

# PACMANN AI

## Learning Machine Learning

Adityo Sanjaya

PACMANN AI Head of Research

Email: [adityosanjaya3.14@gmail.com](mailto:adityosanjaya3.14@gmail.com)

+62 55777490099

This presentation will give you overview on what PACMANN AI learn and our learning style

## I. The Basic

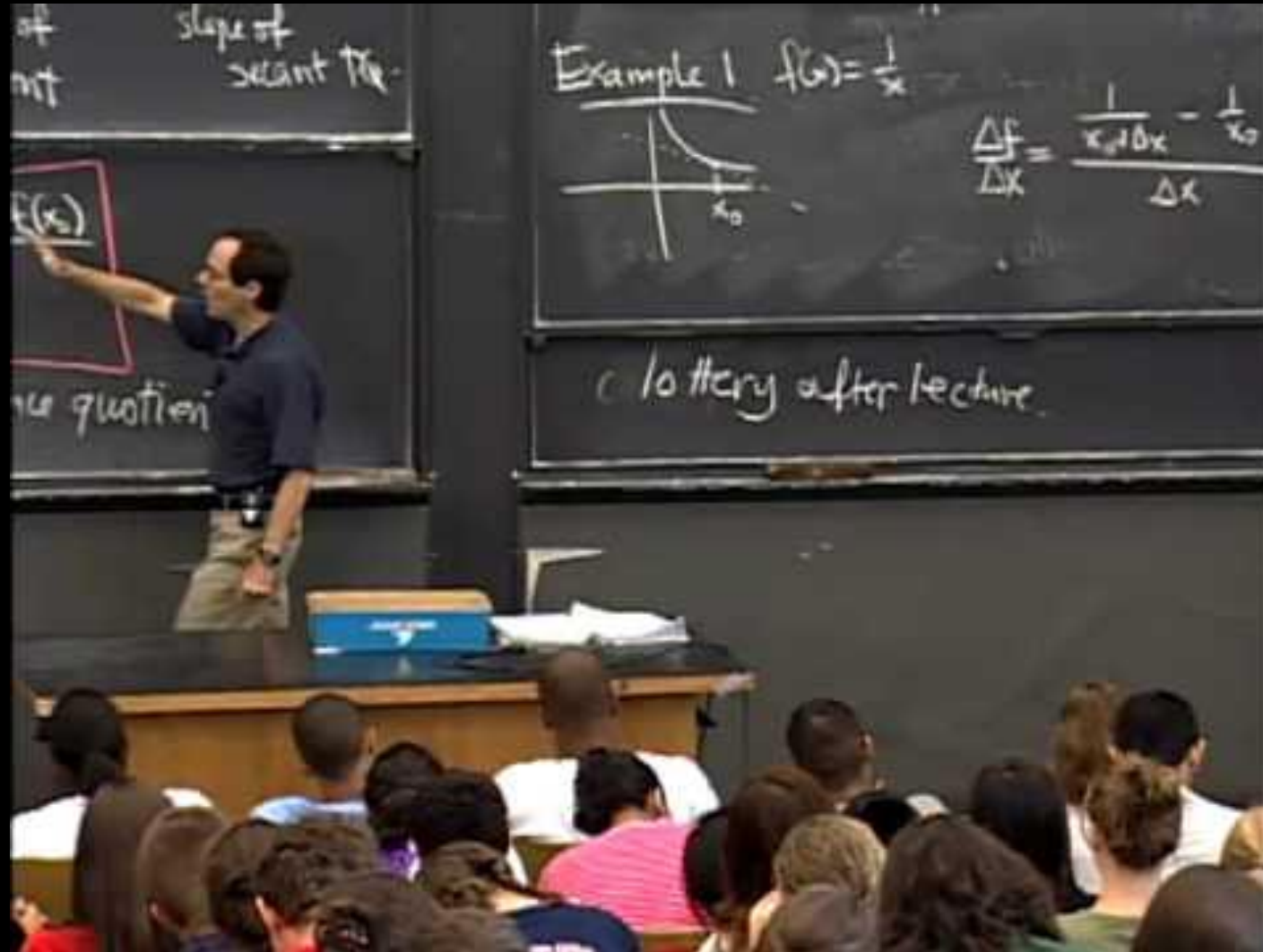
You want to learn the basic math and computer sciences foundation to learn Machine Learning

