FUNDAMENTALS OF MACHINE LEARNING

INTRODUCTION

CSCI3320

Prof. John C.S. Lui, CSE Department, CUHK

Introduction to Machine Learning

Administrative Matter

- **Teacher:** John C.S. Lui
- **Email:** cslui@cse.cuhk.edu.hk
- **URL:** www.cse.cuhk.edu.hk/~cslui
- **URL to the course:** www.cse.cuhk.edu.hk/~cslui/csci3320.html
  - Also available on blackboard (for those registered students)
- **Lecture:**
  - Tuesday, 8:30 am – 10:15 am (LSK LT3)
  - Wed, 8:30 am – 9:15 am (LSK LT1)
  - **Tutorial:** Wed, 9:30 am – 10:15 am (LSK LT1)
- **Teaching Assistants:**
  - **Mr. Jincheng Wang**
  - **Miss Shiyuan Zheng**
  - **Mr. Xuchuang Wang**

Administrative Matter

- **Grades:**
  - **Homework: 5%**
  - **Python/scikit-learn Programming: 35%**
  - **Final Exam: 60%** (*no minimum threshold*)
  - **Student/Faculty Expectations on Teaching and Learning:**
  - http://www.cse.cuhk.edu.hk/~cslui/student_teacher_expectations.pdf
- **Plagiarism:** Plagiarism, an act of passing off the work of others as one's own, is considered a serious offense by the university. If a student is found plagiarizing, his/her case will be reported to the faculty disciplinary committee. The definition of plagiarism includes copying part or all of written assignments, programming exercises, reports, quiz papers, mid-term examinations and final examinations. The penalty applies to both the student who copies the work and the student whose work is being copied (unless the latter can prove his/her work has been copied unwittingly.) Furthermore, inclusion of others' works or results without citation in assignments and reports is also regarded as plagiarism with similar penalty to the offender.

Reading and Homework Assignment

- Finish Reading Chapter 1
- Start Reading Chapter 2

- **Issues to handle:**
  - ***how to handle students who are still in the waiting queue or want to enroll in this course?***
  - ***Decide on make-up lecture on conference leave***
  - ***Decide on the final examination date (see the homepage)***

AI Paradox

- Many difficult problems for human are easy AI
- Many easy problems for human are difficult for AI
- **Narrow intelligence vs. General intelligence**

AI Narrow Intelligence

- Beat Go World Champions
- Recognize facial expression
- Write music
- Diagnose illness
- Comfort senior citizens

AI General Intelligence

- Understand abstract concept
- Explain why
- Be creative even like children
- Tell right from wrong
- Have emotions

- Is Machine Learning part of AI? NO !!!!!

Abundance of Data

- Widespread use of personal computers and communication leads to "big data"
- A better wording, "***data science***"
- We are both producers and consumers of data (*example*)
- Data are not random, they have structures, e.g., customer behavior, housing prices, university ranking, …, etc
- We need "**theories**" and "**algorithms**" and "**systems**" to extract that structure from data so to

  (a) *Understand the process*

  (b) *Make predictions for the future*

Motivation

- To solve a problem, we need an "**algorithm**"
- Input + algorithm => Output
  - Example: sorting, network protocols, etc *(we know the logic)*
- In some instances, we *don't know the algorithm*
  - **Example**: Email spam filters
    - We know the input (character strings)
    - We know the output (yes or no)
- What we lack in knowledge, make up with data
  - *Learn what constitutes a spam*
  - *Want computer to learn (or extract) the algorithm*

What do we mean by "Learn" ?

- **Machine learning** is to program computers to optimize a performance criterion using ***example data*** or ***past experience***.
- There is no need to "learn" to do sorting, for it requires logic
- Learning is used when:
  - Human expertise does not exist (*cure to cancer*),
  - Humans are unable to explain their expertise (*facial or pattern recognition*)
  - Solution changes in time (*edge computing, load balancing*)
  - Solution needs to be adapted to particular cases (*user biometrics*)

What We Talk About When We Say "Learning" ?

