UVA CS 6316 – Fall 2015 Graduate: Machine Learning

Lecture 1: Introduction

Dr. Yanjun Qi

University of Virginia
Department of
Computer Science

Welcome

- CS 6316 Machine Learning
  - MoWe 3:30pm-4:45pm,
  - Mechanical Engr Bldg 341
- http://www.cs.virginia.edu/yanjun/teach/2015f
- Your UVA collab: Course 6316 page

Today

- ❑ **Course Logistics**
- ❑ My background
- ❑ Basics and rough content plan
- ❑ Application and History

Course Staff

- Instructor: Prof. Yanjun Qi
  - QI: /ch ee/
  - You can call me "professor", "professor Jane", "professor Qi";
- TA: Ritambhara Singh <rs3zz@virginia.edu>
- TA office hours: Wed 5pm-6pm @ Rice 504
- My office hours: Thur 5pm-6pm @ Rice 503

Course Logistics

- Course email list has been setup. You should have received emails already !

- Policy, the grade will be calculated as follows:
  - Assignments (50%, **Five** total, each 10%)
  - In-class quizzes (10%, multiple)
  - mid-term (20%)
  - Final project (20%)

Dr. Yanjun Qi / UVA CS 6316 / f15

Course Logistics

- Midterm: late Oct or mid Nov., 75mins in class
- Final project:
  - proposal + report + in-class presentation
- Five assignments (each 10%)
  - Due Sept 16, Sept 30, Oct 14, Nov 4, Nov 28
  - **three** extension days policy (check course website)
- Multiple in-class quizzes (total 10%)
  - About 10 small quizzes
  - Randomly distributed over the whole semester

Course Logistics

- Policy,
  - Homework should be submitted electronically through UVaCollab
  - Homework should be finished individually
  - Due at midnight on the due date
  - In order to pass the course, the average of your midterm and final must also be "pass".

Course Logistics

- Text books for this class is:
  - NONE
- My slides – **if it is not mentioned in my slides, it is not an official topic of the course**

Course Logistics

- **Background Needed**
  - Calculus, Basic linear algebra, Basic probability and Basic Algorithm
  - Statistics is recommended.
  - Students should already have good programming skills, i.e. python is required for all programming assignments
  - We will review "linear algebra" and "probability" in class

Dr. Yanjun Qi / UVA CS 6316 / f15

Today

- ❑ Course Logistics
- ❑ **My background**
- ❑ Basics and rough content plan
- ❑ Application and History

About Me

- Education:
  - PhD from School of Computer Science, Carnegie Mellon University (@ Pittsburgh, PA) in 2008
  - BS in Department of Computer Science, Tsinghua Univ. (@ Beijing, China)
    - My accent **PATTERN** : /l/, /n/,/ou/, /m/
- Research interests:
  - **Machine Learning, Data Mining, Biomedical applications**

About Me

- Five Years' of Industry Research Lab in the past :
  - 2008 summer – 2013 summer, **Research Scientist in IT** industry (Machine Learning Department, NEC Labs America @ Princeton, NJ)
  - 2013 Fall – Present, **Assistant Professor**, Computer Science, UVA

Industry + Academia

Today

- ❑ Course Logistics
- ❑ My background
- ❑ **Basics and Rough content plan**
- ❑ Application and History

OUR DATA-RICH WORLD

- Biomedicine
  - Patient records, brain imaging, MRI & CT scans, …
  - Genomic sequences, bio-structure, drug effect info, …
- Science
  - Historical documents, scanned books, databases from astronomy, environmental data, climate records, …
- Social media
  - Social interactions data, twitter, facebook records, online reviews, …
- Business
  - Stock market transactions, corporate sales, airline traffic, …
- Entertainment
  - Internet images, Hollywood movies, music audio files, …

BIG DATA CHALLENGES

- Data capturing (sensor, smart devices, medical instruments, et al.)
- Data transmission
- Data storage
- Data management
- High performance data processing
- Data visualization
- Data security & privacy (e.g. multiple individuals)
- ……

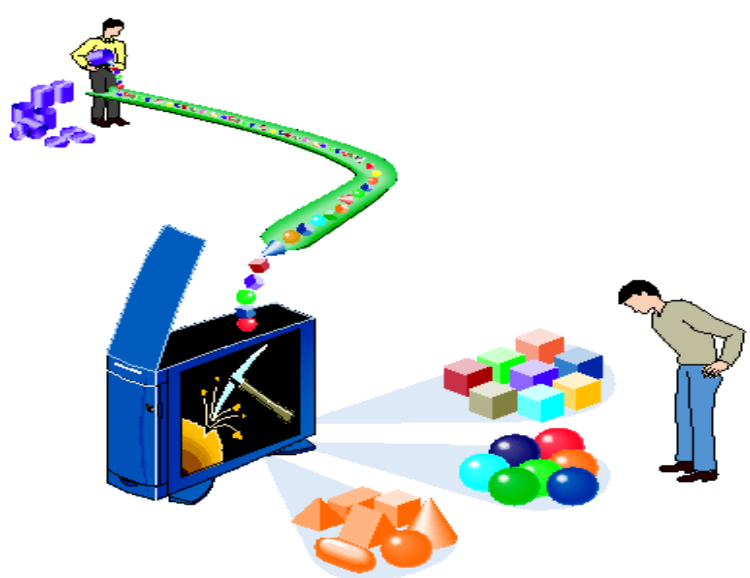
e.g. cloud computing

e.g. HCI

this course

- Data analytics
  - How can we analyze this big data wealth ?
  - E.g. Machine learning and data mining

Drowning in data, Starving for knowledge



BASICS OF MACHINE LEARNING

- "The goal of machine learning is to build computer systems that can learn and adapt from their experience." – Tom Dietterich
- "Experience" in the form of available data examples (also called as instances, samples)
- Available examples are described with properties (data points in feature space X)

e.g. SUPERVISED LEARNING

- Find function to map input space X to output space Y $f : X \longrightarrow Y$
- So that the difference between y and $f(\boldsymbol{x})$ of each example $\boldsymbol{x}$ is small.

e.g.