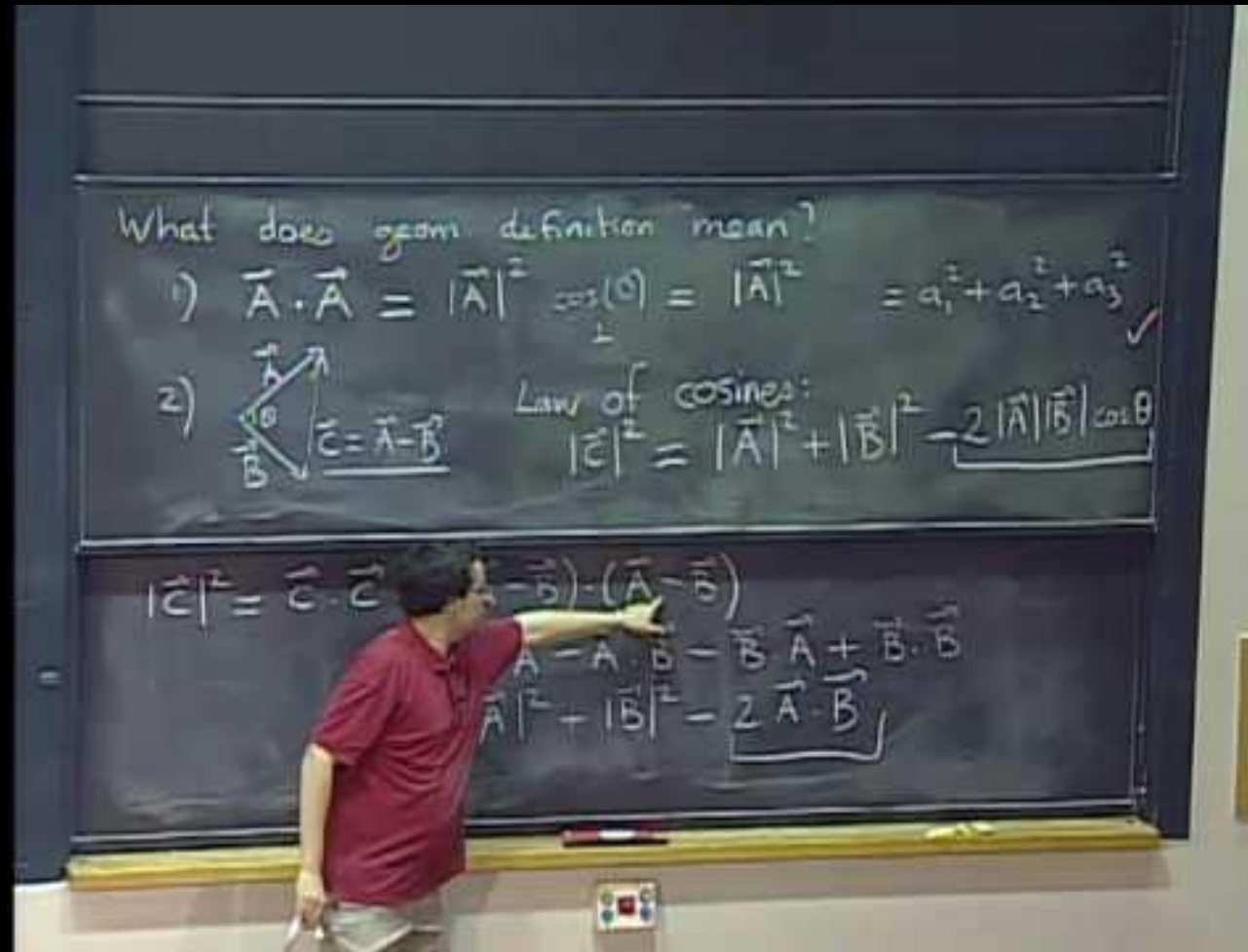


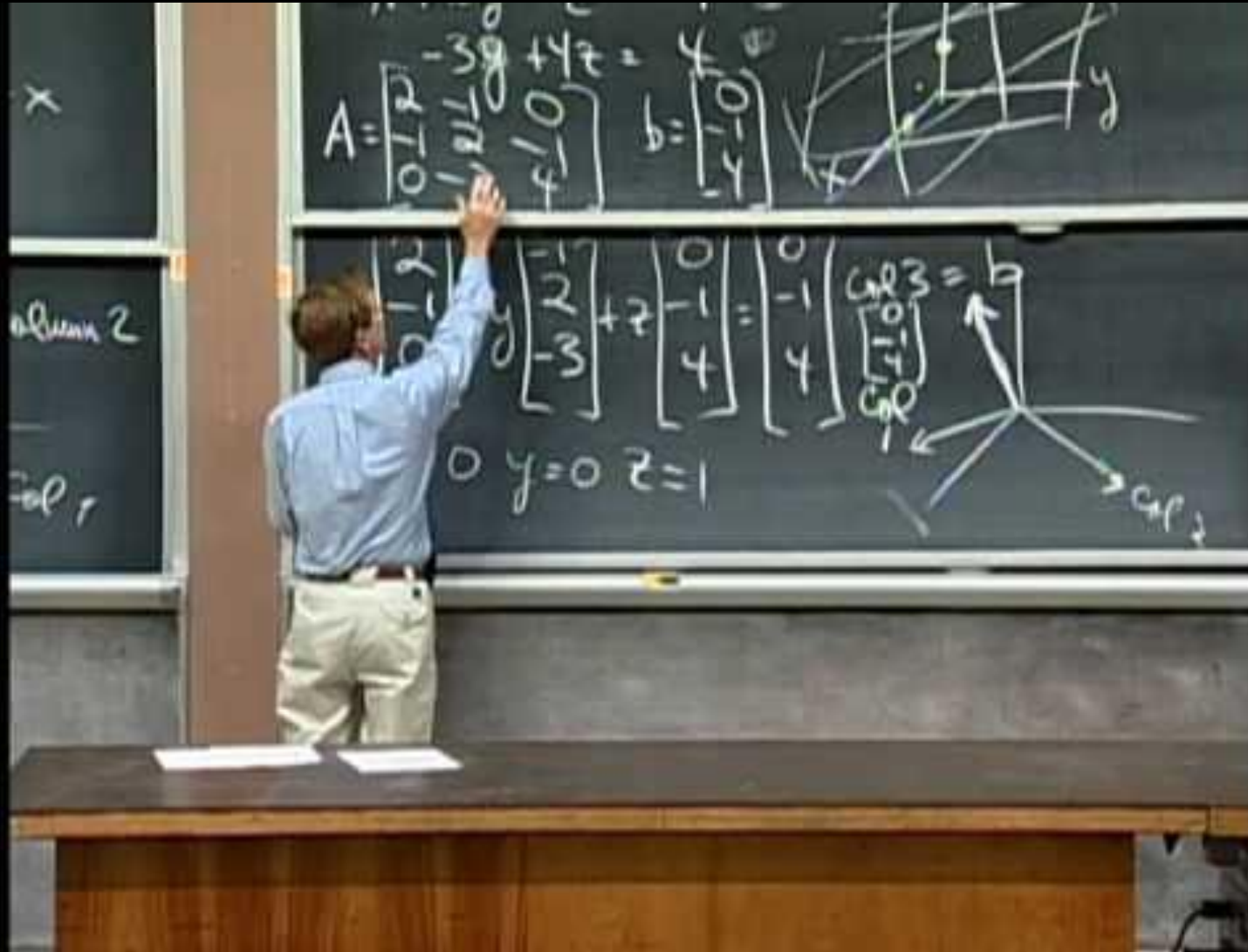
## I. The Noob and tWhiner

- Single Variable Calculus
- Multivariable Calculus
- Linear Algebra
- Probability and Statistics

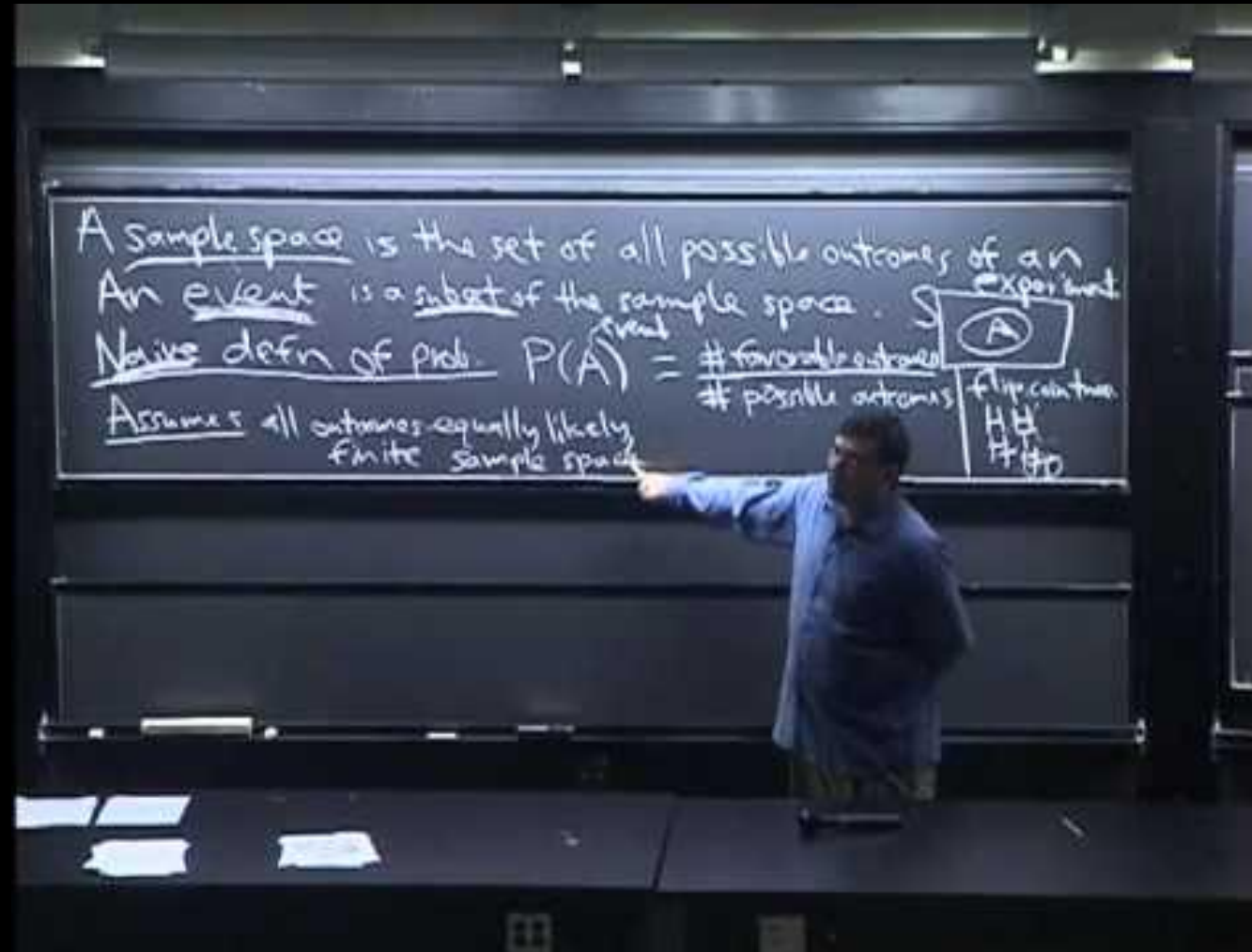
- Introduction to Computer Sciences
- Introduction to Statistical Learning

