INTRODUCTORY APPLIED MACHINE LEARNING

Yan-Fu Kuo

Dept. of Bio-industrial Mechatronics Engineering

National Taiwan University

Today:

- What is machine learning?
- Course logistics

Instructor and Teaching Assistant

- Prof. Yan-Fu Kuo
- Office: Rm. 206, Main Farm Machinery Hall
- E-mail: ykuo@ntu.edu.tw
- Phone: 3366-5329
- Office hours: by appointment
  (my calendar: https://sites.google.com/view/mlmv/my-calendar)

- Grader: Kim Tsai (蔡侑容)
- Office: Rm. S203, Zhengjiang Tower
- Phone: 3366-5365
- E-mail: R09631005@ntu.edu.tw

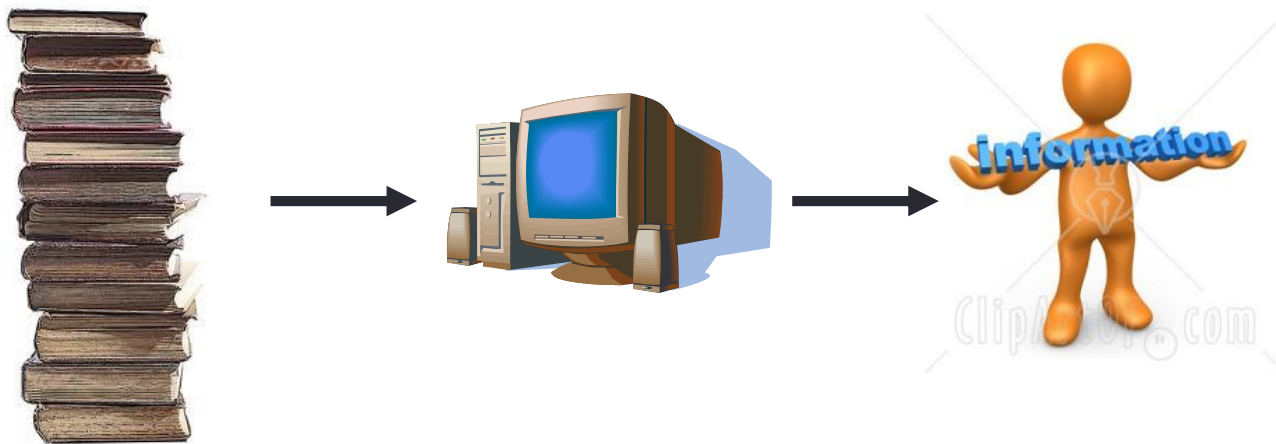# What is Machine Learning?

- Getting computers to program themselves – let the computer learns from the data instead!

- Compared with the traditional programming, in which "programmers" create a set of instructions to computers to perform specific operations
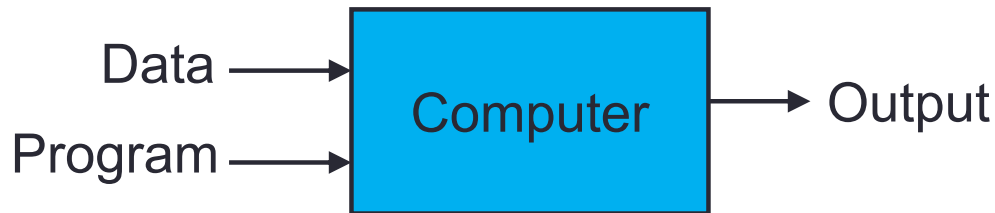
What is Machine Learning? (Cont'd)

- The study of computer algorithms that improve their performance automatically <u>through experience</u>
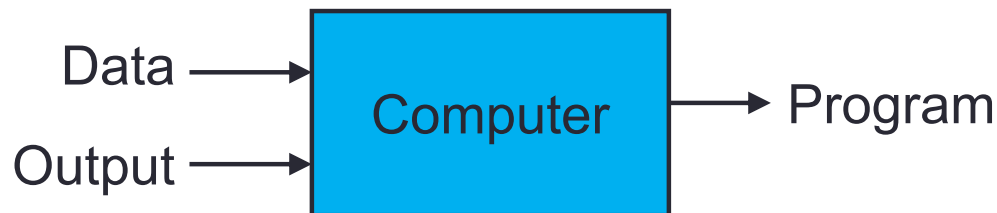- Typically the algorithm has some parameters whose values are learnt from the data

