UVA CS 4774: Machine Learning

Lecture 1: Introduction

Dr. Yanjun Qi

University of Virginia
Department of Computer Science

Roadmap

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

ATT:

- video-recording of my lectures all available on youtube
- Each session includes multiple small modules --- each about 20mins (each into a different recorded video)

Welcome

- Course Website: (link)
- Discussion via Slack Channel (link from TA announcement!)
- Announcement via email (see Course overview)
- Communication with instructor team via:
  - Email and Slack (see Course overview)

We focus on learning fundamental principles, algorithm design and 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, deep learning, generative AI and truthworthy GenAI

- 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**
  - (all on Course Website + Video + Readings + HWs)
- Your UVA CourseSite for Assignments Submissions
- 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

Assessment Breakdown

- See policy in CourseWeb

Quiz

- See policy in CourseWeb

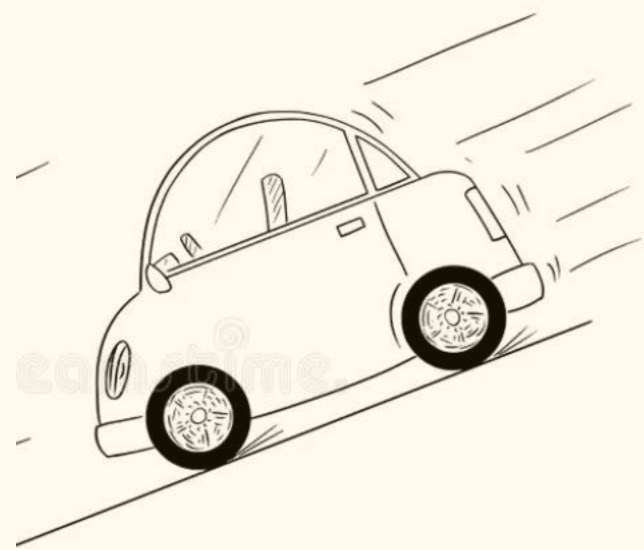
Course Format

- Emphasis on active learning
  - Understand core ML concepts and algorithms
  - Apply models using Python (scikit-learn, Keras)
  - Analyze model performance and ethics
  - Become a ML tool builder!

Let us start building baby Steps!

Learning ML basics

Current GenAI space: growth speed



Thank You

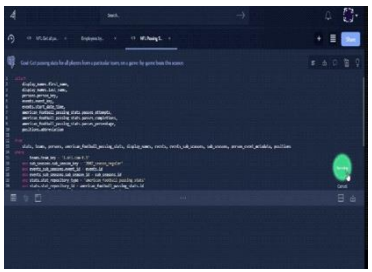
Roadmap

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

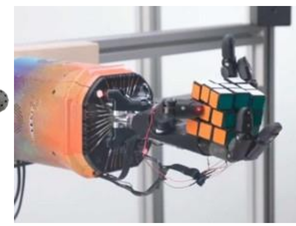
Artificial intelligence (AI)

Computer systems that attempt to model and apply the intelligence of the human mind.

Artificial Intelligence: A Systems Approach: (Computer Science Series)



Code generation
Claude code, Cursor, GitHub Copilot, Devin, Replit,…

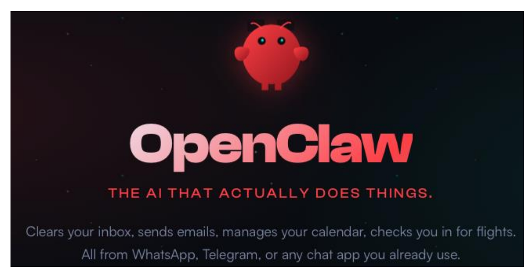


Robotics
Figure AI, Tesla Optimus,…

- Education
- Law
- Finance
- Healthcare
- Cybersecurity

…

Personal assistant
Google Genimi, OpenAI, OpenClaw, ….

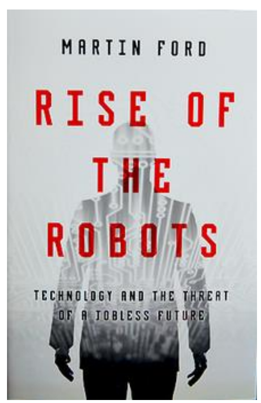


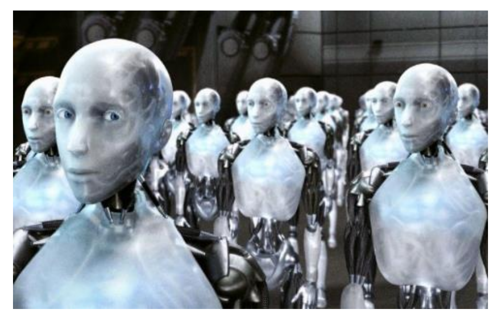
Impact: Good and Bad?

Economic, cultural, social,
health… endless disruption

Labor - McKinsey >50%
of jobs automated

Martin Ford,
Rise of the Robots





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”?

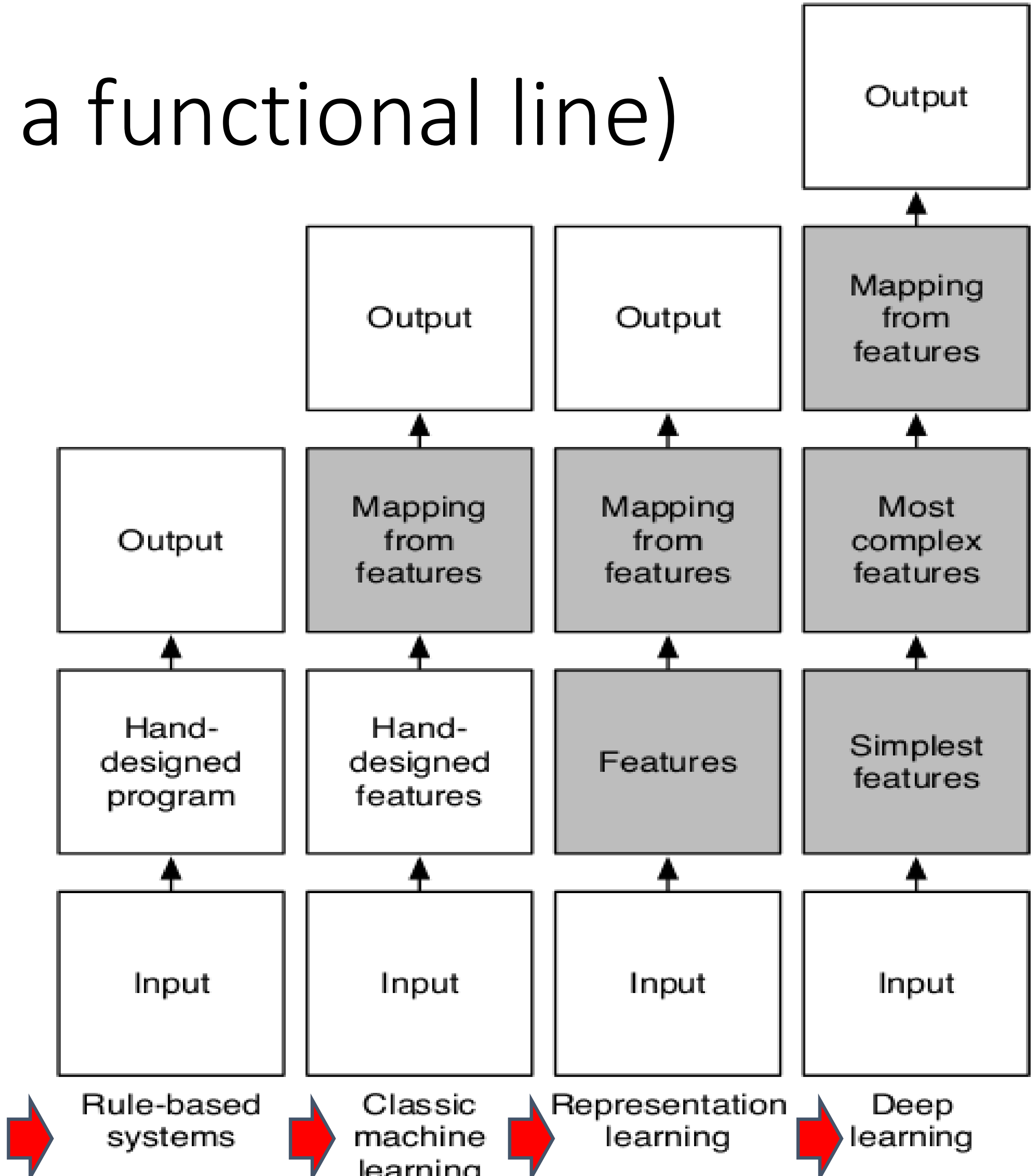
Why is it that we assume humans are intelligent?