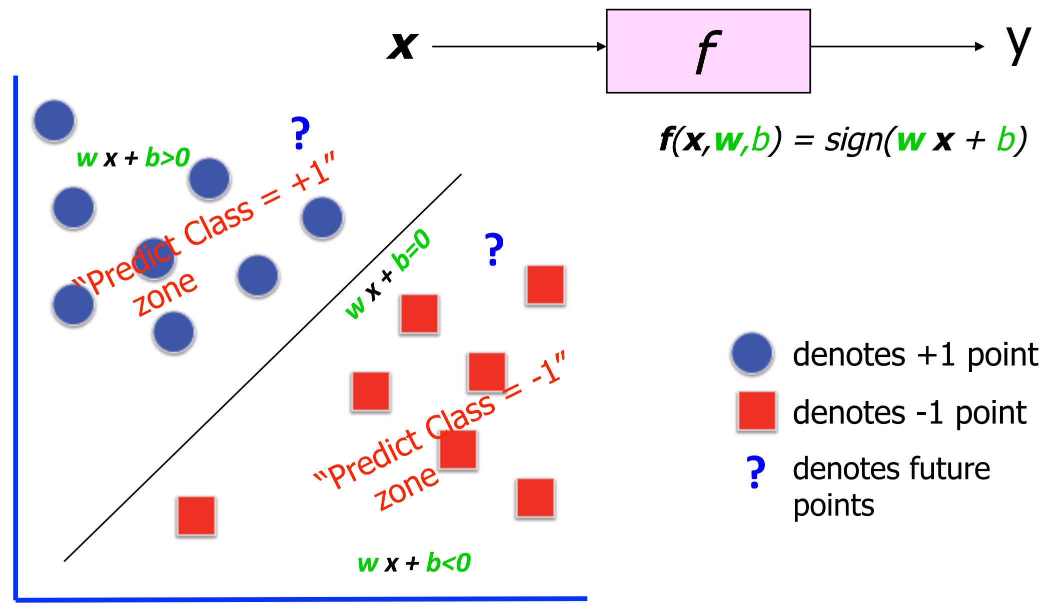
x I believe that this book is not at all helpful since it does not explain thoroughly the material . it just provides the reader with tables and calculations that sometimes are not easily understood …

→ **y** -1

Output Y: {1 / Yes , -1 / No }
e.g. Is this a positive product review ?

Input X : e.g. a piece of English text

e.g. SUPERVISED Linear Binary Classifier



Courtesy slide from Prof. Andrew Moore's tutorial

Basic Concepts

- Training (i.e. learning parameters $\boxed{w,b}$)
  - Training set includes
    - available examples $x_1,\ldots,x_L$
    - available corresponding labels $y_1,\ldots,y_L$
  - Find (w,b) by minimizing loss (i.e. difference between y and $f(x)$ *on available examples in training set*)

$$(W, b) = \underset{W,\, b}{\mathrm{argmin}} \sum_{i=1}^{L} \ell(f(x_i), y_i)$$

Basic Concepts

- Testing (i.e. evaluating performance on "future" points)
  - Difference between true $y_?$ and the predicted $f(x_?)$ on a set of testing examples (i.e. *testing set*)
  - Key: example $x_?$ not in the training set

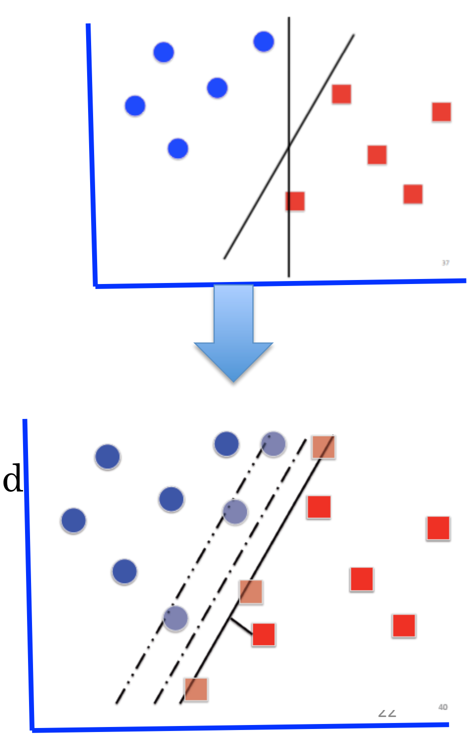
- Generalisation: learn function / hypothesis from past data in order to "explain", "predict", "model" or "control" new data examples

