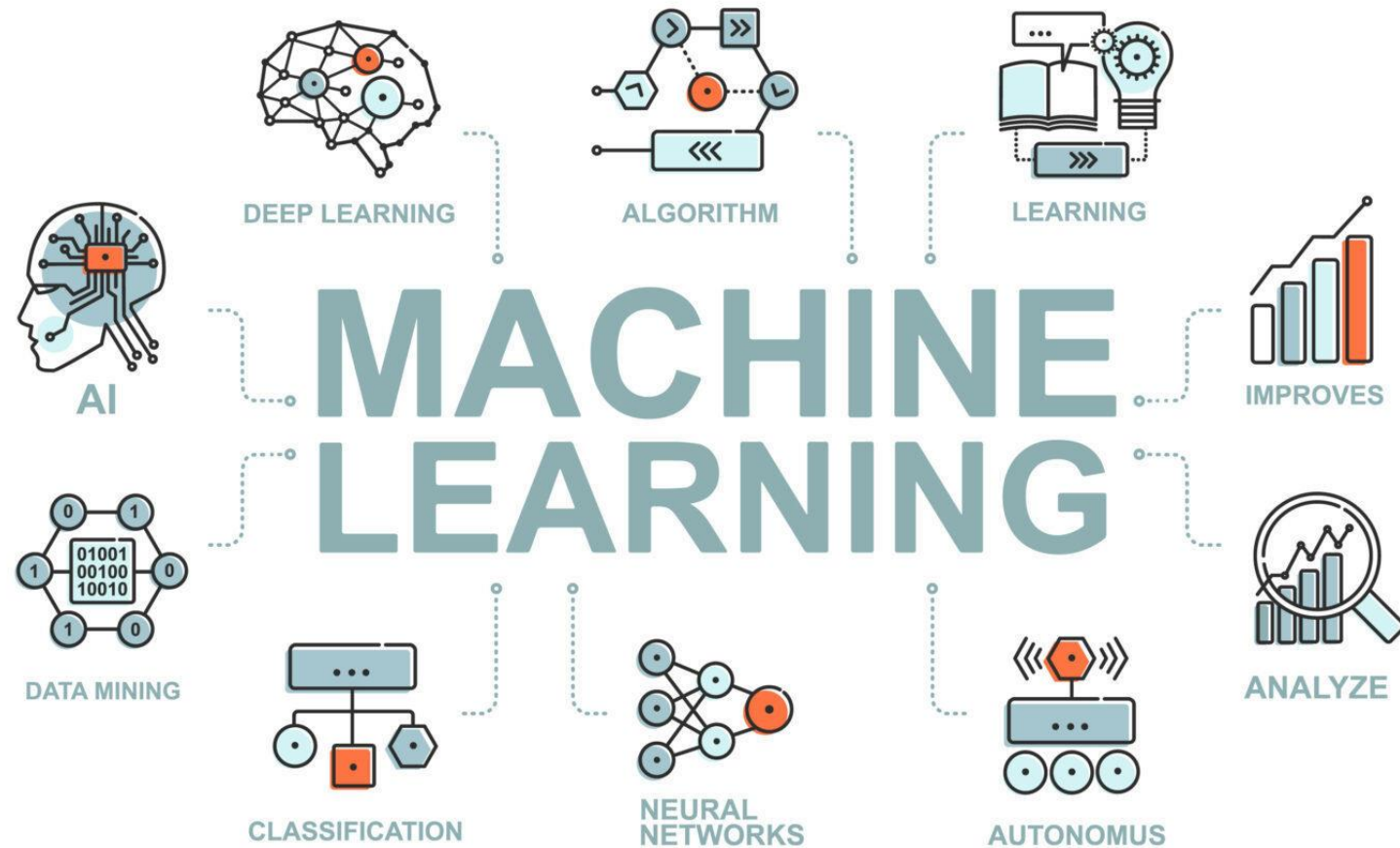# MSBA 7027 Machine Learning Overview

Zhengli Wang

Faculty of Business and Economics
The University of Hong Kong
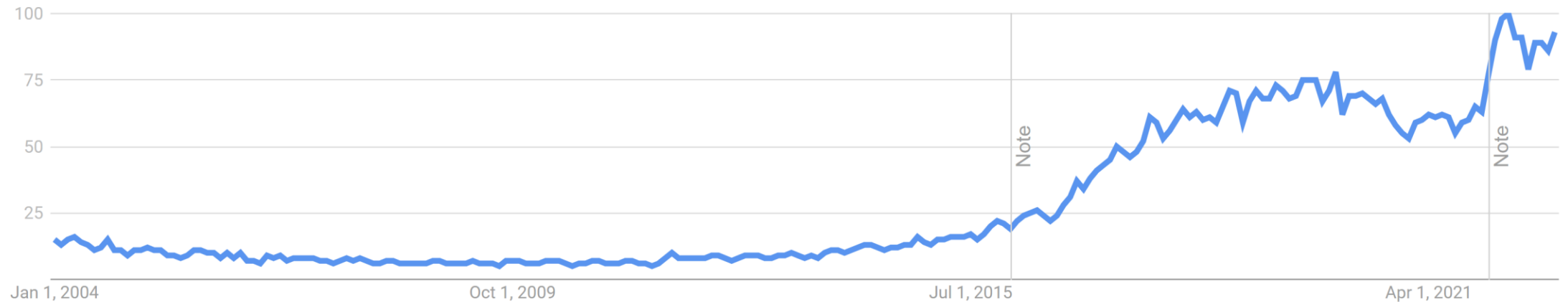2023

Welcome to Machine Learning

Overview of ML

- ML Trend
  - Google, McKinsey, Harvard Business Review, Forbes

- ML Applications in different industries

- What is ML

