Chapter 1 Introduction and Outline

AMA4680
Fu Guanxing
guanxing.fu@polyu.edu.hk

Schedule (1/2)

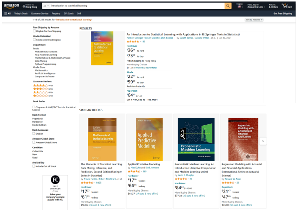
- Lectures
  - Time: 15:30-17:20, Thursday, by instructor
  - Location: QR403
- Tutorials
  - Time: 9:30-10:20, Tuesday, by Mr. ZHOU Xueyu
  - Location: N001
- 1 midterm exam (closed-book): 20%, around Week 8.
- 1 final exam (closed-book): 60 %
- 2 assignments: 10% for each, 20% in total.
- Grade cut-off:
  - A: ≥ 85.0
  - B: ≥ 65.0
  - C: ≥ 50.0
  - D: ≥ 40.0

Schedule (2/2)

- Learning materials (in Blackboard):
  - Slides
  - Textbooks: *An Introduction to Statistical Learning* James, G., Witten, D., Hastie, T., and Tibshirani, R.
    The Elements of Statistical Learning Hastie, T., Tibshirani, R, and Friedman, J.
    ...
  - Videos
- Content: theory of classical machine learning algorithms + python programming.

Introduction of machine learning

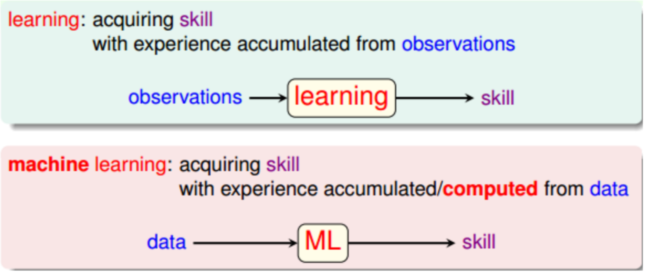
- Tricks of goods display in supermarketsi:
  - high-interest sale items at the level of customers' sight.
  - Children's products and female products are close to each other.
  - ...
- How about online shopping center?



- The technique to let computer **imitate the way that human beings learn** is called machine learning (ML).

What is ML (2/3)

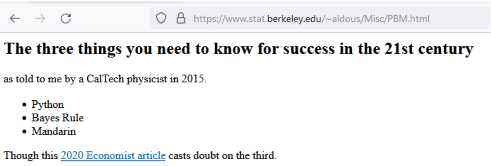
- Human learning vs machine learning:



- How does ML work? Learning from **similarity**!
  - Customers buying *Introduction to Statistical Learning* also buy...(Key words, The same authors...)
  - Then...will be put into the recommendation list by machine!

What is ML (3/3)

- ML is not mysterious.
- People from various field can step into this area.
- A combination of probability, statistics, optimization ... and programming (**Python**, R, Matlab...).
- Why python?
  - Easy to learn.
  - Great library (e.g. Numpy, Matplotlib, **Scikit-learn**, Tensorflow, **PyTorch**...).
  - Growing popularity: most commonly used, and choice of many large incorporations (Google, Facebook, Amazon...)


https://www.stat.berkeley.edu/~aldous/Misc/PBM.html

The three things you need to know for success in the 21st century

as told to me by a CalTech physicist in 2015.

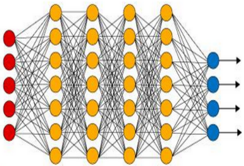
- Python
- Bayes Rule
- Mandarin

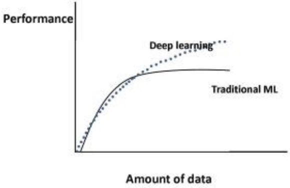
Though this 2020 Economist article casts doubt on the third.

Figure: Webpage of David Aldous.

Machine learning vs Deep learning

- DL is part of ML, based on artificial neural networks.
- In comparison with traditional ML,
  - performance of DL relies more on data.
  - DL depends more on hardwares to do calucations of matrices.
  - DL solves problems from beginning to end directly.
  - DL algorithm is often unexplained.
  - ...





