the scope of human possibilities.

Deep Learning

10 Breakthrough Technologies
2017

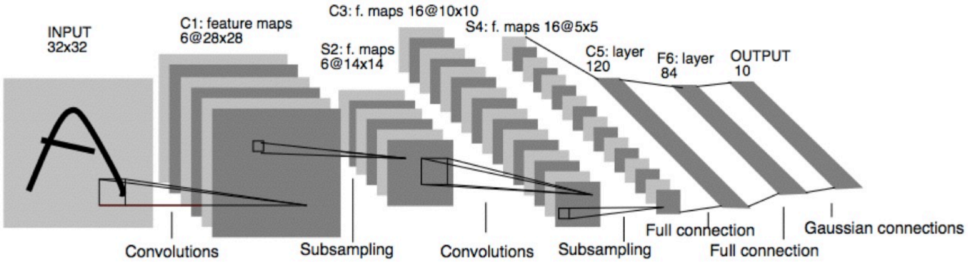
These technologies all have staying power. They will affect the economy and our politics, improve medicine, or influence our culture. Some are unfolding now; others will take a decade or more to develop. But you should know about all of them right now.

Deep Reinforcement Learning



Generative Adversarial Network (GAN)

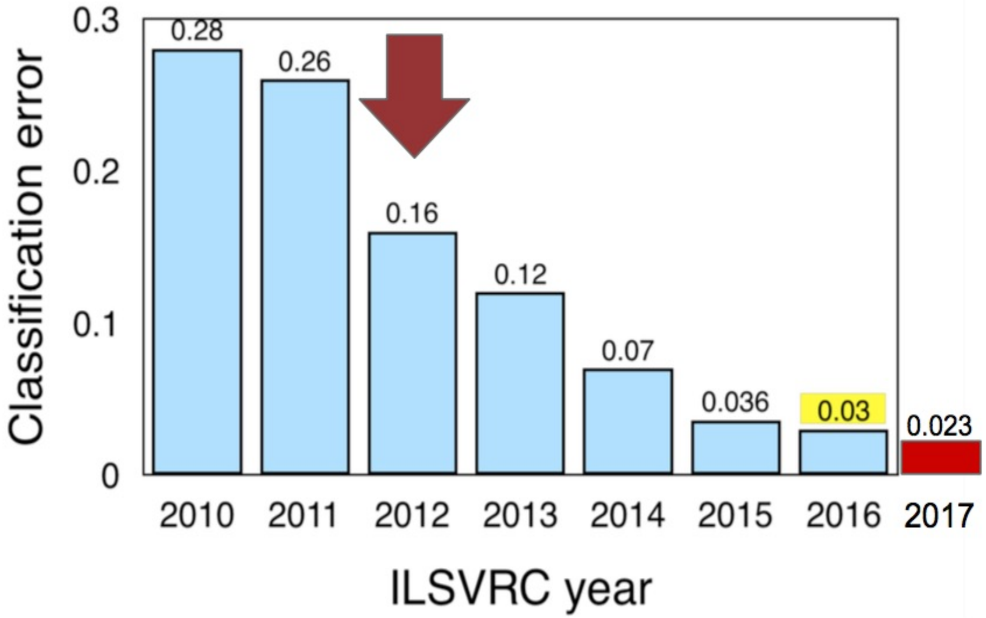
- **1952-1969 Enthusiasm:** Lots of work on neural networks
- **1990s:** Convolutional neural network (CNN) and Recurrent neural network (RNN) were invented



Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86(11): 2278–2324, 1998.

ImageNet Challenge

- 2010-11: hand-crafted computer vision pipelines
- 2012-2016: ConvNets
  - 2012: AlexNet
    - major deep learning success
  - 2013: ZFNet
    - improvements over AlexNet
  - 2014
    - VGGNet: deeper, simpler
    - InceptionNet: deeper, faster
  - 2015
    - ResNet: even deeper
  - 2016
    - ensembled networks
  - 2017
    - Squeeze and Excitation Network

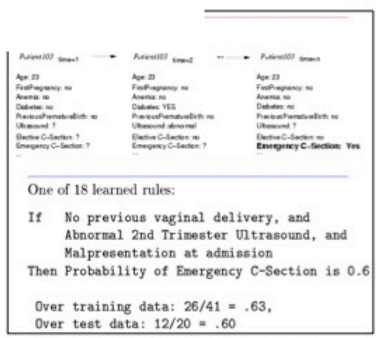


ImageNet Competition:

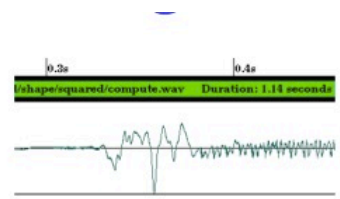
[Training on 1.2 million images [X] vs. 1000 different word labels [Y]]

Adapt from From NIPS 2017 DL Trend Tutorial

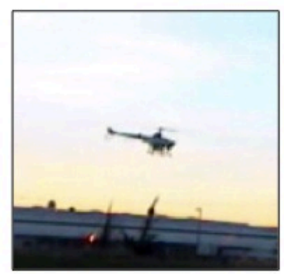
Deep Learning is Changing the World



Mining Databases

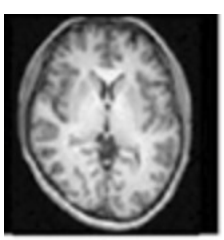


Speech Recognition

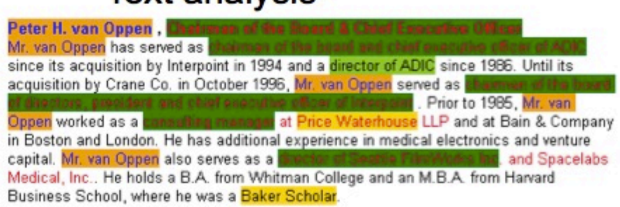


Control learning

Object recognition



Text analysis



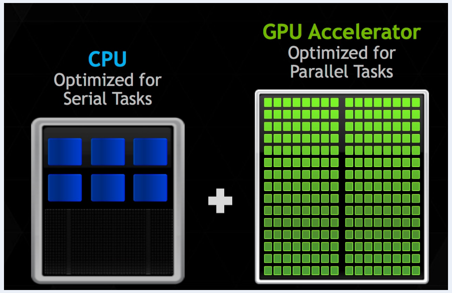
Many more !

