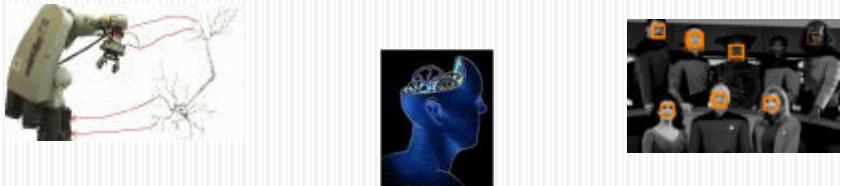




## Welcome to

### *Introduction to Machine Learning!*



1 Introduction to Machine Learning: Introduction

## Course Instructors

**Teacher 1:** ZHANG Min 张敏

- Associate Professor, Information Retrieval Group, DCST
- Tel: 62792595
- Office: FIT 1-506
- Email: [z-m@tsinghua.edu.cn](mailto:z-m@tsinghua.edu.cn) HP: <http://www.thuir.cn/group/~mzhang>

**Teacher 2:** HU Xiaolin 胡晓林

- Associate Professor, Intelligent Multimedia Group, DCST
- Tel: 62795869
- Office: FIT 1-509
- Email: [xlhu@tsinghua.edu.cn](mailto:xlhu@tsinghua.edu.cn)

**TA:** ZHANG Fan 张帆

- Tel: 18671829106 Addr. : FIT 1-506
- Email: [frankyzf94@gmail.com](mailto:frankyzf94@gmail.com)



2 Introduction to Machine Learning: Introduction

## A little bit about Min...



- Experiences
  - Achieved Bachelor and PhD in 1999 and 2003 respectively in DCST, Tsinghua Uni.
  - Now Associate Professor in THUIR group, CST Dept.
  - Visiting scholar/researcher in DFKI (Germany), Kyoto University, NUS, City University of HongKong, and MSRA
- Research Interests:
  - Information Retrieval & Recommendation, user behavior analysis, data mining, machine learning.
- Achievements and Awards
  - Published over 100 papers on important international conferences and journals on related fields (e.g. IR, AI, DM, NLP), including
    - SIGIR, IJCAI, WWW, CIKM, WSDM, JIR, JASIST, ...
  - Multiple Top Performances in TREC and NTCIR during 15 years.
  - Beijing Science and Technology Progress Award, 1<sup>st</sup> prize, 2015
  - State of Ministrition of Radio, Film and Television Science and Technology Innovation Award, 2<sup>nd</sup> prize
  - Guangxi Science and Technology Progress Award, 3<sup>rd</sup> prize
  - Excellent Young Faculty Teaching Award, Tsinghua University

3

Information Retrieval: Introduction

## A little bit about Min...



- Academic Activities and Industry Connections
  - Chief Director of the state key lab of intelligent technology and systems ([AI institute](#))
  - Chief Director of [Tsinghua – MSRA](#) joint research lab on Web and multimedia
  - Associate Editor of Transaction on Information Systems ([TOIS](#)) (CCFA journal) and reviewers for multiple top journals
  - Program Chair of WSDM 2017, AIRS 2016
  - Area Chair, Senior PC or PC for
    - IJCAI, SIGIR, WWW, KDD, WSDM, CIKM, ACL.....
  - Committee members of China Associations
    - Chinese Information Processing Committee (CCF)
    - Information Retrieval Committee (CIPS)
    - Machine Learning Committee (CAAI)
- [The first one](#) to construct Machine Learning course in DCST, THU (since 2003)

4

Information Retrieval: Introduction

# Course Information

- Key points
  - Basic concepts
  - Fundamental learning theory
  - Classical / Important algorithms
    - Basic idea
    - Problem definition
    - Algorithm design and analysis
    - Future work
  - Analysis (problem, features, results)
  - Ability to solve practical problems

5

Introduction to Machine Learning: Introduction

# Study materials

- Slides and your notes
- Main reference (**partially** as textbook) :
  - Tom Mitchell, Machine Learning, McGraw-Hill (China machine press, English version)
  - 中文版：曾华军，张银奎等译，机器学习，机械工业出版社
- Other good references
  - Introduction to machine learning (Ethem Alpaydin)  
机器学习导论(范明等译, 机械工业出版社)
  - Pattern classification (Richard O. Duda, etc)  
模式分类(李宏东等译, 机械工业出版社)
  - 机器学习 (周志华, 清华大学出版社)