- Learning **general models** from data of particular examples
- Data is cheap and abundant
  - *Example: data warehouses, blogs, online social networks;*
  - knowledge is **expensive** and **scarce**
- Example in retail: Customer transactions to consumer behavior:

  People who bought "diaper" also bought "milk or baby formula"
- Build a model that is *a good and useful approximation* to the data

Data Mining: Application of Machine Learning in Databases

- **Retail:** Market basket analysis, Customer relationship management (CRM)
- **Finance:** Credit scoring, fraud detection
- **Manufacturing:** Control, robotics, troubleshooting
- **Medicine:** Medical diagnosis
- **Telecommunications:** Spam filters, intrusion detection
- **Bioinformatics:** Motifs, alignment
- **Web mining:** Search engines
- …

What is Machine Learning?

- Optimize a performance criterion using example data or past experience.
- Have a model with parameters. **Learning** *is the execution of the computer program to* ***optimize*** *the parameters of the model* using the data
- **Role of Statistics:** Inference from a sample
- **Role of Computer science:** Efficient algorithms to
  - Solve the optimization problem
  - Representing and evaluating the model for inference

Learning Problem Definition

- Improve some measure of performance P when we execute some task T via some experience E
- **Example:** Learn to detect frauds in credit card usage
- **Task T:**
  - Want to assign labels (Yes to fraud, No to normal) to credit card transactions
- **Performance measure: P**
  - Accuracy in the label assignment
  - Note that we have two types of errors: false negative; false positive
- **Experience E**
  - Log of credit card transactions with labels of whether a transaction is fraudulent or not.

Applications

- Association
- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning

Learning Associations

- Basket analysis in supermarket:
- "*People who buy X also buy Y*"
- The association rule can be expressed in **conditional probability**

  $P(Y \mid X)$ probability that somebody who buys X also buys Y where X and Y are products/services.

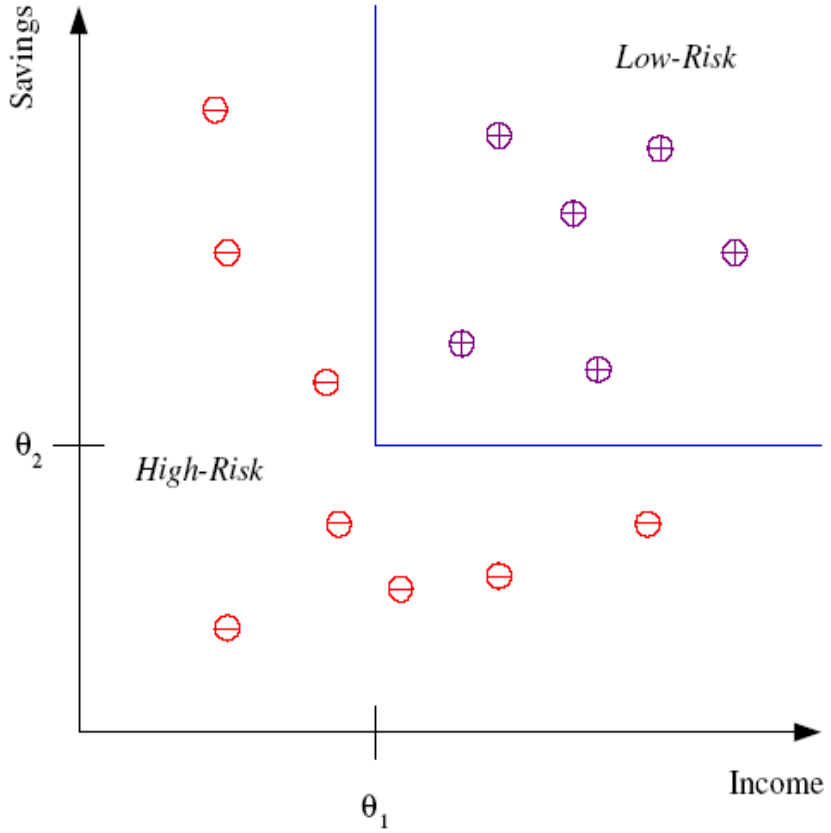
  Example: P (potato chips | beer) = 0.7

  Rule: 70% of people who buy beer also buy chips

  More advance: find P(Y|X,D)