Reason of Recent Deep Learning breakthroughs:

Plenty of Good Quality Data

Text: trillions of words of English + other languages
Visual: billions of images and videos
Knowledge graph: billions of labeled relational triplets
……

Advanced Computer Architecture that fits Deep Learning



GPU delivers:

- Same or better prediction accuracy
- Faster results
- Lower power
- Smaller footprint

Powerful Machine Learning Libraries

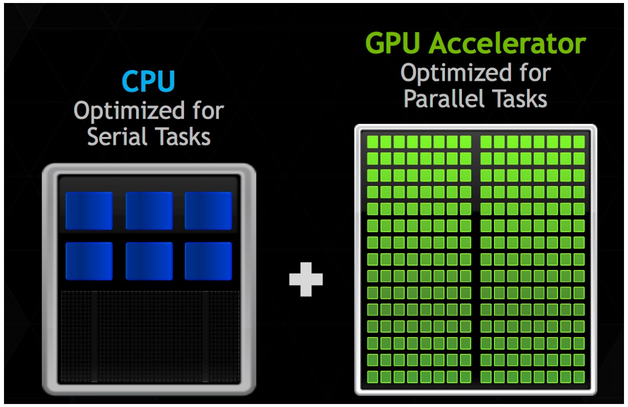
Multiple Well-engineered software libraries

- Easy to learn
- Easy to use
- Easy to extend

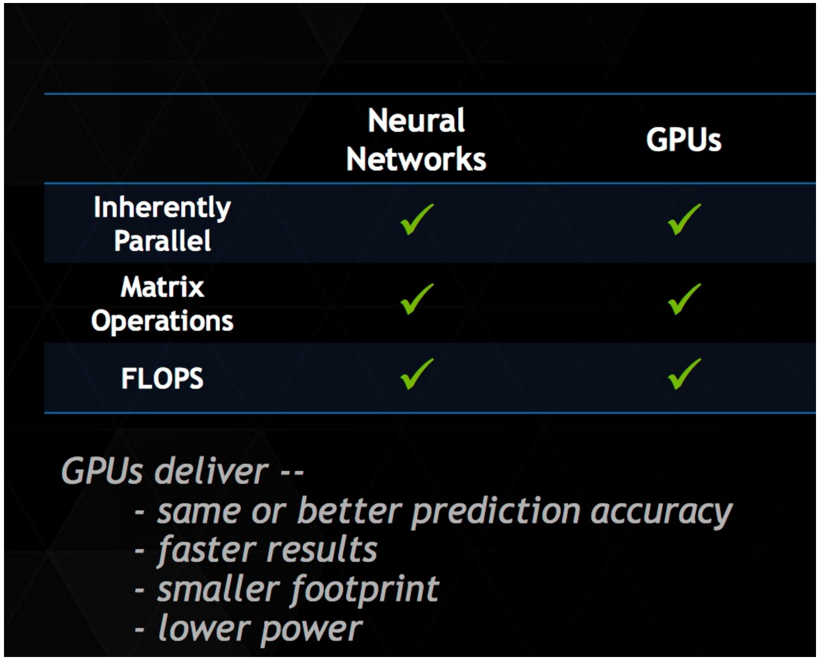
Reason: Plenty of (Labeled) Data

- Text: trillions of words of English + other languages
- Visual: billions of images and videos
- Audio: thousands of hours of speech per day
- User activity: queries, user page clicks, map requests, etc,
- Knowledge graph: billions of labeled relational triplets
- Genomics data:
- Medical Imaging data:

Reason: Advanced Computer Architecture that fits DNNs



http://www.nvidia.com/content/events/geoInt2015/LBrown_DL.pdf

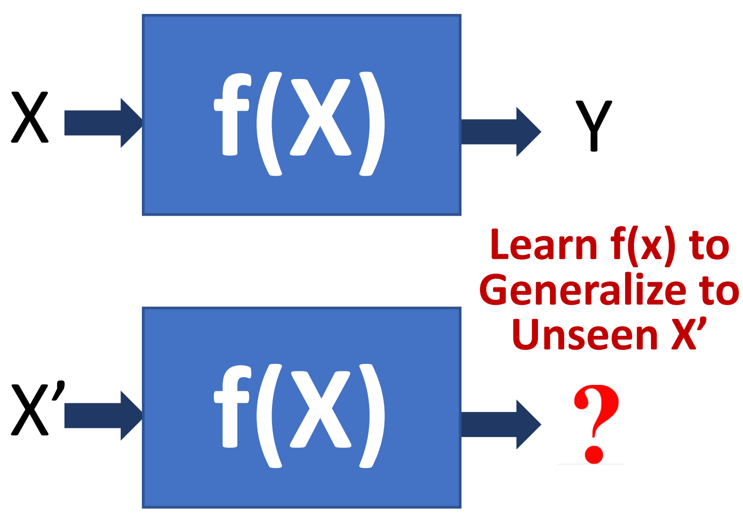


	Neural Networks	GPUs
Inherently Parallel	✓	✓
Matrix Operations	✓	✓
FLOPS	✓	✓

GPUs deliver --
- *same or better prediction accuracy*
- *faster results*
- *smaller footprint*
- *lower power*

