

CS6510  
Applied Machine Learning

# Course Introduction

6 Aug 2016

Vineeth N Balasubramanian



# A few recent quotes

- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet” (Tony Tether, Director, DARPA)
- “Machine learning is the hot new thing” (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, ex-Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, CTO, Sun)
- “Machine learning is today’s discontinuity” (Jerry Yang, ex-CEO, Yahoo)

# What is Machine Learning?

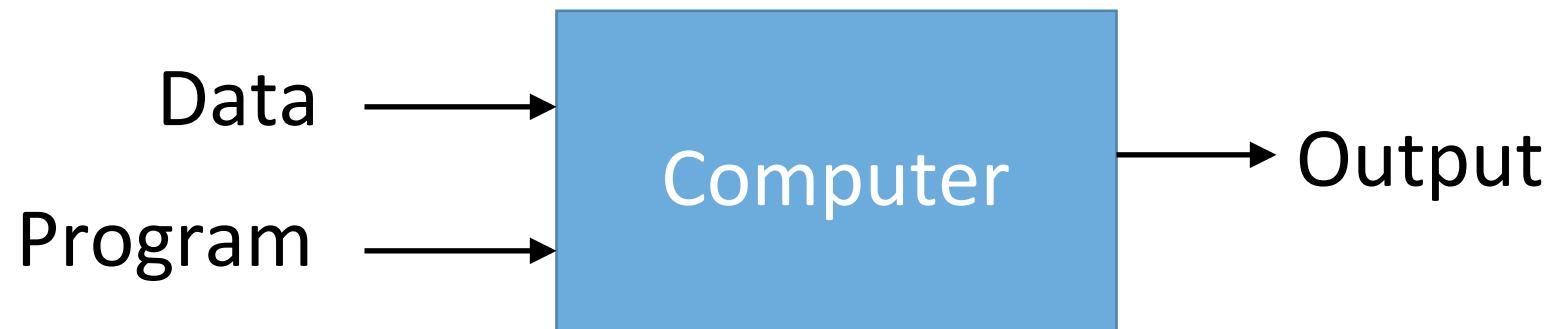


# What is Machine Learning?

- Making predictions or decisions from data
- “Programming computers to optimize a performance criterion using example data or past experience” (Ethem Alpaydin, Machine Learning, 2010)
- “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.” (Tom Mitchell, Machine Learning, 1997)
- “Learning general models from a data of particular examples”
- “Build a model that is *a good and useful approximation* to the data.”

# Today

## Traditional Programming



## Machine Learning



Source: Domingos

# Magic?

No, more like gardening

- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs



Source: Domingos

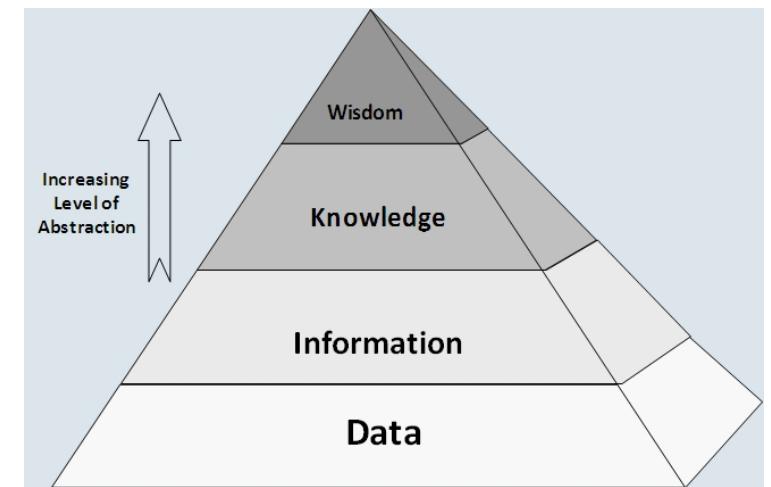
# Related Terms

# Machine Learning, Data Mining, Knowledge Discovery, Artificial Intelligence, Statistical Learning, Pattern Recognition, Computational Learning



# When is Machine Learning Used?

- Human expertise does not exist
  - E.g. navigating on Mars
- Humans are unable to explain their expertise
  - E.g. speech recognition
- Solution changes in time
  - E.g. routing on a computer network
- Solution needs to be adapted to particular cases
  - E.g. user biometrics
- Data is cheap and abundant; knowledge is expensive and scarce



# Applications of Machine Learning

From: cheapsales@buystufffromme.com  
To: ang@cs.stanford.edu  
Subject: Buy now!

Deal of the week! Buy now!  
Rolex w4tchs - \$100  
Med1cine (any kind) - \$50  
Also low cost M0rgages available.