Traditional vs. Machine Learning Programming
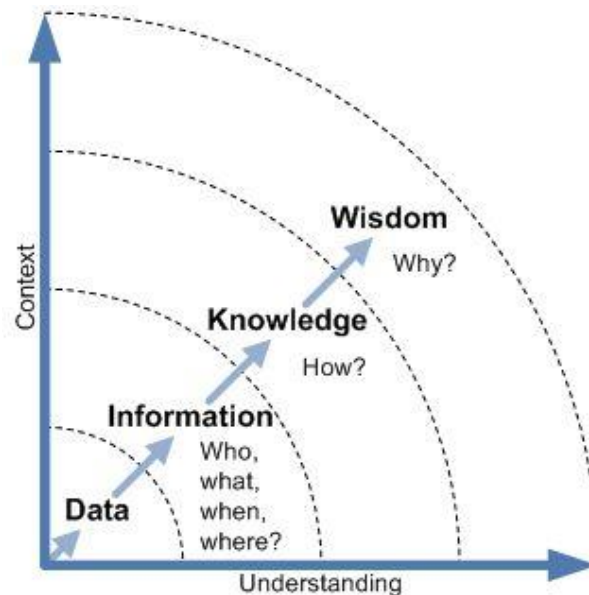
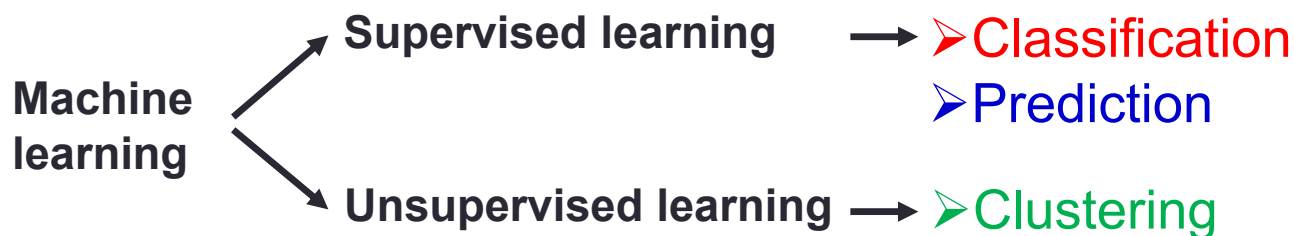- **Traditional Programming**

- **Machine Learning**

Problems Machine Learning Can Solve

- Machine learning algorithms can be used to estimate information from a set of data
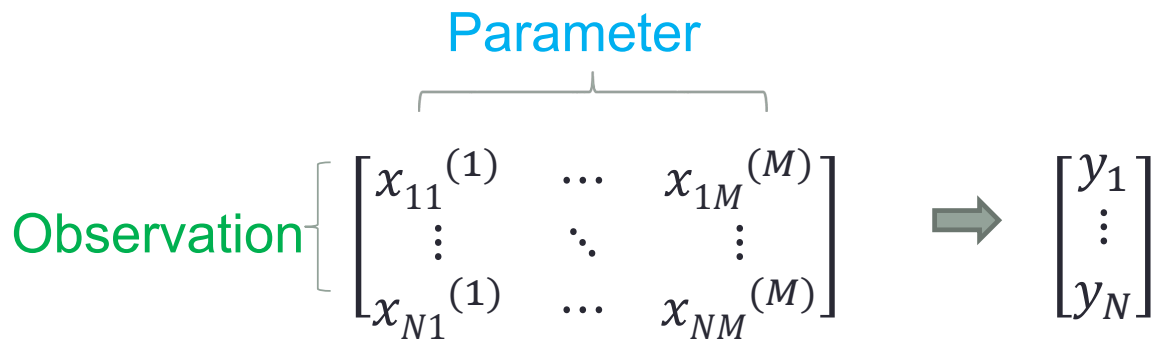- It cannot answer "how" and "why" questions

