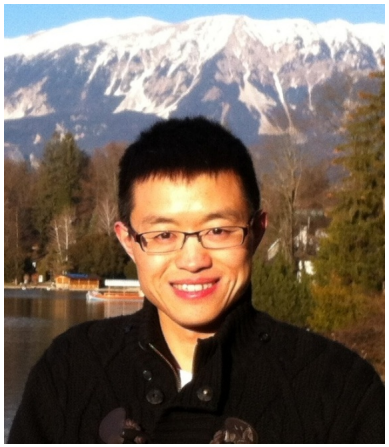
Machine Learning

Jie Tang
Department of Computer Science & Technology
Tsinghua University

A bit about Jie…



- Jie Tang, Tenured associate professor, Department of Computer Science of Tsinghua Univ. My research interests include social networks, data mining, and machine learning.
- I have been visiting scholar at Cornell U. (working with John Hopcroft, Jon Kleinberg), UIUC (working with Jiawei Han), CUHK (with Jeffrey Yu), and HKUST (with Qiong Luo).
- I was awarded with the CCF Young Scientist Award, NSFC Excellent Young Scholar, Newton Advanced Fellowships Award, IBM Innovation Faculty Award, and New Star of Beijing S&T.
- Have published more than 200 paper on major international conf/journals, including KDD (18), IJCAI/AAAI (13), IEEE TKDE (8), ICML, Machine Learning
- #Citation: 6,890 and H-index: 41
- Have a notable system, AMiner.org for academic researcher network analysis. The system has attracted 8.32 million users from 220 countries/regions.
- **Homepage:** http://keg.cs.tsinghua.edu.cn/jietang/

Contact Information

- Jie Tang
  - Software Division, Department of Computer Science, Tsinghua U.
  - Office: Rm 1-308, FIT Building
  - E-mail: jietang@tsinghua.edu.cn
  - Phone: 62788788-20, 13911215746

Teaching Assistant

- TA: Fang Zhang
  - FIT 1-308
  - E-mail: thuzhf@gmail.com
  - Phone: 62788788-20
  - Open hours: 15:00pm-18:00pm Tuesday