ML language (1/2)

- In data collection, for example, Amazon wants to predict whether a given book is interesting to some specific customer or not, some information of books is needed:

  (Author=James,Topic=ML,Publisher=Springer),

  (Author=Witten,Topic=ML,Publisher=Wiley),

  ...

  Each piece of information is called a **sample** or **instance**.
- All samples together are called a **dataset**.
- The process from data to model is called **learning** or **training**. The corresponding data is called **training data**.
- In each sample, "Author", "Topic" and "Publisher" are called **attribute** or **feature**.
- Generally, a dataset with m samples, each sample having n features, is written as $D = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$, where $\boldsymbol{x}_i = (x_{1i}, \cdots, x_{ni})' \in \mathcal{X} \subset \mathbb{R}^n$. Here, $\mathcal{X}$ is called the **sample space**, n is called the **dimensionality** of the sample.

ML language (2/2)

- However, in order to do prediction, we also need the "interest" information for each sample:

  ((Author=James,Topic=ML,Publisher=Springer), yes),

  ((Author=Witten,Topic=ML,Publisher=Wiley), no),

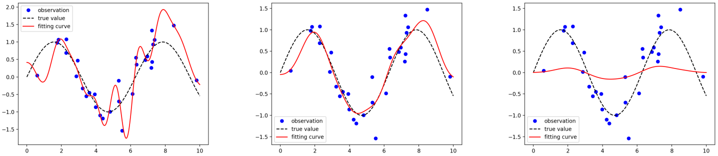
  ...

  Here, "yes" and "no" are called **label**. A sample equipped with a label is called an *example*.
- Generally, use $(\boldsymbol{x}_i, y_i)$ to denote the i-th example, where $y_i \in \mathcal{Y}$ and $\mathcal{Y}$ is called **label space**.

Goal of ML

- **Generalization** refers to the model/machine/computer's ability to adapt to *new data*, (previously unseen data).
- A model's ability to generalize is central to the success of the model.
- The goal of ML is to find a model (by training data) with better generalization.
- Generalization, overfitting and underfitting.
  - If a model is trained too well on training data, it would be unable to generalize: the prediction of new data is inaccurate, even though it makes perfect prediction of training data. This is called overfitting.
  - Underfitting happens if the model is not well trained enough. A model with underfitting is as useless as a model with overfitting; it even does not have accurate prediction of training data.

- Which model has the best generalization?



- Left:

  Middle:

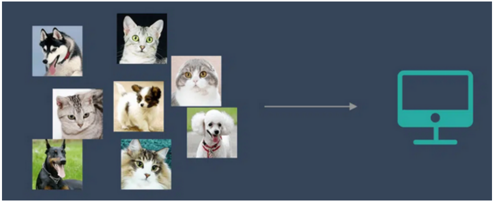
  Right:

- How to avoid overfitting and underfitting: regularization.

Categories of ML (1/2)

- Supervised learning: training data have labels.

- Unsupervised learning: training data have no labels. Computers find potential rule of dataset and divide samples into different groups.



- Reinforcement learning: an agent interacts with unknown environment, takes actions, and learns by a trial-and-error method.

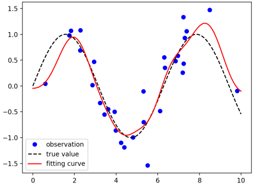
Categories of ML (2/2)

Criteria	Supervised learning	Unsupervised learning	Reinforcement learning
Define	The machine learns by labeled data	The machine learns by unlabeled data without any guidance	An agent interacts with environment, takes actions, to maximize his reward by trial and error
Type of problems	regression and classification	clustering and dimension reduction	stochastic control
Type of data	labeled	unlabeled	no predefined data
Training	external supervision	no supervision	no supervision
Approach	Maps labeled inputs to known outputs	Understand patterns, discover the output	Trial and error

Topics in this course

Supervised learning

- Regression: linear regression, LASSO, **(kernel) ridge regression**.