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)

ChatGPT / GenAI/ Agentic AI

- 1950s~60s: Start / NN popular
- 70s ~ 80s: Rule / Knowledge based
- 90s: CNN / RNN invented
- 2000s: Classic ML like SVM, RF, BN
- 2010s: ImageNet / Deep Learning Breakthroughs
- 2023-Now

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.

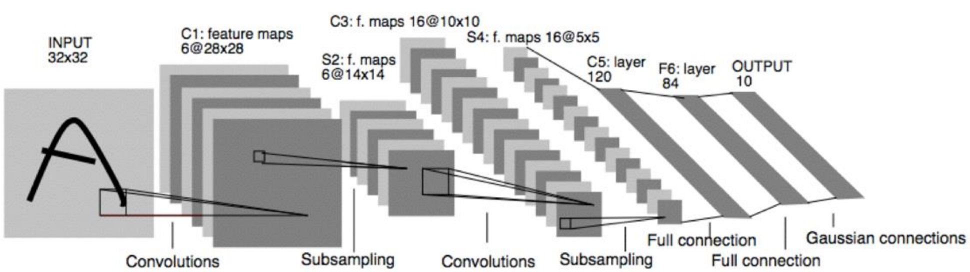
Deep Learning


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)

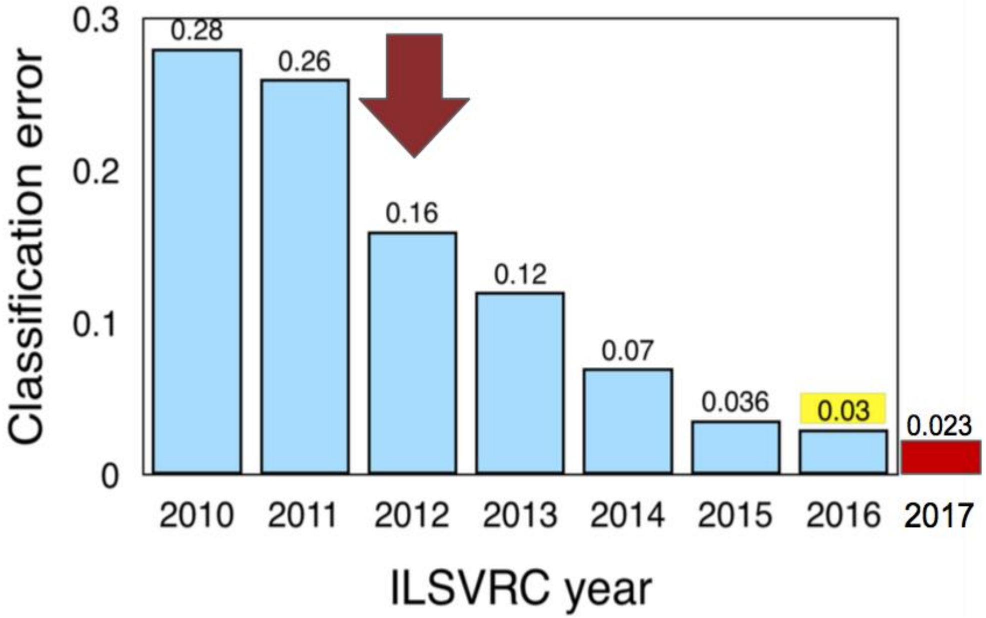
- **1952-1969 Enthusiasm:** Lots of work on neural networks
- **1990s:** Convolutional neural network (CNN) and Recurrent neural network (RNN) were invented



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



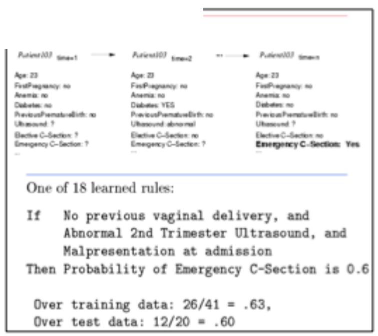
ImageNet Competition:

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

Adapt from From NIPS 2017 DL Trend Tutorial

Deep Learning has changed the World

How may I help you, human?



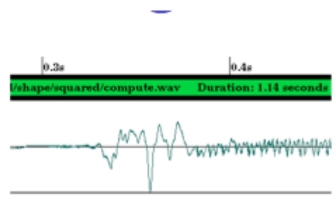
Patient103 time=1	Patient103 time=2	Patient103 time=n
Age: 23	Age: 23	Age: 23
FirstPregnancy: no	FirstPregnancy: no	FirstPregnancy: no
Anemia: no	Anemia: no	Anemia: no
Diabetes: no	Diabetes: YES	Diabetes: no
PreviousPrematureBirth: no	PreviousPrematureBirth: no	PreviousPrematureBirth: no
Ultrasound: ?	Ultrasound: abnormal	Ultrasound: ?
Elective C–Section: ?	Elective C–Section: no	Elective C–Section: no
Emergency C–Section: ?	Emergency C–Section: ?	**Emergency C–Section: Yes**
...	...	...

One of 18 learned rules:

```
If   No previous vaginal delivery, and
     Abnormal 2nd Trimester Ultrasound, and
     Malpresentation at admission
Then Probability of Emergency C-Section is 0.6

 Over training data: 26/41 = .63,
 Over test data: 12/20 = .60
```

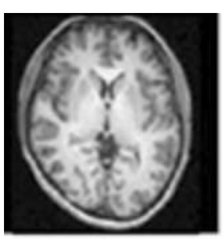
Mining Databases



Speech Recognition

Control learning

Object recognition



Text analysis

Peter H. van Oppen , Chairman of the Board & Chief Executive Officer
Mr. van Oppen has served as chairman of the board and chief executive officer of ADIC since its acquisition by Interpoint in 1994 and a director of ADIC since 1986. Until its acquisition by Crane Co. in October 1996, Mr. van Oppen served as chairman of the board of directors, president and chief executive officer of Interpoint . Prior to 1985, Mr. van Oppen worked as a consulting manager at Price Waterhouse LLP and at Bain & Company in Boston and London. He has additional experience in medical electronics and venture capital. Mr. van Oppen also serves as a director of Seattle FilmWorks Inc. and Spacelabs Medical, Inc.. He holds a B.A. from Whitman College and an M.B.A. from Harvard Business School, where he was a Baker Scholar.

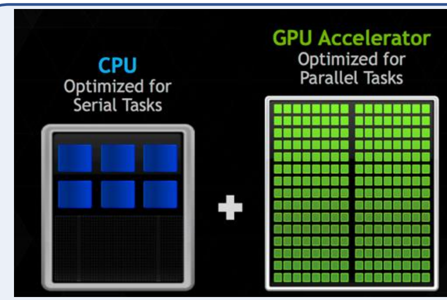
Many more !

Reason of 2010s 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

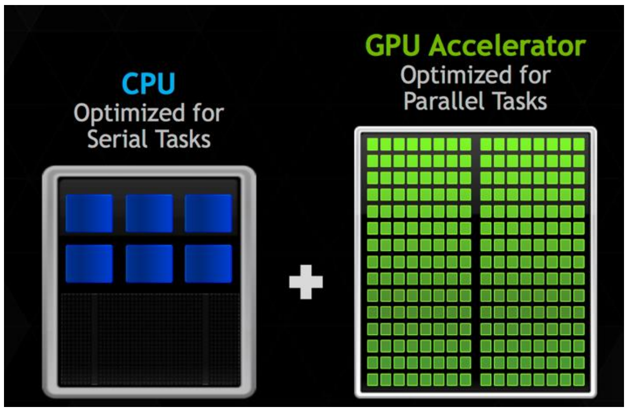
Powerful Machine Learning Libraries/Architecture

Multiple Well-engineered software libraries
- Easy to learn
- Easy to use
- Easy to extend

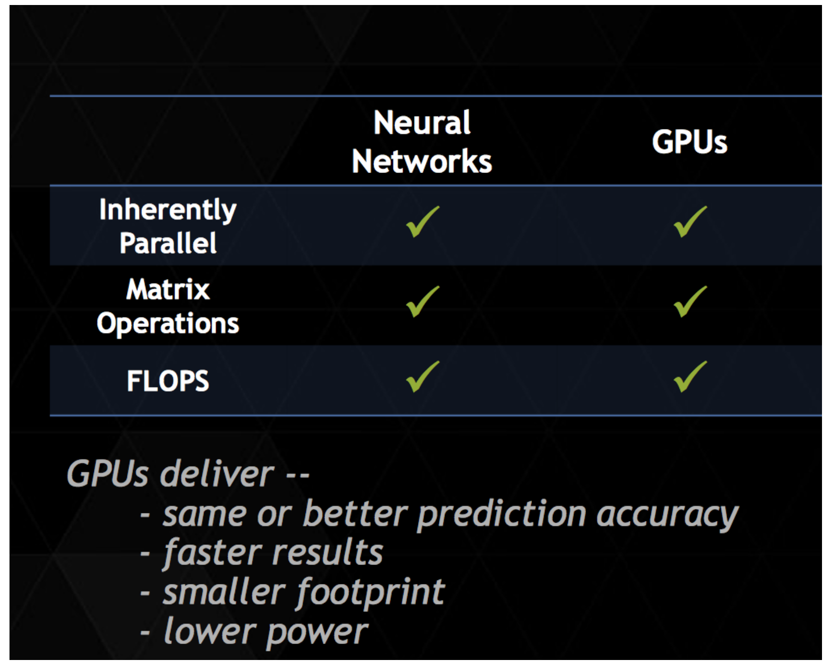
Reason: Plenty of (Labeled) Data

- 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: Powerful Libraries

Multiple Well-engineered software libraries

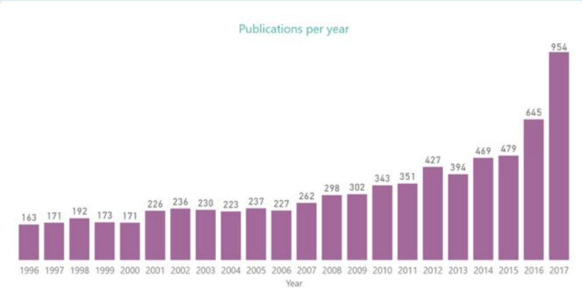
- Easy to learn
- Easy to use
- Easy to extend