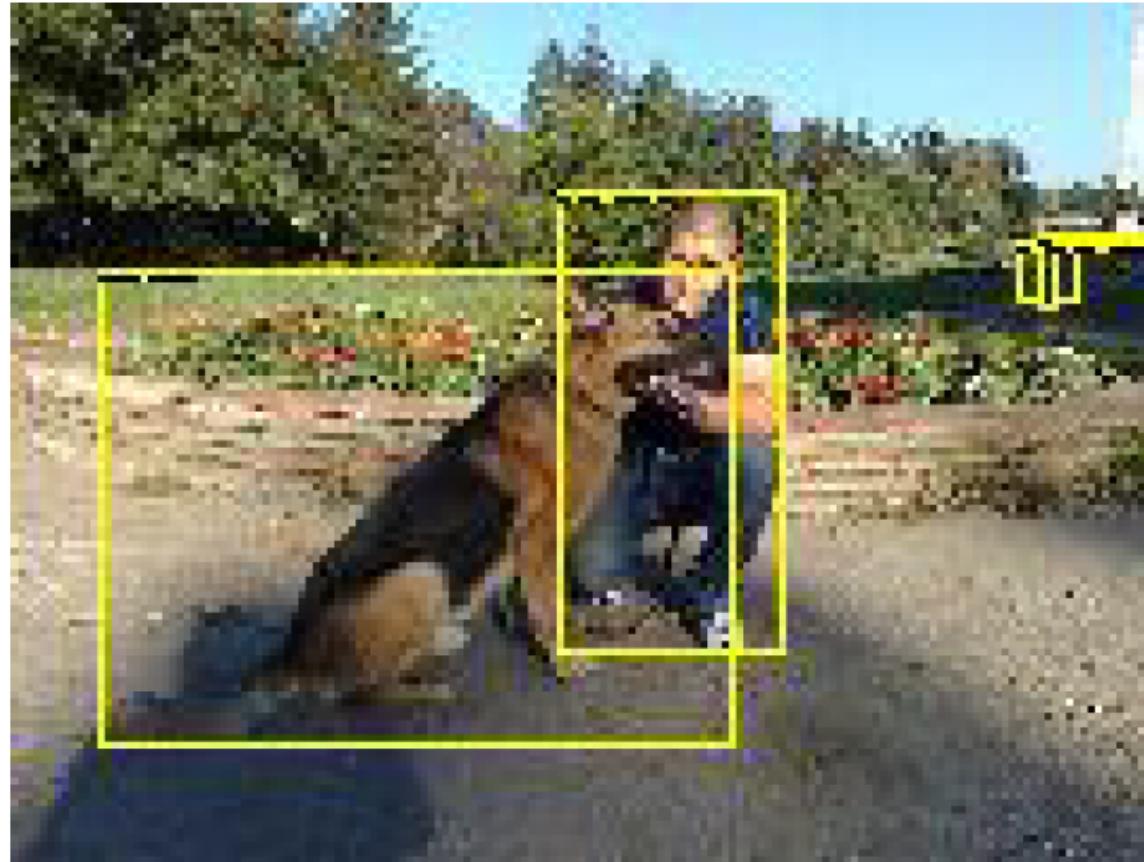
Spam

From: Alfred Ng  
To: ang@cs.stanford.edu  
Subject: Christmas dates?

Hey Andrew,  
Was talking to Mom about plans for Xmas. When do you get off work. Meet Dec 22?  
Alf

Non-spam

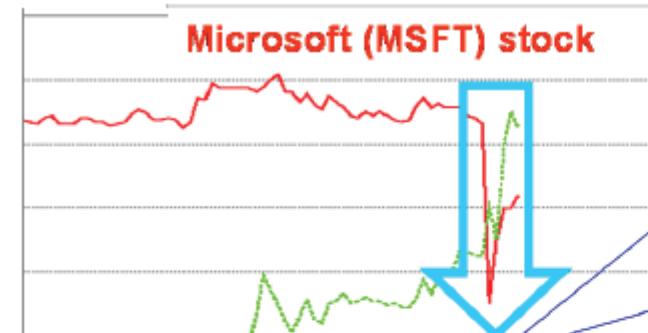
# Applications of Machine Learning



# Applications of Machine Learning

CCAGCTGCATCACAGGAGGCCAGCGAGCAGGTCTGTTCCAAGGGCCTCGAGCCAGTCTG EI  
GAGGTGAAGGACGTCTCCCCAGGAGCCGGTGAGAAGCGCAGTCGGGGCACGGGGATG EI  
TAAATTCTCTGTTAACACCTTCAGACTTATGTGTATGAAGGAGTAGAACCCAAA IE  
AAACTAAAGAATTATTCTTACATTCAGTTTCTGATCATGAAAACGCCAACAAAA IE  
AAAGCAGATCAGCTGTATAAACAGAAAATTATTCGTGGTTCTGTCACTGTGTGATGGT N  
TTGCCCTCAGCATCACCATGAACGGAGAGGCCATGCCCTGCGCTGAGGGCTGCCAGGCCA N

# Applications of Machine Learning



News:



Words like **Jackson** and **antitrust** are more likely in the stories preceding the plunge.

$P(\text{shares}) = 0.074$   
 $P(\text{antitrust}) = 0.009$   
 $P(\text{judge}) = 0.006$   
 $P(\text{trading}) = 0.032$   
 $P(\text{against}) = 0.025$   
 $P(\text{Jackson}) = 0.001$

Software giant Microsoft saw its shares dip a few percentage points this morning after U.S. District Judge Thomas Penfield Jackson issued his "findings of fact" in the government's ongoing antitrust case against the Seattle wealth-creation machine...



$P(\text{shares} \mid \text{MSFT}\downarrow) = 0.071$   
 $P(\text{antitrust} \mid \text{MSFT}\downarrow) = 0.044$   
 $P(\text{judge} \mid \text{MSFT}\downarrow) = 0.039$   
 $P(\text{trading} \mid \text{MSFT}\downarrow) = 0.029$   
 $P(\text{against} \mid \text{MSFT}\downarrow) = 0.027$   
 $P(\text{Jackson} \mid \text{MSFT}\downarrow) = 0.025$

$$P(\text{MSFT}\downarrow \mid \text{Jackson}) = P(\text{Jackson} \mid \text{MSFT}\downarrow) P(\text{MSFT}\downarrow) / P(\text{Jackson})$$

# Applications of Machine Learning

The screenshot shows the Netflix homepage with a red header. At the top right is a "Member Sign In" button. Below the header are four navigation tabs: "Start Your 1 Month Free Trial", "How It Works", "Browse Selection", and "1 Month Free Trial Info". A large banner headline reads: "Unlimited TV episodes & movies instantly over the Internet plus unlimited DVDs by mail!". The main content area is divided into three sections:

- On your TV:** Shows icons for Wii, PS3, and XBOX360, each with a "Learn more >" link. It includes a note: "Connect devices like these to your Netflix account to watch instantly on your TV." An arrow points from these devices to a television screen displaying a movie interface.
- On your computer:** Shows a laptop icon with a "PLAY" button and a "Learn more >" link. It includes the text: "Watch instantly on your computer too!"
- DVDs by mail:** Shows a black mailbox with a red Netflix envelope inside, with a red arrow pointing to it. It includes the text: "Exchange as often as you want. No late fees - ever!"