Classification

- Example: Credit scoring
- Differentiating between low-risk and high-risk customers from their *income* and *savings*



Discriminant: IF *income* > θ_1 AND *savings* > θ_2
THEN low-risk ELSE high-risk

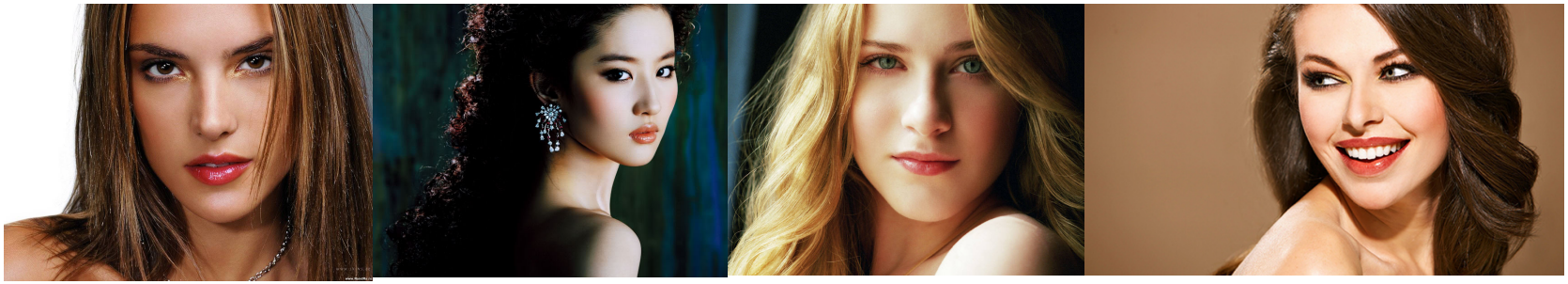
Classification: Applications

- Aka Pattern recognition
- **Face recognition:** Pose, lighting, occlusion (glasses, beard), make-up, hair style
- **Character recognition**: Different handwriting styles.
- **Speech recognition:** Temporal dependency
- **Medical diagnosis:** From symptoms to illnesses
- **Biometrics:** Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc
- **Outlier/novelty detection:** network traffic, computer malware/virus detection