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

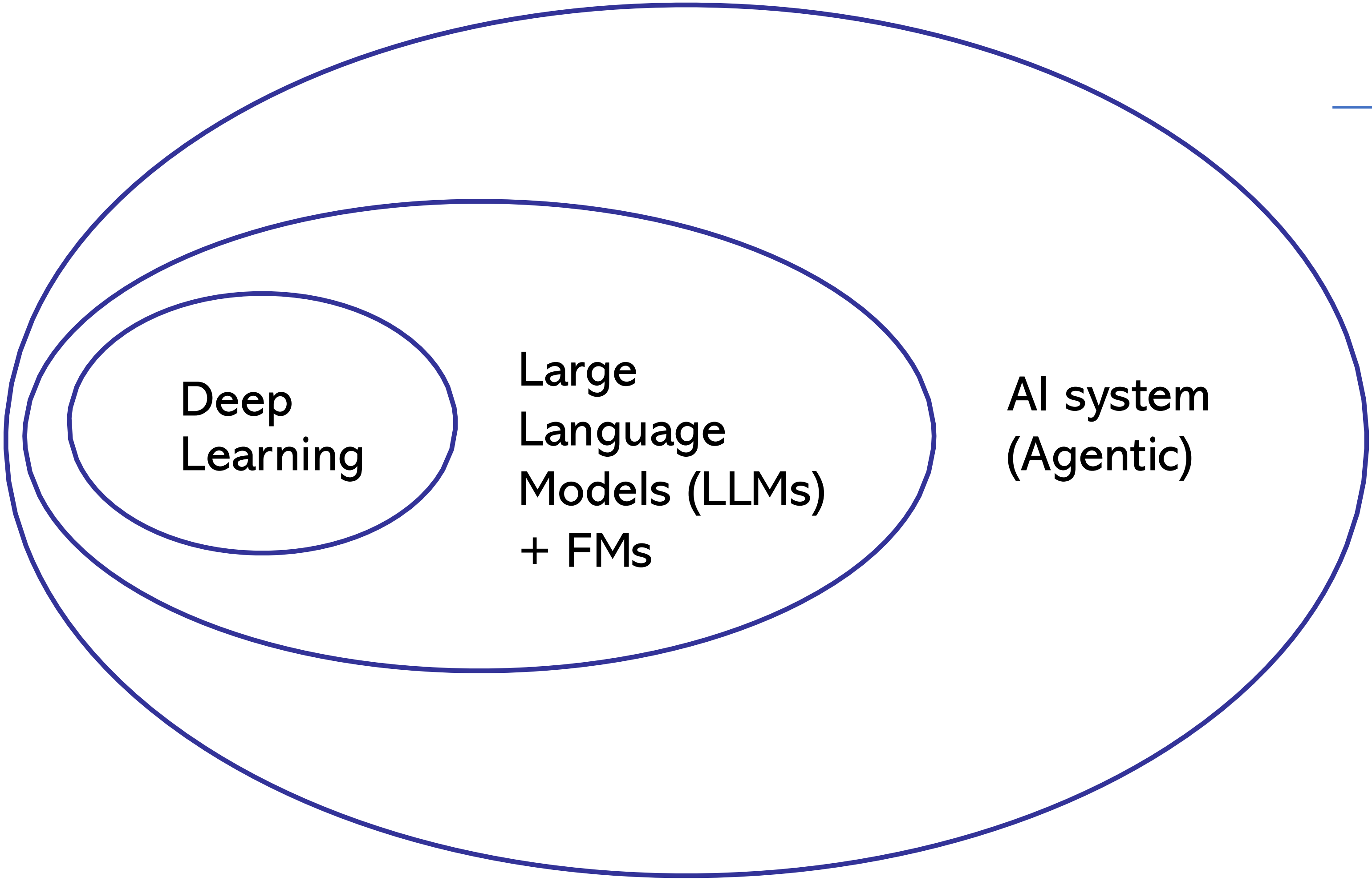


Universal connectivity linked
by a digital mesh

History (on a timeline)

ChatGPT / GenAI/ Agentic AI

- **1950s~60s**: Start / NN popular
- **70s ~ 80s**: Rule / Knowledge based
- **90s**: CNN / RNN invented
- **2000s**: Classic ML like SVM, RF, BN
- **2010s**: ImageNet / Deep Learning Breakthroughs
- **2023-Now**


Deep
Learning
Large
Language
Models (LLMs)
+ FMs
AI system
(Agentic)

Recent Advances (2023–Now) on Generative AI

What is Generative AI?

- AI systems that *create* content (text, images, audio, video, code)
- Powered by foundation models (transformers, diffusion, LLMs)
- Distinct from discriminative AI: not just classification, but generation

Key Applications

- Text & Chat: LLMs (GPT, Claude, Llama, Mistral)
- Images: Diffusion models (Stable Diffusion, MidJourney, DALL·E)
- Video: Gen-2, Sora, Pika Labs
- Code: GitHub Copilot, Cursor, Claude Code
- Multi-modal: ChatGPT-4o, Gemini, OpenAI o1 models

What makes this possible:

- Scaling law!

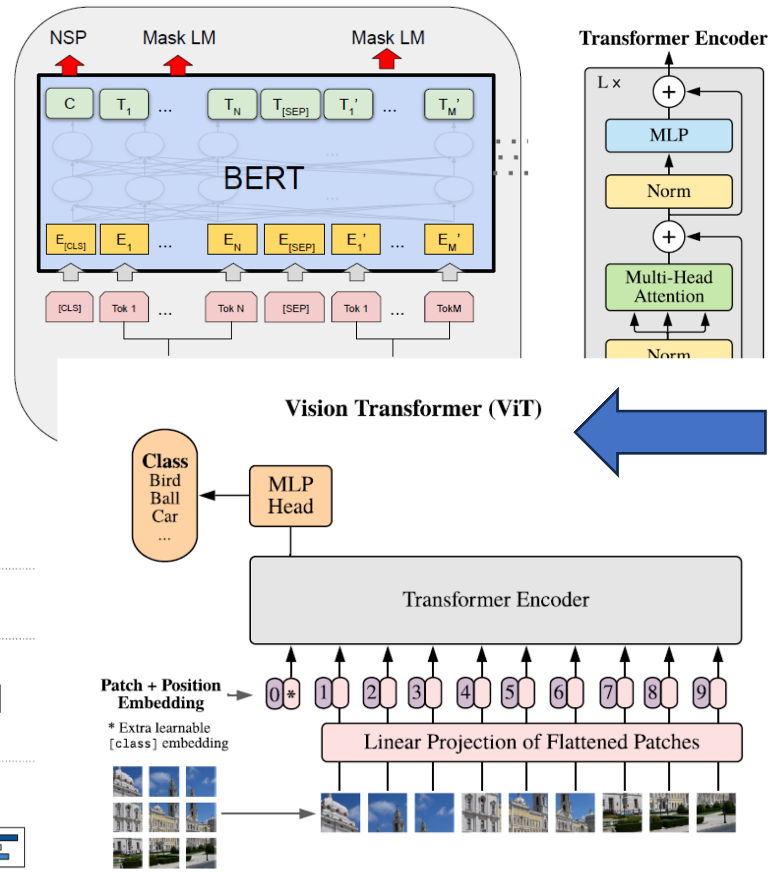
Plenty of Good Quality Data

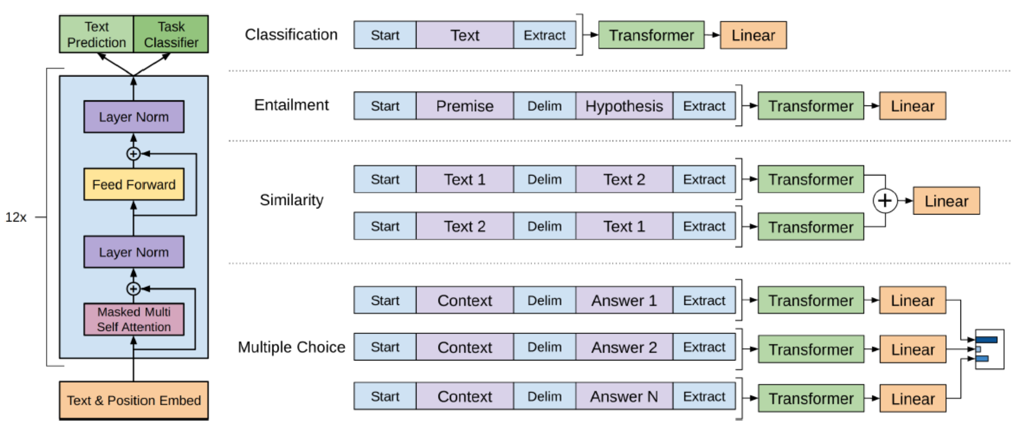
Advanced Computer Architecture that fits Deep Learning