Below these sections, there's a note: "Watch as often as you want, anytime you want." and a link: "See other devices that stream instantly from Netflix".

**FAQs**

- How does Netflix work?
- What is the selection like?
- How much does it cost?
- How many DVDs can I rent during my Free Trial?
- How fast will I get my DVDs?
- How long is the free trial?

**How does Netflix work?**

**Rent what you want**

Simply point and click to add movies & TV episodes to your list. Get DVDs by mail plus instantly watch movies (some new releases) & TV episodes (including current season) online on your PC or Mac or streamed instantly from Netflix over the Internet right to your TV via a Netflix ready device.

**Start Your 1 Month Free Trial**

Free trial offer details.

Email

Confirm Email

Password  4-10 characters

Confirm Password

# More ML Applications

- Science (Astronomy, neuroscience, medical imaging, bio-informatics)
- Environment (energy, climate, weather, resources)
- Retail (Intelligent stock control, demographic store placement)
- Manufacturing (Intelligent control, automated monitoring, detection methods)
- Security (Intelligent smoke alarms, fraud detection)
- Marketing (promotions, ...)
- Management (Scheduling, timetabling)
- Finance (credit scoring, risk analysis...)
- Web data (information retrieval, information extraction, ...)

# More Recent ML Applications

- AlphaGo!
- Automating Employee Access Control
- Identifying whales in ocean based on audio recordings
- Predict wait times for patients in emergency rooms
- Extract heart failure diagnosis criteria from free-text physician notes
- Predicting hospital readmissions
- Is (s)he a psychopath?

Source: <http://www.forbes.com/sites/85broads/2014/01/06/six-novel-machine-learning-applications/#6b6f9a9e67bf>



6-Aug-16

CS6510 - Applied Machine Learning

15

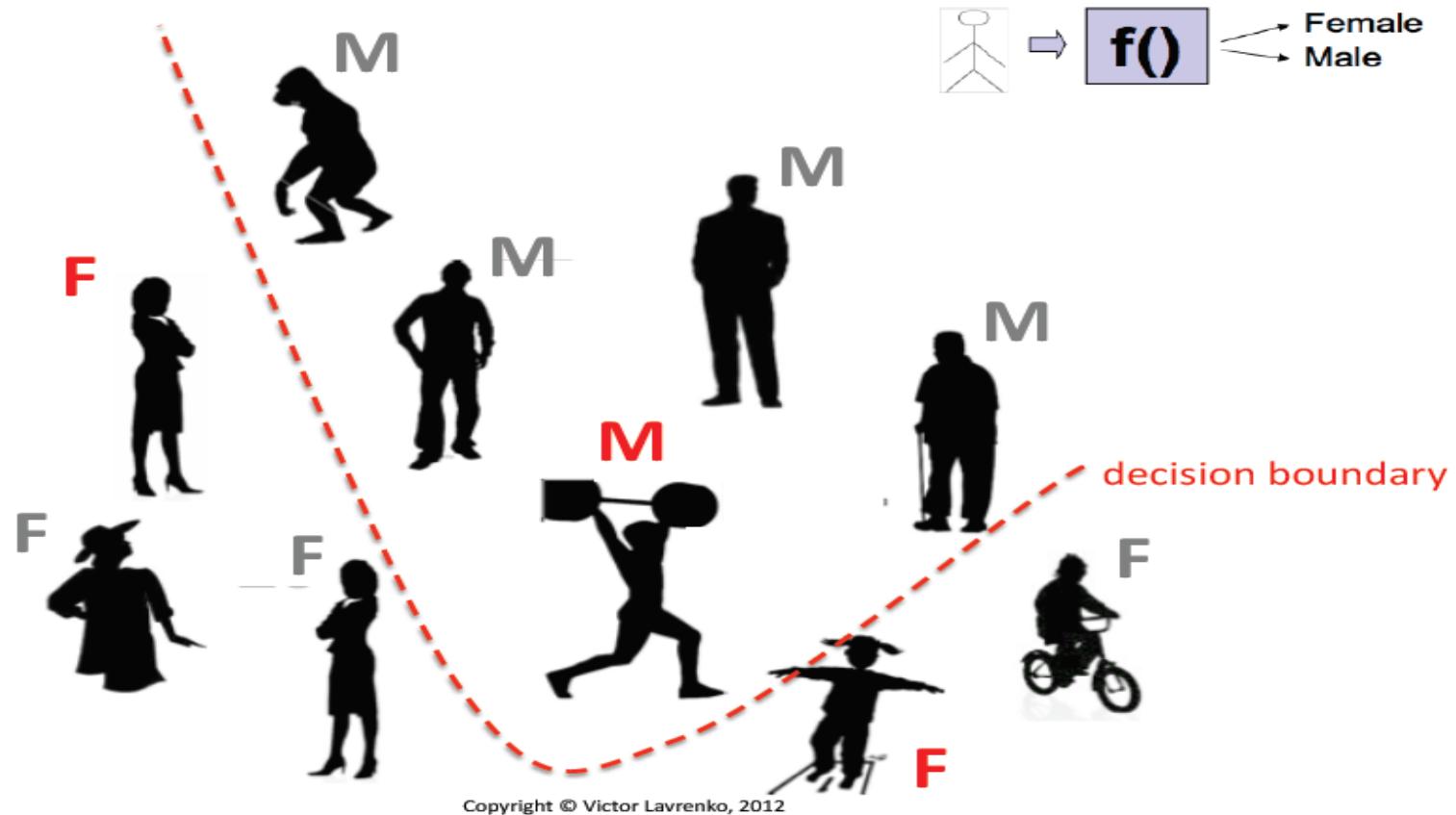
# When are ML algorithms not needed?