6

Introduction to Machine Learning: Introduction

# Topics

- **Introduction**
- **Decision Tree learning**
- **Bayesian learning (Naïve Bayes)**
- **Hidden Markov Model**
- **Instance-based learning: KNN, distance-weighted NN**
- **SVM and kernel-based learning**
- **Unsupervised learning, clustering**
- **Probabilistic graphical models (brief introduction)**
- **Ensemble learning (weighted majority, bagging, Boosting)**
- **Parameter estimation: EM (optional)**
- **Learning theory (1- Learning bias, evaluation, 2- PAC, mistake bound)**
- **Deep Learning (brief overview)**
- **Transfer Learning (overview, optional)**
- **Brainstorming – invited lecture (optional)**
- **Overview of machine learning**

7

Introduction to Machine Learning: Introduction

## Not involve (due to course redundancy, time or/and difficulty issues)

- Neural Network **Basics** (NN)
  - Have been introduced in AI course
  - Another course in Postgraduate course
  - But the advanced NN will be introduced in brief overview of Deep Learning section of this course
- Evolutionary Computation
  - Genetic Algorithms in AI courses (undergraduate or graduate)
- **Insights** of Probabilistic Graphical Models
  - Better for graduate students
  - But the overview and basic concepts will be introduced in this course
- Active learning, Semi-supervised learning
  - Better for graduate students

8

Introduction to Machine Learning: Introduction

## Coffee Time



- A short but **important** part of this course
- Application topics
  - Problem understanding and design
  - Features selection
  - Analysis and discussion
- And ...
  - Good books / Big idea / Gurus
  - Fancy new techniques

9

Introduction to Machine Learning: Introduction

## Grading

### (Subject to modifications)

- 3 Homework assignments ( $\sim 20\%$ )
- 2 Experiments ( $\sim 40\%$ )
- Project and Examination (choose one of the following) ( $\sim 40\%$ )
  - Final exam
  - Or Project (no predefined topic)
- Course activities (**bonus**)

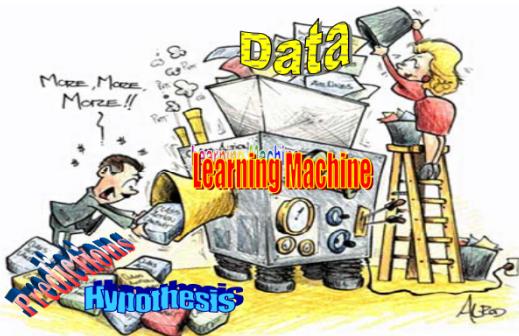
10

Introduction to Machine Learning: Introduction

# Topic 1. Introduction

11

Introduction to Machine Learning: Introduction



Application background of machine learning

12

Introduction to Machine Learning: Introduction

## App. background for machine learning

- Data mining
  - Using historical data to improve decisions
  - .....

13

Introduction to Machine Learning: Introduction

## App. background : Data Mining (1)

- Business Intelligence
- Example 1 Supermarket



+



?



+



- Example 2: location (P&G products, Wal-mart, ...)



: Introduction



14

Int

## App. background : Data Mining (2)

- Credit risk analysis

Customer 103 (time = t0)	Customer 103 (time = t1)	Customer 103 (time = tn)
Years of credit: 9	Years of credit: 9	Years of credit: 9
Loan balance: \$2,400	Loan balance: \$3,250	Loan balance: \$4,500
Income: \$52K	Income: ?	Income: ?
Own House: Yes	Own House: Yes	Own House: Yes
Delinquent accts: 2	Delinquent accts: 2	Delinquent accts: 3
Max billing cycles late: 3	Max billing cycles late: 4	Max billing cycles late: 6
<b>Profitable customer? :</b>	<b>Profitable customer? :</b>	<b>Profitable customer? :</b>
?	?	NO

15

Introduction to Machine Learning: Introduction

## App. background for machine learning

- Data mining
  - Using historical data to improve decisions
- **Self customizing programs**
  - Email filtering (and organization)
  - Applications that learn user interests (e.g. news readers, forum, social community ...)
  - Recommendation in e-commerce
  - .....