Powerful Learning Libraries/Architecture

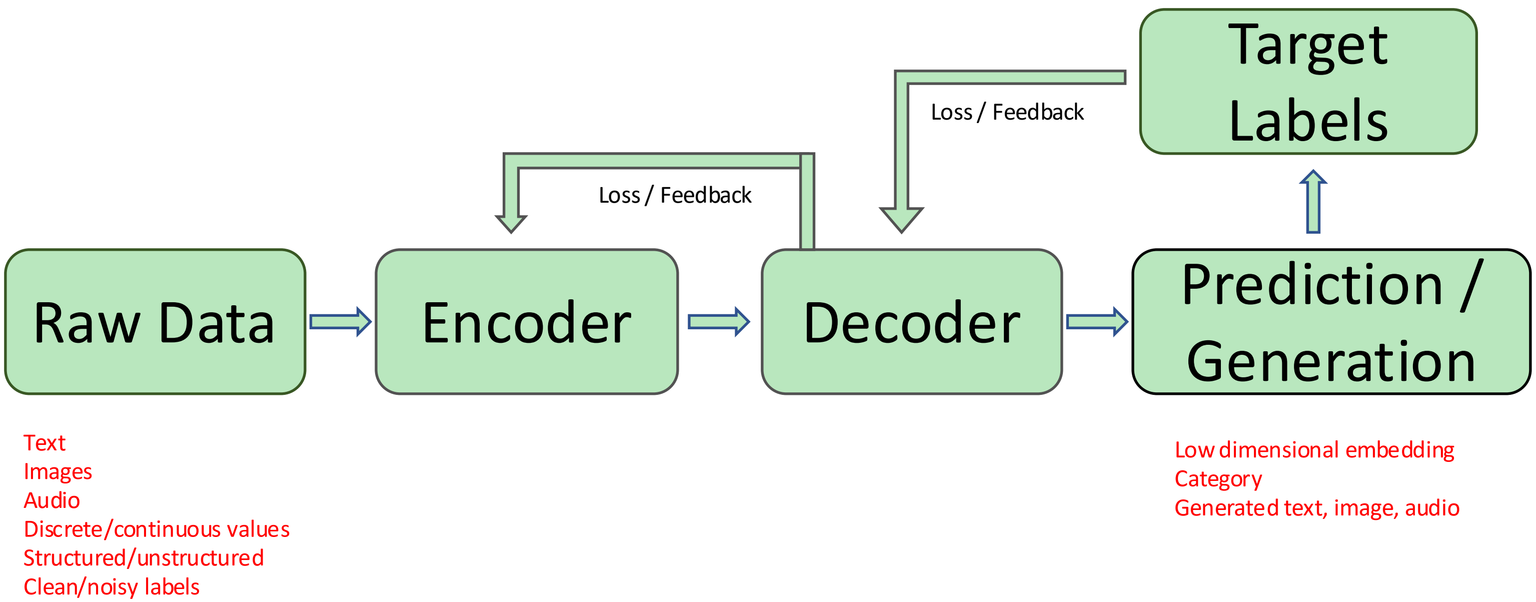
Transformer Models (to cover)

Transformers are efficient, multi-modal data processors





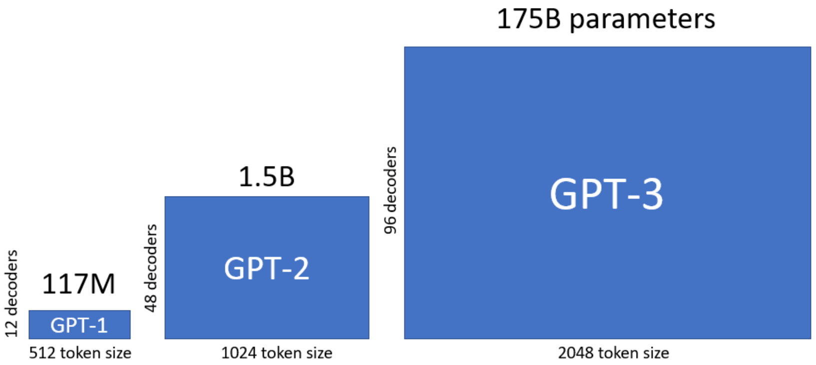
Deep Learning (to cover):



GPT: **G**enerative **P**retraining **T**ransformer Models for Language
CLIP: **C**ontrastive **L**anguage-**I**mage **P**retraining for Vision
BERT: **B**idirectional **Encoder R**epresentations from **Transformers**.

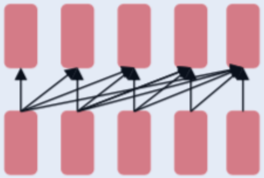
Recent LLM (to cover?)

- GPT1 / 2/ 3
- Emergent Abilities of Large Language Models
- Scaling Instruction-Finetuned Language Models
- On the Opportunities and Risks of Foundation Models



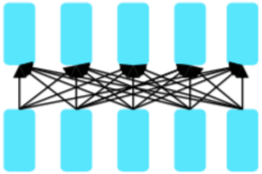
https://medium.com/@YanAIx/step-by-step-into-gpt-70bc4a5d8714

Background: Pretraining for three types of architectures



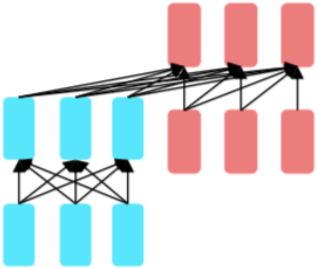
Decoders

- Nice to generate from; can't condition on future words
- **Examples:** GPT-2, GPT-3, LaMDA



Encoders

- Gets bidirectional context – can condition on future!
- Wait, how do we pretrain them?
- **Examples:** BERT and its many variants, e.g. RoBERTa