- When the relationships between all system variables (input, output, and hidden) is completely understood!
- This is NOT the case for almost any real system!

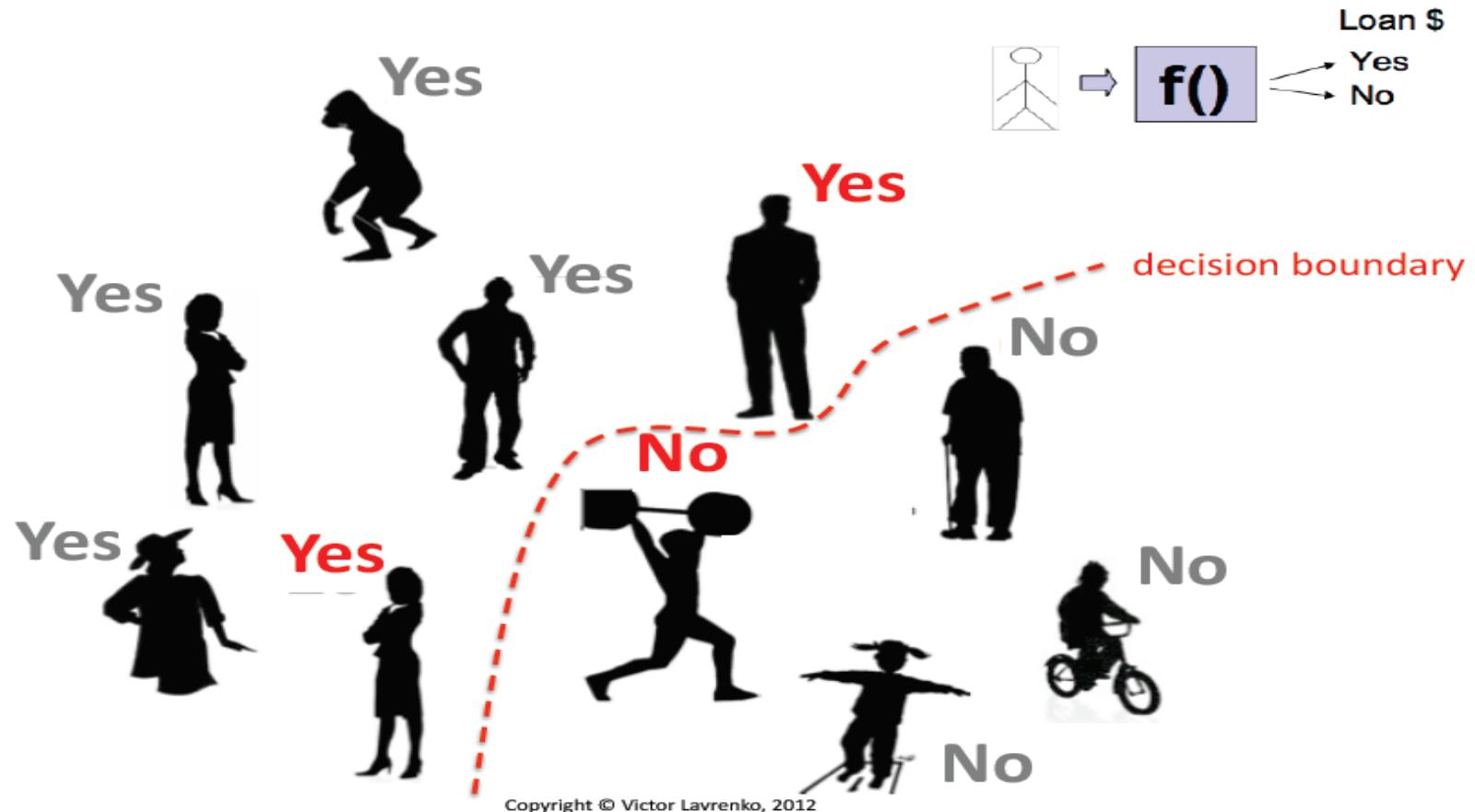
# Overview of ML

- Supervised learning
  - Predict an output  $y$  when given an input  $x$
  - For categorical  $y$  : classification.
  - For real-valued  $y$  : regression.
- Unsupervised learning
  - Create an internal representation of the input, e.g. clustering, dimensionality
  - This is important in machine learning as getting labels is often difficult and expensive
- Other areas of ML
  - Learning to predict structured objects (e.g., graphs, trees)
  - Reinforcement learning (learning from “rewards”)
  - Semi-supervised learning (combines supervised + unsupervised)