Basic Concepts

- Loss function
  - e.g. hinge loss for binary classification task

$$\sum_{i=1}^{L} \ell(f(x_i), y_i) = \sum_{i=1}^{L} \max(0, 1 - y_i f(x_i)).$$

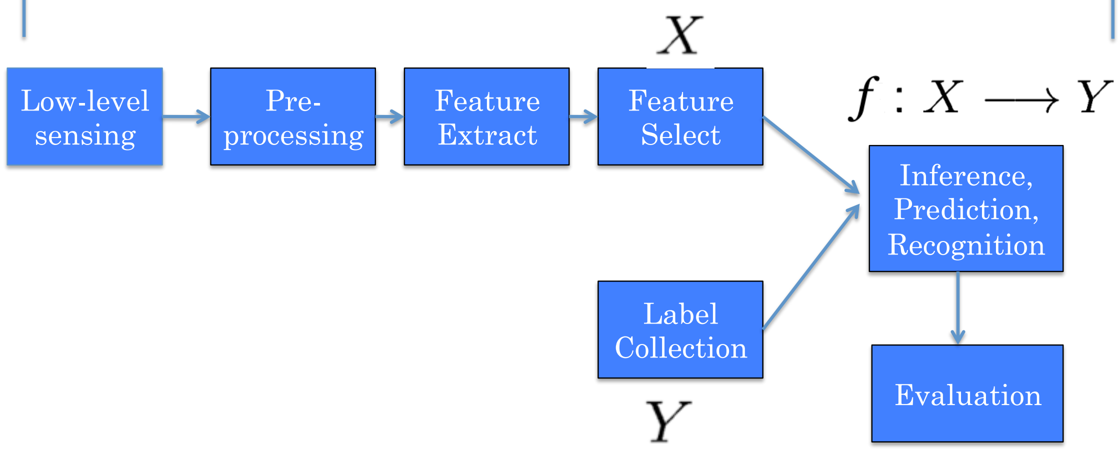
  - e.g. pairwise ranking loss for ranking task (i.e. ordering examples by preference)



- Regularization
  - E.g. additional information added on loss function to control model

$$C \sum_{i=1}^{L} \ell(f(x_i), y_i) + \frac{1}{2}\|w\|^2,$$

TYPICAL MACHINE LEARNING SYSTEM



“Big Data” Challenges for Machine Learning

LARGE-SCALE

HIGH-COMPLEXITY

✓Large size of samples
✓High dimensional features

Not the focus, will be covered in advanced-level course

Large-Scale Machine Learning: SIZE MATTERS

LARGE-SCALE

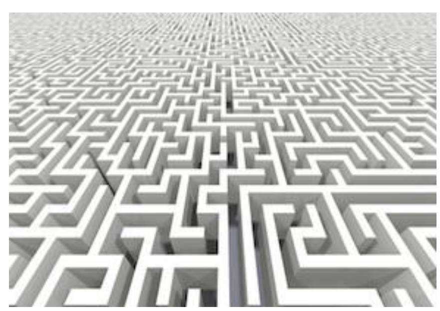
- One thousand data instances
- One million data instances
- One billion data instances
- One trillion data instances

Those are not different numbers, those are different mindsets !!!

BIG DATA CHALLENGES FOR MACHINE LEARNING

LARGE-SCALE

Highly Complex

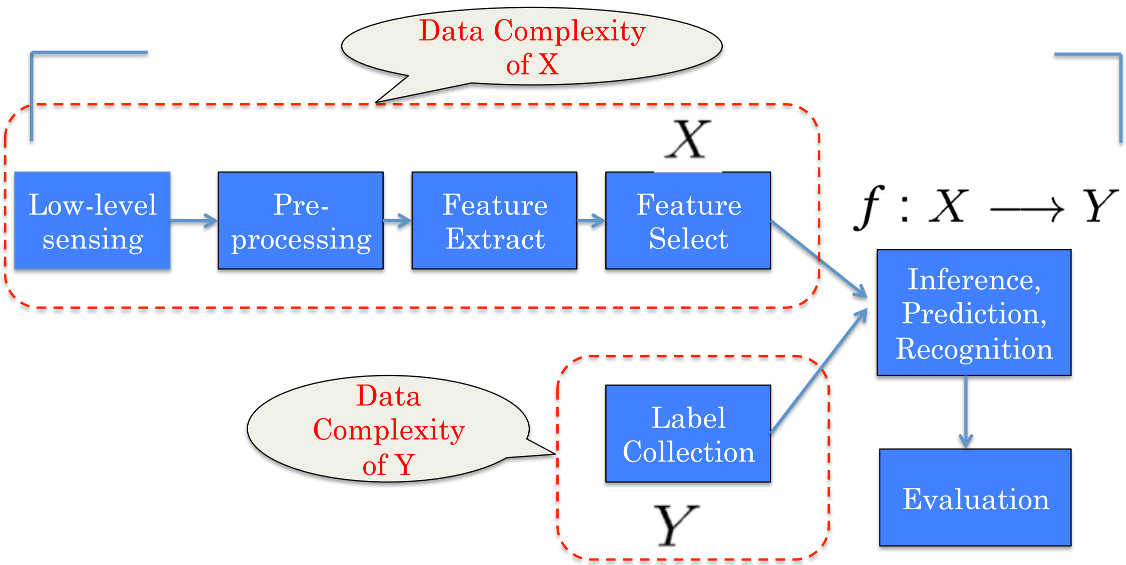


Most of this course

The situations / variations of both X (feature, representation) and Y (labels) are complex !

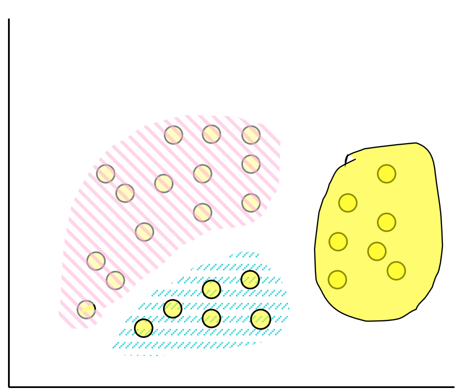
✓Complexity of X
✓Complexity of Y

TYPICAL MACHINE LEARNING SYSTEM



UNSUPERVISED LEARNING : [COMPLEXITY OF Y]