- Class Logistics

Machine Learning Trend: Google

Google Trends
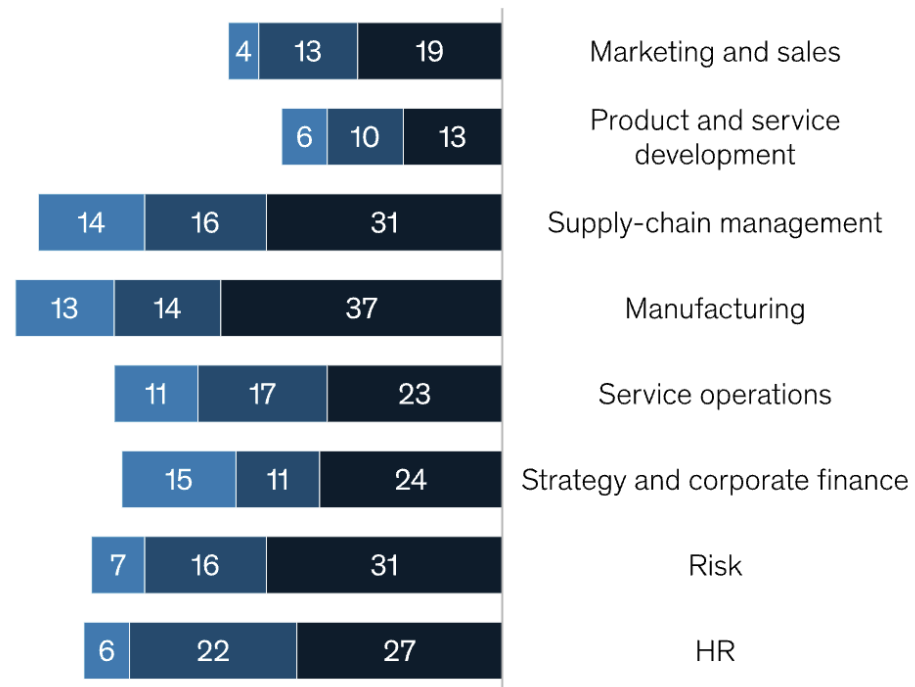Term: Machine Learning
Time Range: 2004 – 2022

Fundamental Reason of Popularity

ML/AI Decreases Cost & Increases Revenue

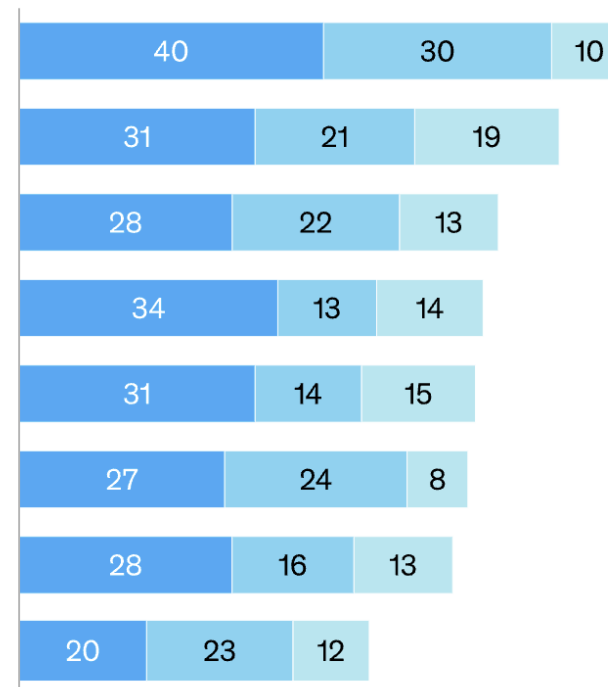**Cost decrease and revenue increase from AI adoption, by function,[1]** % of respondents[2]