16

Introduction to Machine Learning: Introduction

## App. background for machine learning

- Data mining
  - Using historical data to improve decisions
- **Self customizing programs**
  - Email filtering (and organization)
  - Applications that learn user interests (e.g. news readers, forum, social community ...)
  - Recommendation in e-commerce
  - .....



17

Introduction to Machine Learning: Introduction

## App. Background: Learning users interests

提示		lowy	Feb 9	● 北京地区影院观影指南（影院信息介绍）101026
提示		lowy	Mar 2	● 北京地区影院观影指南（影院分布及优惠）101026
提示		tenniszbl	Dec 21	● 北京地区IMAX影院及IMAX／3D入门贴 1.15更新
提示		iame	May 18	● <a href="#">Movie</a> 版简明FAQ
提示		deliver	Jan 19	● [通知]《将爱》接受打分

[刷新] [第一页] [上一页] [下一页] [最后一页] 跳转到 第  篇/共 15826 篇

**movie** 来这个版的人常去的其他版面: [[NetResources](#)] [[TV](#)] [[Picture](#)] [[Virus](#)] [[Graduate](#)] [[PieLove](#)] [[Career](#)]

**MilitaryMovie** [[进版画面](#)] [[版内投票](#)] [[发文模板](#)]

来这个版的人常去的其他版面: [[MilitaryTech](#)] [[Taiwan](#)] [[NetResources](#)] [[Movie](#)] [[History](#)] [[HistoryClub](#)] [[Aero](#)]

**MovieClub** [[进版画面](#)] [[版内投票](#)] [[发文模板](#)]

来这个版的人常去的其他版面: [[Movie](#)] [[NetResources](#)] [[NR\\_Request](#)] [[LordOfTheRings](#)] [[DVDZone](#)] [[TV](#)] [[PopMusic](#)]

提示		jigod	Apr 3	● 在 TeX 版如何提问
提示		hitluckyfox	Mar 15	● 本版 ftp
提示		LittleLeo	Mar 26	● 清华大学学位论文模板

[刷新] [第一页] [上一页] [下一页] [最后一页] 跳转到 第  篇/共 1759 篇

**tex** 来这个版的人常去的其他版面: [[LinuxApp](#)] [[Emacs](#)] [[NetResources](#)] [[MathTools](#)] [[Virus](#)] [[NewSoftware](#)] [[LinuxDev](#)]

18 Introduction to Machine Learning: Introduction

## App. background for machine learning

- Data mining
  - Using historical data to improve decisions
- **Self customizing programs**
  - Email filtering (and organization)
  - Applications that learn user interests (e.g. news readers, forum, social community ...)
  - Recommendation in e-commerce
  - .....



19

Introduction to Machine Learning: Introduction

## App. Background: Recommendation in Social Networks

**猜你喜欢**  
在这里，我们会根据你的爱好标签、公司/学校资料等为你推荐你最有可能感兴趣的人；你也可以敲敲键盘，用搜索框找到更多感兴趣的人。  
----这里的人也许每天都在变，每天都期待你来看一看： )

**特别推荐** | **相同标签** | **同一学校** | **同一公司** | **可能在我附近** | **人气热门**

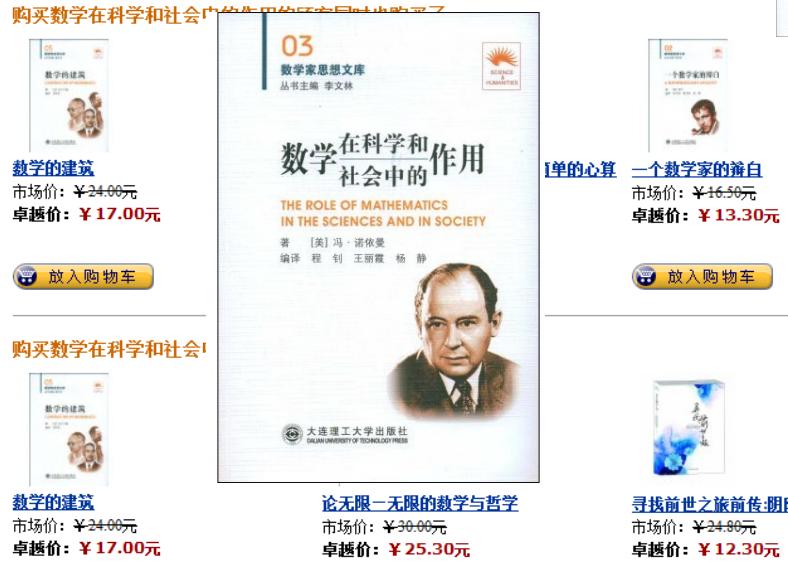
**某 id** | **北京** | 粉丝**431人** | **+加关注**