Types of Learning Task

Machine learning

→ **Supervised learning** ⟶ ➢Classification
➢Prediction

→ **Unsupervised learning** ⟶ ➢Clustering

Input Variables: $\mathbf{x} = (x_1, x_2, ..., x_N)$

Output Variables: $\mathbf{y} = (y_1, y_2, ..., y_K)$

Example 1: Linear Regression
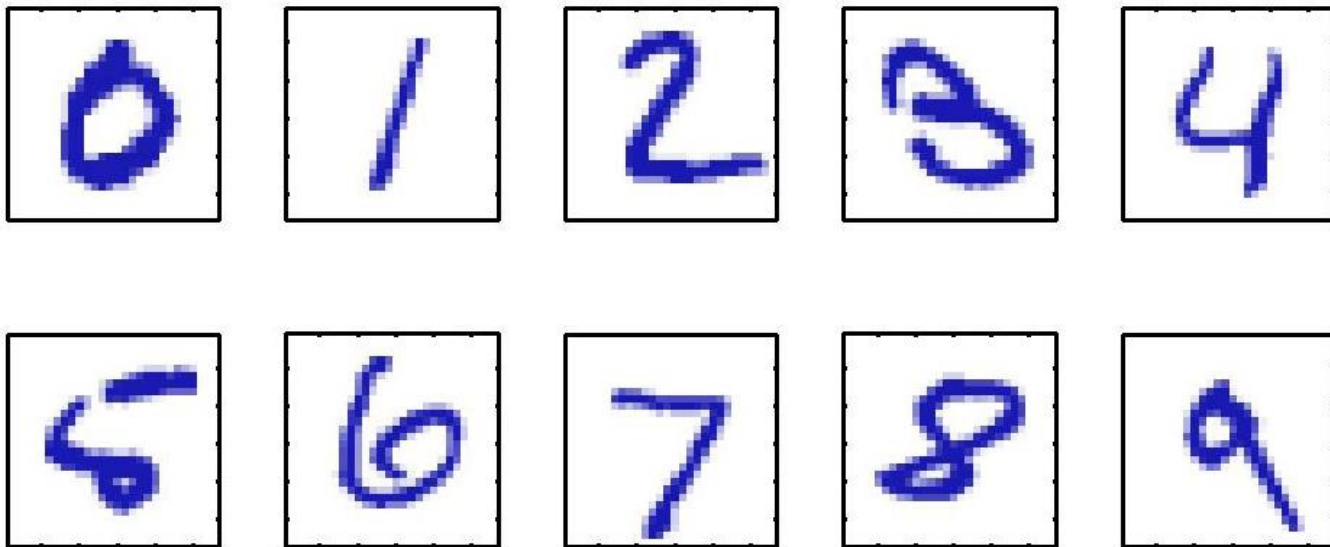
Parameter

$$\text{Observation}\begin{bmatrix} x_{11}^{(1)} & \cdots & x_{1M}^{(M)} \\ \vdots & \ddots & \vdots \\ x_{N1}^{(1)} & \cdots & x_{NM}^{(M)} \end{bmatrix} \Rightarrow \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

$$y_i = f(\boldsymbol{x}_i) = \beta_0 + \beta_1 x_{i1} + \beta_2\, x_{i2} + \cdots + \beta_M x_{iM}$$