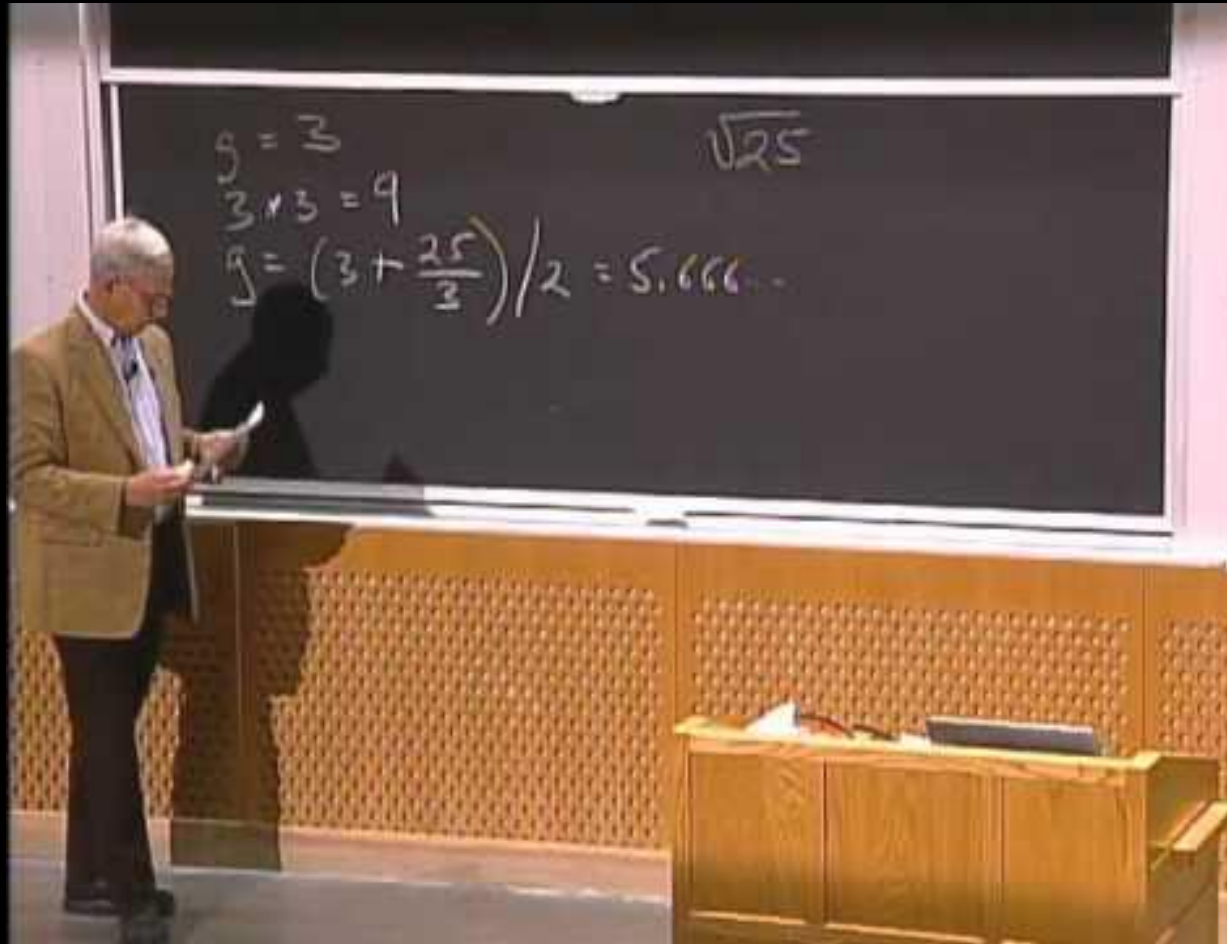








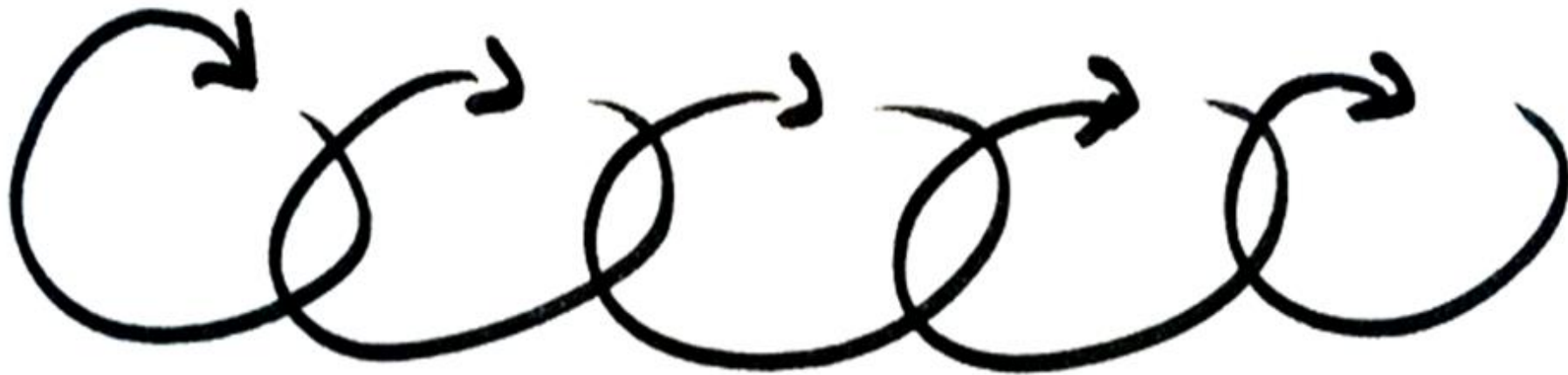




## Introduction to Statistical Learning (This Course)



## The Strategy



## The Strategy

### **Bootstrapping**

*“Relying entirely on one’s efforts and resources”*

dictionary.com







## The Strategy





## The Strategy






## The Strategy


### Welcome to Kaggle Competitions

Challenge yourself with real-world machine learning problems




#### New to Data Science?

Get started with a tutorial on our most popular competition for beginners, [Titanic: Machine Learning from Disaster](#).



#### Build a Model

Get the data & use whatever tools or methods you prefer to make predictions.



#### Make a Submission

Upload your prediction file for real-time scoring & a spot on the leaderboard.


✕ Dismiss

11 active competitions

Sort By Prize

Active All Entered

All Categories




### Data Science Bowl 2017

Can you improve lung cancer detection?

**Featured** · 2 months to go · 335 kernels

**\$1,000,000**  
949 teams



### The Nature Conservancy Fisheries Monitoring

Can you detect and classify species of fish?