我的好友中: [aaa](#), [bbb](#), [ccc](#), [ddd](#), [eee](#), [fff](#), [ggg](#)/等12人也与他互相关注  
我关注的人中: [甲](#), [乙](#), [丙](#), [丁](#), [戊](#), [己](#), [庚](#), [辛](#) 等  
15人也关注了他

20

Introduction to Machine Learning: Introduction

## App Background: e-commerce recommendation



21

放入购物车

放入购物车

简单的心算 一个数学家的美白  
市场价格: ¥16.50元  
卓越价: ¥13.30元

放入购物车

寻找前世之旅前传: 阴阳师物语  
市场价格: ¥24.00元  
卓越价: ¥12.30元

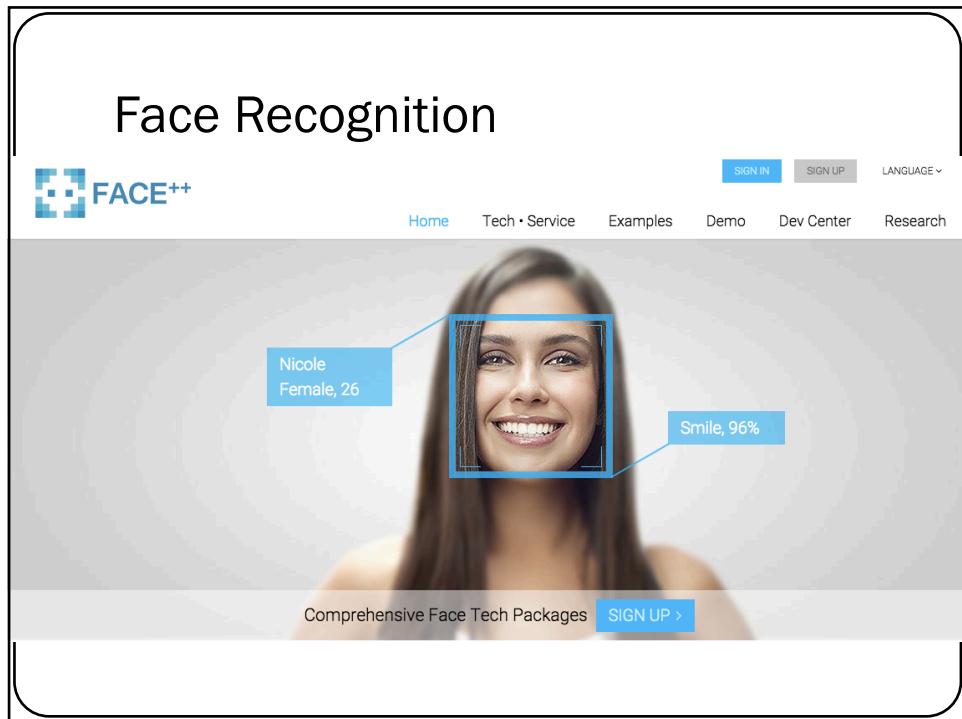
放入购物车

## App. background for machine learning

- Data mining
  - Using historical data to improve decisions
- Self customizing programs
  - Email filtering (and organization)
  - Applications that learn user interests  
(e.g. news readers, forum, social community ...)
  - Recommendation in e-commerce
  - .....
- Software applications we can't program by hand
  - Face/Speech/Handwritten recognition (Pattern Recognition)
  - Autonomous driving
  - Ranking in information retrieval (Search Engine)
  - .....