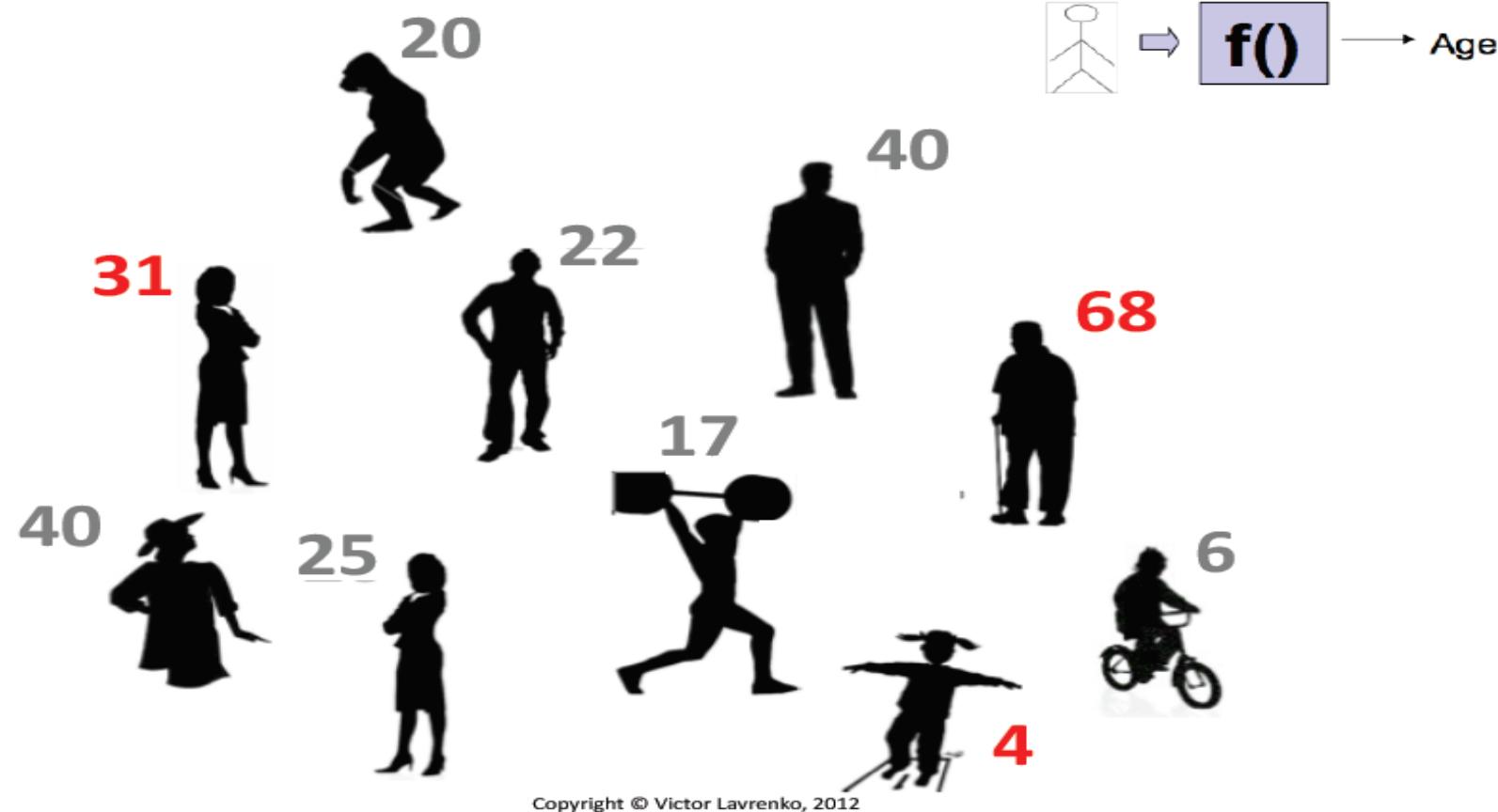
# Classification (Supervised Learning)



# Classification (Supervised Learning)



# Regression (Supervised Learning)



# Ranking (Supervised Learning)

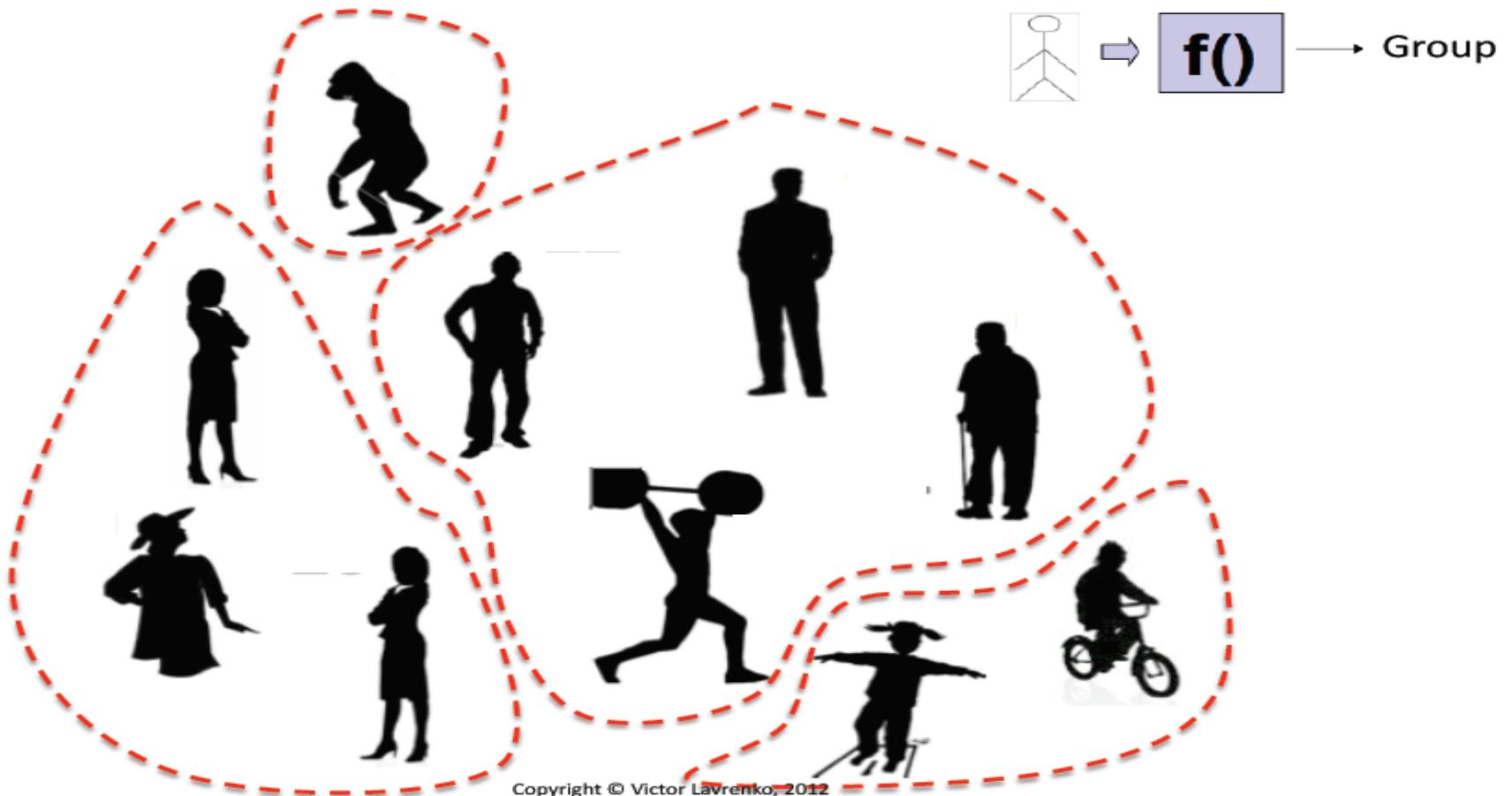
Given a query and a set of web pages, rank them according to relevance