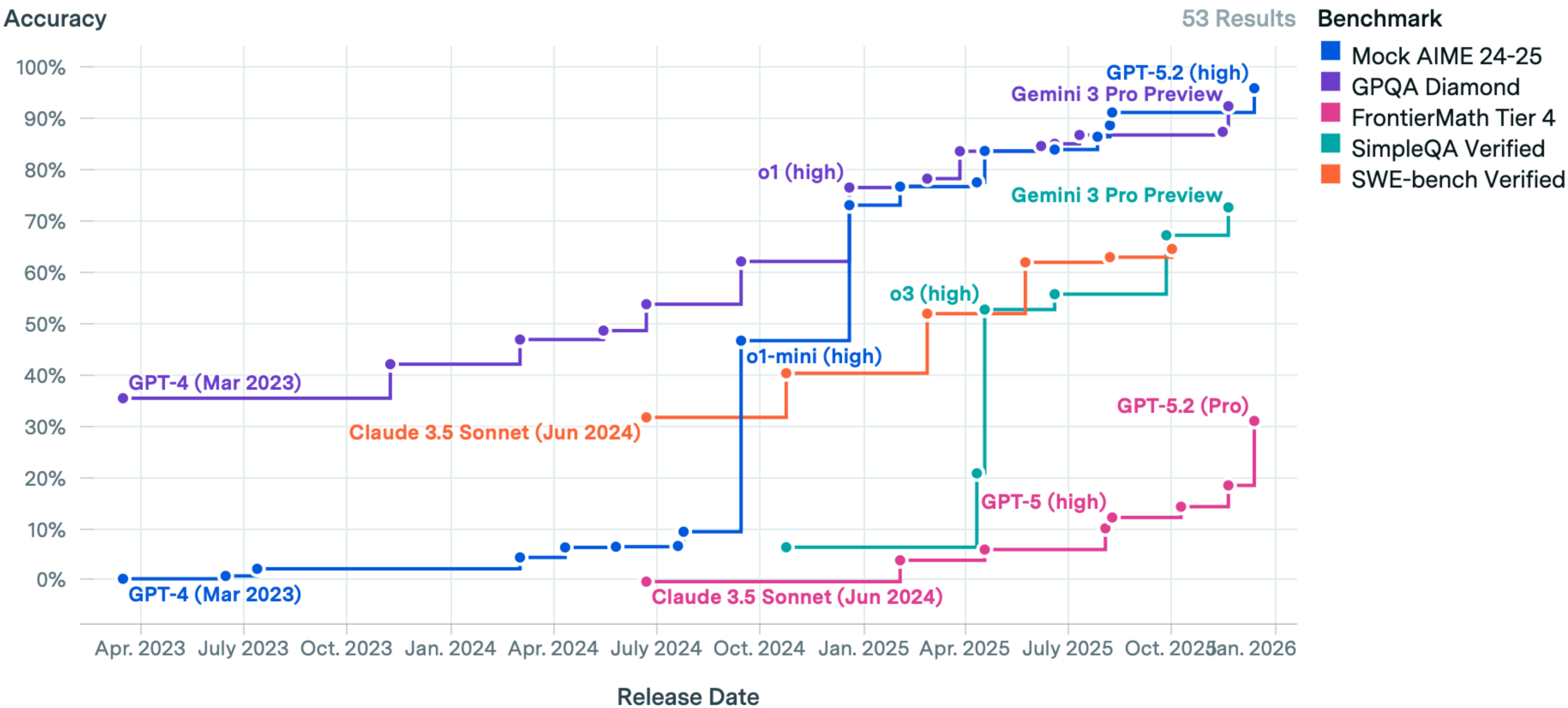
Encoder-Decoders

- Good parts of decoders and encoders?
- What's the best way to pretrain them?
- **Examples:** T5, Meena

General Deep learning in AI Trends

- To Complex tasks
  - E.g., generating slides from an outline, summarizing and reporting information from diverse sources
- Integrating into physical devices
  - E.g., Robots
- Multimodal and broadly
  - Use vision, language, audio, and broader knowledge like DB, as input or outputs
- Complex learning systems
  - Integrate predictive/generative
  - Integrate retrieval of private memories or data
  - Integrate with planning, task decomposition, and prioritization

Frontier performance across benchmarks



epoch.ai

https://epoch.ai/benchmarks

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

GPT-3

Large natural-language computer models that learn to write and speak are a big step toward AI that can better understand and interact with the world.

10 Breakthrough Technologies, 2022

AI for Protein Folding

2025

10 Breakthrough Technologies

Generative AI Search

Small LLM

Fast Learning Robots

General Lessons for Excellence

- Good breath in fundamentals is key
- Strength in particular targeted topics help standing out

Thank You

Roadmap (Module 3)

- Course Logistics
- History and Now
- Machine Learning Basics

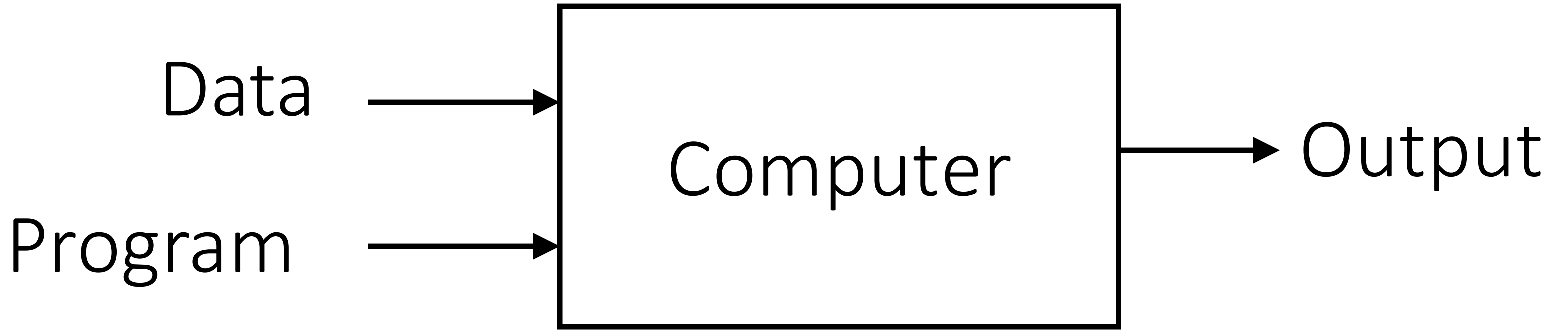
BASICS OF MACHINE LEARNING

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

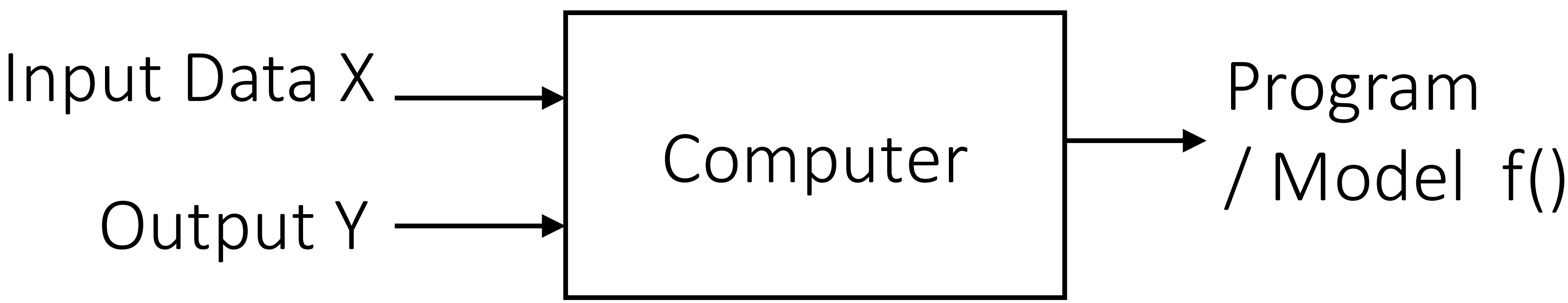
- "Experience" in the form of available data examples (also called as instances, samples)