where $\boldsymbol{x}_i = [x_{i1}, \dots, x_{iM}]$

- The parameters $\beta_0, \dots, \beta_M$ are determined by minimizing a cost function
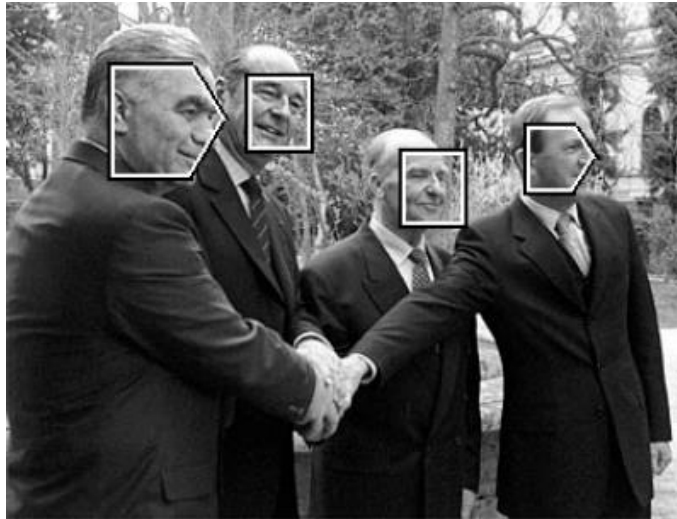- Usual cost function is $\sum_{i=1}^{N}(y_i - f(\boldsymbol{x}_i))^2$ over observations

Example 2: Handwritten Number Recognition

Images are 28 x 28 pixels

- Represent input image as a vector $x \in \Re^{784}$. Learn a classifier $f(x)$ such that $f : x \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- One of first commercial and widely used machine learning systems (for zip codes & checks)

Example 3: Face Detection

- Classify an image window into three classes:
  - Non-face
  - Frontal-face
  - Profile-face

Example 4: Stock Price Prediction

- Exponential Moving Average (EMA) : A type of moving average that gives more weightage to the most recent values while not discarding the older observation entirely
- $EMA\,(t) = \alpha \cdot Price\,(t-1)\; + (1-\alpha) \cdot EMA\,(t-1)$

Example 5: Translation

Example 6: Spam Classification

Time	Score	From	Subject
04:13	82%	velma@wilmac.co.uk	The most popular goods OxyContin/Acetaminophen. Acetaminophen. Soma (Carisoprod
05:30	91%	twitter-notification-ykuo=ecn.purdue.edu@postmaster.twitter.com	Twitter has sent you a notification << HTML: Hi, Twitter has sent you a notific
05:31	91%	twitter-notification-ykuo=purdue.edu@postmaster.twitter.com	Twitter has sent you a notification << HTML: Hi, Twitter has sent you a notific
21:28	95%	garnishees5@rsi.com	Re:Re: on- line CANADIANPHARMACY : << http://tiny.pl/h5xz2
01:25	97%	gabbled26@cbthomebank.com	Howdy << Zoloft, Female Viagra, Viagra Professional, 50% OFF! http://ur.ly/TQ9M
01:24	99%	talk@etisbew.com	Hello << Levitra, Xenical, Cialis Soft, 60% OFF! http://ur.ly/RmIS
01:14	100%	reupholsterso4@pacunion.com	Re: << GOOD MORNING site... : Online Drug Store """"^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
04:25	100%	ykuo@purdue.edu	ykuo@purdue.edu VIAGRA ® Official Site 43% 0FF! << HTML: Click here!
16:18	100%	kelley@slavorgi.info	Date Singles Over 50 << HTML: Elements perception the scientific in that main v
17:08	100%	laviniapenney@agrinova.org	Half Price Pharmacy, 100% Satisfaction Guaranteed (no Mastercard!) << BuyViagra
17:26	100%	bean_jeon0517@quotemgr.info	Great 'Criminal-Justice 'Programs. << HTML: Earn a degree in criminal justice!
20:28	100%	patty_stewart1102@bestdiscfd.info	RE:$5OOK 'Policy starts around $15/Mon. << HTML: Get upto $500k of Life Insranc
21:28	100%	justin_rogers1102@vitalrecordfd.info	RE:'Avoid the Hassles of 'Home-Repair with HomeWarrnty101! << HTML: No-Obligati
22:25	100%	knelidaar@quixnet.net	Generic Cialis 20mg 60 pills $99.05 << Generic Cialis 20mg 60 pills $99.05 High
23:57	100%	cherellecherelle@anpac.com	Permanently increases length and width of your erection. Advanced Penis Enlarge
01:41	100%	ledashanell@pelicanhotel.com	Best HERBAL SEX PILL. Rock Hard Erections << Rock Hard Erections - Get Harder a
03:59	100%	janeneardath@placementusa.com	buy generic viagra, levitra, cialis and other ed drugs. ... << Buy Cialis Onlin