**Featured** · 2 months to go · 233 kernels

**\$150,000**  
1,367 teams

## II. The Learner

- Convex Optimization

- Machine Learning
- Machine Learning
- Machine Learning
- Machine Learning
- Machine Learning

You want to learn all general machine learning course

## Machine Learning

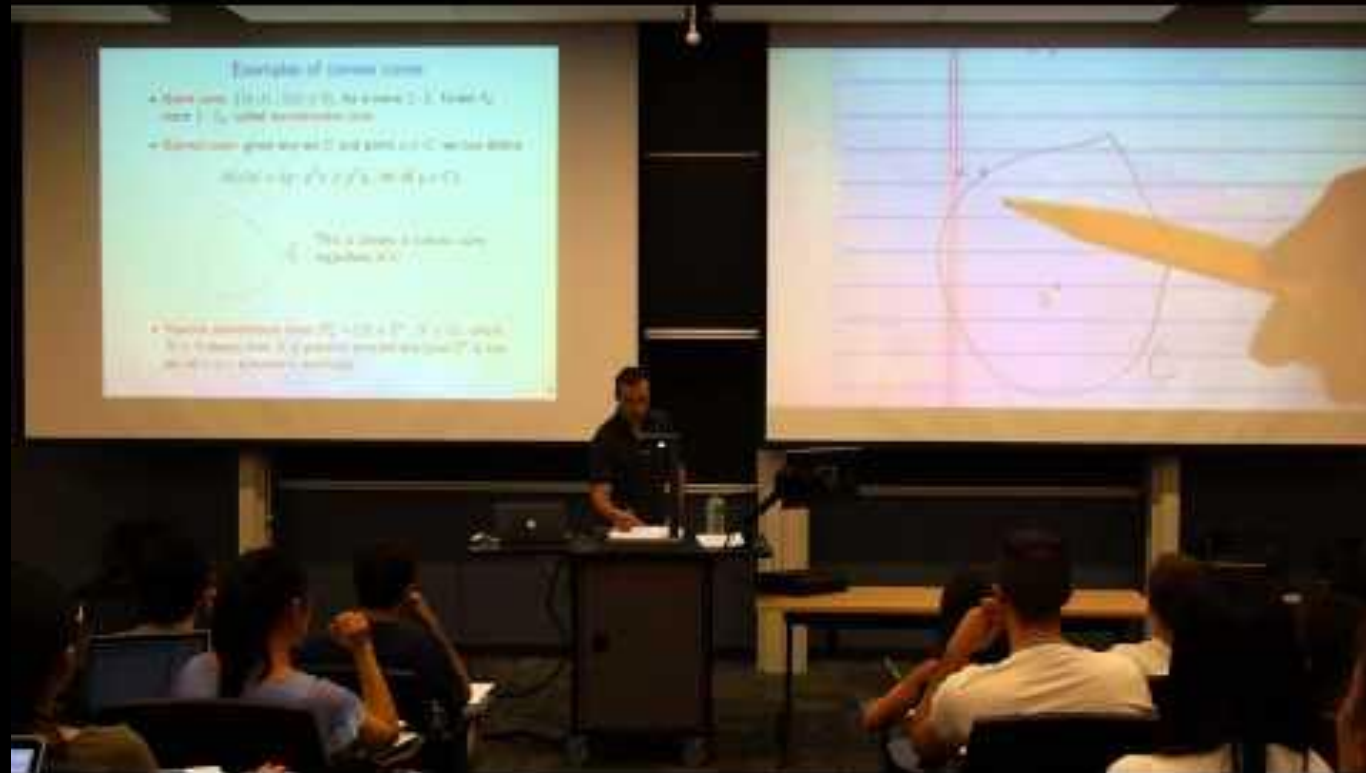
- Grew out of work in AI
- New capability for computers



Andrew Ng

Machine Learning Abu Mostafa, CalTech





Machine Learning Nando de Freitas, Oxford PhD Level

The image is a screenshot of a presentation slide displayed on a computer screen. The slide is titled "CPSC540" and "Machine Learning". It features two brain icons: one on the left with a colorful, abstract pattern inside, and one on the right with a blue, anatomical pattern. The slide is presented by Nando de Freitas at the University of British Columbia in January 2013. The presentation is titled "CPSC 540" and "Machine Learning". The slide is displayed in a window titled "CPSC 540 Machine Learning". The window is part of a presentation software interface, likely Beamer, showing a list of slides on the left and a navigation bar at the top. The system tray at the bottom shows the time as 12:41 PM on January 8, 2013.