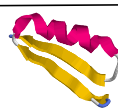
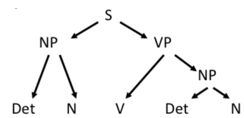
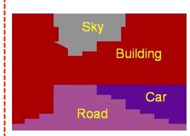
- No labels are provided (e.g. No Y provided)
- Find patterns from unlabeled data, e.g. clustering



e.g. clustering => to find "natural" grouping of instances given un-labeled data

STRUCTURAL OUTPUT LEARNING : [COMPLEXITY OF Y]

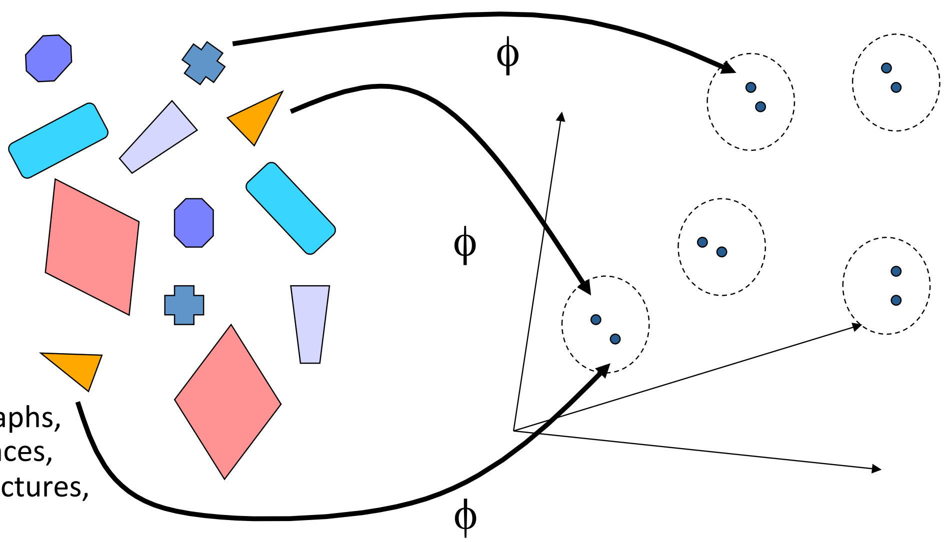
- Many prediction tasks involve output labels having structured correlations or constraints among instances

Structured Dependency between Examples	Sequence	Tree	Grid
Input X	APAFSVSPASGACGPECA…	The dog chased the cat	
Output Y	 CCEEEEECCCCCHHHCCC…		

Many more possible structures between y_i , e.g. spatial , temporal, relational …

STRUCTURAL INPUT : Kernel Methods [COMPLEXITY OF X]

Vector vs. Relational data

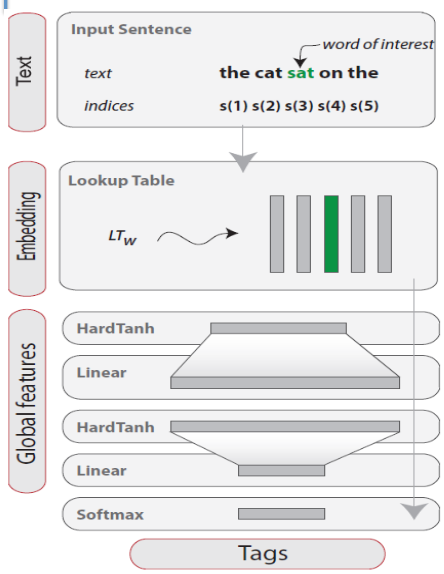


e.g. Graphs, Sequences, 3D structures,

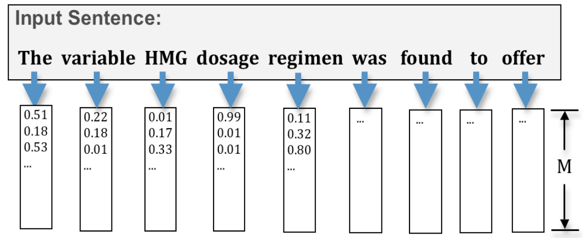
Original Space Feature Space

MORE RECENT: FEATURE LEARNING
[COMPLEXITY OF X]

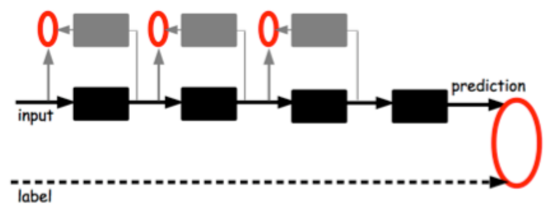
Deep Learning



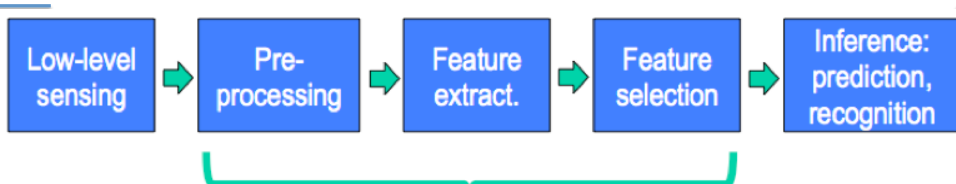
Supervised Embedding



Layer-wise Pretraining



DEEP LEARNING / FEATURE LEARNING :
[COMPLEXITY OF X]