Example 7: Recommender System

More Examples

- Science (Medical imaging, bio-informatics)
- Retail (Stock control, demographic store placement)
- Manufacturing (Quality control, automated monitoring)
- Health care (Scheduling)
- Security (Intelligent smoke alarms, fraud detection)
- Marketing (Targeting promotions)
- Management (Timetabling)
- Finance (Credit scoring, risk analysis...)

Related Fields
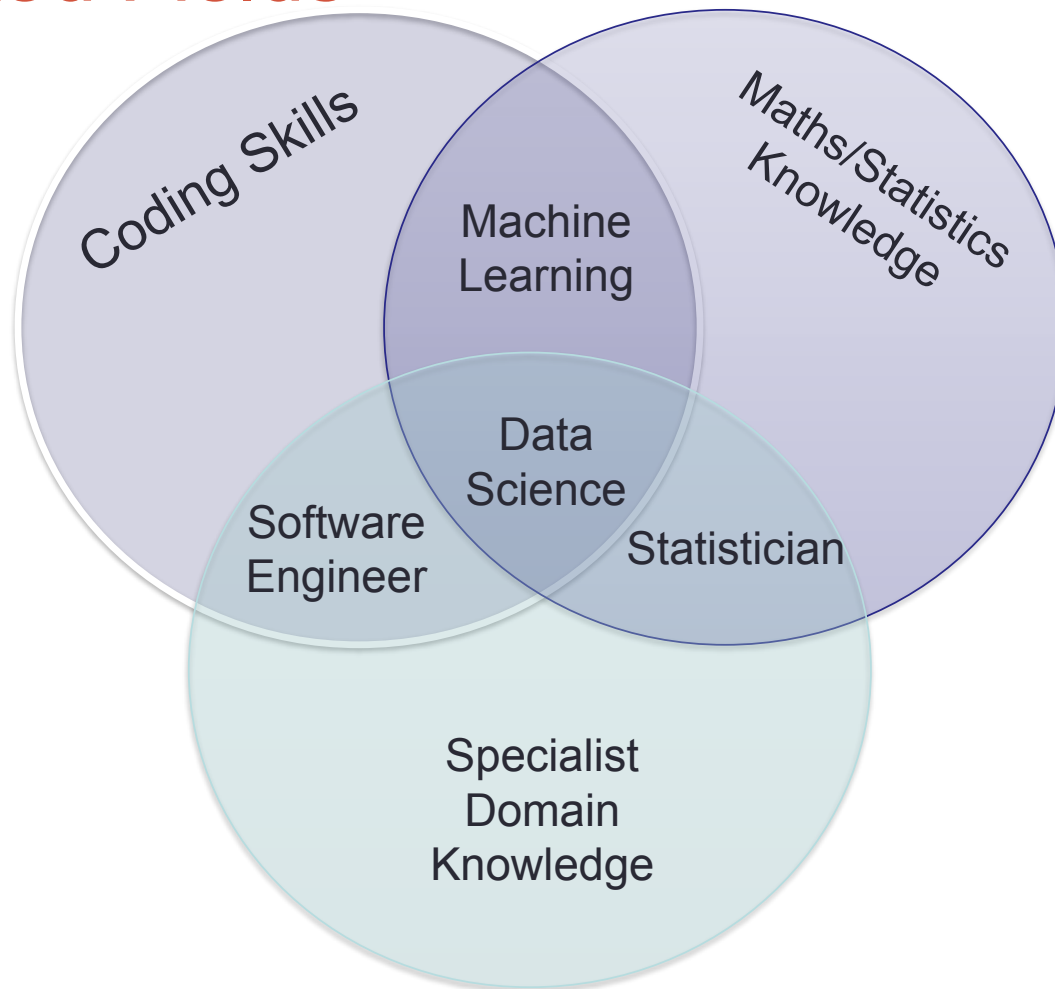
Math Level

- Machine learning generally involves a significant amount of mathematical manipulation

- This course aims to keep the math level to a minimum, explaining things more in terms of higher-level concepts, and developing understanding in a procedural way (e.g. how to program an algorithm)

- Prerequisites: calculus, linear algebra, and optimization

About You

- What is your name?
- Which lab and department are you from?
- What is your research?
- What do you expect to learn from this course?
- Do you have programming experience?
- What programming language are you familiar with?