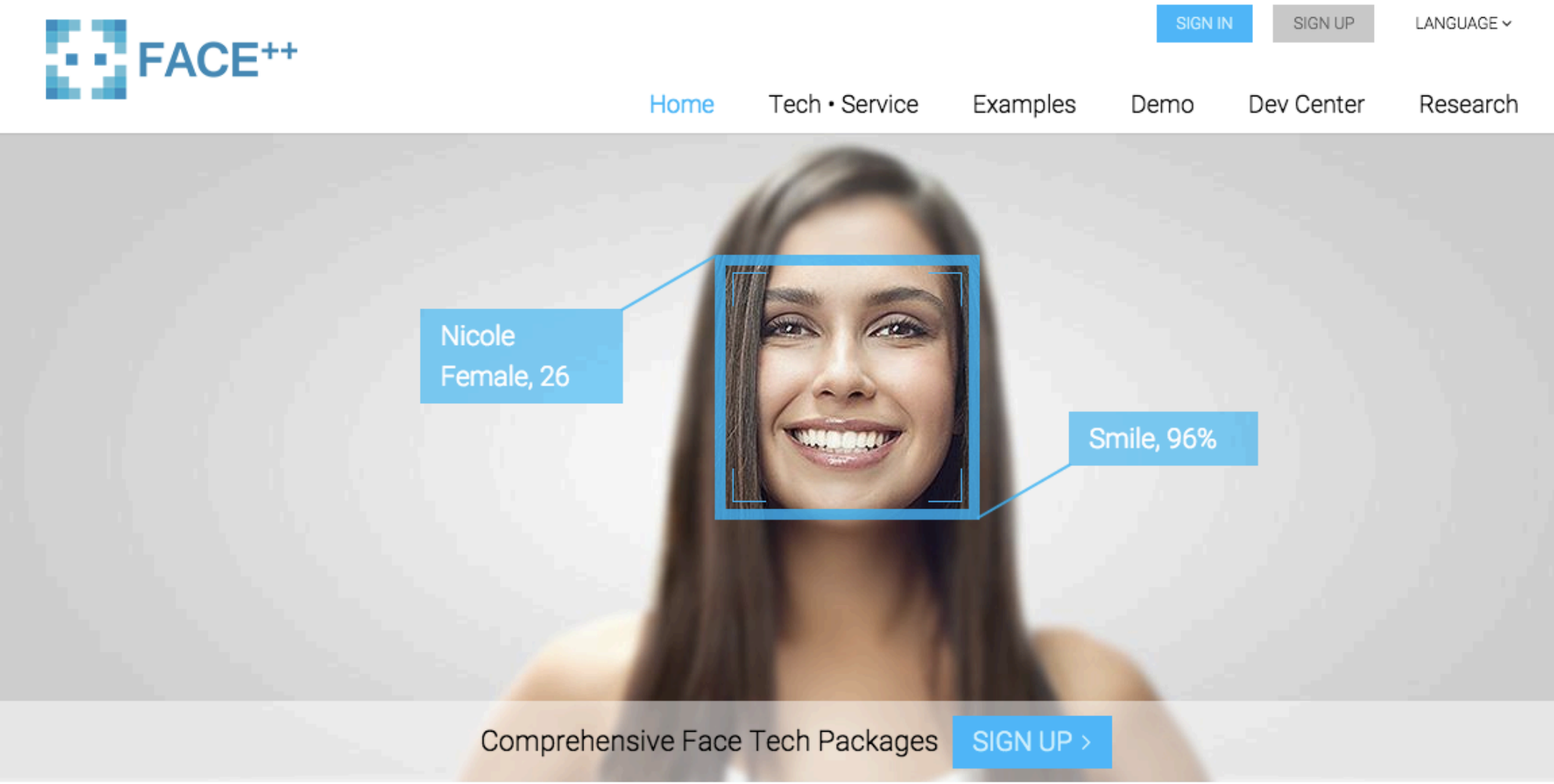
Resources

- Conferences:
  - Theory: NIPS, COLT, STOC/FOCS
  - Algorithm: ICML, KDD, UAI, IJCAI/AAAI
  - App: SIGIR, WWW, ACL
- Journals:
  - JMLR, JAIR, MLJ, ACM TKDD, IEEE TKDE

Why Machine Learning?

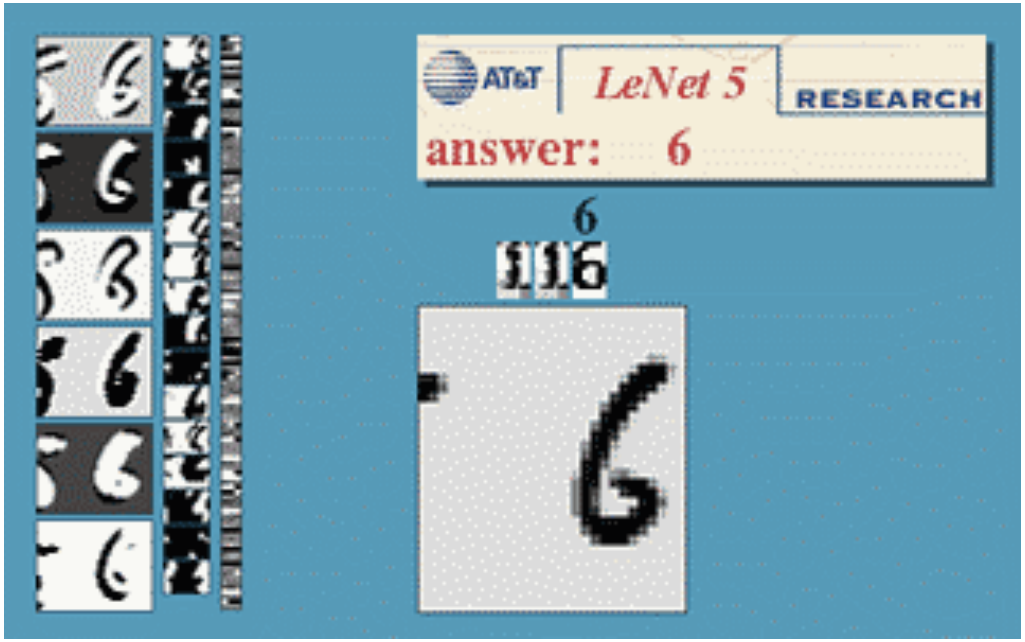
Jie Tang
Department of Computer Science & Technology
Tsinghua University

Face Recognition



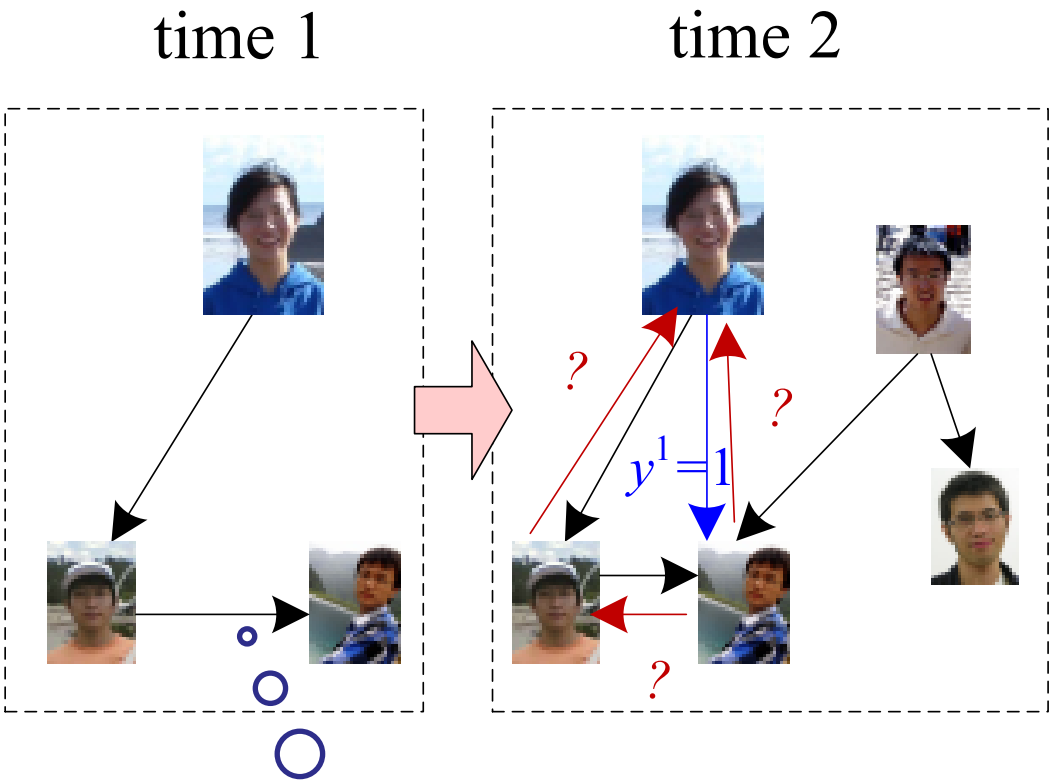
Handwritten Recognition

- Convolutional neural networks for handwritten recognition



- http://yann.lecun.com/exdb/lenet/index.html

Follow back Prediction



When you follow a friend in Twitter, how likely he will follow back?