Average cost decrease

■ Decrease by ≥20%	■ Decrease by 10–19%	■ Decrease by <10%

Average revenue increase

■ Increase by ≤5%	■ Increase by 6–10%	■ Increase by >10%

Function	Cost ≥20%	Cost 10–19%	Cost <10%	Revenue ≤5%	Revenue 6–10%	Revenue >10%
Marketing and sales	4	13	19	40	30	10
Product and service development	6	10	13	31	21	19
Supply-chain management	14	16	31	28	22	13
Manufacturing	13	14	37	34	13	14
Service operations	11	17	23	31	14	15
Strategy and corporate finance	15	11	24	27	24	8
Risk	7	16	31	28	16	13
HR	6	22	27	20	23	12

Source: Mckinsey

5

Firms' Reaction

- 76% of organization: Prioritize AI/ML over other IT initiatives

- 83% of enterprises: Have increased their budgets for AI and machine learning year-over-year

- 83% of IT leaders: AI & ML is transforming customer engagement and 69% say it is transforming their business.

Source: Forbes

If you master ML, how you can contribute

- Improving customer loyalty and retention.
- Automating finance.
- Detecting fraud.
- Smoother supply chains

Harvard Business Review

Analytics And Data Science | 8 Ways Machine Learning Is Improving Co...

Analytics And Data Science

8 Ways Machine Learning Is Improving Companies' Work Processes
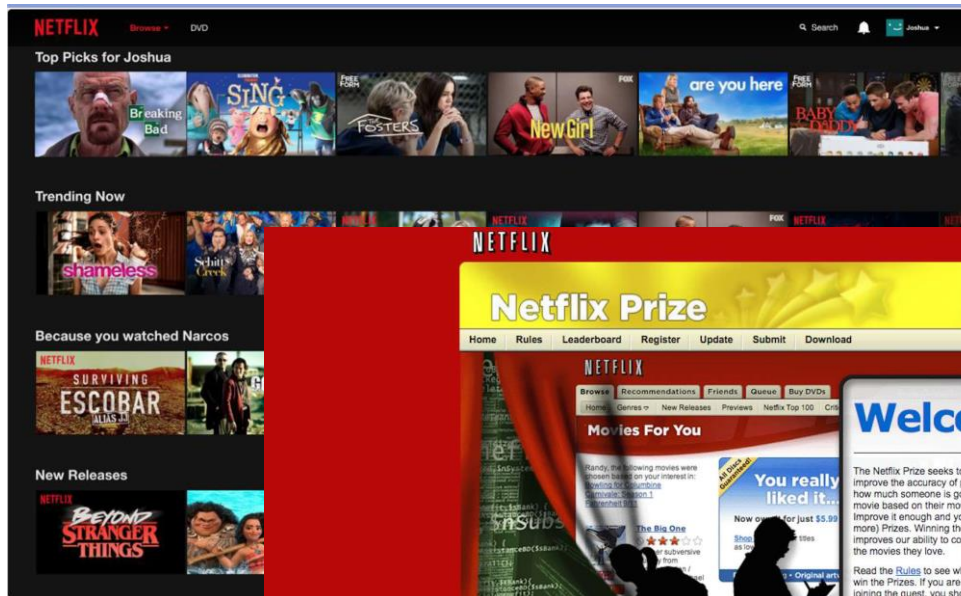
Source: Harvard Business Review

Machine Learning Applications

Across Industries

- Online Services
- Finance
- Retail
- Healthcare
- Supply Chain/Logistics
- Real Estate
- Others (Education, Manufacturing, etc.)