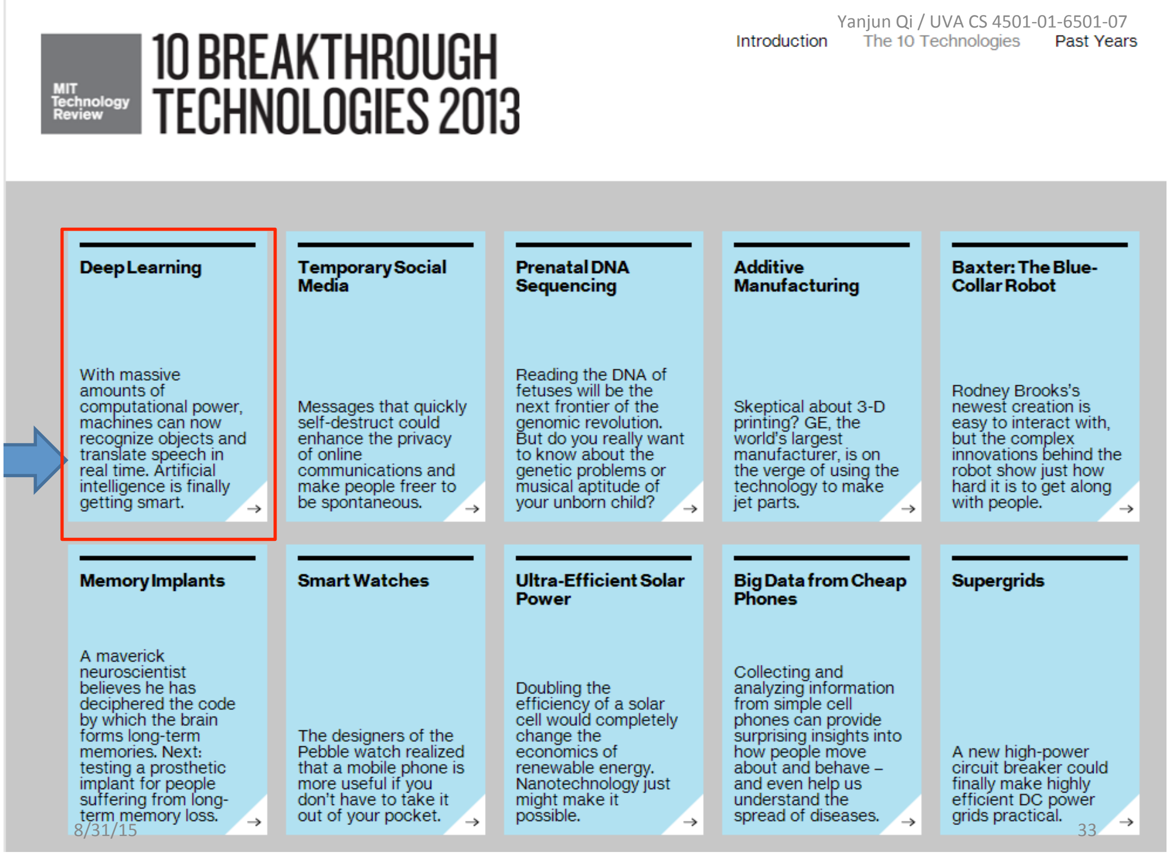


Feature Engineering

- ✓ Most critical for accuracy
- ✓ Account for most of the computation for testing
- ✓ Most time-consuming in development cycle
- ✓ Often hand-craft and task dependent in practice

Feature Learning

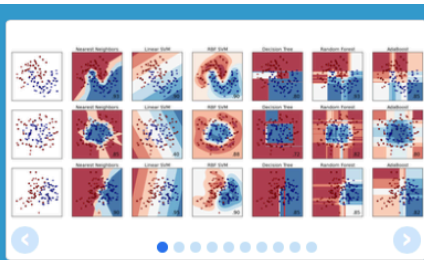
- ✓ Easily adaptable to new similar tasks
- ✓ Layerwise representation
- ✓ Layer-by-layer unsupervised training
- ✓ Layer-by-layer supervised training



Course Content Plan ➔ Five major sections of this course

- ❑ Regression (supervised)
- ❑ Classification (supervised)
- ❑ Unsupervised models
- ❑ Learning theory
- ❑ Graphical models

http://scikit-learn.org/



scikit-learn

Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification

Identifying to which set of categories a new observation belong to.

Applications: Spam detection, Image recognition.
Algorithms: *SVM*, *nearest neighbors*, *random forest*, ... — *Examples*

Regression

Predicting a continuous value for a new example.

Applications: Drug response, Stock prices.
Algorithms: *SVR*, *ridge regression*, *Lasso*, ... — *Examples*

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes
Algorithms: *k-Means*, *spectral clustering*, *mean-shift*, ... — *Examples*

Dimensionality reduction

Reducing the number of random variables to consider.

Applications: Visualization, Increased efficiency
Algorithms: *PCA*, *feature selection*, *non-negative matrix factorization*. — *Examples*

Model selection

Comparing, validating and choosing parameters and models.

Goal: Improved accuracy via parameter tuning
Modules: *grid search*, *cross validation*, *metrics*. — *Examples*