- 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

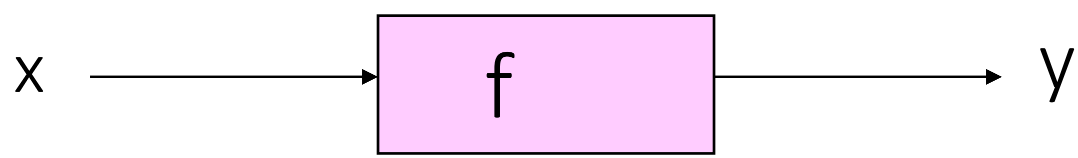
y

-1

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

e.g., SUPERVISED Linear Binary Classifier

- Now let us check out a VERY SIMPLE case of



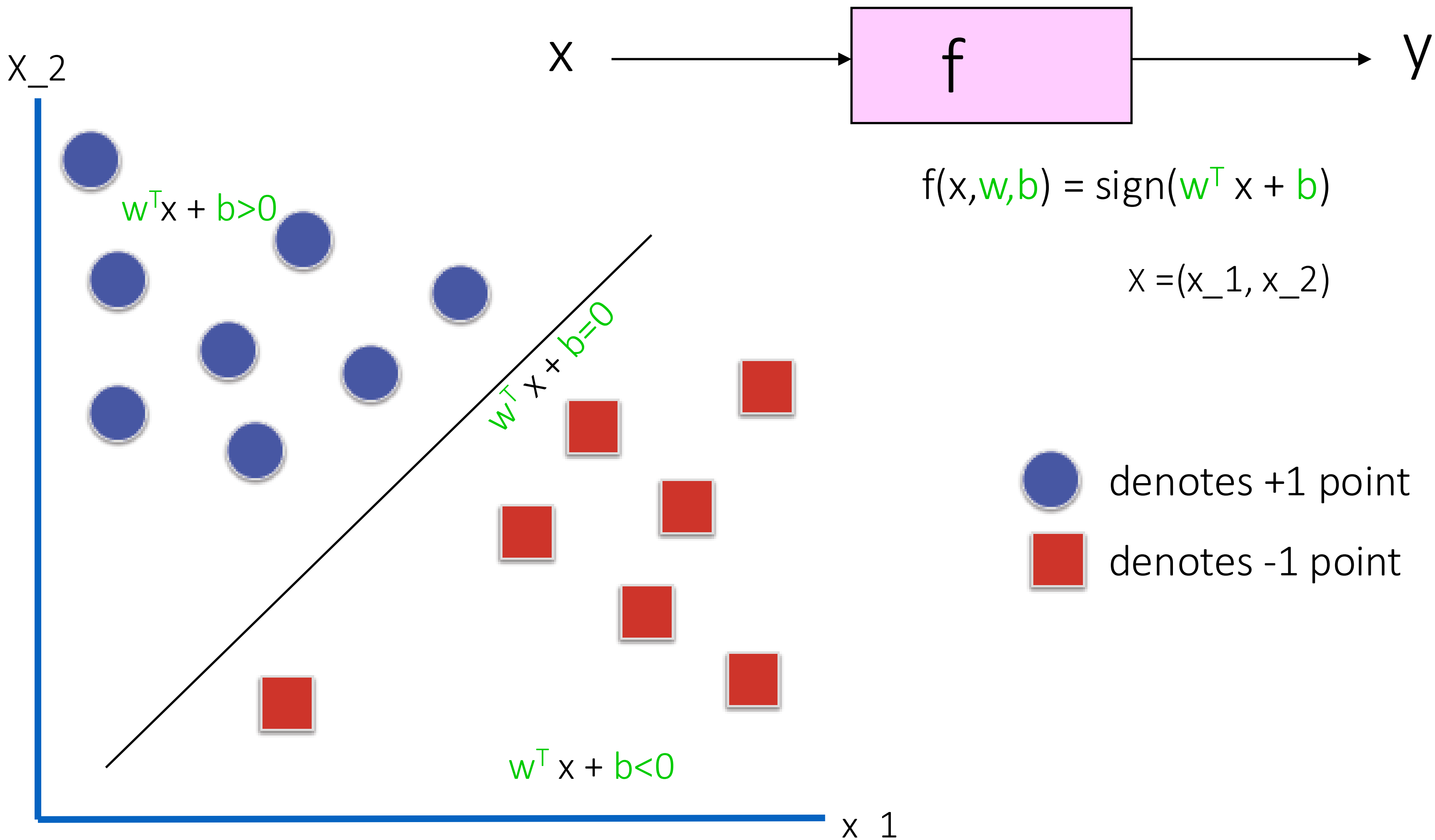
e.g.: Binary y / Linear f / X as R^2

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

$\mathbf{X}^T$ =(x_1, x_2)

$\mathbf{w}^T$ =(w_1, w_2)

e.g., SUPERVISED Linear Binary Classifier

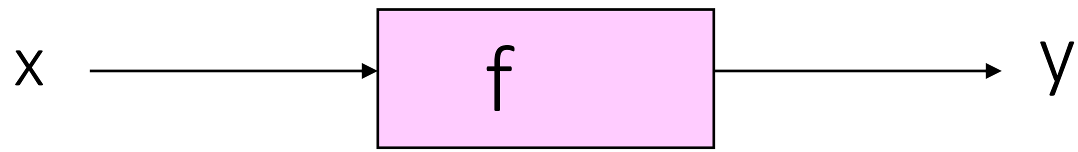


$f(x,w,b) = \mathrm{sign}(w^T x + b)$

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

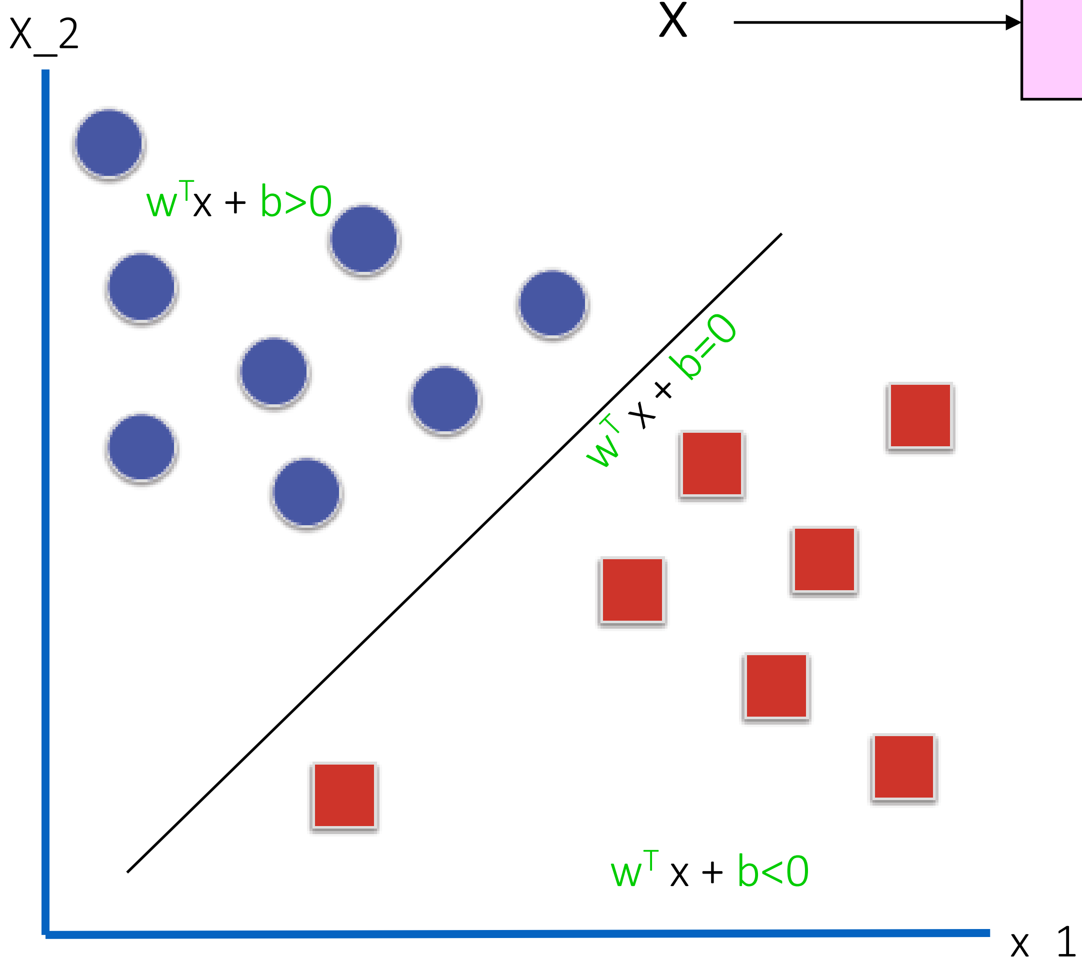
Courtesy slide from Prof. Andrew Moore's tutorial

e.g., SUPERVISED Linear Binary Classifier



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

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



denotes +1 point

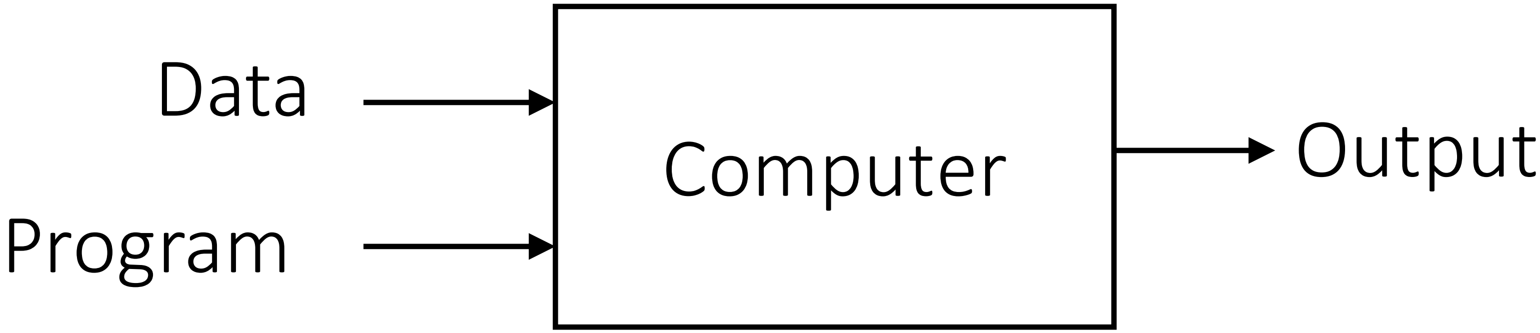
denotes -1 point

p #input variables = 2

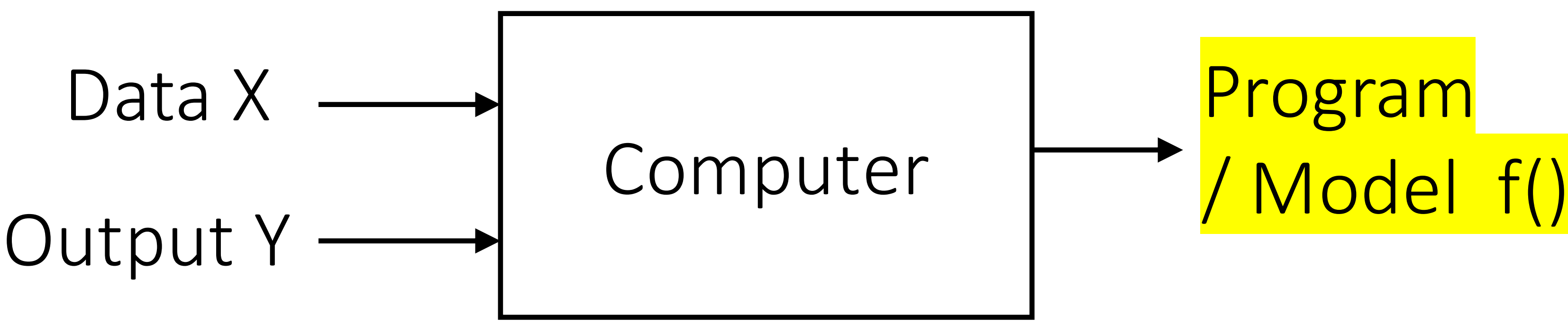
n #observations = 15

Courtesy slide from Prof. Andrew Moore's tutorial

Traditional Programming



Machine Learning (training phase)