IE from Web Page (IJCAI'07)

October 14, 2002, 4:00 a.m. PT

For years, Microsoft Corporation CEO Bill Gates railed against the economic philosophy of open-source software with Orwellian fervor, denouncing its communal licensing as a "cancer" that stifled technological innovation.

Today, Microsoft claims to "love" the open-source concept, by which software code is made public to encourage improvement and development by outside programmers. Gates himself says Microsoft will gladly disclose its crown jewels--the coveted code behind the Windows operating system--to select customers.

"We can be open source. We love the concept of shared source," said Bill Veghte, a Microsoft VP. "That's a super-important shift for us in terms of code access."

Richard Stallman, founder of the Free Software Foundation, countered saying…

* Microsoft Corporation
CEO
Bill Gates

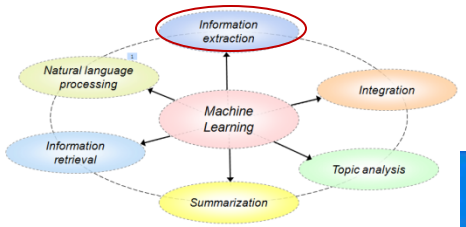
* Microsoft
Gates

* Microsoft
Bill Veghte

* Microsoft
VP

Richard Stallman
founder
Free Software Foundation

NAME	TITLE	ORGANIZATION
Bill Gates	CEO	Microsoft
Bill Veghte	VP	Microsoft
Richard Stallman	founder	Free Soft..



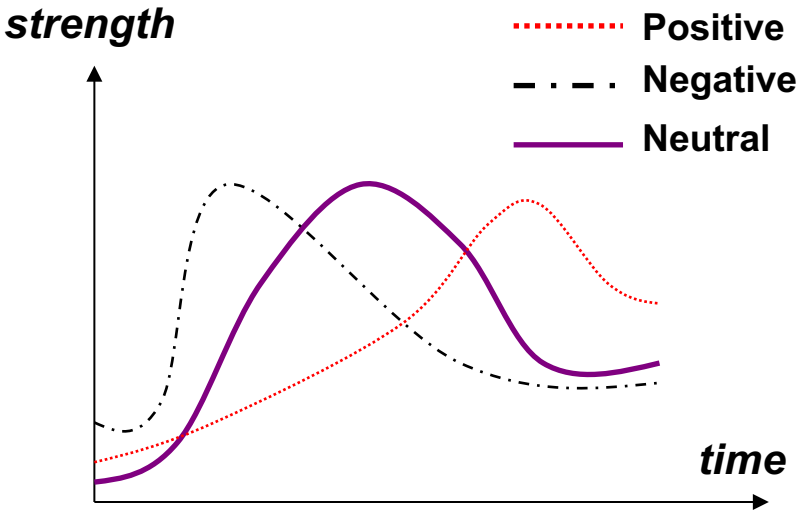
Sentiment Summary (WWW'08)

Topic-sentiment summary

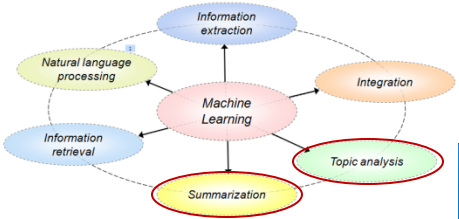
Query: *Dell Laptop*

	positive	negative	neutral
Topic 1 (*Price*)	• it is the best site and they show Dell coupon code as early as possible	• Even though Dell's price is cheaper, we still don't want it. • ……	• mac pro vs. dell precision: a price comparis.. • DELL is trading at $24.66
Topic 2 (*Battery*)	• One thing I really like about this Dell battery is the Express Charge feature.	• my Dell battery sucks • Stupid Dell laptop battery • ……	• i still want a free battery from dell.. • ……