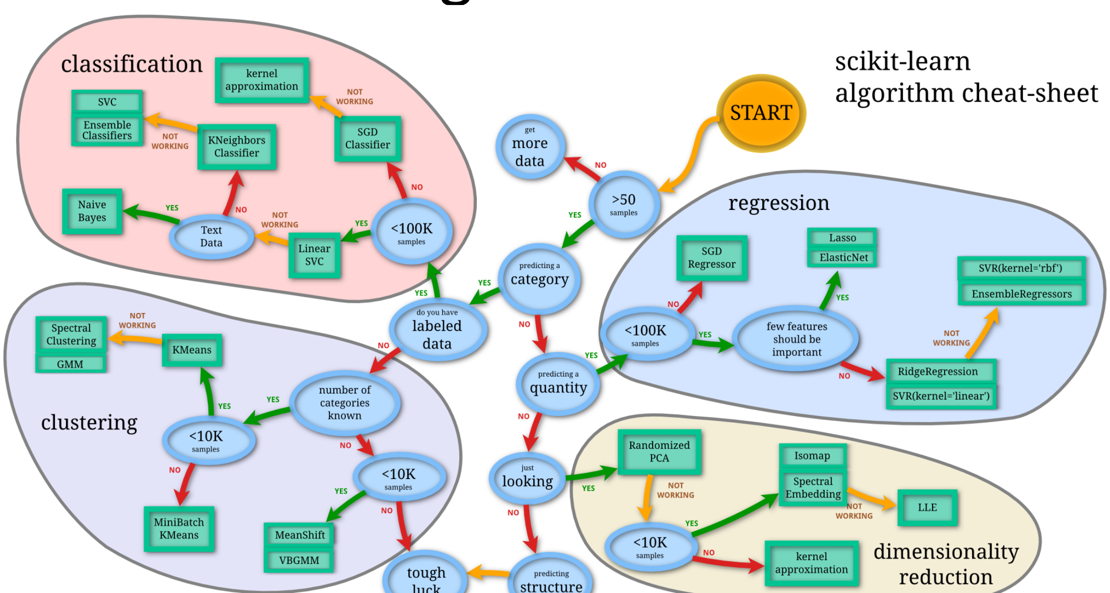
Preprocessing

Feature extraction and normalization.

Application: Transforming input data such as text for use with machine learning algorithms.
Modules: *preprocessing*, *feature extraction*. — *Examples*