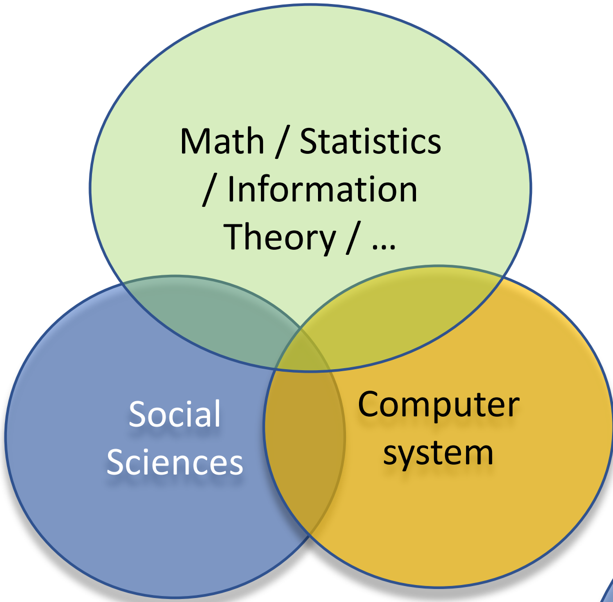
Reason: Data-Driven Machine Learning Algorithms and Platforms

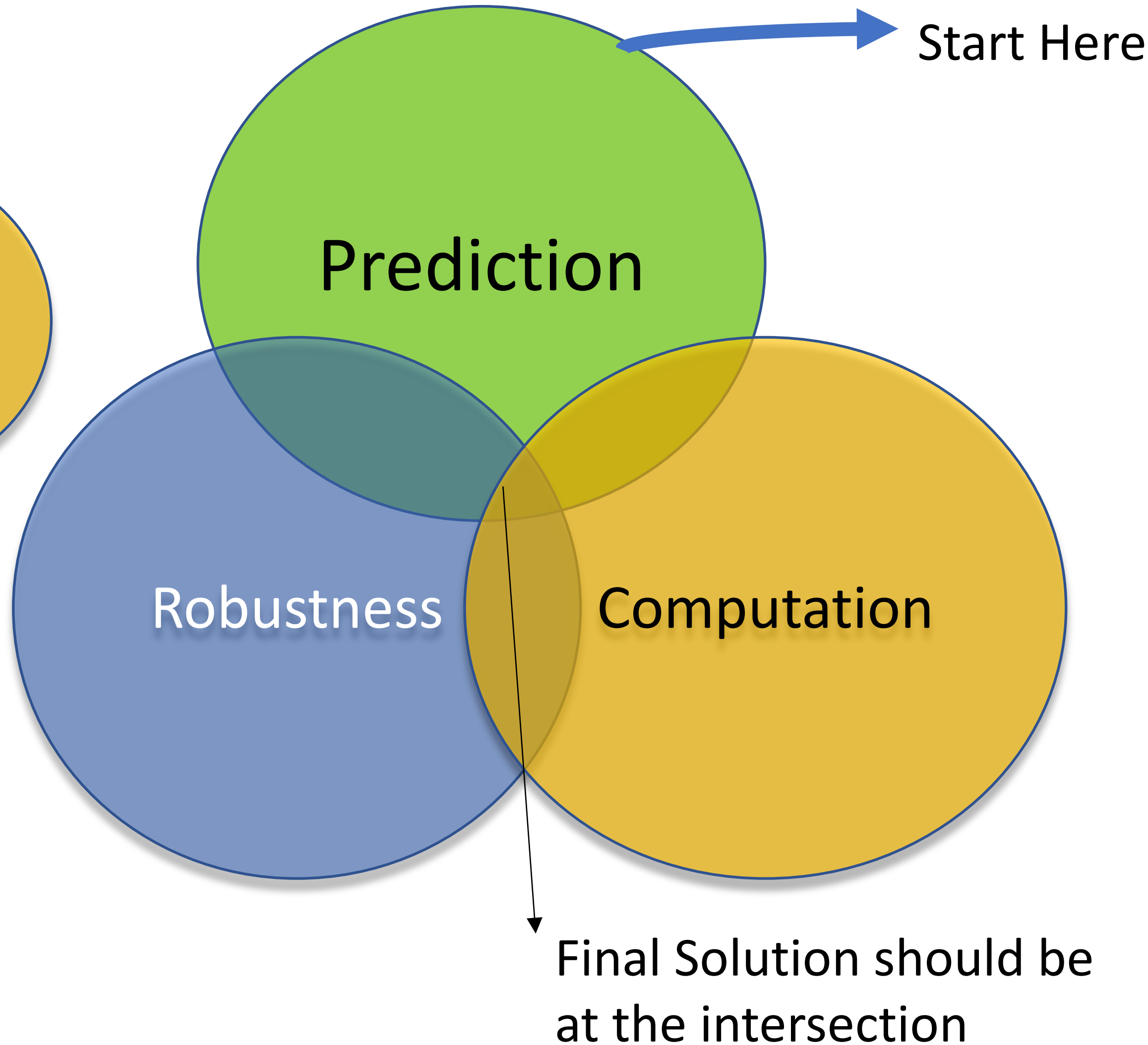
- Inductive reasoning
  - Generalizations from observed data to unseen data



- Computer systems that can learn and adapt from their experience (data)
- Well-engineered software architectures to build upon
- Provide prediction accuracy
- Create software that improves over time

Three Aimed Features of ML Models





Future?

ML + Digital Data Platforms: Unprecedented Era

Hyper time compression
new disruptive innovations

Extreme convergence
of multiple domains

Exponential accelerating automation
– smart sensors and the billion IoT devices

**Universal connectivity linked
by a digital mesh**

General Lessons for Excellence

- Good breath in fundamentals is key

- Strength in particular targeted topics help standing out

Highly Recommend
Two Extra-curriculum books:

1. Book: By Dr. Domingos: Master Algorithm

So How Do Computers Discover New Knowledge?