Face Recognition

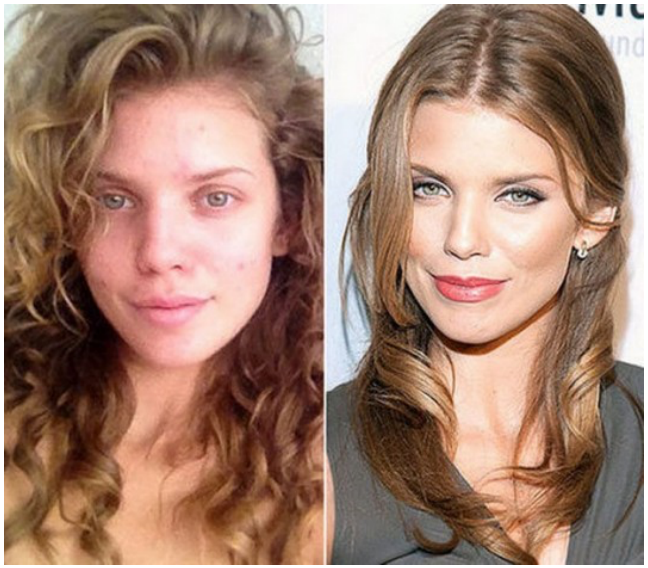
Training examples of a person

Test images



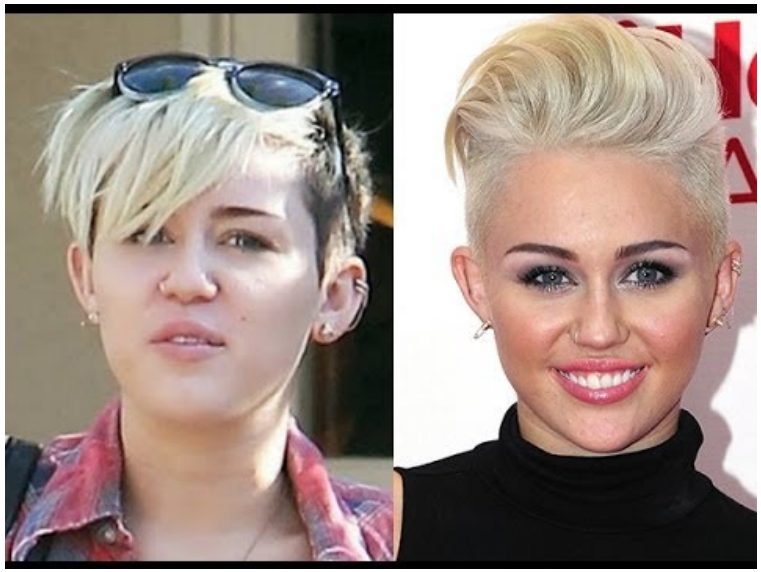
Face Recognition

AnnaLynne McCord



Madonna

Miley Cyrus



Michael Jackson

Regression

Regression: a statistical technique for estimating relationship among variables

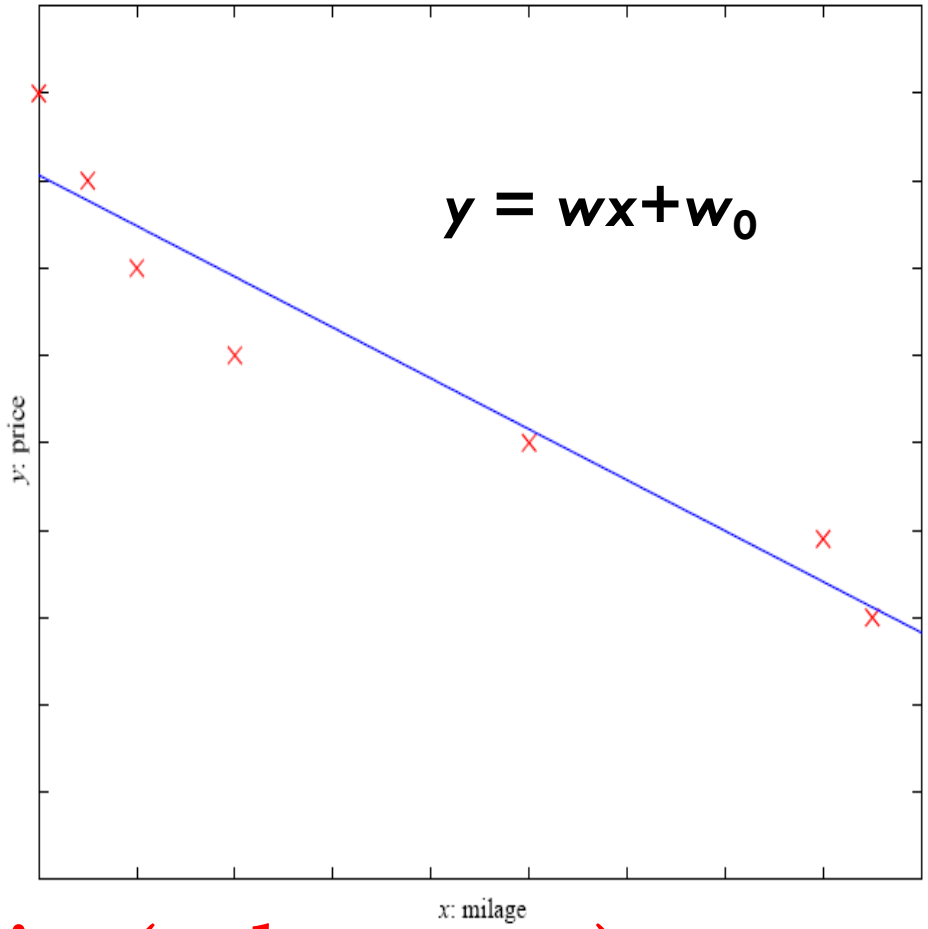
- Example: Price of a used car
- x : car attributes