- Dataset $\{\boldsymbol{x}_i, y_i\}_{i=1}^m$, where y_i is **continuous-valued**.
- The regression task is to find some function $f\colon \mathcal{X} \to \mathcal{Y}$ such that the predicted output of new sample $\boldsymbol{x}$ is close enough to its true value.

- Classification: Decision tree, naive Bayes classifier, artificial neural network, support vector machine, logistic regression.

- Dataset $\{\boldsymbol{x}_i, y_i\}_{i=1}^m$, where y_i is **discrete-valued**. e.g. $\mathcal{Y} = \{-1, +1\}$.
- The classification task is to find some function $f\colon \mathcal{X} \to \mathcal{Y}$ such that the predicted label of new sample $\boldsymbol{x}$ has a large confidence to be the right one.

Unsupervised learning

- Dimensionality reduce.
  - From high-dimensional data to low-dimensional data, so that the meaningful properties of the original data is not lost.
  - PCA; refer to AMA4602.
- Clustering: K-means, Agglomerative clustering, GMM clustering...
  - Given a set of data points, each having a set of features, and a similarity measure among them. Find clusters such that
    - Data points in one cluster are more similar to each other;
    - Data points in different clusters are less similar to each other;
  - Some similarity measures:
    - Euclidean distance $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$.
    - Hamming distance: $d(\mathbf{x}, \mathbf{y}) = \#\{i : x_i \neq y_i\}$, the number of components of $\mathbf{x}$ and $\mathbf{y}$ that are not equal.

Reinforcement learning

- Known environment: Markov decision process. An MDP is specified by
  - A set of states $X = \{1, 2, \cdots, n\}$
  - A set of actions $A = \{1, 2, \cdots, m\}$
  - Transition probabilities

$$P(x'|x,a) = \textsf{Prob}(\text{Next state } = x'|\text{Action } a \text{ in state } x)$$

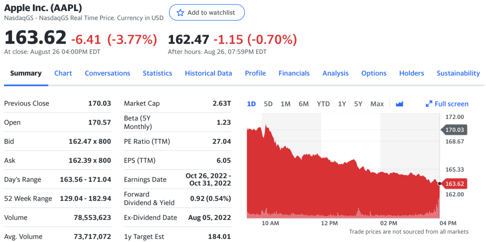
  - A reward function $r(x, a)$.

  The goal of the agent is to maximize the accumulative reward $\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(x_t, a_t)\right]$ by choosing $\{a_t\}$.
- Unknown environment: tradeoff between exploitation and exploration

$$\pi_\epsilon(x) = \begin{cases} \text{optimal action with probability } 1-\epsilon, \\ \text{uniformly selected action in } A \text{ with probability } \epsilon. \end{cases}$$

Further applications of ML

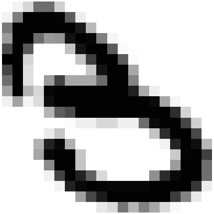
- Stock price prediction.



- A **regression** problem.

Selected applications of ML, *cont'd*

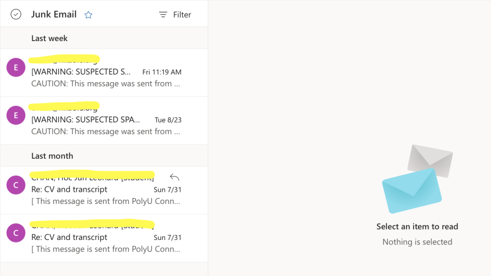
- Image recognition.



- A (multi-class) **classification** problem.

Selected applications of ML, *cont'd*

- Junk email filter.



- A **clustering** problem.

Selected applications of ML, *cont'd*

- Game design (AlphaGo)

- Reinforcement learning.

Learning objective of Chapter 1

1. Understand the goal of ML: generalization.
2. Overfitting vs. underfitting.
3. Understand the difference between supervised learning, unsupervised learning and reinforcement learning.

Requirement of this course

- Probability
- Statistics
- Linear algebra
- Analysis
- ...