Machine Learning Applications

Online Services: Recommendation

Machine Learning Applications

Online Services: Product Bundling

Source: https://www.amazon.com

Dynamic Pricing

https://www.price2spy.com/blog/dynamic-pricing-explained-benefits-strategies-and-examples/

Machine Learning Applications

Finance: Fraud Detection   Trading Advisory / Algo Trading

ORDER   MACHINE LEARNING MODEL   FRAUD RISK ESTIMATE   ACCEPT OR REJECT

Machine Learning Applications

Retail: Customer Churn Modeling

Customer Segmentation & Market Research

Source: https://cio-wiki.org/wiki/Customer_Churn
Source: https://gritglobal.io/blog/5-common-customer-segmentation-mistakes/

Machine Learning Applications

Healthcare: Disease Identification / Diagnosis

Disease Prediction

Source: https://ai.myesr.org/education/
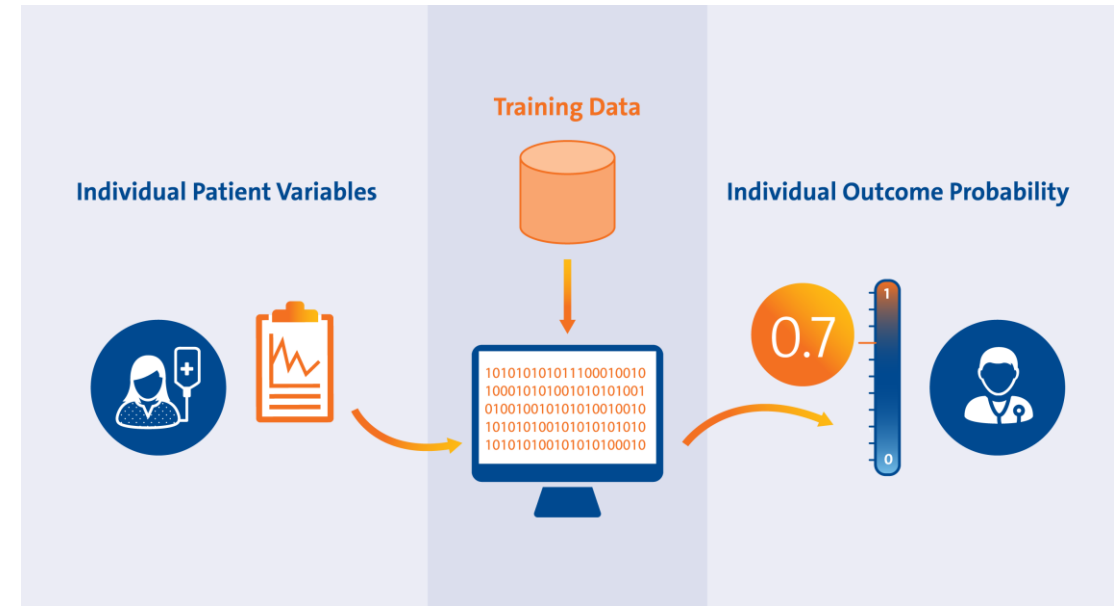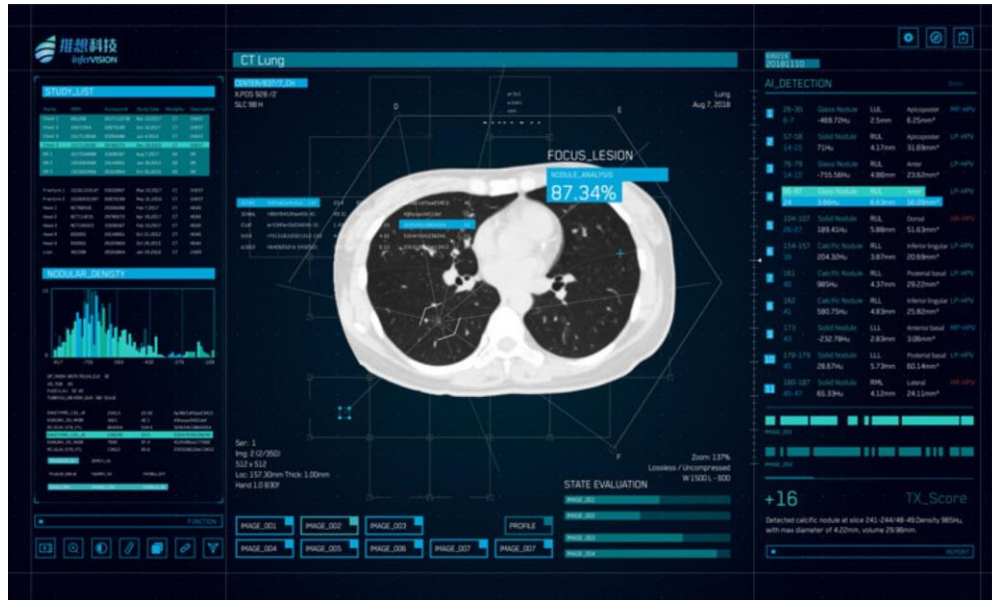
Machine Learning Applications