22

Introduction to Machine Learning: Introduction



**Speech recognition**

**Handwritten recognition**

<http://yann.lecun.com/exdb/lenet/index.html>

Introduction to Machine Learning: Introduction

24

## App. background for machine learning

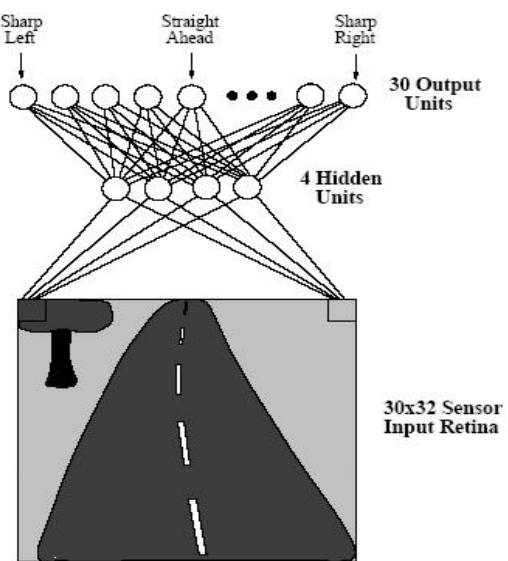
- Data mining
  - Using historical data to improve decisions
- Self customizing programs
  - Email filtering (and organization)
  - Applications that learns user interests  
(e.g. news readers, forum, social community ...)
  - Recommendation in e-commerce
  - .....
- Software applications we can't program by hand
  - Face/Speech/Handwritten recognition (Pattern Recognition)
  - Autonomous driving
  - Ranking in information retrieval (Search Engine)
  - .....

25

Introduction to Machine Learning: Introduction

## Example: Autonomous driving (CMU)

- ALVINN (1989~1996, CMU) 70mph, highway



26

Introduction to Machine Learning: Introduction

## Example: Autonomous driving (THU)



## Example: Autonomous driving: Google Driverless Car



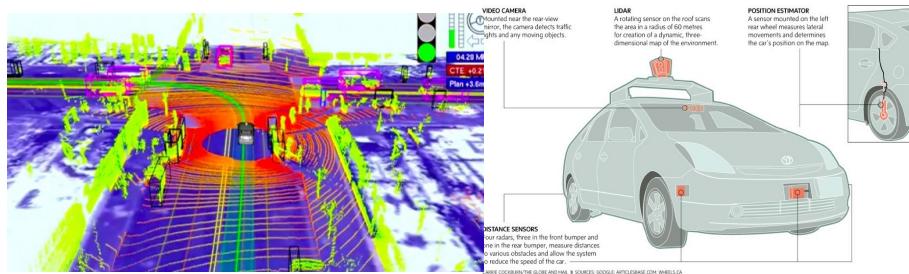
- 2009 Google's self-driving cars project
- Dec. 2016 Google transitioned the project into a new company called *Waymo*
- Plans to make self-driving cars available to the public in 2020



28

Introduction to Machine Learning: Introduction

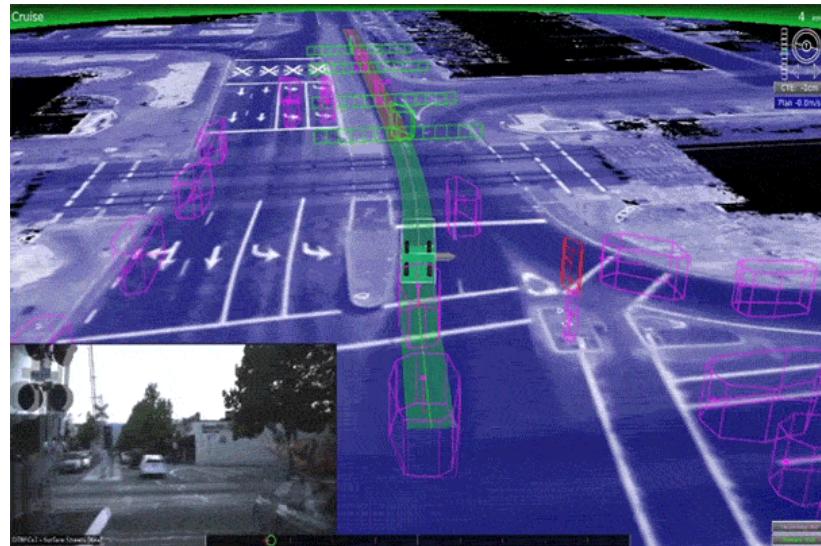
## Google driverless car



The system combines information gathered from Google Street View with artificial intelligence software that combines input from **video cameras** inside the car, a **LIDAR sensor (Velodyne 64-beam laser HDL-64E)** on top of the vehicle, **radar sensors** on the front of the vehicle and a **position sensor** attached to one of the rear wheels that helps locate the car's position on the map