  y : price

  $y = g(x \mid \theta)$

  $g(\ )$ model,

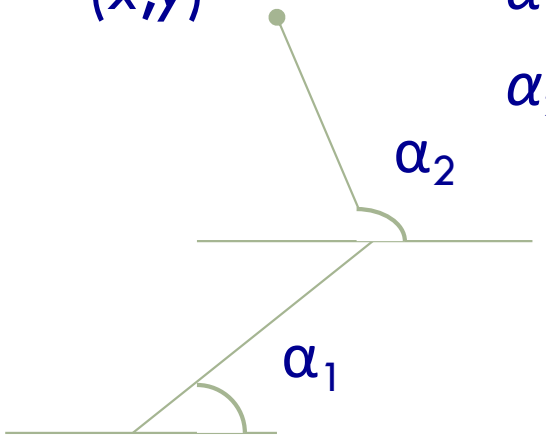
  θ parameters



Find w_0 and w to optimize (reduce error)

Regression Applications

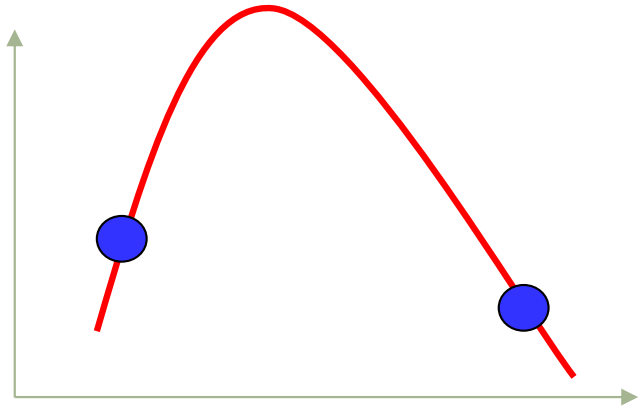
- Navigating a car: Angle of the steering
- Kinematics of a robot arm



$\alpha_1 = g_1(x,y)$

$\alpha_2 = g_2(x,y)$

- Response surface design



Supervised Learning: Uses

- Both *classification* and *regression* are "**supervised learning**", or learn the function from "**labeled data**"
- **Prediction of future cases:** Use the rule to predict the output for future inputs
- **Knowledge extraction:** The rule is easy to understand
- **Compression:** The rule is simpler than the data it explains
- **Outlier detection**: Exceptions that are not covered by the rule, e.g., fraud

Supervised Learning

- Inputs x are vectors, or some complex objects
  - Documents
  - DNA sequences
  - Graphs

- Outputs are binary or multiclass labels (K)
  - Multi-label (label is one of the K classes)
  - Ranking