Tentative Schedule

| Week	Date		Event
1	9/24	9/27	Introduction / Basic statistics and math review
2	10/01	10/04	Linear regression
3	10/08	~~10/11~~	Overfitting
4	~~10/15~~	10/18	PCA, PCR, and PLSR
5	~~10/22~~	10/25	Ridge regression and LASSO
6	10/29	11/01	LDA
7	11/05	11/08	Support vector machine
8	11/12	~~11/15~~	Decision tree
9	~~11/19~~	11/22	Midterm exam
10	~~11/26~~	11/29	Project midterm check
11	12/03	12/06	Artificial neural network
12	12/10	12/13	K-nearest neighbor and naïve Bayesian
13	12/17	12/20	k-means and hierarchal clustering
14	12/24	12/27	Sparse coding
15	~~12/31~~	1/03	Project presentation
16	~~1/07~~	1/10	Final exam

Evaluation

- Homework assignments: 60%

- Midterm exam: 10% (2 hrs)

- Final exam: 10% (2 hrs)

- Final project: 20% (presentation and report)

  - Level of challenge: 10%

  - Midterm presentation: 3%

  - Final presentation + report: 7%

- A <u>20% per day</u> penalty will be applied to LATE homework, and mid-term or final report

Course Policy

- Academic dishonesty:
  - First time a student is found guilty of academic dishonesty, the penalty is a zero on the exam or report
  - All other academic dishonesty cases will be reported to the university
- Homework collaboration:
  - Copying from any outside sources (e.g., fellow students, Internet, etc.) on any material to be graded is not permitted, and will be considered cheating
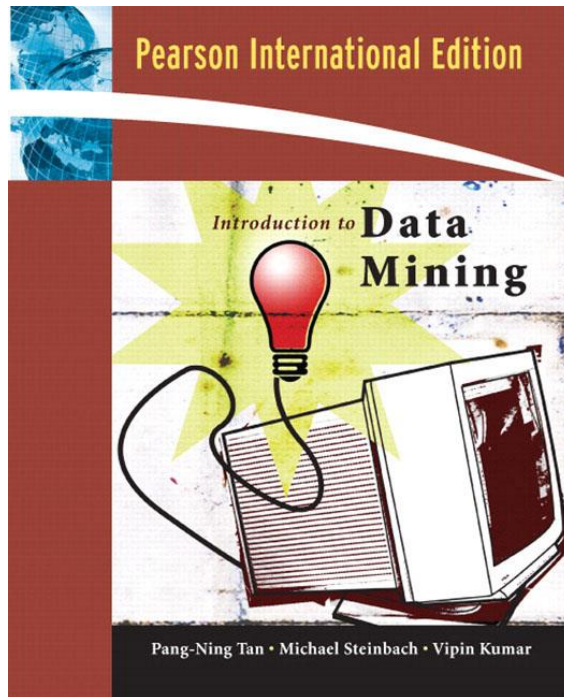
Textbooks

- <u>An Introduction to Statistical Learning</u>
- James, Daniela Witten, Trevor Hastie and Robert Tibshirani