Supply Chain/Logistics: Demand Estimation   Delivery Prediction

https://www.forbes.com/sites/louiscolumbus/2018/06/11/10-ways-machine-learning-is-revolutionizing-supply-chain-management/?sh=787ad2303e37
https://medium.com/@ODSC/ai-as-the-ultimate-disrupter-in-logistics-how-to-manage-last-mile-costs-c4874e8f2ea0
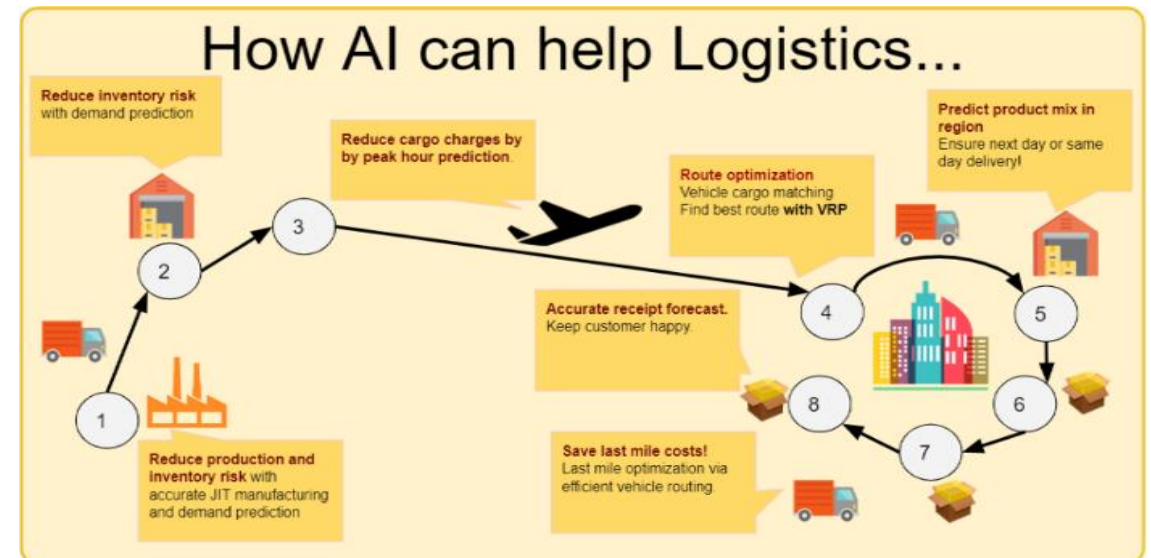
Machine Learning Applications