1. **Symbolists**--Fill in gaps in existing knowledge
2. **Connectionists**--Emulate the brain
3. **Evolutionists**--Simulate evolution
4. **Bayesians**--Systematically reduce uncertainty
5. **Analogizers**--Notice similarities between old and new

SRC: Pedro Domingos ACM Webinar Nov 2015
http://learning.acm.org/multimedia.cfm

Highly Recommend
Two Extra-curriculum books

- 2. Book: [Homo Deus- A Brief History of Tomorrow]
  - https://www.goodreads.com/book/show/31138556-homo-deus
  - “Homo Deus explores the projects, dreams and nightmares that will shape the twenty-first century—from overcoming death to creating artificial life. It asks the fundamental questions: Where do we go from here? And how will we protect this fragile world from our own destructive powers? This is the next stage of evolution. This is Homo Deus.””


Thank You

Roadmap

- Course Logistics
- History and Now
- ML Basics

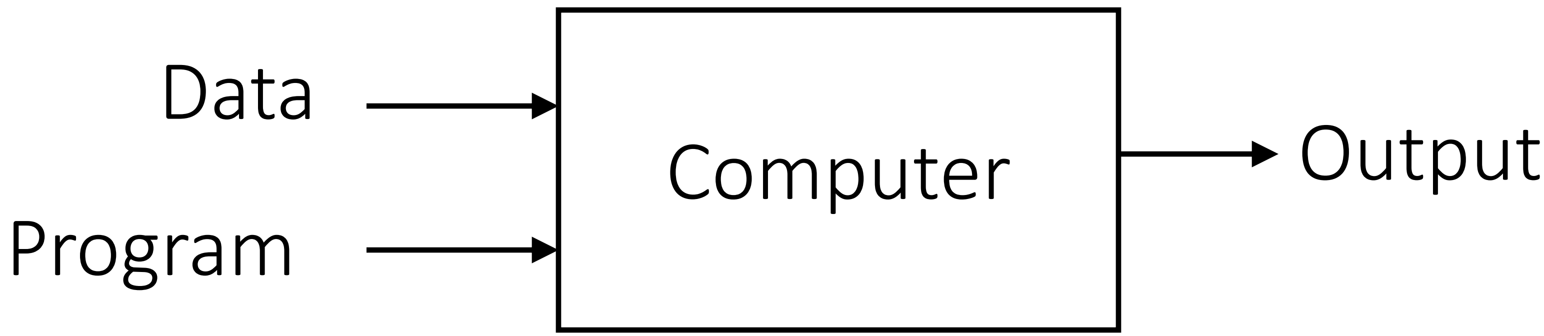
BASICS OF MACHINE LEARNING

- "The goal of machine learning is to build computer systems that can learn and adapt from their experience." – Tom Dietterich

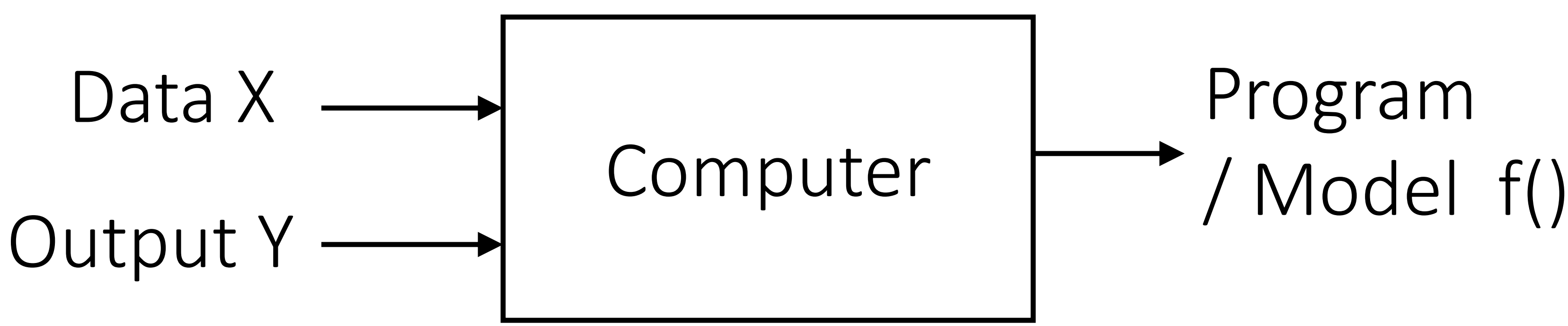
- "Experience" in the form of available data examples (also called as instances, samples)

- Available examples are described with properties (data points in feature space X)

Traditional Programming



Machine Learning (training phase)



e.g. SUPERVISED LEARNING

- Find function to map input space X to output space Y

$$f : X \longrightarrow Y$$

- So that the difference between y and f(x) of each example x is small.

e.g.

x

I believe that this book is not at all helpful since it does not explain thoroughly the material . it just provides the reader with tables and calculations that sometimes are not easily understood …

Input X : e.g. a piece of English text

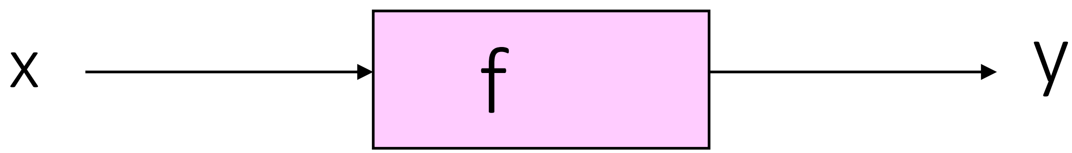
y

-1

Output Y: {1 / Yes , -1 / No }
e.g. Is this a positive product review ?