CPSC540

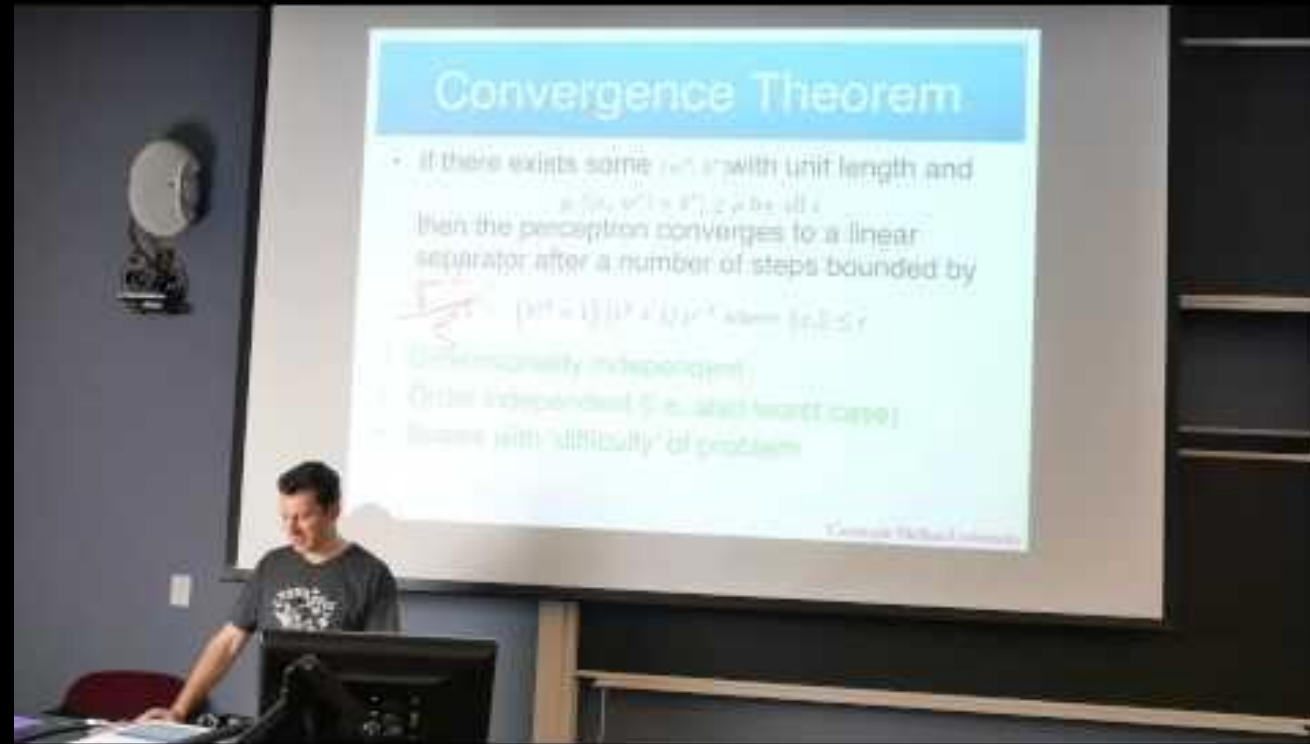
Machine Learning

UBC Nando de Freitas January 2013 University of British Columbia

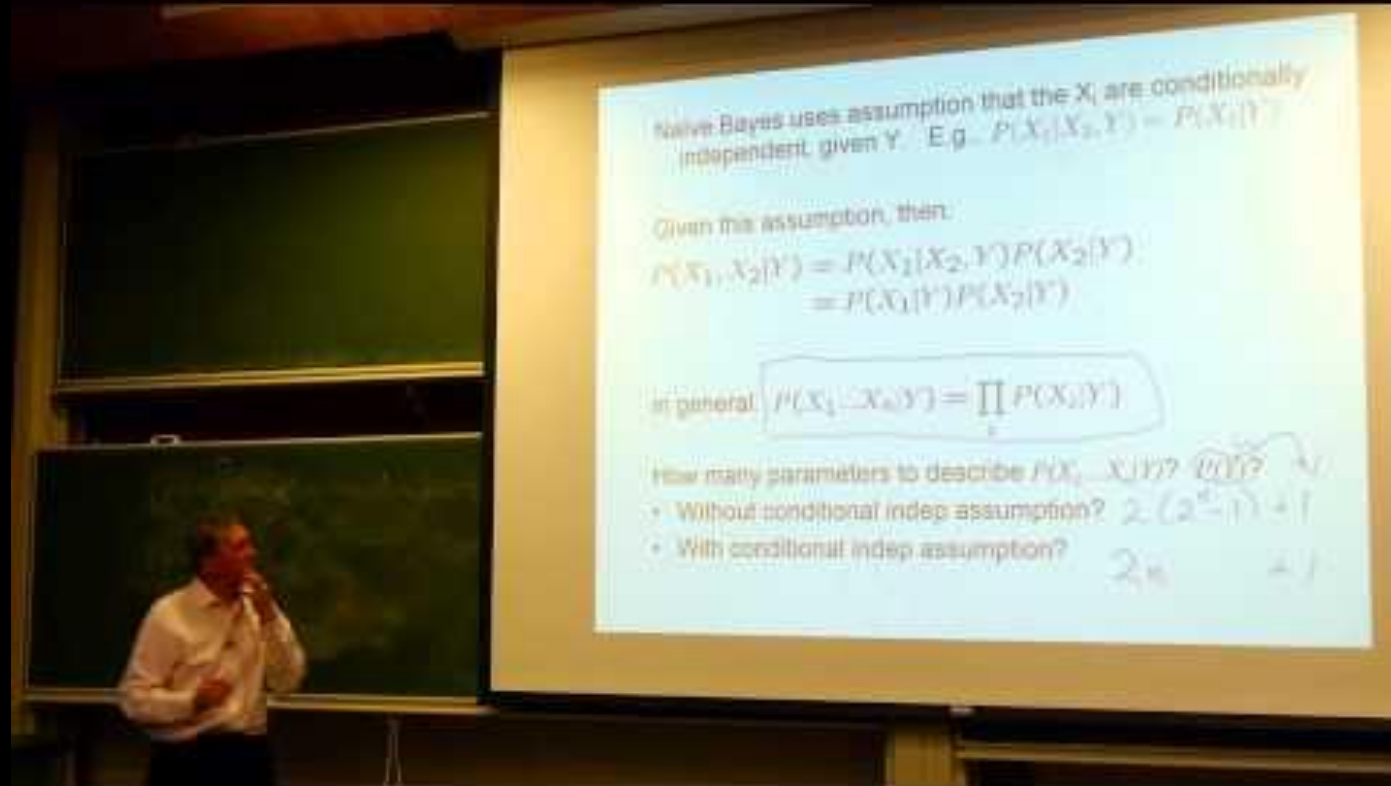
UBC Computer Science

CPSC 540 Nando De Freitas

Jan 08 2013 12:41







Naive Bayes uses assumption that the  $X_i$  are conditionally independent, given  $Y$ . E.g.  $P(X_1|X_2, Y) = P(X_1|Y)$

Given this assumption, then:

$$P(X_1, X_2|Y) = P(X_1|X_2, Y)P(X_2|Y) \\ = P(X_1|Y)P(X_2|Y)$$

in general:  $P(X_1, \dots, X_n|Y) = \prod_i P(X_i|Y)$

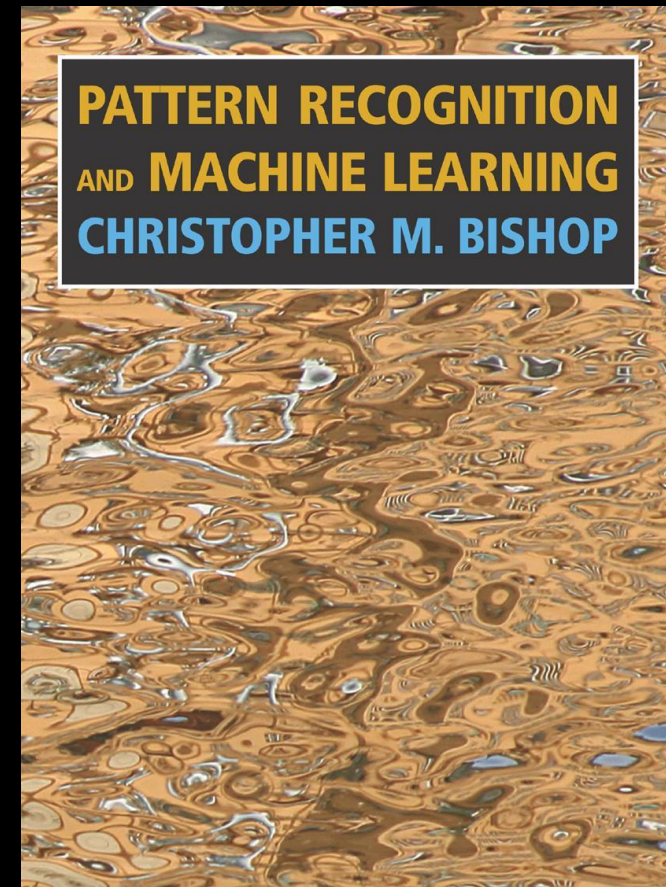
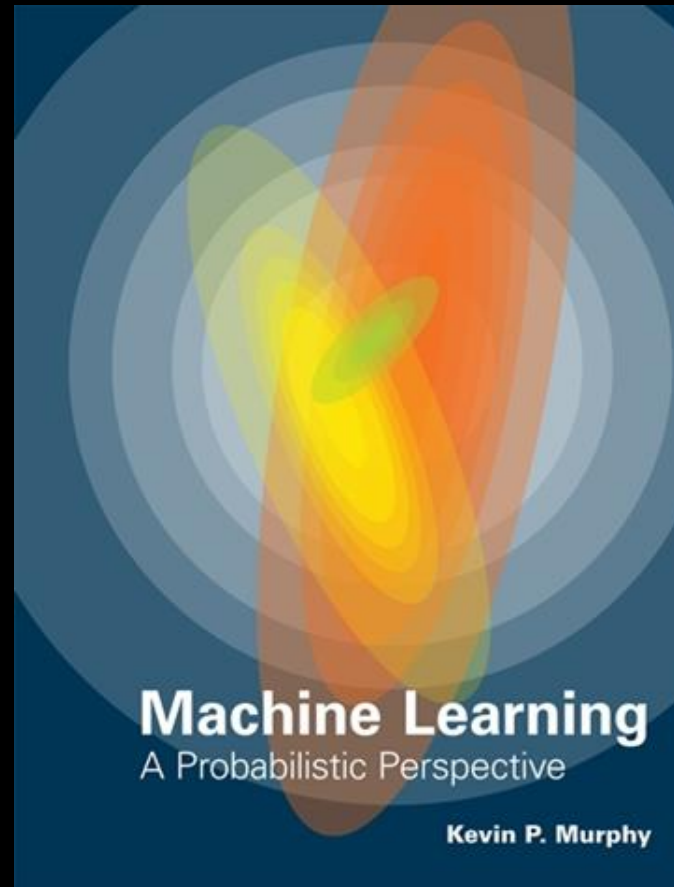
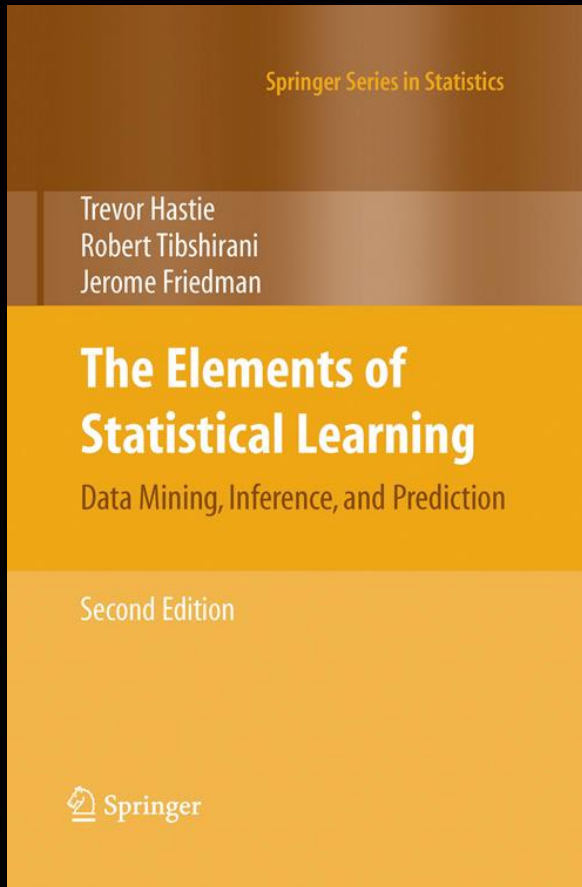
How many parameters to describe  $P(X_1, \dots, X_n|Y)$ ?  $2^{n+1}$