Real Estate: Housing Price Prediction

Source: https://github.com/RohitLearner/House-Prices-Visualization-Prediction
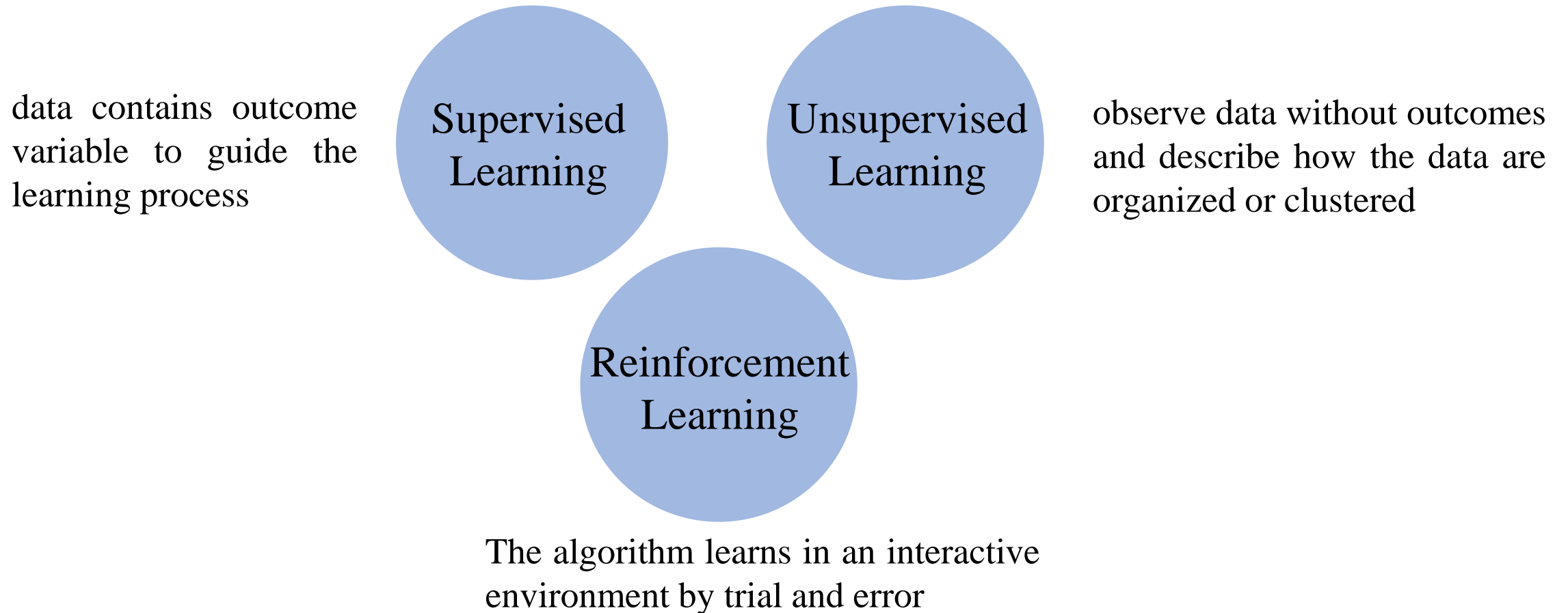
What is Machine Learning?

Making the computer able to learn without being explicitly programmed.

data contains outcome variable to guide the learning process

Supervised Learning

Unsupervised Learning

observe data without outcomes and describe how the data are organized or clustered

Reinforcement Learning

The algorithm learns in an interactive environment by trial and error

Supervised Learning

Given: A dataset that contains n samples

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

Goal: given $x^{(n+1)}$, predict $y^{(n+1)}$

- **Regression:** Outcomes are continuous values

  - Tomorrow's temperature / Stock price / House price

  - Linear Regression, Ridge/LASSO Regression

- **Classification:** Outcomes are categorical values

  - Customer Purchase or No purchase / Diabetes or Not / Spam or No Spam (Email)

  - Logistic Regression, LDA/QDA, Linear SVM

Unsupervised Learning

Given: A dataset with no labels

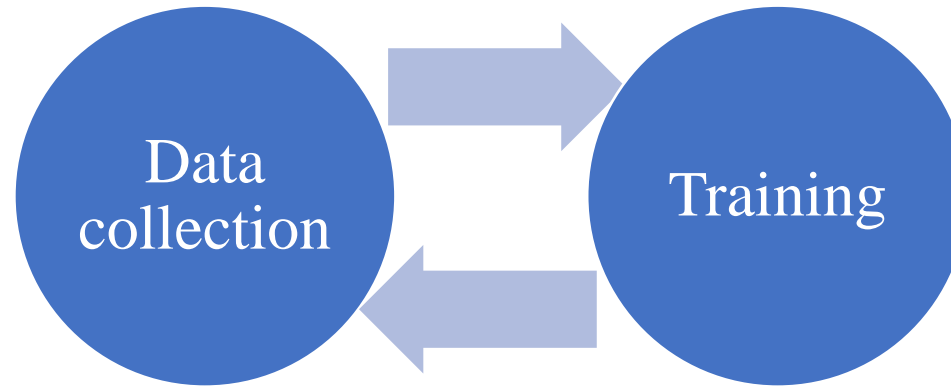$$x^{(1)}, x^{(2)}, \ldots, x^{(n)}$$

Goal: find interesting structures / patterns in the data

- PCA
- Clustering
  - K-means
  - Hierarchical

Reinforcement Learning
The algorithm collects data interactively

Try the strategy and collect feedback

Data collection

Training

Improve the strategy based on the feedback

ALPHAGO

Most relevant in business: Supervised Learning

This is what we will focus on in this course

In MSBA7002

Linear/Logistic Regression
Linear SVM
LDA/QDA
PCA
Clustering

Topics in This Course (MSBA7027)

K nearest neighbors (KNN)
Nonlinear Regression: Splines
Nonlinear SVM
Machine Learning System Design
Tree, Bagging, Random Forest,
Gradient Boosting Machine
Stacking
Interpretable Machine Learning
Recommender System