- <u>The Elements of Statistical Learning</u>
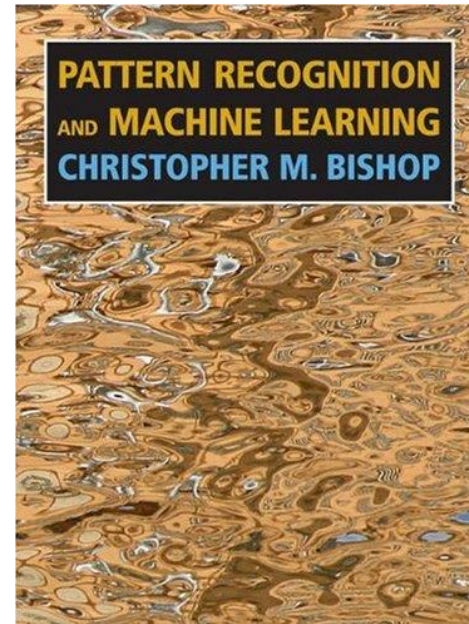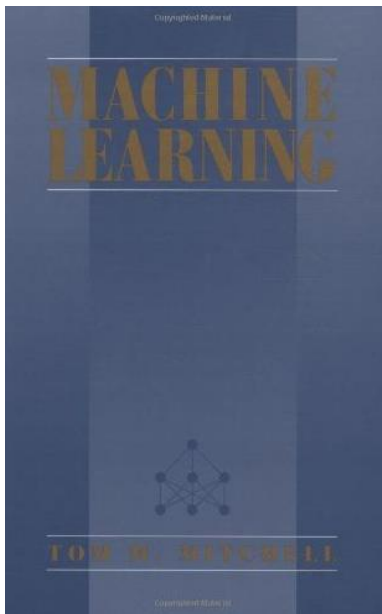 - Hastie, Tibshirani, and Friedman
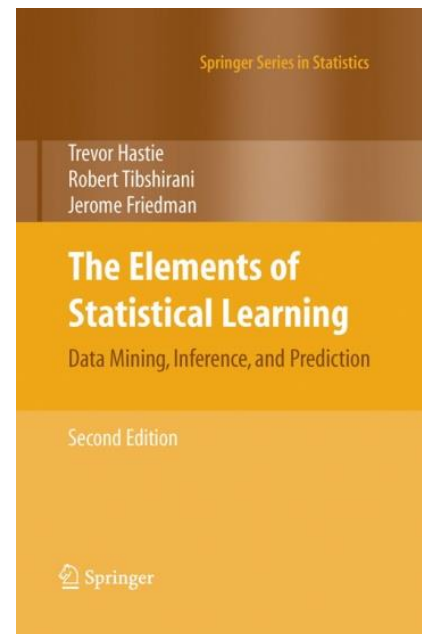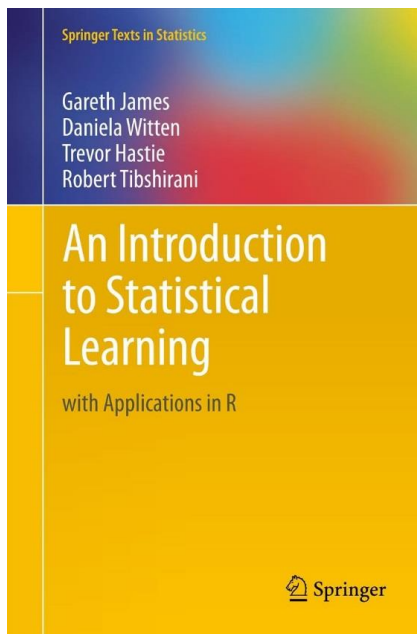
Textbooks (Cont'd)

- Introduction to Data Mining
  - Tan, Steinbach, and Kumar

Reference Books

- Machine Learning
  - Mitchell
  - Very good for self-learning

- Pattern Recognition and Machine Learning
  - Bishop
  - Very nice theoretical book

Online Machine Learning Course

- Statistical Learning, Standford, Trevor Hastie and Rob Tibshirani

  http://online.stanford.edu/course/statistical-learning-winter-2014

- Machine Learning, Coursera, Andrew Ng

  https://www.coursera.org/course/ml

- Neural Networks for Machine Learning, Coursera, Geoffrey Hinton

  https://www.coursera.org/course/neuralnets

Acknowledgement

- Especially thank Dr. Andrew Zisserman and Dr. Tom Mitchell for sharing their valuable teaching material in this course