SUPERVISED Linear Binary Classifier

- Now let us check out a VERY SIMPLE case of

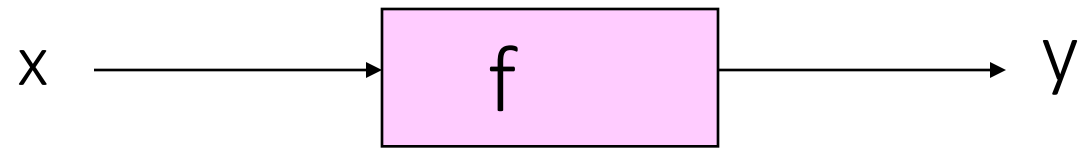


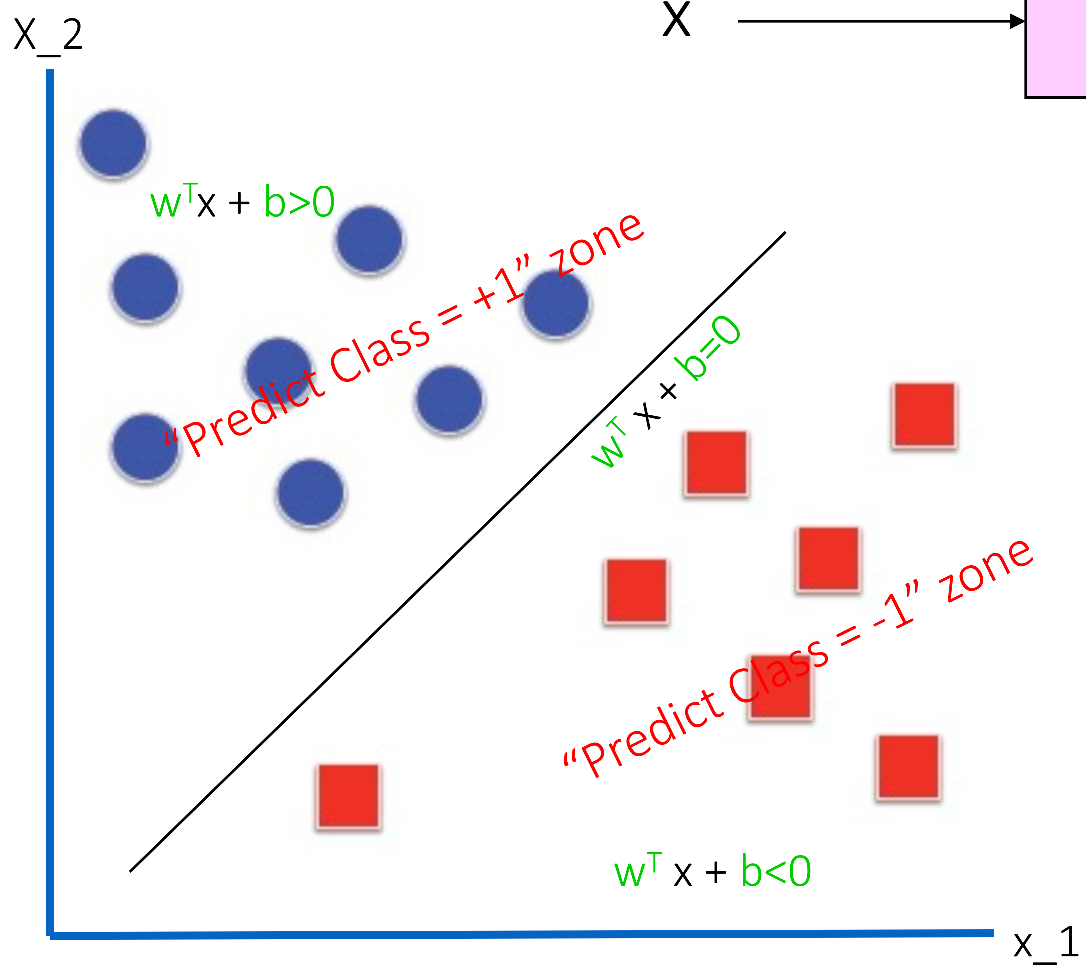
e.g.: Binary y / Linear f / X as R^2

$$f(x,w,b) = \mathrm{sign}(w^T x + b)$$

X =(x_1, x_2)