Class Logistics

Pre-requisite

MSBA7002 Business Statistics

Linear algebra

Statistics and probability

Programming experiences with R or Python

Tentative Schedule

Topic	Content	Lecture
1	ML Overview, Class Logistics	~1-4
2	K-nearest neighbors	
3	Nonlinear Methods in Regression: Splines	
4	Nonlinear Methods in SVM	
5	Machine Learning System Design	~5-7
6	Decision Tree, Bagging, Random forest	
7	Gradient Boosting Machines	
8	Stacking / Ensemble Learning	~8-10
9	Interpretable Machine Learning	
10	Recommender Systems	
11	Preview of Deep Learning	

Tentative Schedule

TAs:
Zhao Yuwei (Vera): zhaoyw16@hku.hk
Hu Yue: huyue777@connect.hku.hk
Zhu Zhuofu: zhuzhuofu60@gmail.com

If you have any questions, please email the TAs for fast response.

If you need to talk to the instructor, please do so before or during class break / during instructor's office hour.

Tentative Schedule

Two Tutorials (Optional):
1st Tutorial (after Lec 4):
Dec 18, 2023 (Mon) 7:30 - 9:30pm or Dec 20, 2023 (Wed) 7:30 - 9:30pm

2nd Tutorial (after Lec 8):
Jan 13, 2024 (Sat) 7:30 - 9:30pm or Jan 15, 2024 (Mon) 7:30 - 9:30pm

Weekly Office Hours: Tuesday 3:30-6:00pm (tentative)

Note: 1st class for Section A/B: Dec 4; 1st class for Section C: Dec 6

Programming Language

- Why we use R in this course
    - **Smooth transition** and learning experience from MSBA 7002

    - **Used by the best data scientists in the world.**
        - Most popular tool among professional ML practitioners

    - **Powerful: breadth of techniques it offers in third-party packages.**
        - Has more techniques than any other platforms

    - **Advanced: state-of-the-art**
        - New algorithms developed & released by academics