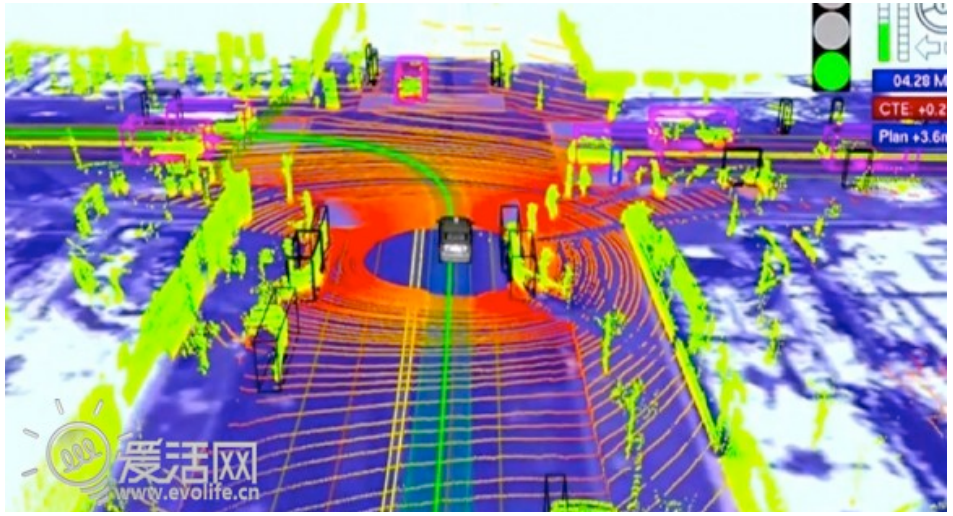
Topic-sentiment dynamics (Topic = *Price*)

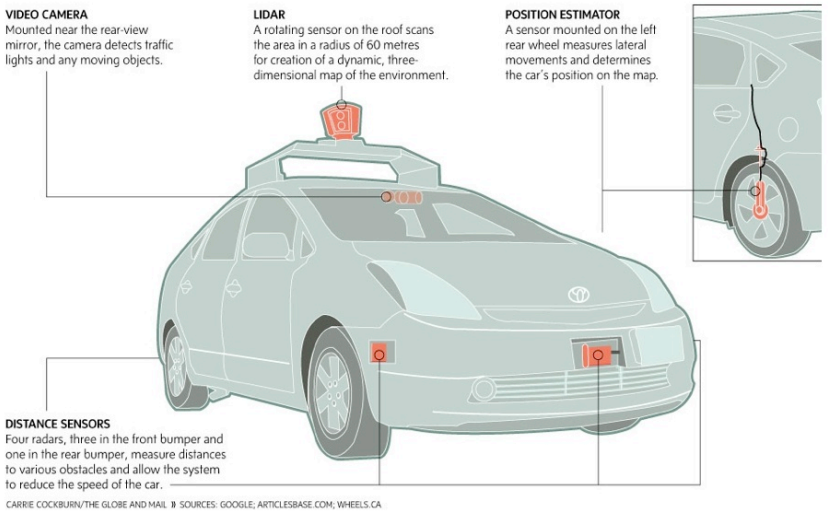


Unsupervised/Semi-supervised discovery of topics and different sentiments of the topics



BIG DATA—Google driverless car





Google's self-driving cars last month hit a 480,000-miles-driven milestone.

--2012/09

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 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

http://www.fiercebigdata.com/story/self-driving-cars-create-1gb-data-second/2013-07-22

IPhone Siri

Siri understands what you say, knows what you mean, and even talks back.

Siri (pronounced /ˈsɪri/) is an intelligent personal assistant and knowledge navigator which works as an application for Apple's iOS. The application uses a natural language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of Web services. Apple claims that the software adapts to the user's individual preferences over time and personalizes results, and performing tasks such as finding recommendations for nearby restaurants, or getting directions

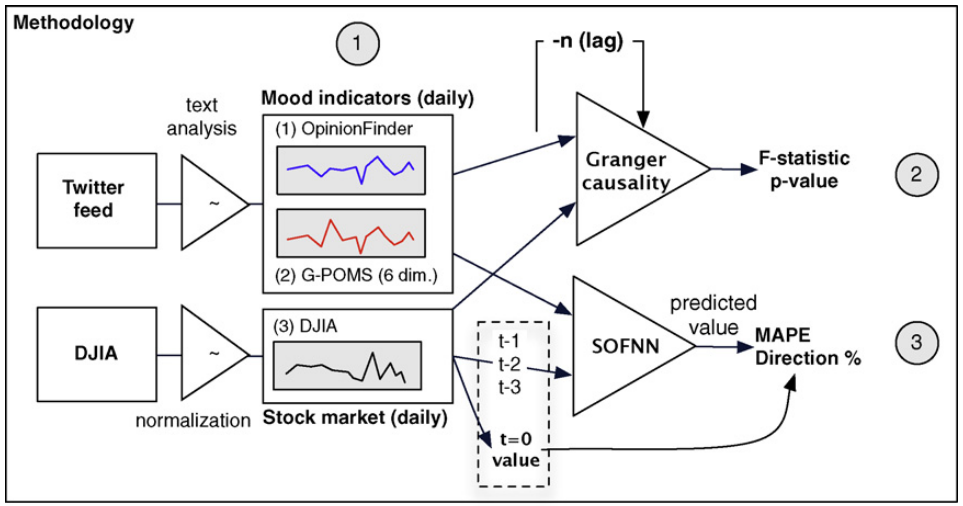
IBM Watson

In 2011, as a test of its abilities, Watson competed on the quiz show Jeopardy!, in the show's only human-versus-machine match-up to date

Watson is a Question answering (QA) computing system built by IBM. IBM describes it as "an application of advanced Natural Language Processing, Information Retrieval, Knowledge Representation and Reasoning, and Machine Learning technologies to the field of open domain question answering" which is "built on IBM's *DeepQA* technology for hypothesis generation, massive evidence gathering, analysis, and scoring

Watson had access to 200 million pages of structured and unstructured content consuming four terabytes of disk storage including the full text of Wikipedia, but was not connected to the Internet during the game. For each clue, Watson's three most probable responses were displayed on the television screen. Watson consistently outperformed its human opponents on the game's signaling device.