Google search results for "machine learning":

- Machine learning - Wikipedia, the free encyclopedia**  
en.wikipedia.org/wiki/Machine\_learning ▾  
Machine learning, a branch of artificial intelligence, concerns the construction and study of systems that can learn from data. For example, a machine learning ...  
Artificial intelligence - Supervised learning - List of machine learning ... - Weka  
Franck Demontcourt +1'd this
- CS 229: Machine Learning**  
cs229.stanford.edu/ ▾  
Check out this year's awesome projects at Fall 2012 Projects. Come check out the cool new projects during the CS229 Poster Session this Thursday December ...  
You've visited this page 2 times. Last visit: 8/14/13
- Machine Learning | Coursera**  
https://www.coursera.org/course/ml ▾  
Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving ...  
Franck Demontcourt and 3 other people +1'd this
- Machine Learning Department - Carnegie Mellon University**  
www.ml.cmu.edu/ ▾  
Large group with projects in robot learning, data mining for manufacturing and in multimedia databases, causal inference, and disclosure limitation.
- Machine Learning - MIT OpenCourseWare**  
ocw.mit.edu › Courses › Electrical Engineering and Computer Science ▾  
6.867 is an introductory course on machine learning which gives an overview of many concepts, techniques, and algorithms in machine learning, beginning with ...

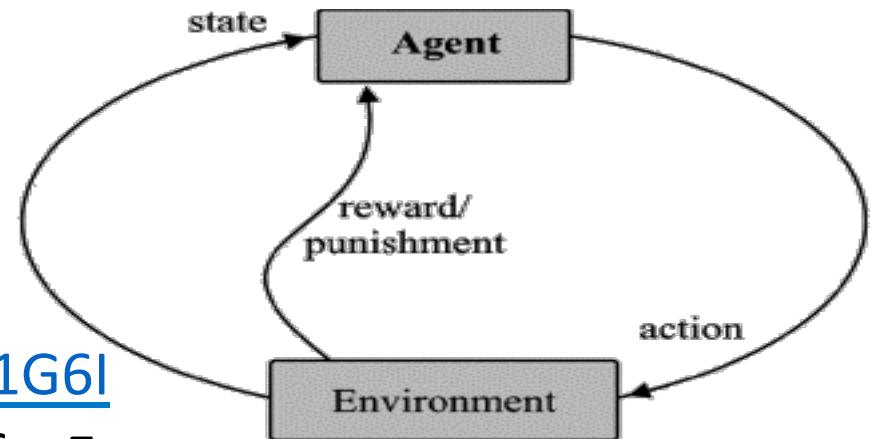
- Other applications
  - User preference, e.g. Netflix “My List” -- movie queue ranking
  - Flight search (search in general)
  - ...