Basic Concepts

- Training (i.e. learning parameters w,b)
  - Training set includes "experience"
    - 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) = argmin L(,)=

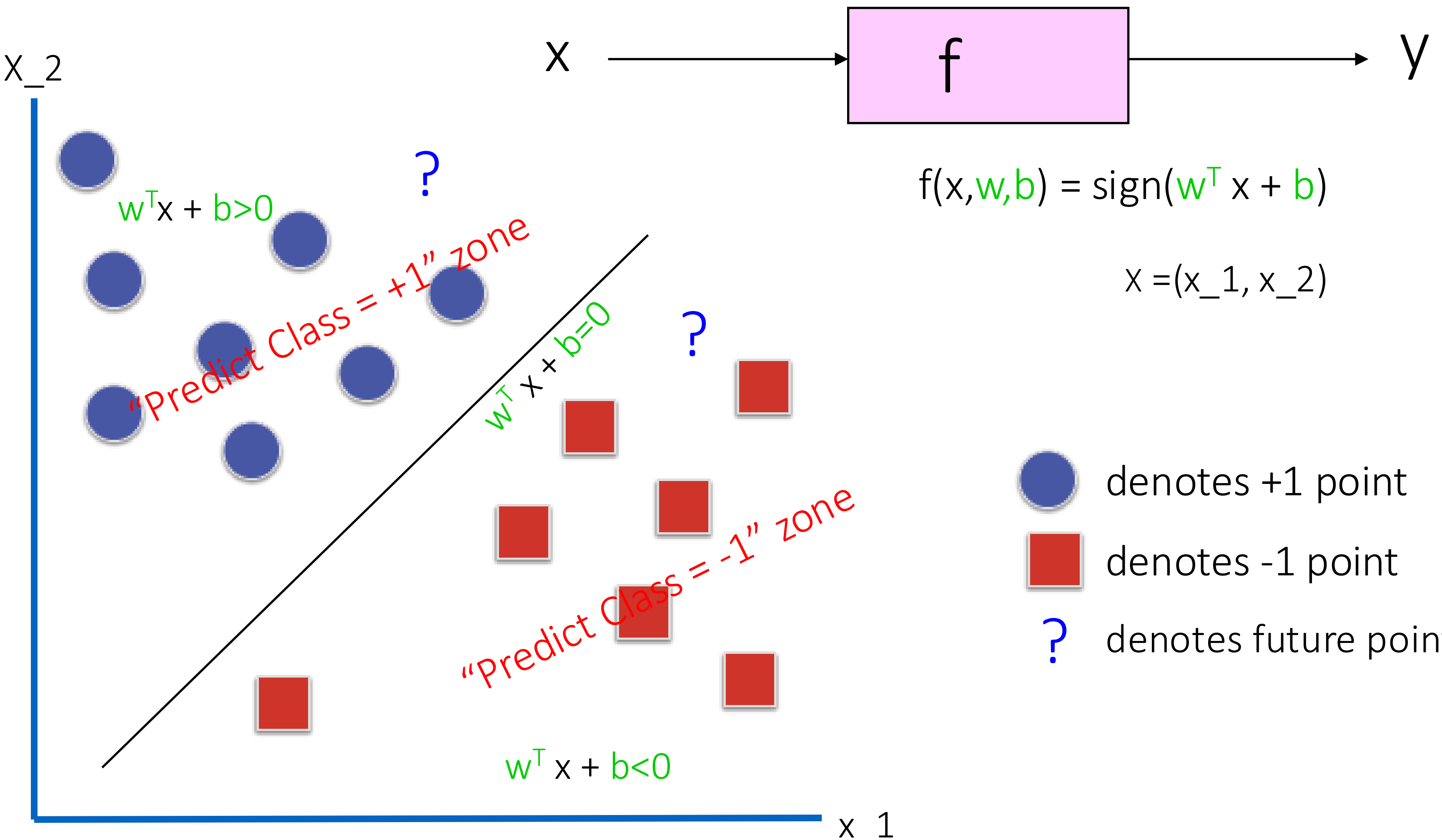
Basic Concepts

- Training (i.e. learning parameters w,b)
  - Training set includes "experience"
    - available examples $x_1, \ldots, x_n$
    - available corresponding labels $y_1, \ldots, y_n$