Stock Prediction using Twitter Mood



The paper investigates whether public sentiment, as expressed in large-scale collections of daily Twitter posts, can indeed be used to predict the stock market.

Two tools are used to measure variations in the public mood from Twitter's tweets from February 28, 2008 to December 19, 2008. The resulting public mood time series are correlated to the Dow Jones Industrial Average (DJIA) to assess their ability to predict changes in the DJIA over time.

The results indicate that the prediction ccuracy of standard stock market prediction models is significantly improved when certain mood dimensions are included.

Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science, Volume 2, Issue 1, March 2011, Pages 1–8

What is Machine Learning?

Data Structures

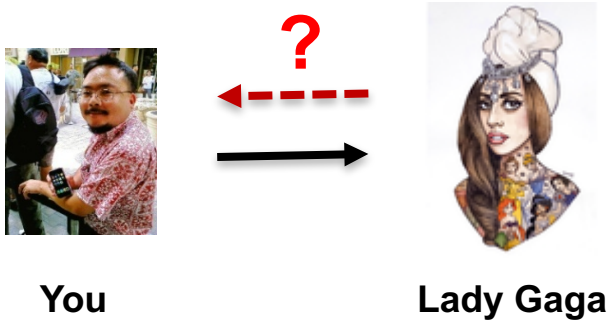
- An example with *p* features

$$\boldsymbol{x}_i=[x_{i1}, x_{i2},\ldots, x_{ip}]$$

- Data (set) matrix

$$\begin{bmatrix} x_{11} & \cdots & x_{1f} & \cdots & x_{1p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{i1} & \cdots & x_{if} & \cdots & x_{ip} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N1} & \cdots & x_{Nf} & \cdots & x_{Np} \end{bmatrix}$$

Example: Follow-back Prediction



When a follow b, we have a new instance $\mathbf{x}_i$;

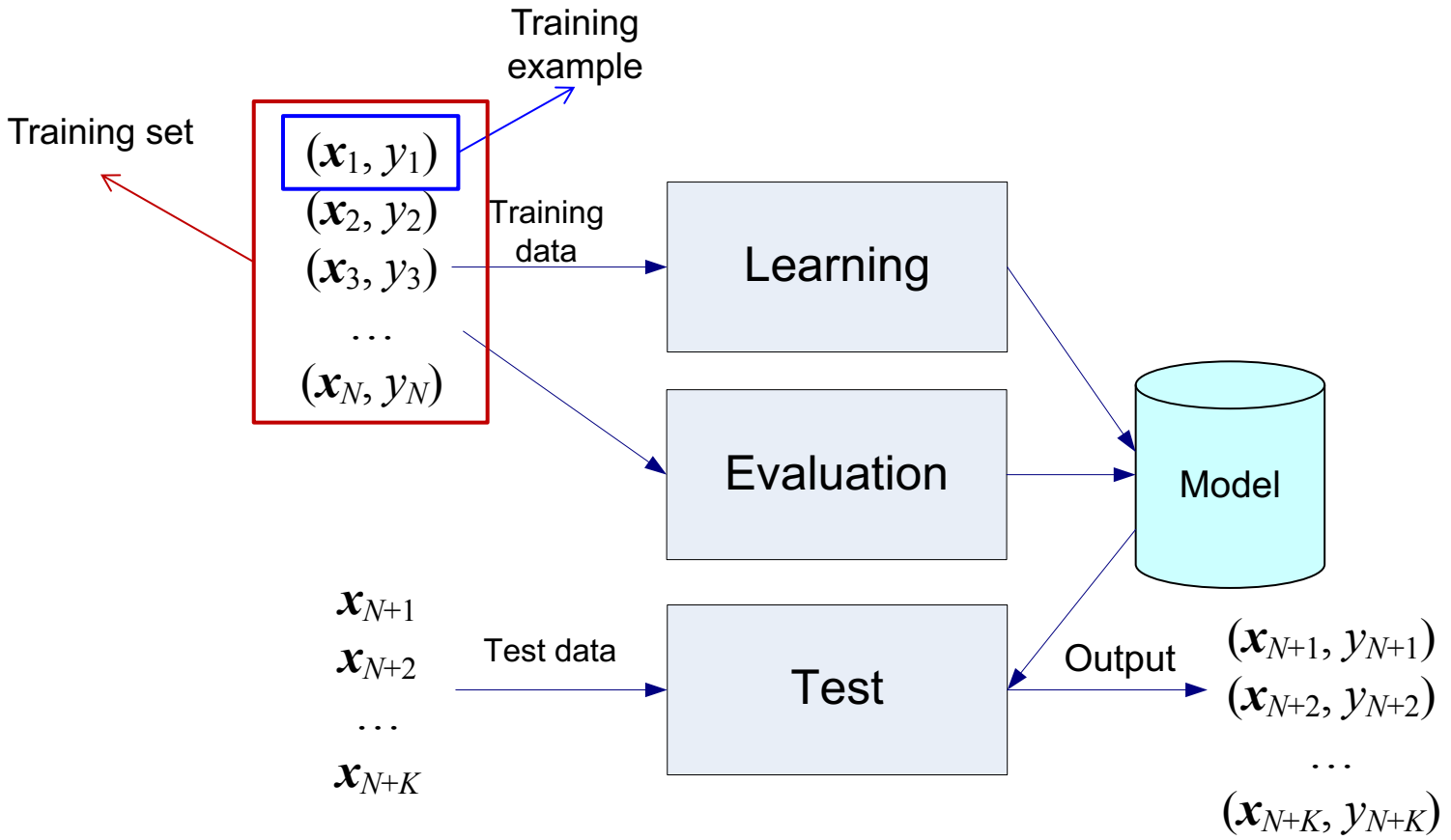
If b immediately follow back a, we say y_i=1;

otherwise y_i=0;

$$L(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}(f(\mathbf{x}_i,\mathbf{w}) - y_i)^2$$