### Self-driving cars to create 1GB of data per second

How Google's Self-Driving Car Works <http://www.tudou.com/programs/view/Hez52EEZQuE/>  
<http://www.fiercebigdata.com/story/self-driving-cars-create-1gb-data-second/2013-07-22>



<http://www.extremetech.com/extreme/189486-how-googles-self-driving-cars-detect-and-avoid-obstacles>

## App. background for machine learning

- Data mining
  - Using historical data to improve decisions
- Self customizing programs
  - Newsreader that learns user interests
  - Email filtering
  - Recommendation in e-commerce
  - .....
- Software applications we can't program by hand
  - Speech recognition/OCR (Pattern Recognition)
  - Autonomous driving
  - Ranking in information retrieval (Search Engine)
  - .....

31

Introduction to Machine Learning: Introduction

## Example: information retrieval



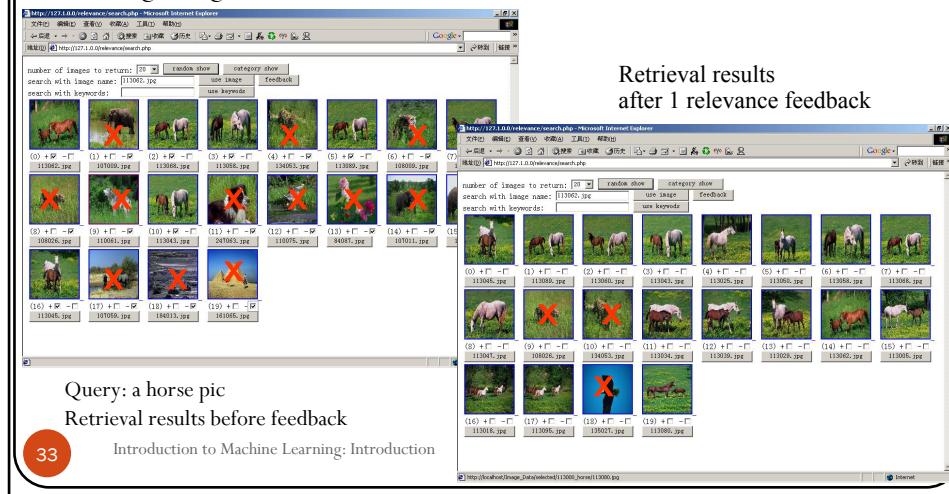
- Ranking: ~1000 parameters
- Learning to rank (new research branch)

32

Introduction to Machine Learning: Introduction

## Example: information retrieval (cont.)

- Relevance feedback
  - e.g. Image retrieval, video retrieval, .....



## App. background for machine learning

- Data mining
  - Using historical data to improve decisions
- Self customizing programs
  - Newsreader that learns user interests
  - Email filtering
  - Recommendation in e-commerce
  - .....
- Software applications we can't program by hand
  - Speech recognition/OCR (Pattern Recognition)
  - Autonomous driving
  - Ranking in information retrieval (Search Engine)
  - .....

Now you should have had ideas on  
the differences between **machine  
learning** and **data mining/pattern  
recognition/neural networks/...**

34 Introduction to Machine Learning: Introduction



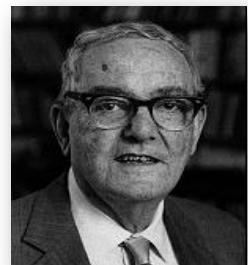
## What's machine learning ?