http://scikit-learn.org/stable/tutorial/machine_learning_map/

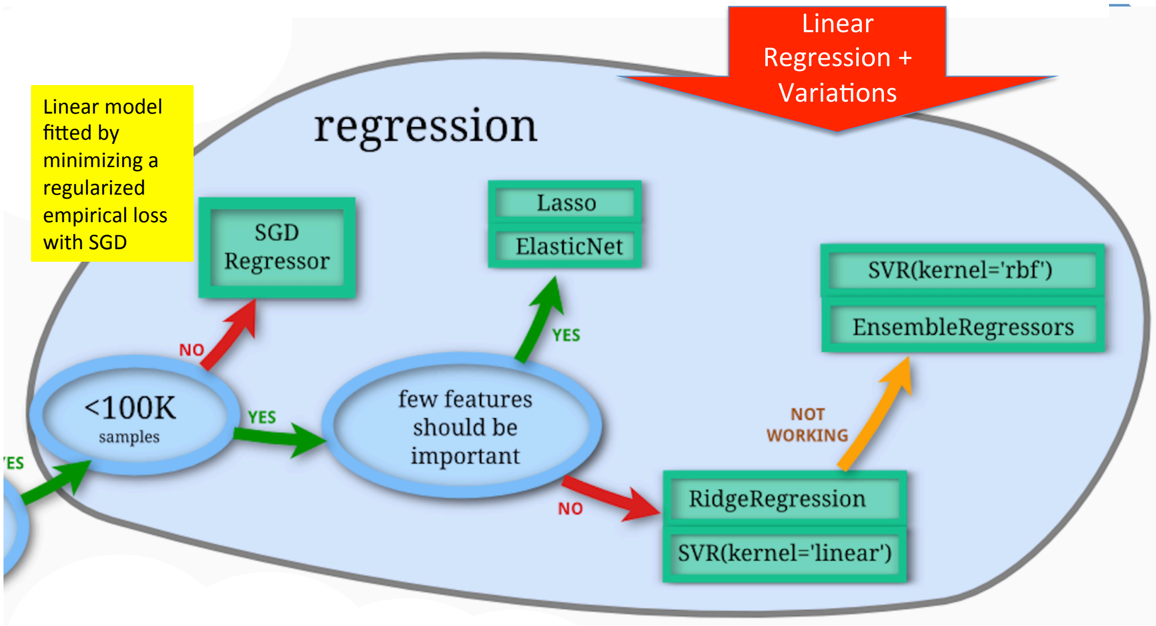
Scikit-learn algorithm cheat-sheet



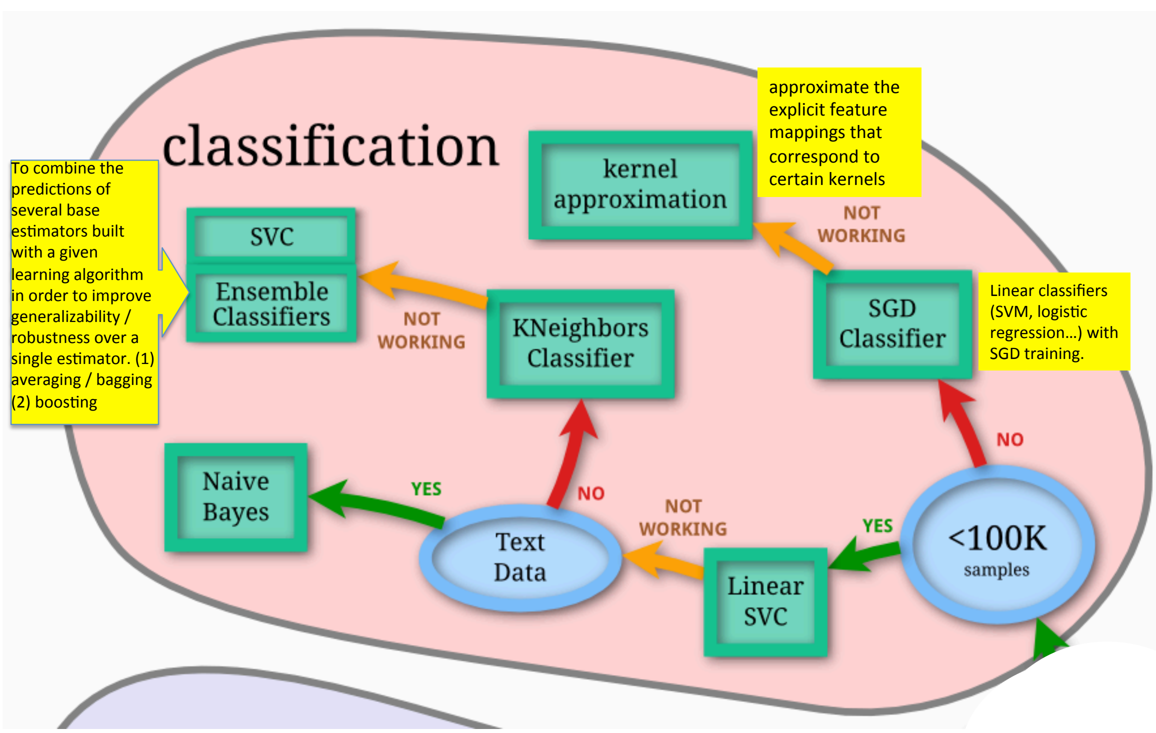
Back

scikit learn

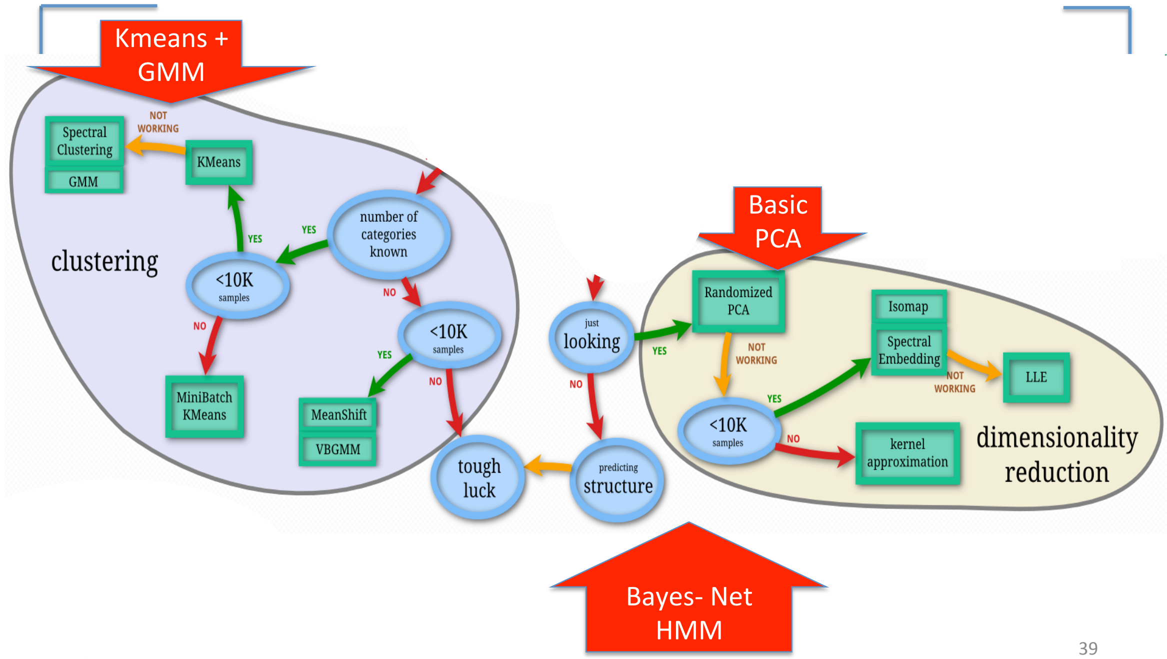
Scikit-learn : Regression



Scikit-learn : Classification



Unsupervised Models



Today

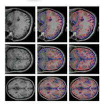
- ❑ Course Logistics
- ❑ My background
- ❑ Basics and rough content plan
- ❑ **Application and History**

What can we do with the data wealth?
➔ REAL-WORLD IMPACT

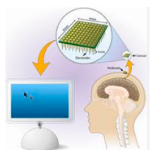
Transportation Data

Genomic Data

Medical Images



Brain computer interaction (BCI)



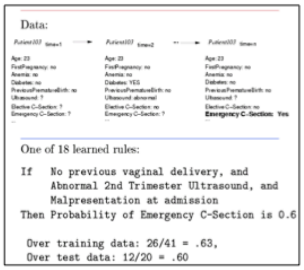
Device sensor data

- Business efficiencies
- Scientific breakthroughs
- Improve quality-of-life:
  - healthcare,
  - energy saving / generation,
  - environmental disasters,
  - nursing home,
  - transportation,
  - …

When to use Machine Learning (Adapt to / learn from data) ?

- 1. Extract knowledge from data
  - Relationships and correlations can be hidden within large amounts of data
  - The amount of knowledge available about certain tasks is simply too large for explicit encoding (e.g. rules) by humans
- 2. Learn tasks that are difficult to formalise
  - Hard to be defined well, except by examples
- 3. Create software that improves over time
  - New knowledge is constantly being discovered.
  - Rule or human encoding-based system is difficult to continuously re-design “by hand”.