$\mathbf{x}_i$={#fans of a, #fans of b, #common friends between a and b}

Supervised Learning

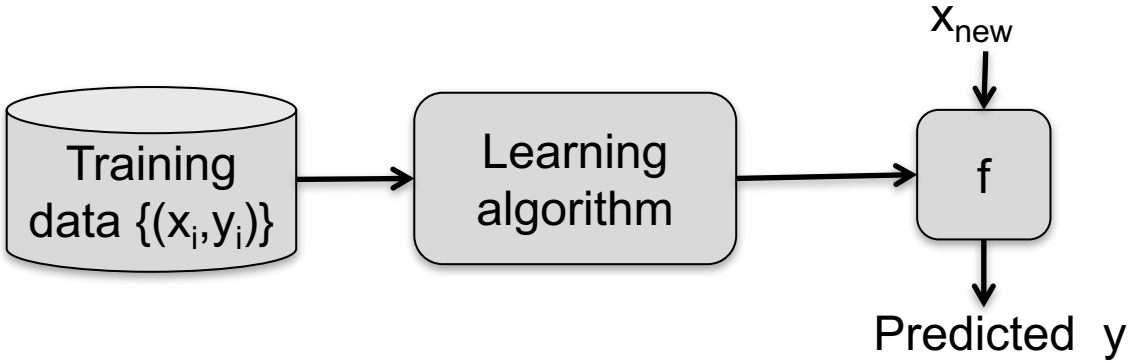


* x_i is a vector and x_{ik} represents the k-th feature of example x_i

Supervised Learning

- Given a training set $S=\{(x_1, y_1),(x_2, y_2),\ldots,(x_N, y_N)\}$, and $x_i \in X=R^m$, $i=1,2,\ldots,N$
- To learn a function $f(x)$, which can best fit the data

$$y = f(x)$$



Regression: When y is continuous (i.e., $y_i \in Y=R$), we call the learning problem as regression.

Classification: When y can take on only a small number of discrete values (e.g., the binary case $y_i \in Y=\{1,-1\}$), we call the learning problem as classification.

Netflix Prize

Training data

user	movie	date	score
1	21	5/7/02	1
1	213	8/2/04	5
2	345	3/6/01	4
2	123	5/1/05	4
2	768	7/15/02	3
3	76	1/22/01	5
4	45	8/3/00	4
5	568	9/10/05	1
5	342	3/5/03	2
5	234	12/28/00	2
6	76	8/11/02	5
6	56	6/15/03	4

Test data

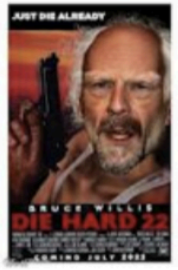
user	movie	date	score
1	62	1/6/05	?
1	96	9/13/04	?
2	7	8/18/05	?
2	3	11/22/05	?
3	47	6/13/02	?
3	15	8/12/01	?
4	41	9/1/00	?
4	28	8/27/05	?
5	93	4/4/05	?
5	74	7/16/03	?
6	69	2/14/04	?
6	83	10/3/03	?