# Clustering (Unsupervised Learning)



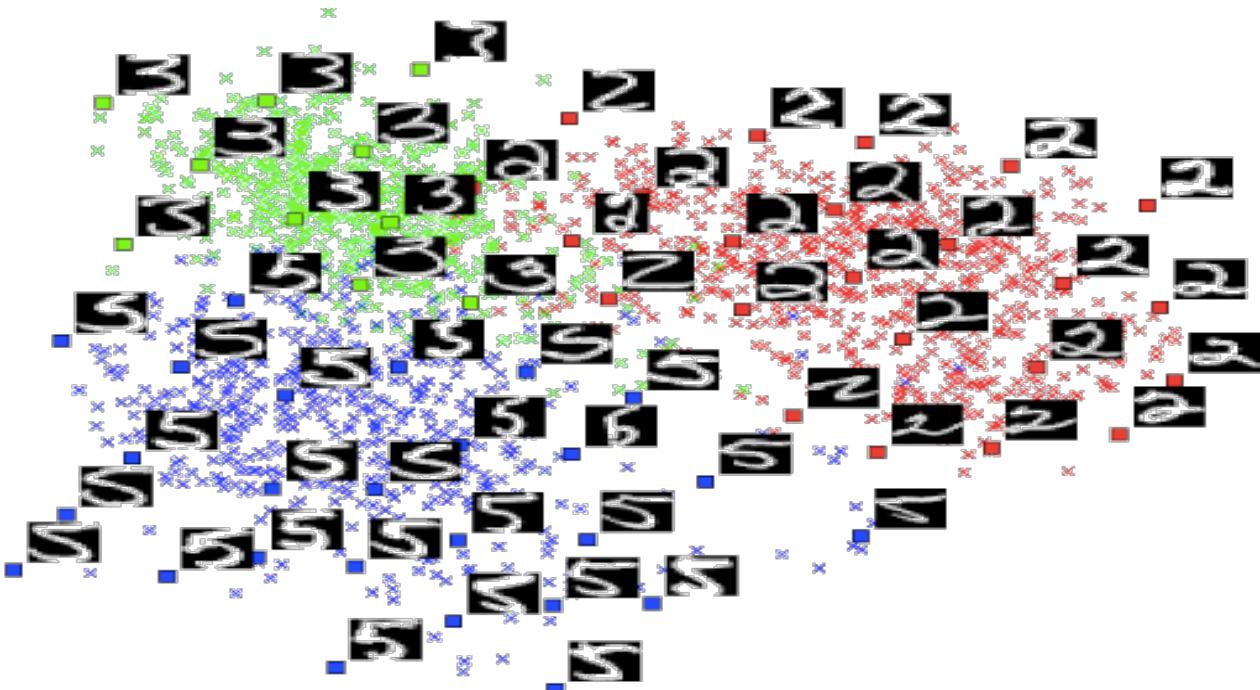
# Reinforcement Learning

- Learning a policy: A **sequence** of outputs
- No supervised output but delayed reward
  - E.g. Game playing
  - E.g. Robot in a maze
- Multiple agents, partial observability, ...
- Examples:
  - <https://www.youtube.com/watch?v=DCjbk4m1G6I>
  - <https://www.youtube.com/watch?v=VCdxqn0fcnE>



# Dimensionality Reduction

- Large sample size is required for high-dimensional data
- Query accuracy and efficiency degrade rapidly as the dimension increases
- Strategies
  - Feature reduction
  - Feature selection
  - Manifold learning
  - Kernel learning



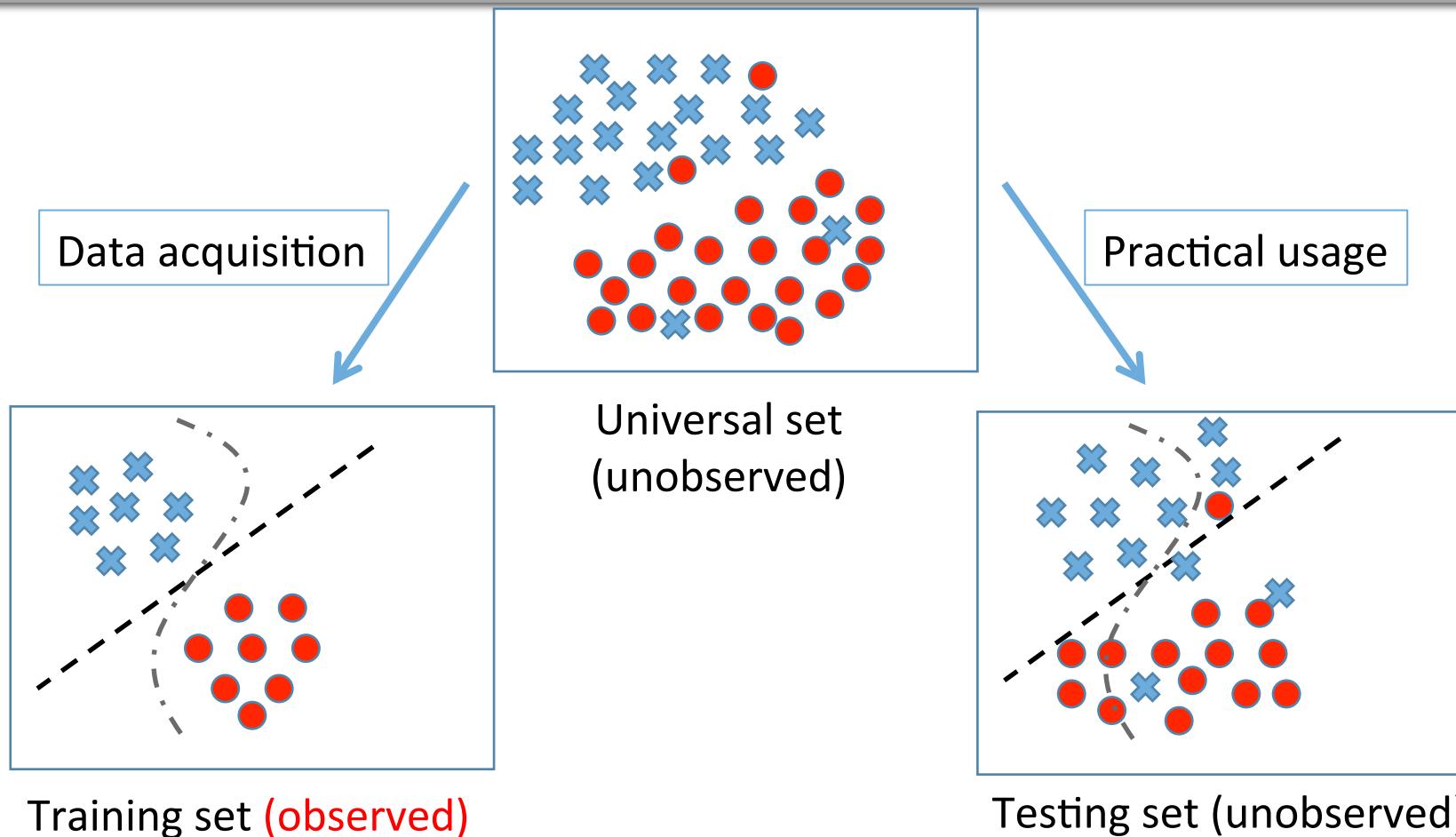
# ML Problems

	<i>Supervised Learning</i>	<i>Unsupervised Learning</i>
<i>Discrete</i>	classification or categorization	clustering
<i>Continuous</i>	regression	dimensionality reduction