    - **Free & open-source**

- Note: Python is also allowed in this course.
    - But R is encouraged because we will be doing many demonstrations using R
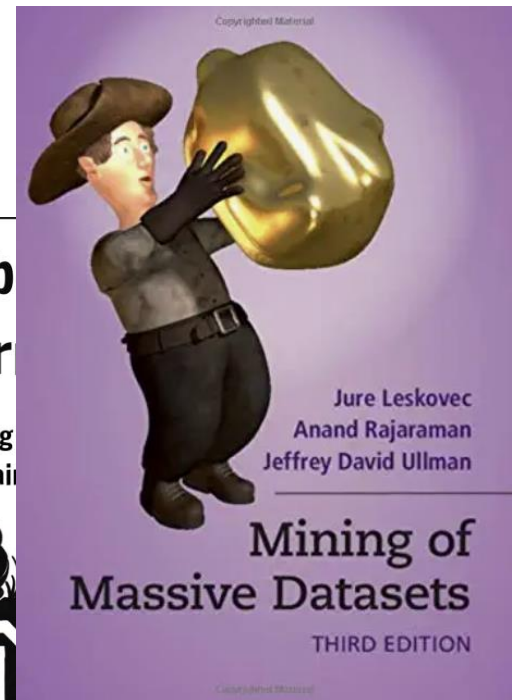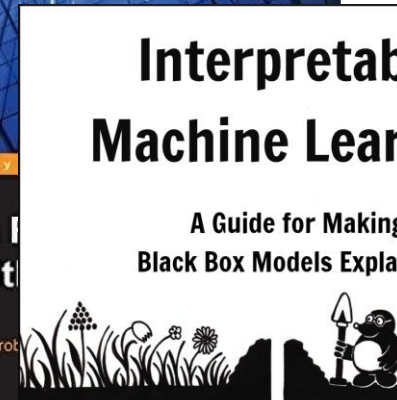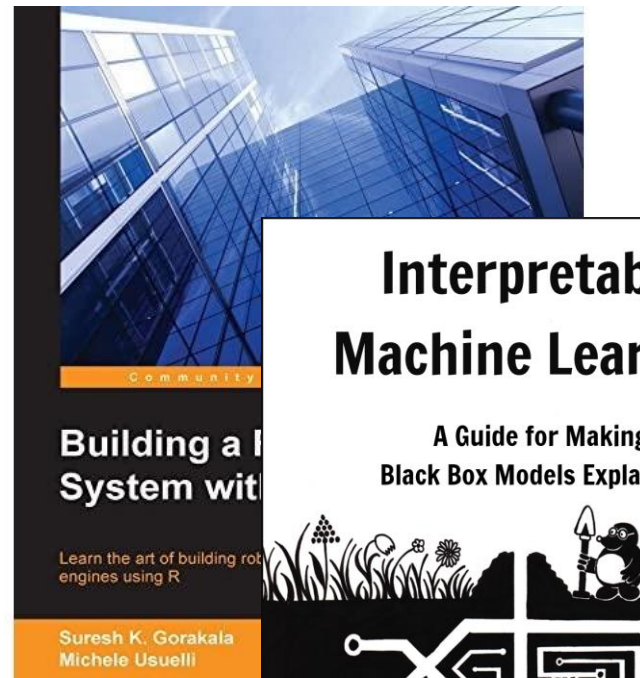
Reference Textbooks
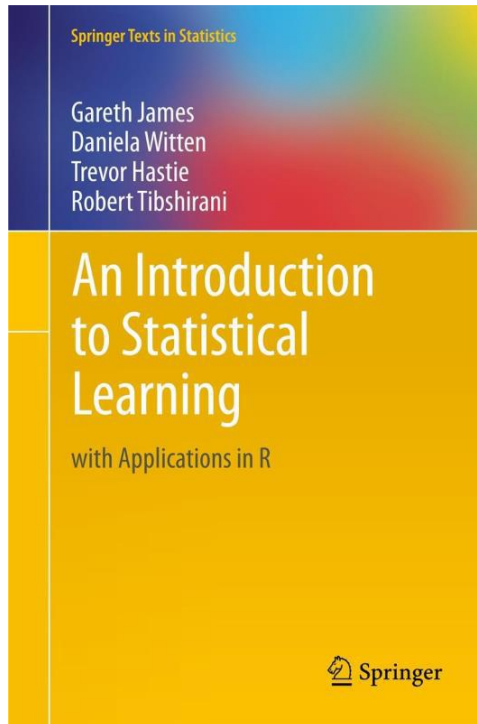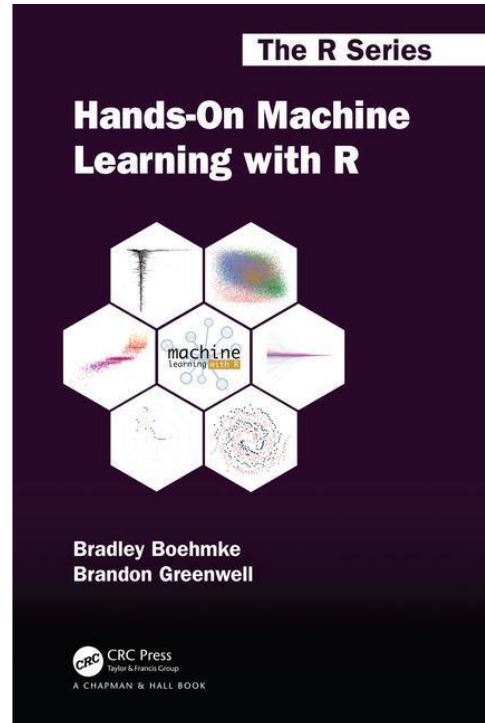
ISLR   HMLR

Grades

Attendance & Participation   20%
Assignment 1   20%   Due Dec 20, released by the week of Dec 11
Assignment 2   20%   Due Jan 20, released by the week of Jan 1
Final Exam   40%   On Jan 24, tentative 6:30PM – 9:00PM

The TAs and the instructor reserve the rights to award bonus points for students who show impressive participation in class

Assignments: General Rules

- Plagiarism will NOT be tolerated, it will automatically result in an F grade

- Don't look at solutions or code that are not your own; everything you submit should be your own work

- Don't share your solution code with others; however discussing ideas or general strategies is fine. Indicate in your submissions anyone you discussed/worked with

Expectation for this course

- Lots of useful tools / material

- Will be challenging
- Lots of coding involved
  - Self-study, adapt to new methods / new packages
  - Patience in debugging

- To facilitate learning, course designed to
  - Contain basic & optional content
  - Will do a lot of coding demonstrations in class

- Be mentally prepared
- Most Important: Review material after every lecture, **put in your best effort** – You will learn lots of useful material and you won't regret putting in the effort!

Welcome to MSBA 7027

Learn Stuff & Have Fun!