Rating Prediction

- Least mean squares prediction error
  - Easy to define

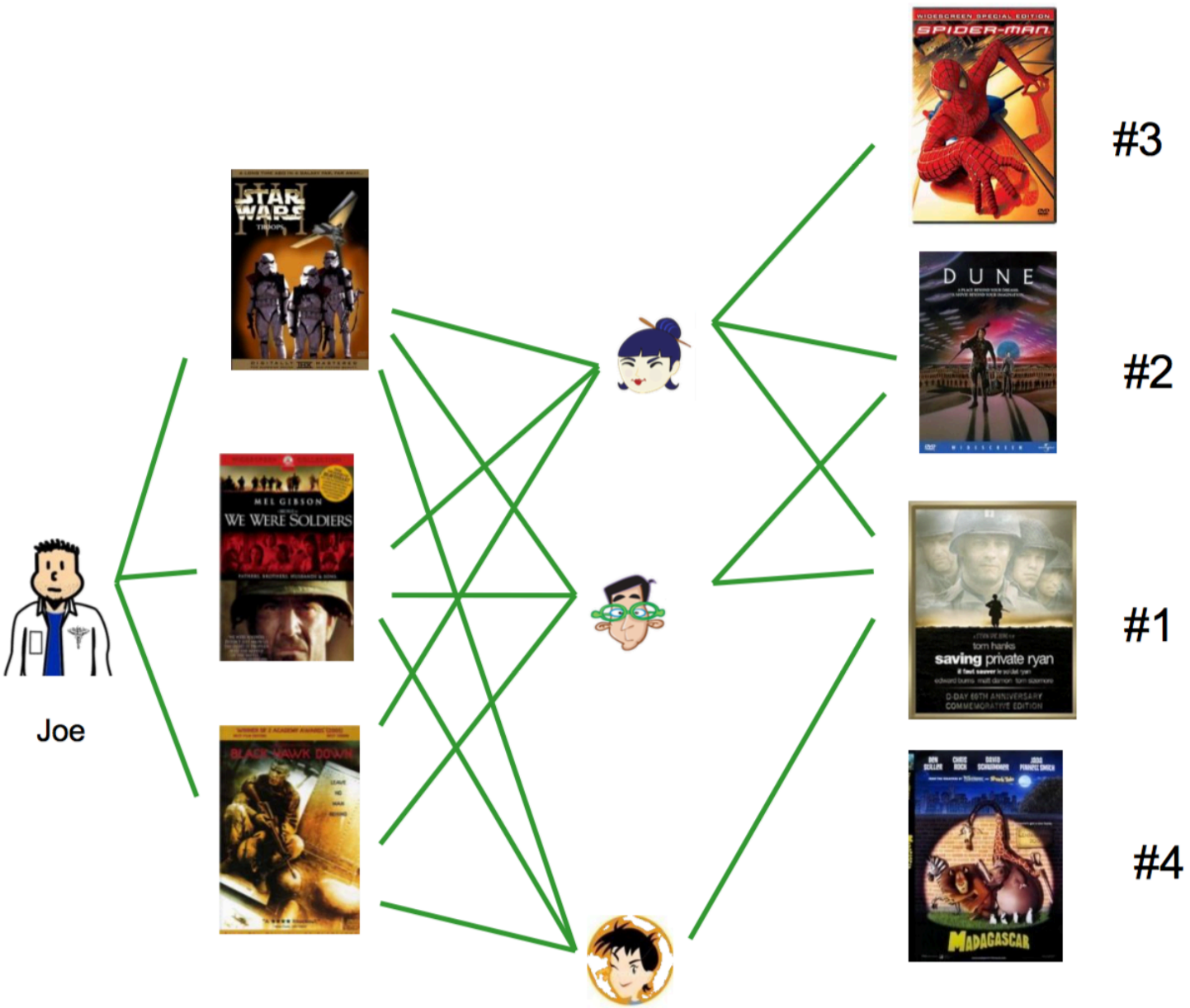
$$\mathrm{rmse}(S) = \sqrt{|S|^{-1} \sum_{(i,u)\in S} (\hat{r}_{ui} - r_{ui})^2}$$

  - Wrong measure for composing sessions!



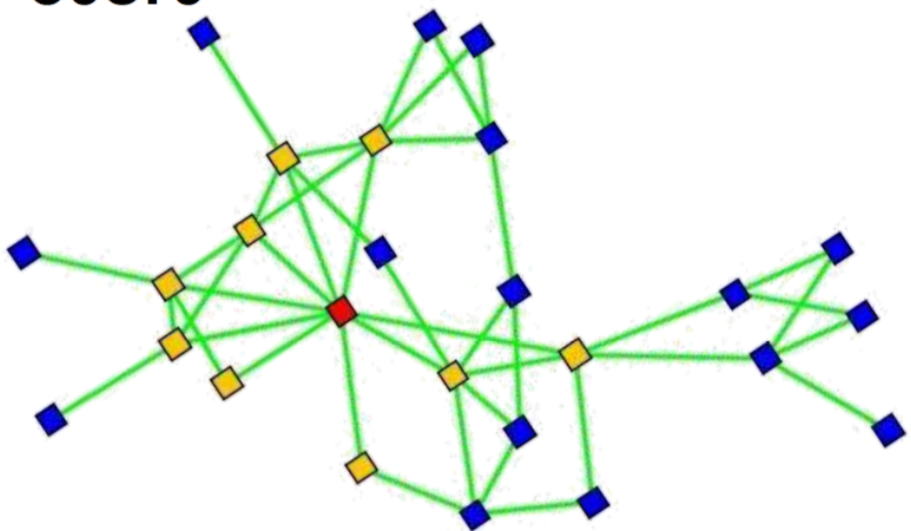

  - Consistent (in large sample size limit this will converge to minimizer)