# ML in Practice

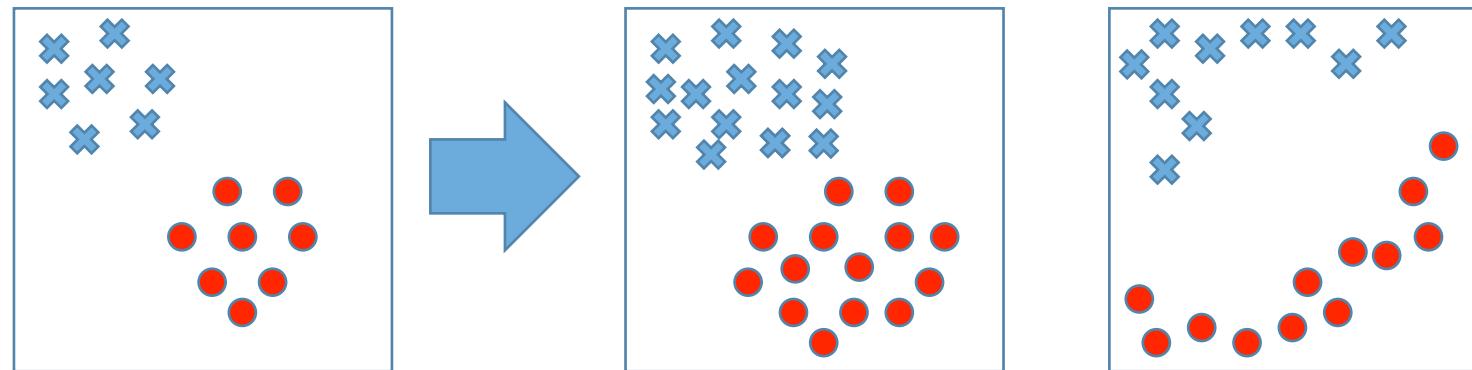
- Understanding domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learning models
- Interpreting results
- Consolidating and deploying discovered knowledge
- Loop

# Training and Testing ML Models



# Training and Testing ML Models

- Training is the process of making the system able to learn.
- No free lunch rule:
  - Training set and testing set come from the same distribution
  - Need to make some assumptions or bias



# Types of Models

- Inductive vs Transductive Learning
- Online vs Offline Learning
- Generative vs Discriminative Models
- Parametric vs Non-Parametric Models

# ML Datasets

- UCI Repository: <http://www.ics.uci.edu/~mlearn/MLRepository.html>
- Statlib: <http://lib.stat.cmu.edu/>

# ML Resources

- MOOCs
  - Coursera
- Conferences/Journals
  - JMLR, Machine Learning, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Pattern Analysis and Machine Intelligence, Annals of Statistics
  - ICML, NIPS, ACM SIG-KDD, IJCAI, AAAI, ICDM

# Mathematical Basis

- Functions, Logarithms and Exponentials
- Vectors, Dot Products, Orthogonality
- Matrices, Matrix Operations, Linear Transformations, Eigendecomposition
- Calculus, Differentiation, Integration
- Probability and Statistics
- Functional Analysis, Hilbert Spaces

# Course Details

- **Timings/Location:**
  - Sat 9:30 am – 12:30 pm
  - Until mid-October for EMDS students
  - Block-A LH1 (LH2 occasionally, if required)
- **Instructor:** Vineeth N B
  - Email: [vineethnb@iith.ac.in](mailto:vineethnb@iith.ac.in)
  - Office: Block-E, 324
- **TAs**
  - Arghya Pal, Supriya Pandhare, Mausam Jain, Akilesh B, Sahil Manocha, Hrishikesh Vaidya
- **Web:**
  - <http://www.iith.ac.in/~vineethnb/teaching/f2016/cs6510-aml.htm>
  - We will soon have a Google Classroom portal for class management

# Course Topics

- Introduction to Machine Learning: Overview and Applications
- Classification Methods
- Machine Learning System Design
- Regression Methods
- Clustering Methods
- Dimensionality Reduction Methods
- Feature Selection Methods
- Anomaly Detection Methods
- Advanced Methods Methods
  - Graphical Models, Deep Learning, New Settings (Active Learning, Transfer Learning, Structured Prediction, Multitask Learning, Multiple Instance Learning)

# Topics not to be covered (likely)

- Deep Learning (not in its “depth”, at least)
- Bayesian Networks (not in depth)
- Reinforcement Learning
- Learning Theory

# Course Evaluation Rubric

- 15%: Quizzes
- 25%: Homework (Theory + Programming)
- 30%: Competitive Coding Assignments (Challenge style)
- 30%: End-semester examination

- 5 grace days to start with – please use them wisely. May not apply to some deadlines, which will be pointed out.
- Best 80% (approx, tentative) of quizzes will be considered
- Not attempting the end-semester exam will result in an FR.

# Programming

- Python
- Libraries
  - Numpy, Scipy – numerical/scientific computing, linear algebra
  - Matplotlib – for plotting
  - Scikitlearn – for machine learning

# References

- Key References
  - Pattern Recognition and Machine Learning, by Christopher Bishop
  - Articles/Blogs/Papers/MOOCs online
- Other Recommended References
  - R. Duda, P. Hart & D. Stork, Pattern Classification (2nd ed.)
  - T. Mitchell, Machine Learning,

Appropriate reading materials for lectures will be posted

# Homework

- Go through remedial videos for math foundations at:
  - [https://www.youtube.com/channel/UC7gOYDYEgXG1ylH\\_rc2LgOw/playlists](https://www.youtube.com/channel/UC7gOYDYEgXG1ylH_rc2LgOw/playlists)
- Programming
  - Learn Python
    - <https://try.jupyter.org/>
    - <https://docs.python.org/3/tutorial/>
    - Video Tutorials: <https://www.youtube.com/watch?v=cpPG0bKHYKc>
  - Note of caution: Python 2.7 vs Python 3.4
    - [http://sebastianraschka.com/Articles/2014\\_python\\_2\\_3\\_key\\_diff.html](http://sebastianraschka.com/Articles/2014_python_2_3_key_diff.html)
  - Play with Numpy