- Without conditional indep assumption?  $2(2^n - 1) + 1$
- With conditional indep assumption?  $2n + 1$

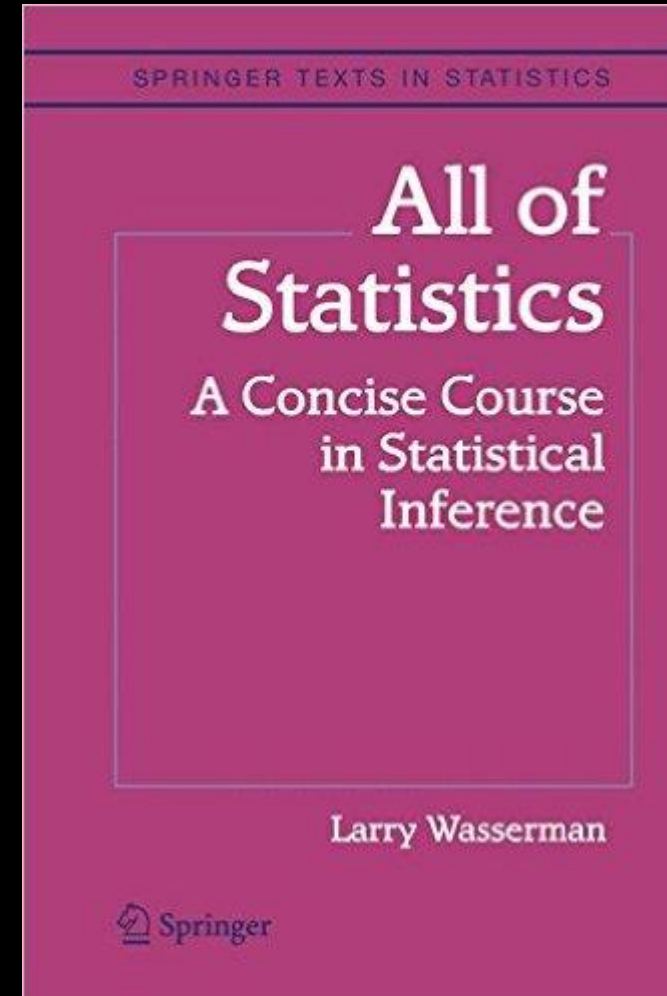
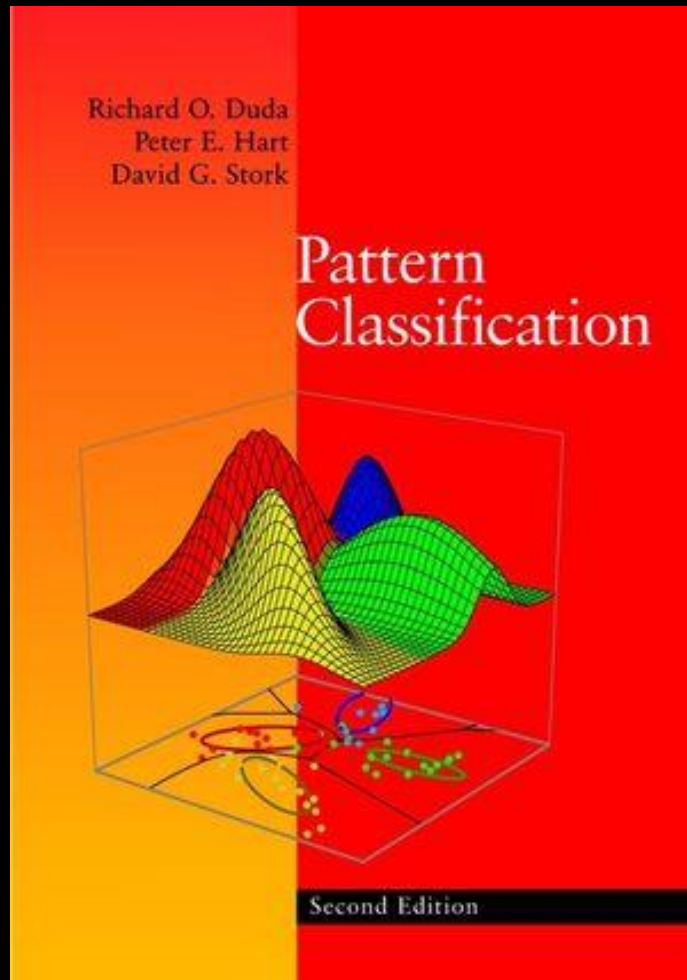
Advance Introduction to Machine Learning Alex Smola, PhD  
Level



## The Strategy



## The Strategy





## The Strategy

**SVM**

SVM - log reg.  
 SVM - Boosting  
 Regularization  
 SVM - multi-class

$(e_i, y_i)$   
 $e_i = d(x_i, M)$   
 $\frac{1}{1+e^2}$   
 $w^T x + b_0 = 1$   
 $w^T x + b_0 = 0$   
 $w^T x + b_0 = -1$

Kernel  $\rightarrow \sum \epsilon_i \leq C$   
 SVM Regression  $\epsilon_i \geq 1$   
 Convex Nonconvex

min  $\frac{1}{2} \|w\|^2 + C \sum \epsilon_i$   
 $\rightarrow \|w\|_2 = 1$  or  $\|w\|_2 = C$   
 $\epsilon = 201$  pos

Solution must satisfy KKT condition

$\alpha_i = 0 \leftarrow w = \sum_{i=1}^n (\alpha_i y_i x_i), \sum \alpha_i y_i = 0, \alpha_i \geq 0, \|w\|_2 = 1$