Basic Idea

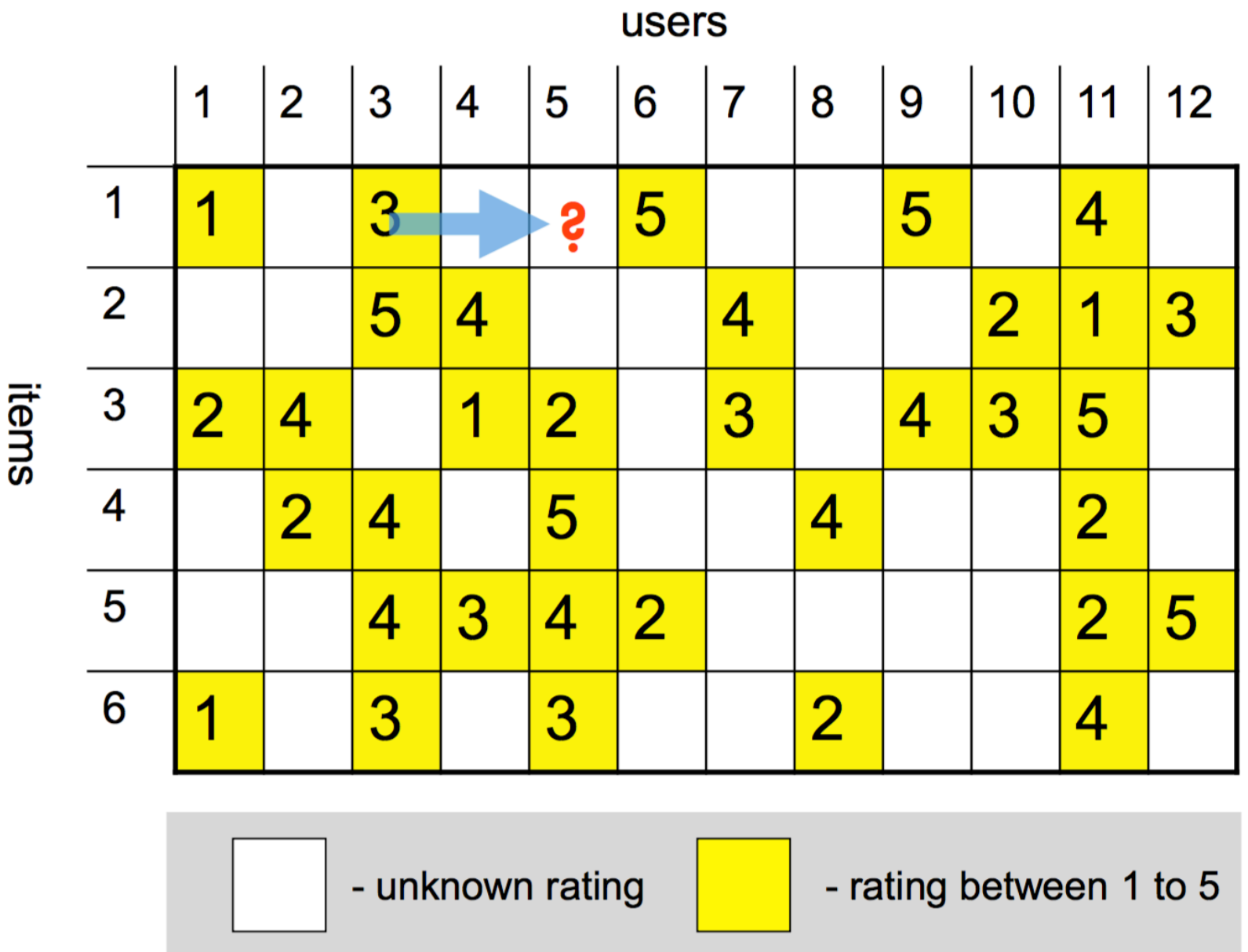


Basic Idea

- (user,user) similarity to recommend items
  - good if item base is smaller than user base
  - good if item base changes rapidly
  - traverse bipartite similarity graph
- (item,item) similarity to recommend new items that were also liked by the same users
  - good if the user base is small is small
- Oldest known CF method



Intuitions


users
items
1 2 3 4 5 6 7 8 9 10 11 12
1 1 3 ? 5 5 4
2 5 4 4 2 1 3
3 2 4 1 2 3 4 3 5
4 2 4 5 4 2
5 4 3 4 2 2 5
6 1 3 3 2 4
- unknown rating
- rating between 1 to 5

The State of Machine Learning

Classification models

Decision tree 
Bayesian classifier
Perceptron
Neural networks

Maximal Margin

SVM (Vapnik) 



Maximal margin sequential learning

M3N network (Ben Taskar) 

Deep learning

Deep belief networks




Topic models

PLSI, LDA, etc.




Sequential learning

HMM 

Sequential learning

MEMM, CRF, voted perceptron



Graphical models

Factor graph, Exponential model




For more from http://aminer.org/

Prerequisites

- Programming and other basic CS skills
  - C, C++
  - Java, C#, .NET
  - Perl, Python
- Theories
  - Data representation (e.g., vector space model, language model)
  - Basic probability theory (e.g., likelihood, conditional probability, posterior probability, Bayes)
  - Basic linear algebra (e.g., linear transformations, eigenvalues, least-squares best fit)

Potential achievements

- Able to understand the underlying principles of classical ML algorithms
- Able to apply right ML algorithms to the applications at you hand
- Able to design effective ML algorithms to solve new problems

Handbooks

- Handbooks:
  - Christopher M. Bishop. *Pattern Recognition and Machine Learning*, Springer, 2007.
  - Yoshua Bengio, Ian J. Goodfellow, and Aaron Courville. Deep Learning. 2015.
  - Daphne Koller and Nir Friedman. *Probabilistic Graphical Models.* MIT Press, 2009
- Additional references
  - John Hopcroft. Computer Science Theory for the Information Age. 2011.
  - Michael I. Jordan. *An Introduction to Probabilistic Graphical Models.* University of California, Berkeley. June 30, 2003.
  - Martin J. Wainwright and Michael I. Jordan. *Graphical Models, Exponential Families, and Variational Inference*, Foundations and Trends in Machine Learning, V1 (1-2), 2008.
  - Trevor Hastie, Robert Tibshirani, Jerome Friedman. *Elements of Statistical Learning*. Springer, 2003.
  - Yoshua Bengio. *Learning Deep Architectures for AI*. Foundations and Trends in Machine Learning, V2 (1), 2009.
  - David J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

Thanks!

HP: http://keg.cs.tsinghua.edu.cn/jietang/