MACHINE LEARNING IS CHANGING THE WORLD

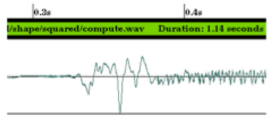

Data:

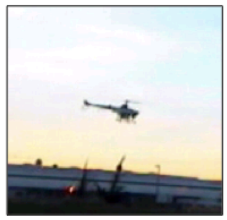
One of 18 learned rules:

If No previous vaginal delivery, and
Abnormal 2nd Trimester Ultrasound, and
Malpresentation at admission
Then Probability of Emergency C-Section is 0.6

Over training data: 26/41 = .63,
Over test data: 12/20 = .60

Mining Databases


Speech Recognition


Control learning


Object recognition

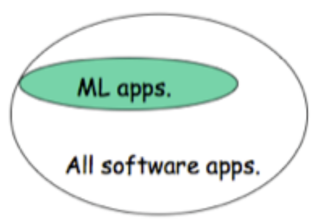
Text analysis

Peter H. van Oppen , Chairman of the Board & Chief Executive Officer
Mr. van Oppen has served as chairman of the board and chief executive officer of ADIC since its acquisition by Interpoint in 1994 and a director of ADIC since 1986. Until its acquisition by Crane Co. in October 1996, Mr. van Oppen served as chairman of the board of directors, president and chief executive officer of Interpoint . Prior to 1985, Mr. van Oppen worked as a consulting manager at Price Waterhouse LLP and at Bain & Company in Boston and London. He has additional experience in medical electronics and venture capital. Mr. van Oppen also serves as a director of Seattle FilmWorks Inc. and Spacelabs Medical, Inc.. He holds a B.A. from Whitman College and an M.B.A. from Harvard Business School, where he was a Baker Scholar.

Many more !

MACHINE LEARNING IN COMPUTER SCIENCE

- Machine learning is already the preferred approach for
  - Speech recognition, natural language processing
  - Computer vision
  - Medical outcome analysis
  - Robot control …



- Why growing ?
  - Improved machine learning algorithm
  - Increased data capture, new sensors, networking
  - Systems/Software too complex to control manually
  - Demand to self-customization for user, environment, ….

RELATED DISCIPLINES

- Artificial Intelligence
- Data Mining
- Probability and Statistics
- Information theory
- Numerical optimization
- Computational complexity theory
- Control theory (adaptive)
- Psychology (developmental, cognitive)
- Neurobiology
- Linguistics
- Philosophy

What are the goals of AI research?

Artifacts that THINK like HUMANS	Artifacts that THINK RATIONALLY
Artifacts that ACT like HUMANS	Artifacts that ACT RATIONALLY

From: M.A. Papalaskar

How can we build more intelligent computer / machine ?

- Able to
  - **perceive the world**
  - **understand the world**

- This needs
  - Basic speech capabilities
  - Basic vision capabilities
  - Language/semantic understanding
  - User behavior / emotion understanding
  - Able to think ??

How can we build more intelligent computer / machine ?

R2-D2 and C-3PO

@ Star Wars – 1977

to serve human beings, and fluent in "over six million forms of communication"

How can we build more intelligent computer / machine ?

Jeopardy Game
➔ Requires a Broad Knowledge Base

IBM Watson

➔ an artificial intelligence computer system capable of answering questions posed in natural language developed in IBM's DeepQA project.

IBM WATSON

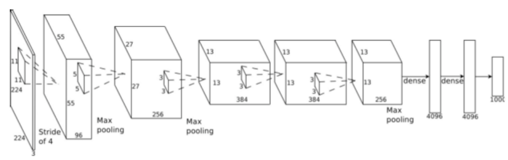
Dr. Yanjun Qi / UVA CS 6316 / f15

How can we build more intelligent computer / machine ?

Apple Siri ➔ an intelligent personal assistant and knowledge navigator

How may I help you, human?

How can we build more intelligent computer / machine ? : Objective Recognition / Image Labeling



72%, 2010

74%, 2011

85%, 2012
89%, 2013

Deep Convolution Neural Network (CNN) won (as Best systems) on "very large-scale" ImageNet competition 2012 / 2013 / 2014 (**training on 1.2 million images [X] vs.1000 different word labels [Y]**)

- 2013, Google Acquired Deep Neural Networks Company headed by Utoronto "Deep Learning" Professor Hinton
- 2013, Facebook Built New Artificial Intelligence Lab headed by NYU "Deep Learning" Professor LeCun

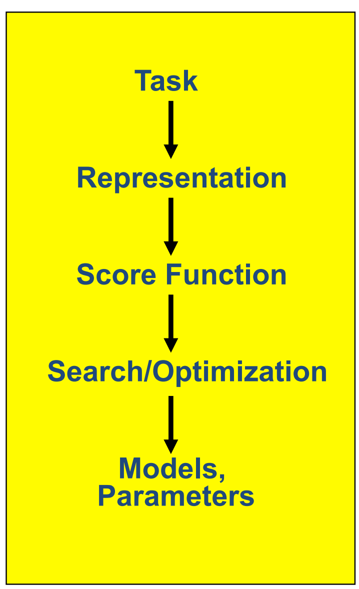
Detour: planned programming assignments

- HW3: Semantic language understanding (sentiment classification on movie review text)
- HW4: Visual object recognition (labeling images about handwritten digits)
- HW5: Audio speech recognition (HMM based speech recognition task)

Today Recap

- ❑ Course Logistics
- ❑ My background
- ❑ Basics and rough content plan
- ❑ Application and History

Next lesson: Machine Learning in a Nutshell



ML grew out of work in AI

Optimize a performance criterion using example data or past experience,

Aiming to generalize to unseen data

Next lesson: Review of linear algebra and basic calculus

References

- Prof. Andrew Moore's tutorials
- Prof. Raymond J. Mooney's slides
- Prof. Alexander Gray's slides
- Prof. Eric Xing's slides
- http://scikit-learn.org/
- Hastie, Trevor, et al. The elements of statistical learning. Vol. 2. No. 1. New York: Springer, 2009.
- Prof. M.A. Papalaskar's slides