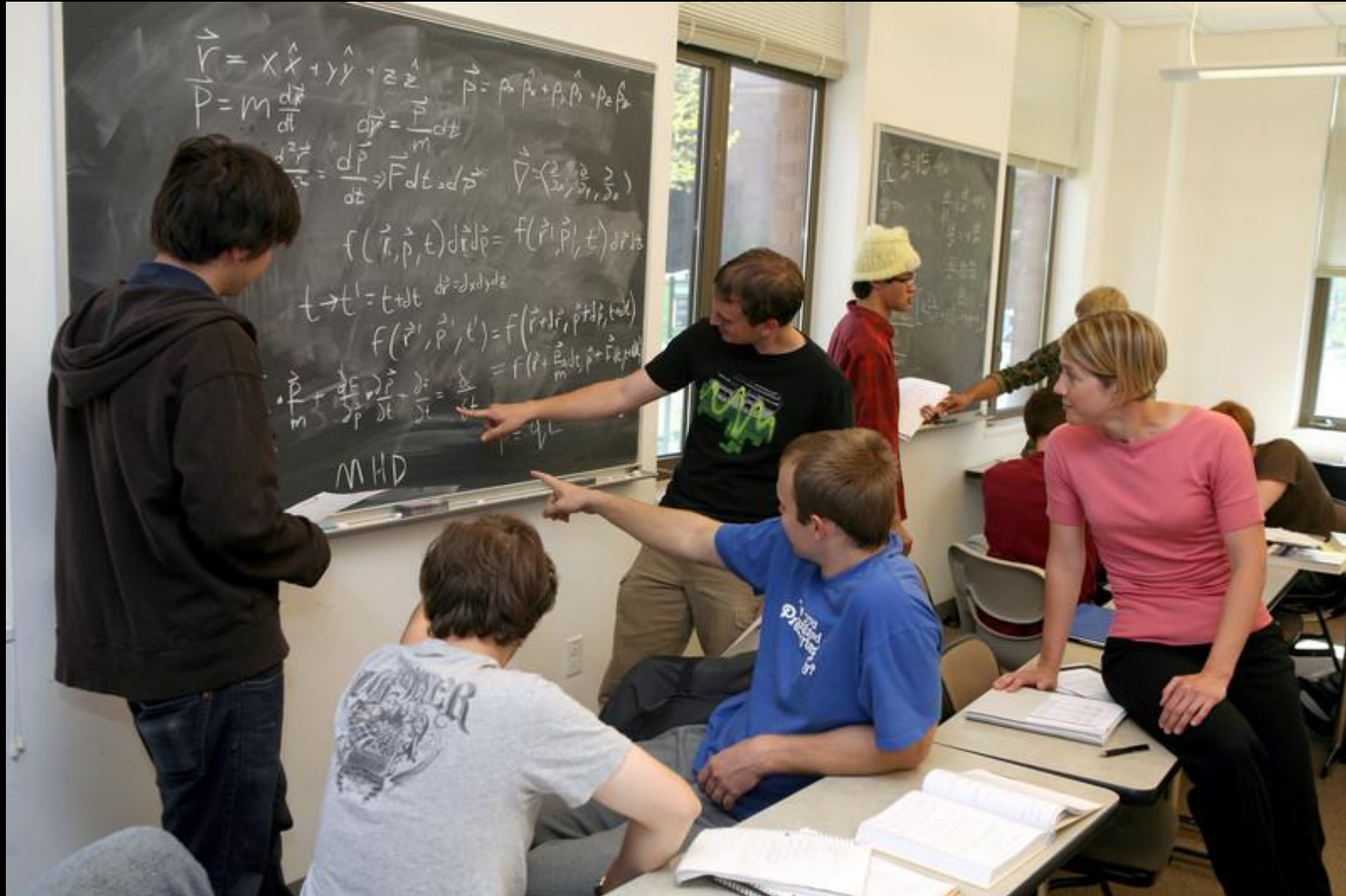
$\alpha_i \geq 0 \leftarrow \alpha_i [y_i (x^T w + b_0) - 1] = 0$   
 $[1 - y_i (x^T w + b_0)]$   
 $> 1$

BUANG  
 SAMP  
 DITEMPAT

## The Strategy



## The Strategy





## The Strategy

A graph is created on the fly

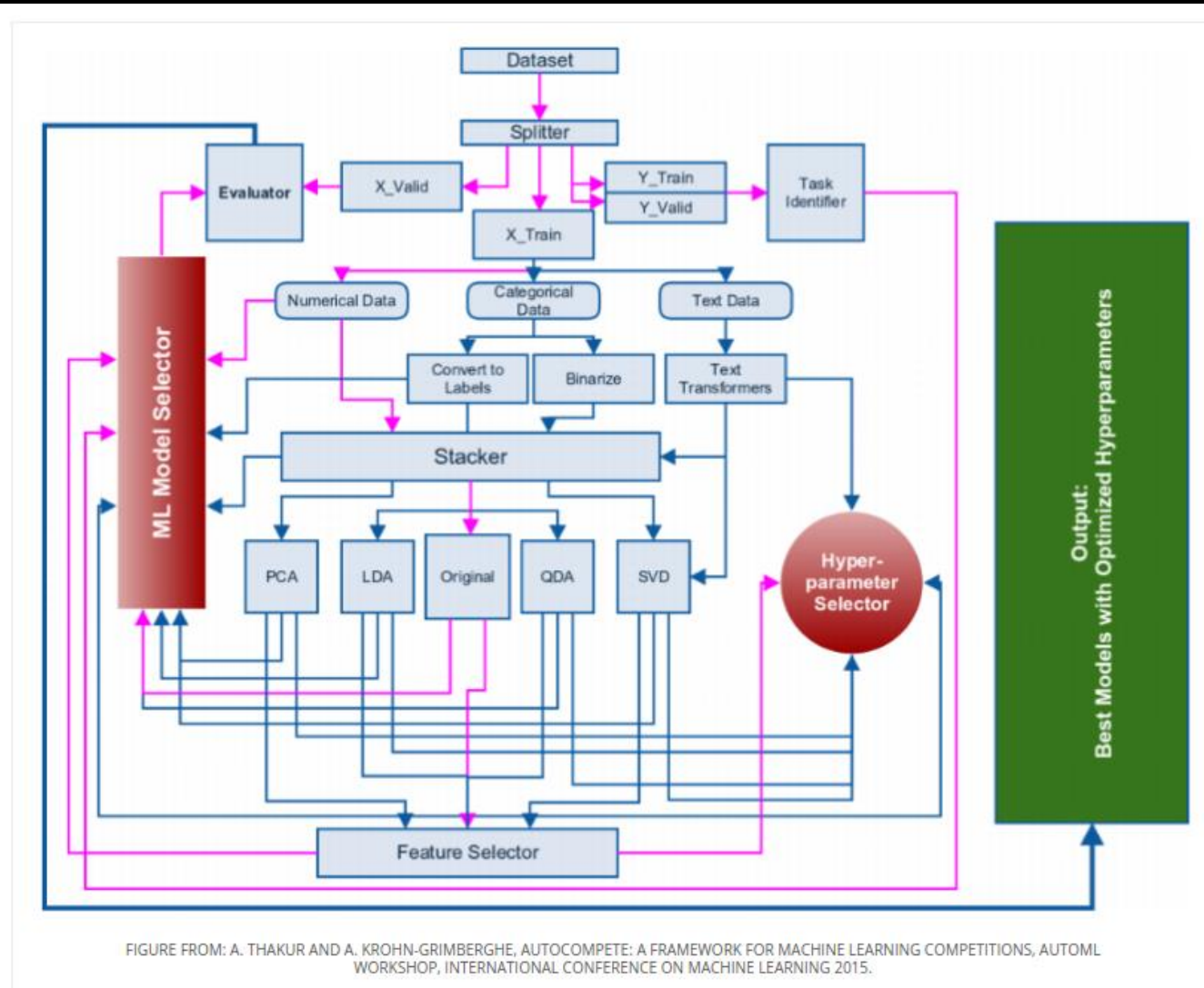
```
from torch.autograd import Variable

x = Variable(torch.randn(1, 10))
prev_h = Variable(torch.randn(1, 20))
W_h = Variable(torch.randn(20, 20))
W_x = Variable(torch.randn(20, 10))
```