35

Introduction to Machine Learning: Introduction

## What's machine learning (1)

- "Learning denotes **changes** in a system that ...  
enable a system to **do the same task more  
efficiently** the next time." -- Herbert Simon
- Herbert Simon (1916 – 2001)
  - Won the **A. M. Turing Award** for his work on  
**computer science** in 1975 (One of the founders of AI)
  - Won the **Nobel Prize** in **economics** in 1978
  - Won the **National Medal of Science** in 1986
  - In 1993, he was awarded **the American Psychological Association's award** for outstanding  
lifetime contributions to **psychology**.
  - In 1994, he became one of foreign scientists inducted into the Chinese Academy of  
Sciences



36

Introduction to Machine Learning: Introduction

## What's machine learning (2)

- "Learning is **constructing** or **modifying** **representations** of what is **being experienced**."  
--Ryszard S. Michalski
- Ryszard S. Michalski (1937-2007)  
<http://www.mli.gmu.edu/michalski/>
  - Michalski, Ryszard S. and Kodratoff, Y. Machine Learning, an AI approach 1990
  - Co-founder: Machine learning research field
  - Co-founder: Machine Learning (Journal)
  - Co-founder: ICML
  - Founder: conceptual clustering, constructive induction, variable-valued logic, variable-precision logic



Introduction to Machine Learning: Introduction

37

## What's machine learning (3)

- Learning = **improving** with **experience** at some **task**
- Tom M. Mitchell (CMU)
  - 1973 MIT S.B. ; 1979 Ph.D. Stanford Uni.
  - Chair: AAAI
  - Co-founder: Machine Learning (Journal)
  - Co-founder : ICML
  - IJCAI Computers and Thought Award, 1983
- T (Task)
- E (Experience)      Learning:  
change / construct or modify / improve
- P (Performance)



Introduction to Machine Learning: Introduction

38

## What's machine learning (examples)

- Learn to play checkers

- T: play checkers
- P : % of games won
- E : e.g. play against self



- Handwriting recognition

- T: recognition characters
- P: classification precision
- E: character sets with known class



39

Introduction to Machine Learning: Introduction



A general machine learning  
system design

40

Introduction to Machine Learning: Introduction

## Designing a learning system e.g. Learning to play checkers

- What experience?

41

Introduction to Machine Learning: Introduction

## System design – Training Experience

- Is training experience **representative** of performance goal?
  - Will the data (games) seen in training represents those seen in practice?
  - play against self vs. play against a master
- Unexpected results
  - Example 1: moon or banana?
  - Example 2: digits recognition with NN
  - Example 3: IBM Watson's answer on the question of "an American city" at Jeopardy – Toronto?
  - *What's wrong?*
  - Due to **training data bias: data, training procedure, features ...**



42

Introduction to Machine Learning: Introduction

## Designing a learning system e.g. Learning to play checkers

- What experience?
  - Be aware of [training data bias: data, training procedure, features](#)
- [What exactly should be learned?](#)

43

Introduction to Machine Learning: Introduction

## What exactly should be learned?

- Going to learn a [Target Function](#) (target concept)
- But of what type? i.e. what is the [concept class](#)?
- Possible definition for Target Function  $V(b)$ 
  - If  $b$  is a [final](#) board state that is [won](#), then  $V(b) = 100$
  - If  $b$  is a [final](#) board state that is [lost](#), then  $V(b) = -100$
  - If  $b$  is a [final](#) board state that is [drawn](#), then  $V(b) = 0$
  - If  $b$  is [not a final](#) board state, then  $V(b) = V(b')$ 
    - Where  $b'$  is **the best final board state** that can be achieved starting from  $b$  and playing optimally until the end of the game.
- Gives correct results, but is [not operational](#) → Not efficiently computable
- Typically limit ourselves to an [approximation](#)  $\hat{V}$  of  $V$ , a.k.a. [Hypothesis](#)
  - e.g.  $V(b) \leftarrow \hat{V}(\text{Successor}(b))$

44

Introduction to Machine Learning: Introduction

## Designing a learning system e.g. Learning to play checkers

- What experience?
  - Be aware of **training data bias**: data, training procedure, features
- What exactly should be learned?
  - Correct vs. operational: Approximation, aka. Hypothesis
- How shall it be represented?

45

Introduction to Machine Learning: Introduction

## Representing (the **Hypothesis Class** $\hat{V}$ )

- Possible representation
  - Tables of all states
  - Collection of rules
  - Polynomial function of board features
  - Neural network
  - .....
- **Expressiveness** of type of function must be chosen carefully
  - **Goodness** of approximation vs. **data** requirements

(To be continued...)

46

Introduction to Machine Learning: Introduction