SUPERVISED Linear Binary Classifier





$f(x,w,b) = \text{sign}(w^T x + b)$

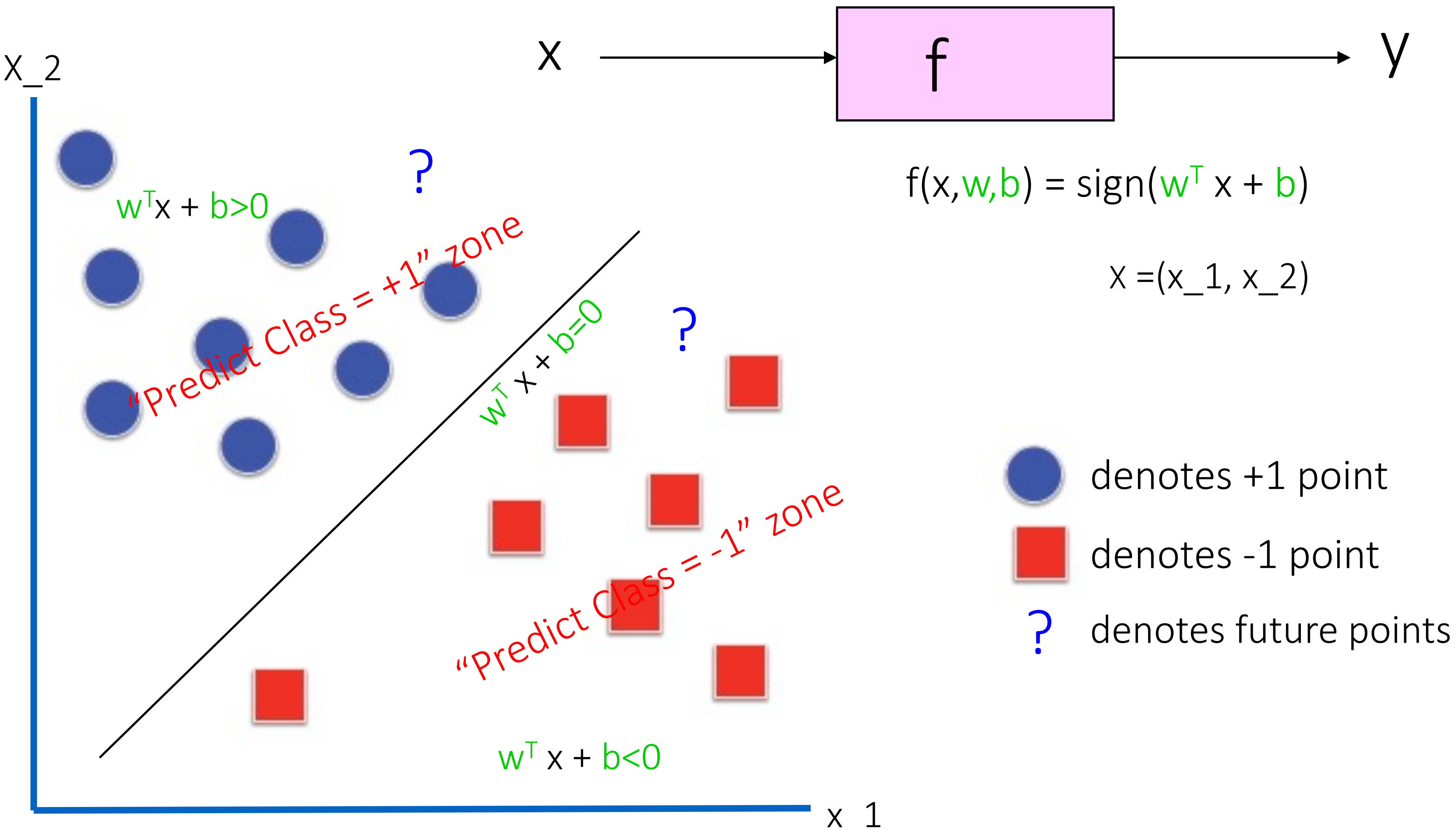
$X = (x\_1, x\_2)$


denotes +1 point


denotes -1 point

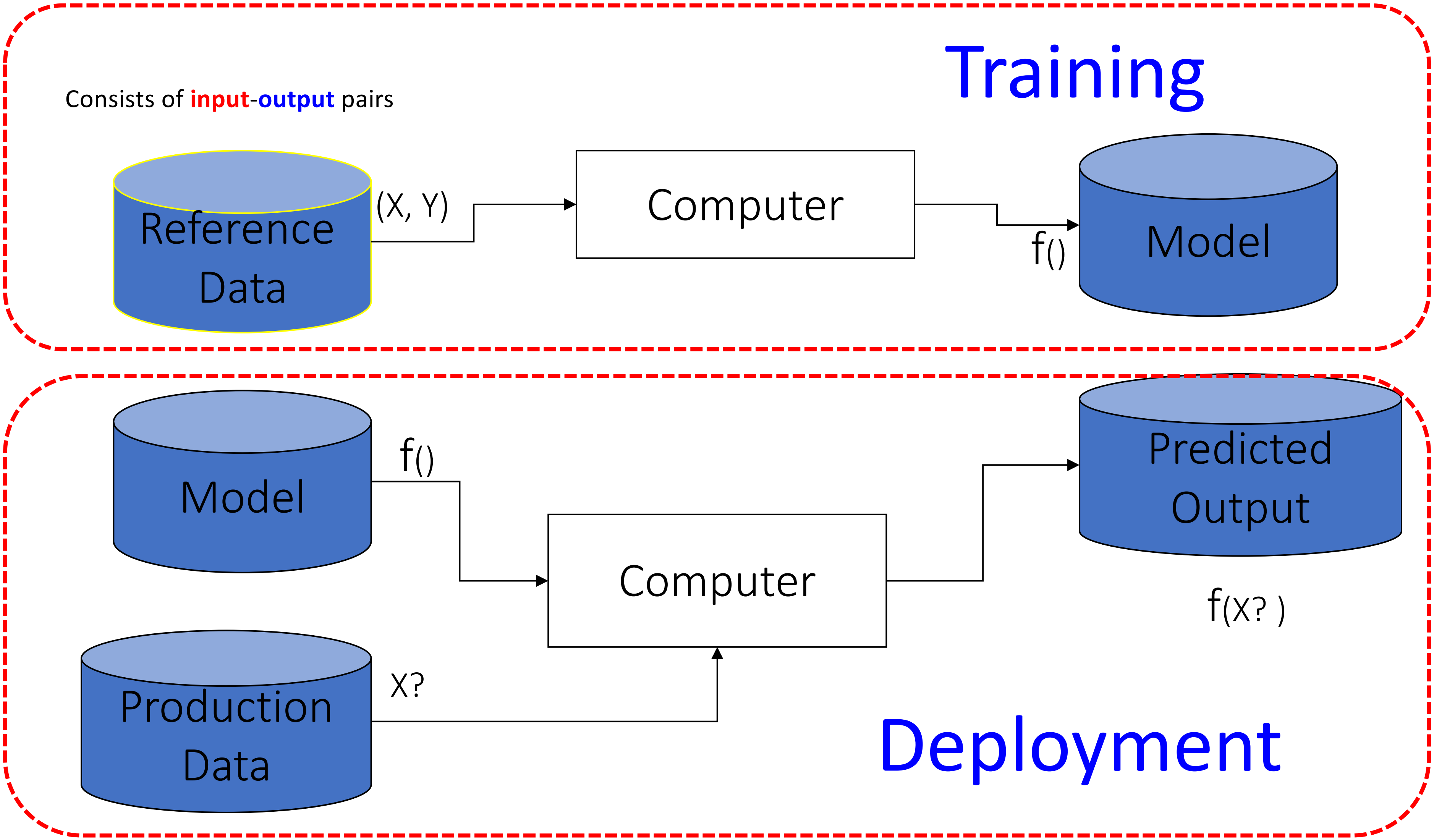
Courtesy slide from Prof. Andrew Moore's tutorial