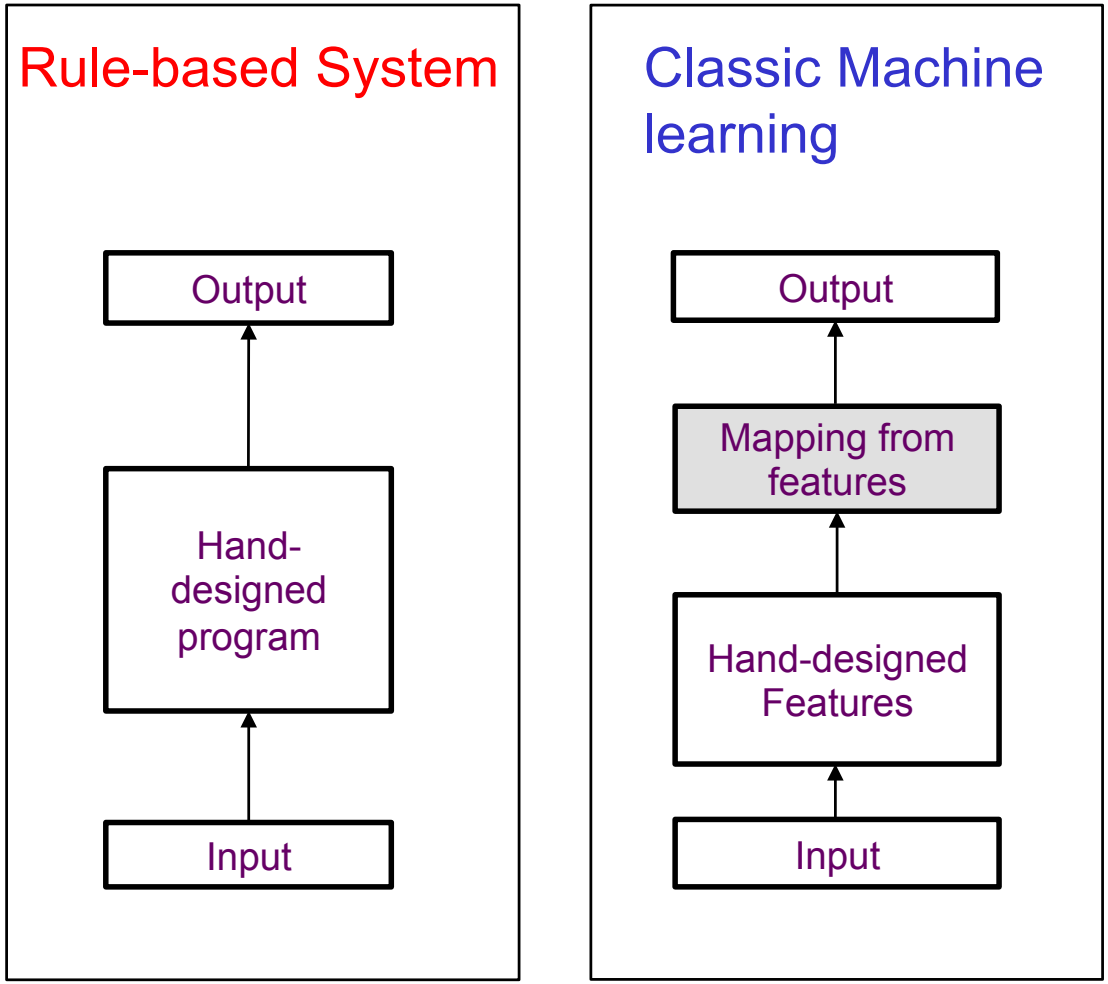
Unsupervised Learning

- Learning **without** "labeled data"
- Learning "what normally happens"
- **Clustering**: Grouping similar instances
- **Example applications**
  - **Customer segmentation** in CRM
  - **Image compression**: Color quantization
  - **Bioinformatics**: Learning motifs
  - **Online social networks**: Triangles, community detection
  - **Recommender Systems**: suggest products for users

Reinforcement Learning

- **Learning to have a good policy:** A sequence of outputs
- No supervised output but delayed reward
- **Examples:**
  - Credit assignment problem
  - Game playing
  - Robot in a maze
  - Multiple agents, partial observability, ...
- **Theory:**
  - Game theory
  - Markov decision processes (MDP)
  - Multi-arm bandits

Two Approaches of "AI"



This is the module we can use ML to learn from data

Background Needed

- Python Programming
- Statistics
- Linear Algebra (e.g., matrix theory)
- Probability
- Optimization theory

Resources: Datasets

- UCI Repository: http://www.ics.uci.edu/~mlearn/MLRepository.html
- UCI KDD Archive: http://kdd.ics.uci.edu/summary.data.application.html
- Statlib: http://lib.stat.cmu.edu/
- Delve: http://www.cs.utoronto.ca/~delve

Resources: Journals

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- Neural Computation
- Neural Networks
- IEEE Trans on Neural Networks and Learning Systems
- IEEE Trans on Pattern Analysis and Machine Intelligence
- Journals on Statistics/Data Mining/Signal Processing/Natural Language Processing/Bioinformatics/...

Resources: Conferences

- International Conference on Machine Learning (ICML)
- European Conference on Machine Learning (ECML)
- Neural Information Processing Systems (NIPS)
- Uncertainty in Artificial Intelligence (UAI)
- Computational Learning Theory (COLT)
- ACM KDD
- International Conference on Artificial Neural Networks (ICANN)
- International Conference on AI & Statistics (AISTATS)
- International Conference on Pattern Recognition (ICPR)
- IEEE International Conference on Data Mining (ICDM)
- …

Reading and Homework Assignment

- Finish Reading Chapter 1
- Start Reading Chapter 2
- Do Homework 1