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

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

SUPERVISED Linear Binary Classifier



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

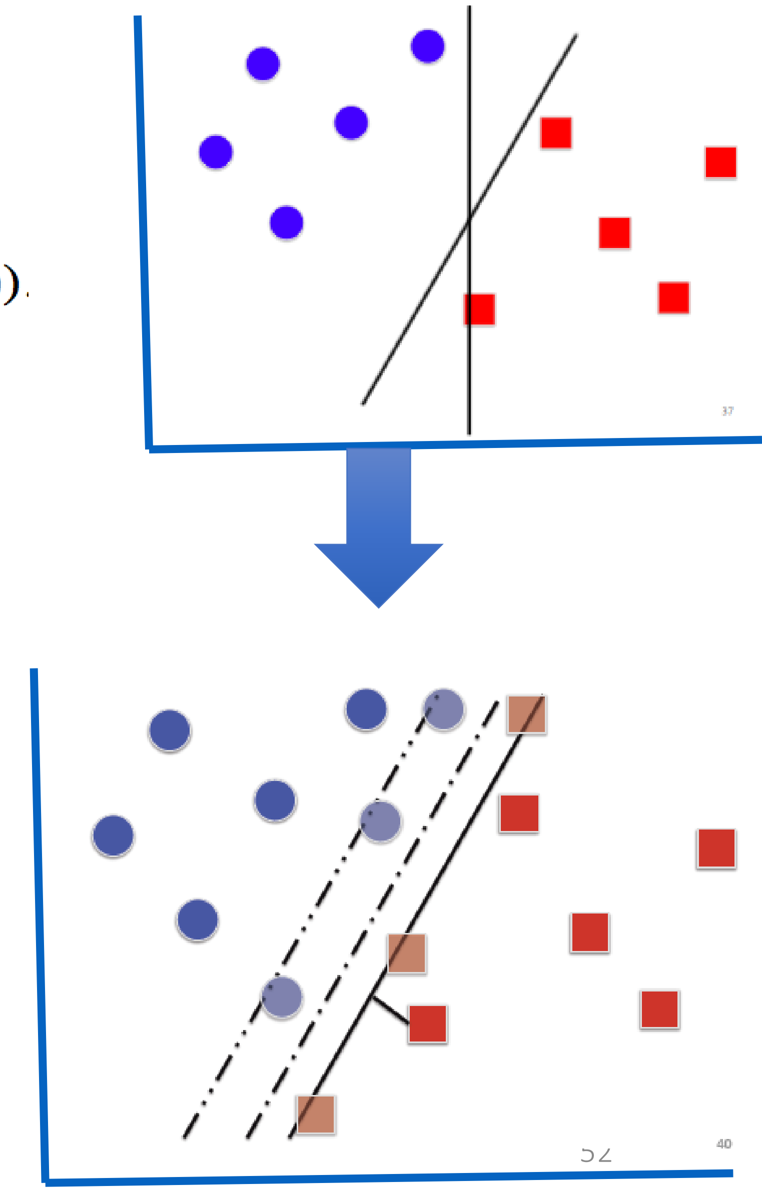
$$X = (x\_1, x\_2)$$

Courtesy slide from Prof. Andrew Moore's tutorial

Basic Concepts

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

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

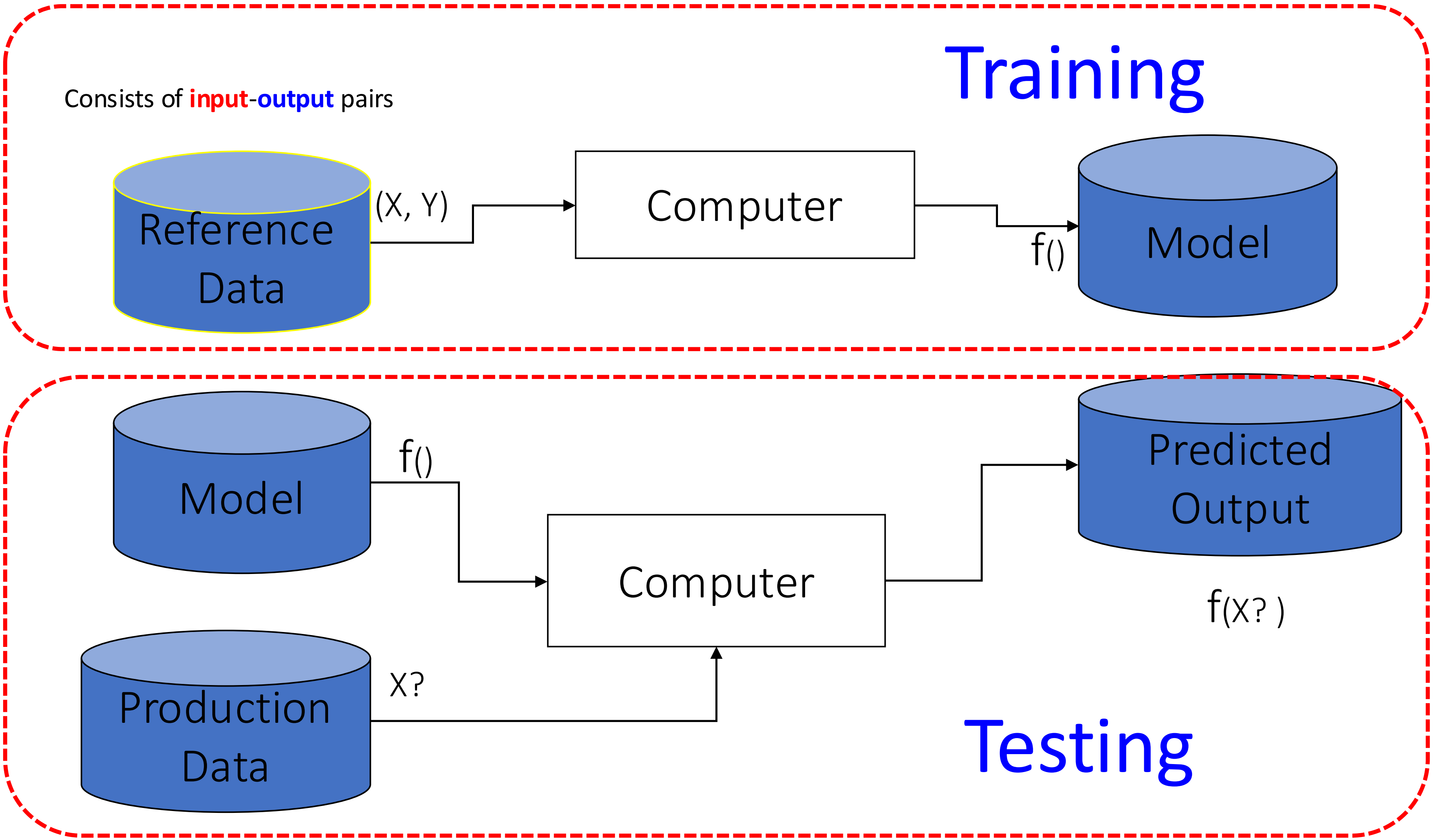


- Regularization
  - E.g. additional information added

on loss function to control f

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

Two Modes of Machine Learning

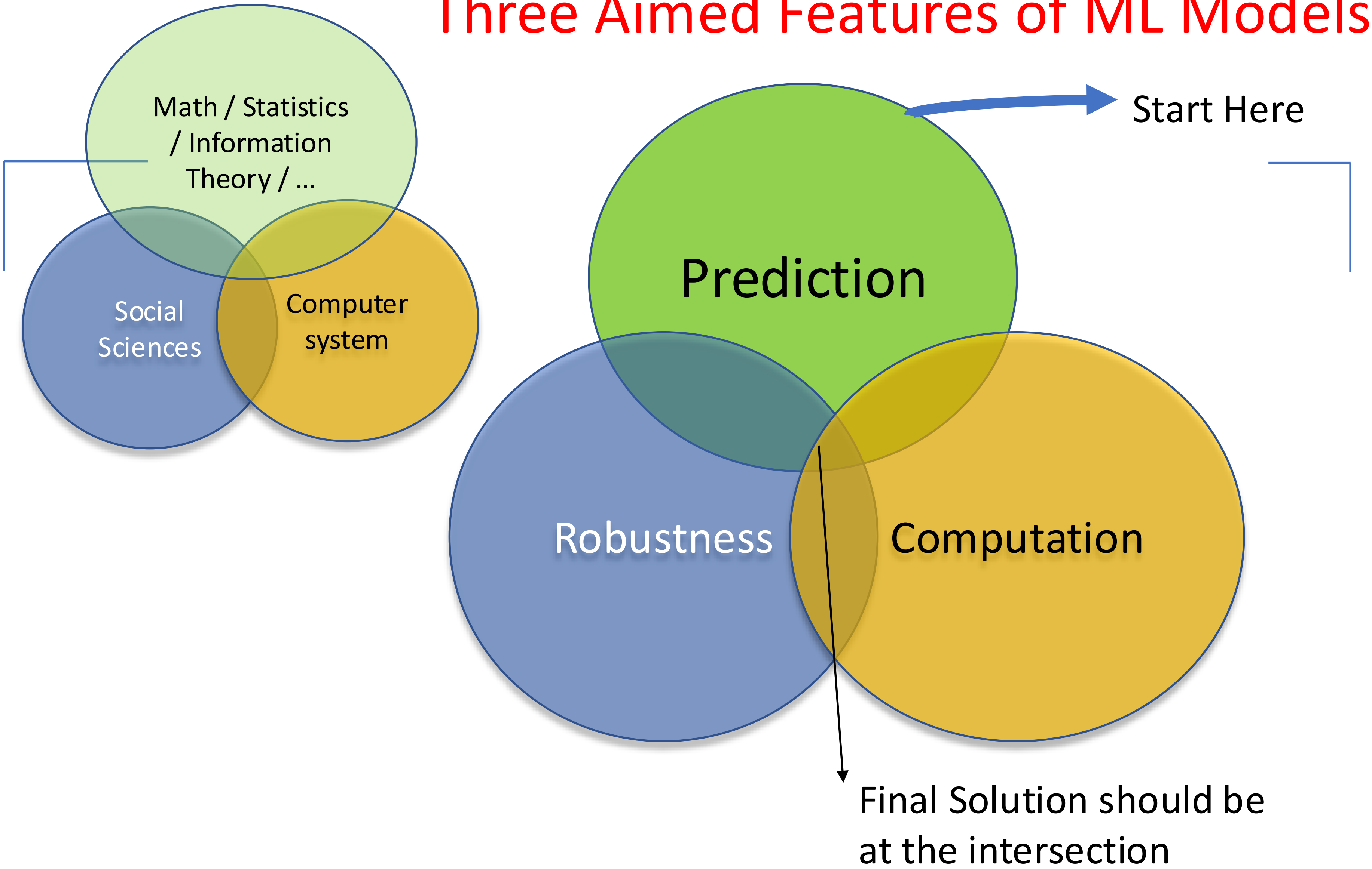


Basic Concepts

- 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

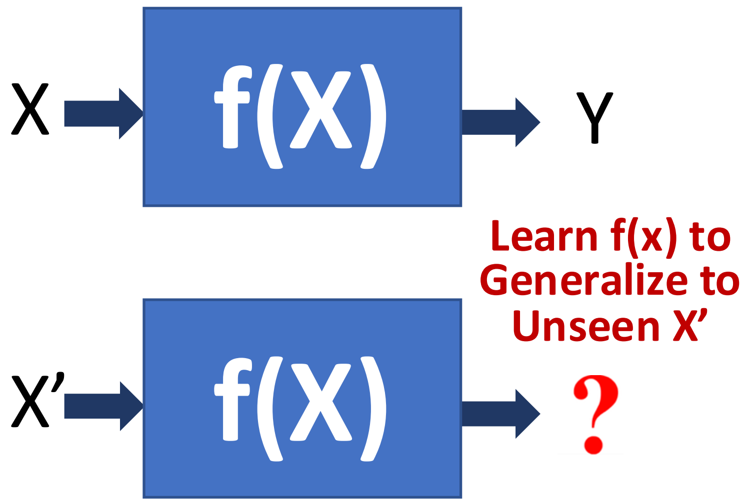
- Generalisation: learn function / hypothesis from past data in order to "explain", "predict", "model" or "control" new data examples

Three Aimed Features of ML Models



Summary: Data-Driven Machine Learning Algorithms and Platforms

- Generalizations from observed data to unseen data
  - (Inductive reasoning!!!)



- Computer systems that can learn and adapt from their experience (data)
- Can Create software that improves over time

Next Session:
Machine Learning in a Nutshell

Data
↓
Task
↓
Representation
↓
Score Function
↓
Search/Optimization
↓
Models, Parameters

Hyperparameter, Metrics

ML grew out of work in AI

Optimize a performance criterion using example data or past experience,

Aiming to generalize to unseen data


Thank You

References

- Hastie, Trevor, et al. The elements of statistical learning. Vol. 2. No. 1. New York: Springer, 2009.

- 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/