SUPERVISED Linear Binary Classifier

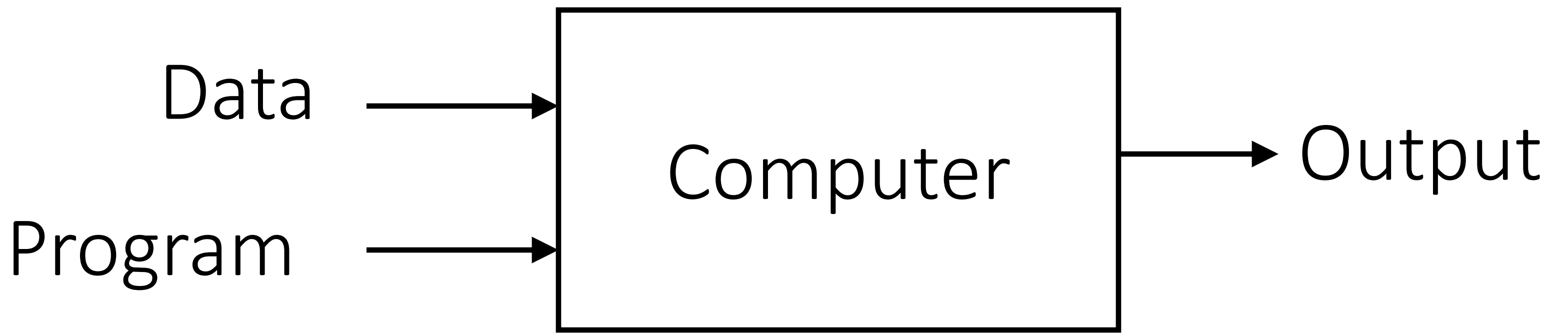


Courtesy slide from Prof. Andrew Moore's tutorial

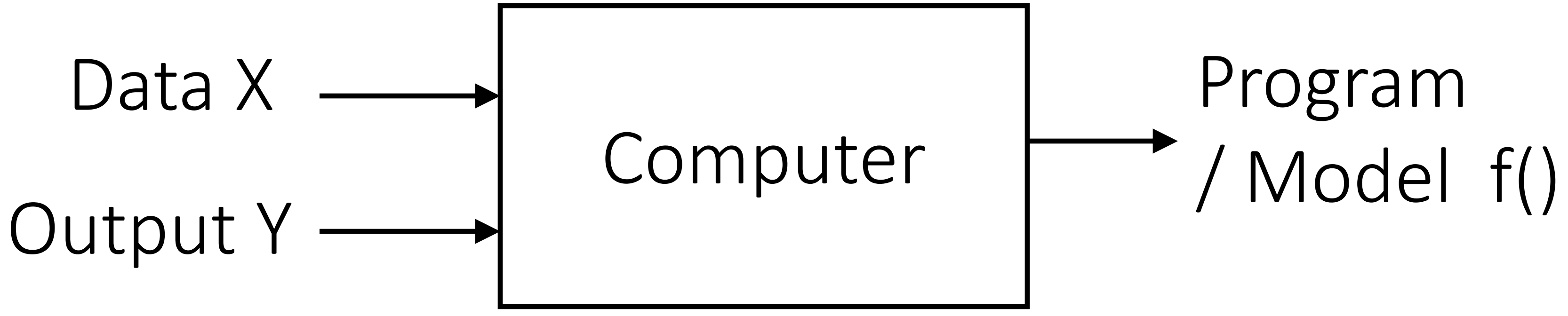
Two Modes of Machine Learning



Traditional Programming



Machine Learning (training phase)



Basic Concepts

- Training (i.e. learning parameters w,b)
  - Training set includes
    - available examples $x_1,\ldots,x_L$
    - available corresponding labels $y_1,\ldots,y_L$

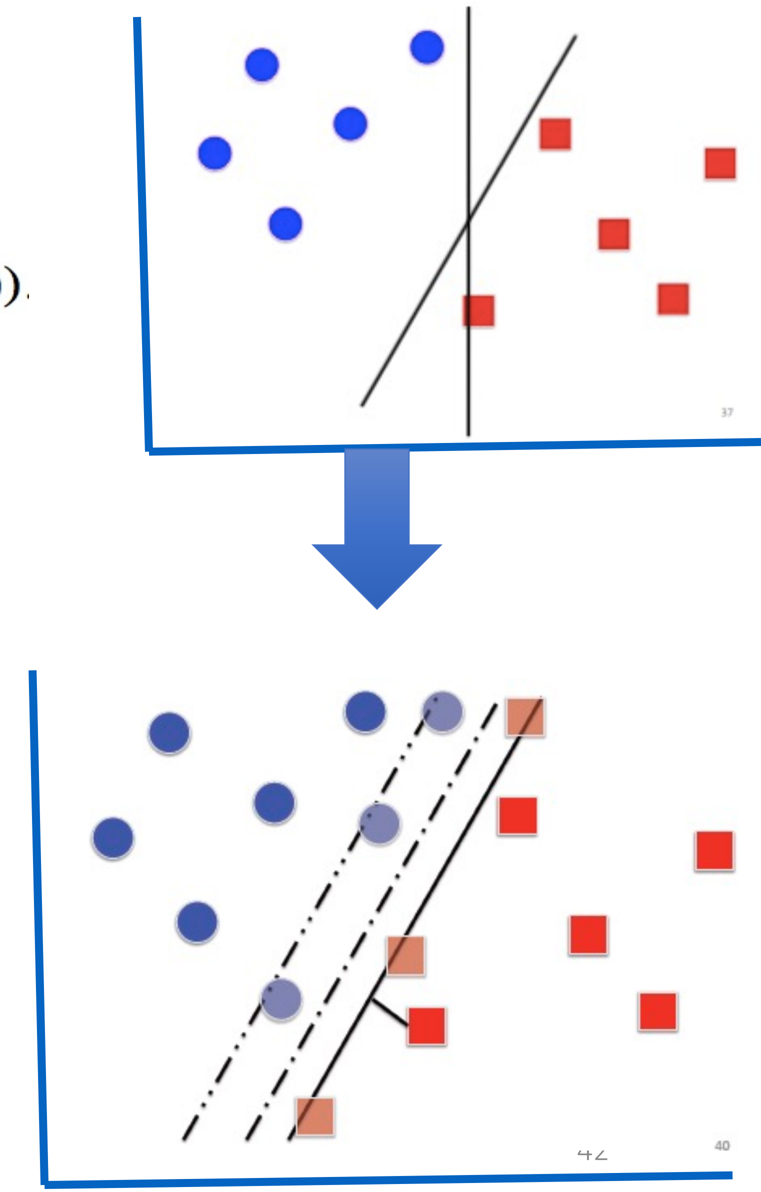
- Find (w,b) by minimizing loss / Cost function L()
  - (i.e. difference between y and f(x) on available examples in training set)

$$(W, b) = \underset{W,\, b}{\operatorname{argmin}} \sum_{i=1}^{L} \ell(f(x_i), y_i)$$

Basic Concepts

- Loss function
  - e.g. hinge loss for binary classification task

$$\sum_{i=1}^{L} \ell(f(x_i), y_i) = \sum_{i=1}^{L} \max(0, 1 - y_i f(x_i)).$$



- Regularization
  - E.g. additional information added

on loss function to control f

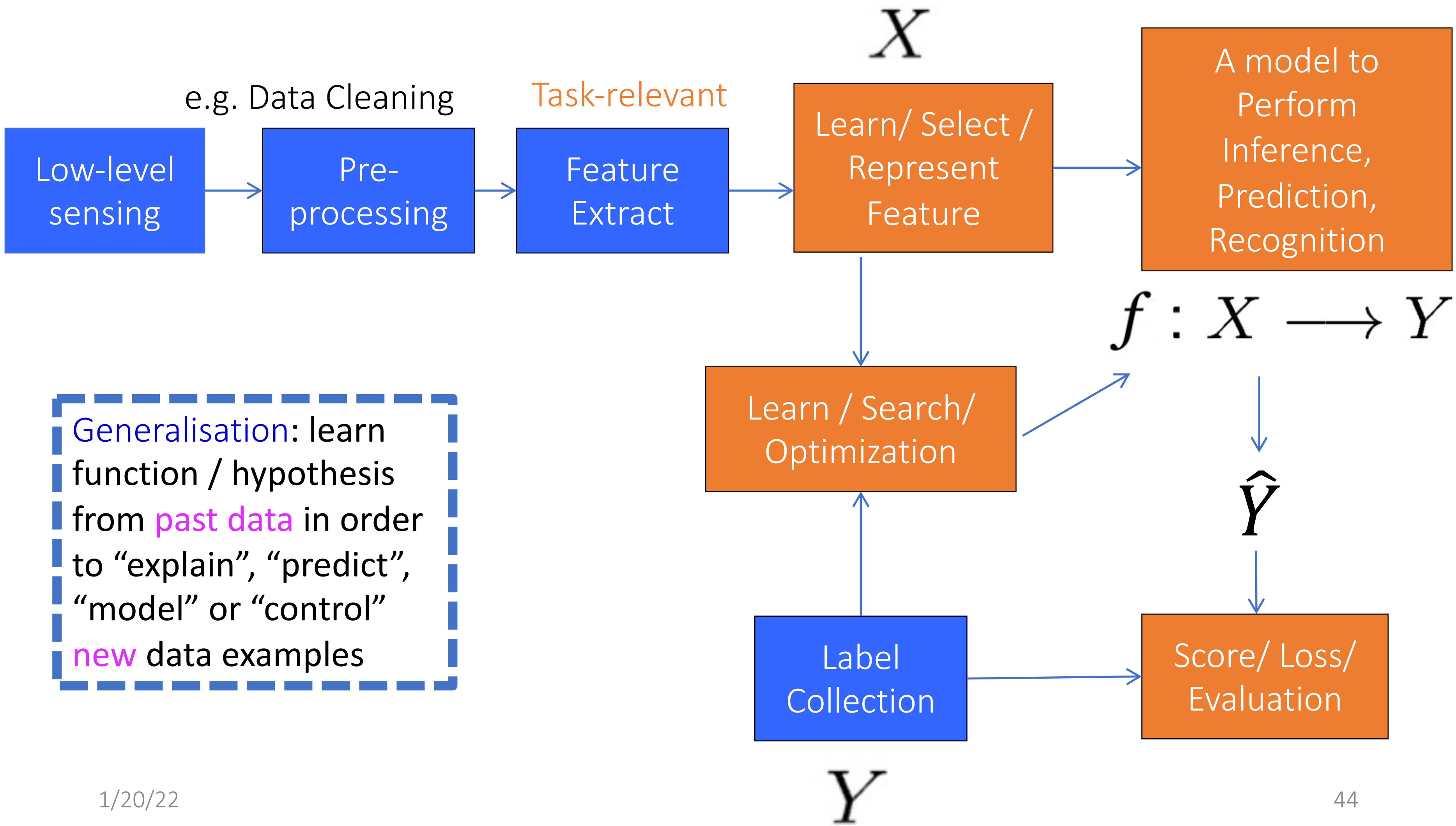
$$C \sum_{i=1}^{L} \ell(f(x_i), y_i) + \frac{1}{2}\|w\|^2.$$

Basic Concepts

- Testing (i.e. evaluating performance on "future" points)
  - Difference between true $y_?$ and the predicted $f(x_?)$ on a set of testing examples (i.e. testing set)
  - Key: example $x_?$ not in the training set

- Generalisation: learn function / hypothesis from past data in order to "explain", "predict", "model" or "control" new data examples

A Typical Machine Learning Application's Pipeline



Generalisation: learn function / hypothesis from past data in order to "explain", "predict", "model" or "control" new data examples


Thank You

References

- Prof. Andrew Moore's tutorials
- Prof. Raymond J. Mooney's slides
- Prof. Alexander Gray's slides
- Prof. Eric Xing's slides
- http://scikit-learn.org/
- Hastie, Trevor, et al. The elements of statistical learning. Vol. 2. No. 1. New York: Springer, 2009.
- Prof. M.A. Papalaskar's slides