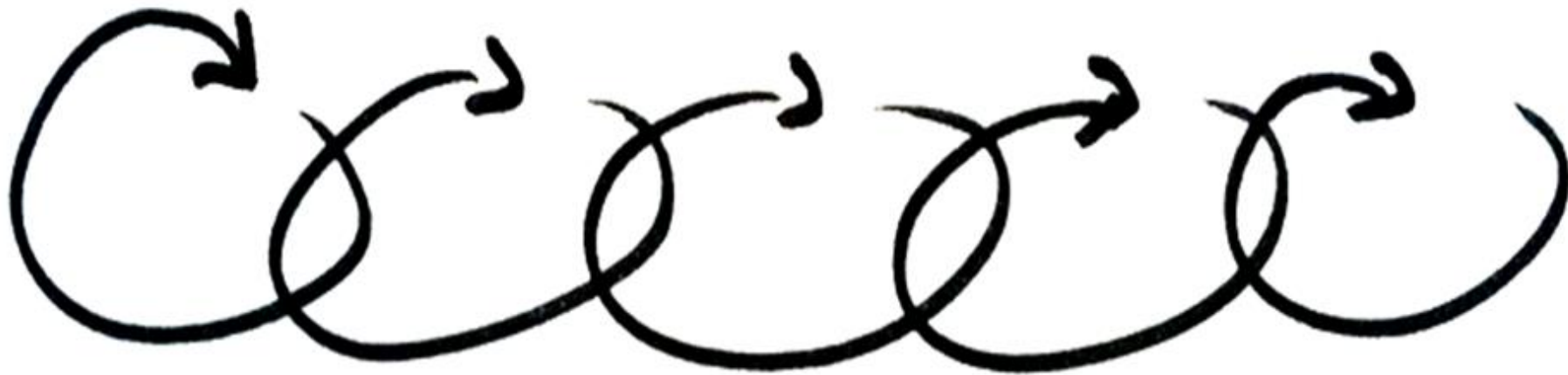




## The Strategy



## The Strategy





## III. The Theorist

- Theoretical Machine Learning
- Statistical Machine Learning
- Regularization Machine Learning
- Intermediate Statistics

You want to learn machine learning theory, so you can sleep well



Machine Learning Shai David, Waterloo PhD  
level



## Regularization Machine Learning Poggio, MIT PhD level


### Final Project

This final project can be

- a Wikipedia entry or
- problems for chapters of the textbook of the class or
- contributions to GURLS (GURLS: a Toolbox for Regularized Least Squares Learning) or
- a research project.

For the Wikipedia article we suggest to post 1-2 pages (short) using Wikipedia standard format (of course).

For the research project (either Application or Theory) you should use the template on the Web site.



The Center for Brains,  
Minds & Machines

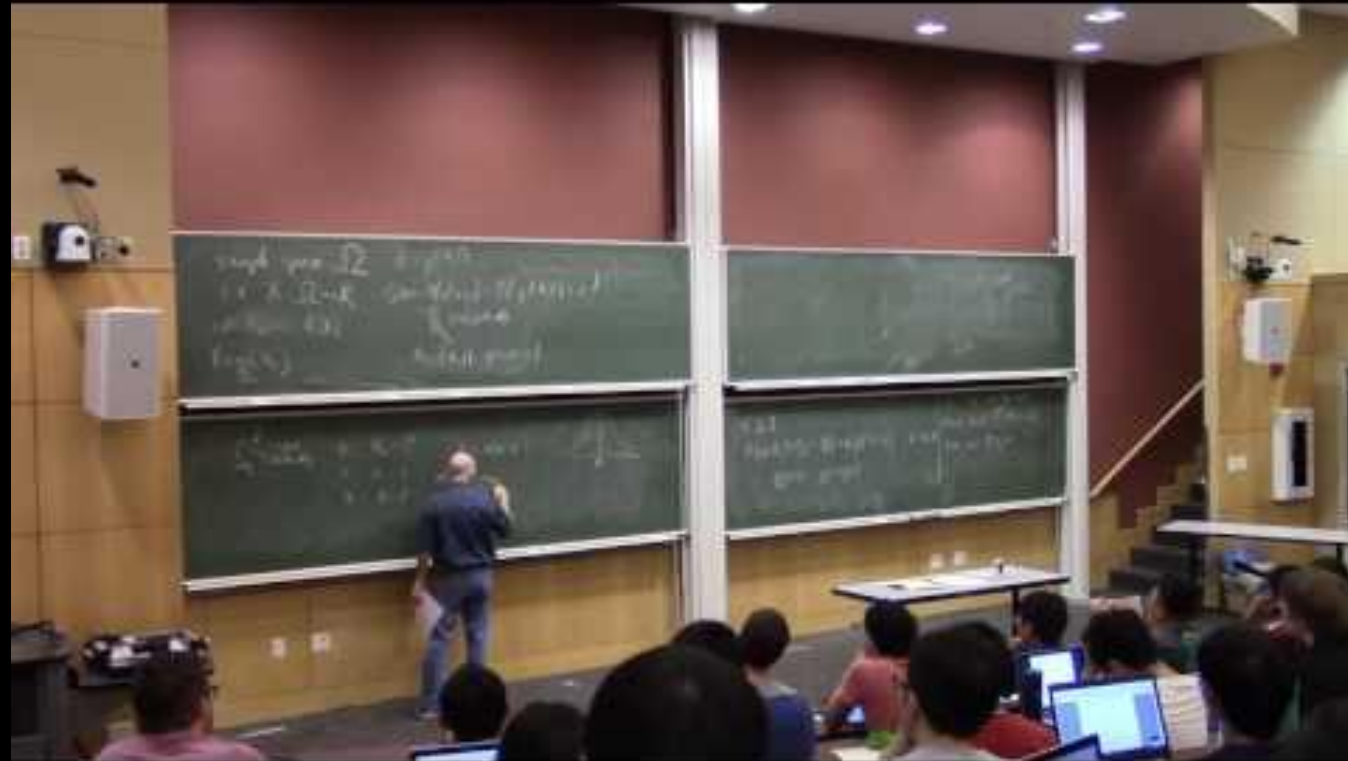
9.520 - Statistical Learning  
Theory and Applications

The Course at  
a Glance

Prof. Tommaso Poggio

September 9, 2015

Intermediate Statistics, Larry Wasserman Carnegie Mellon PhD level

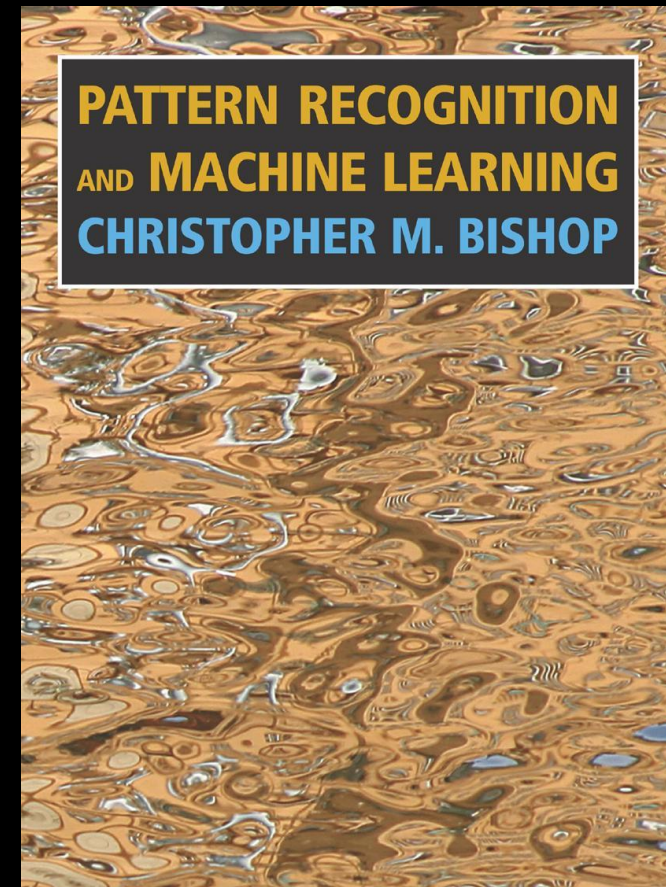
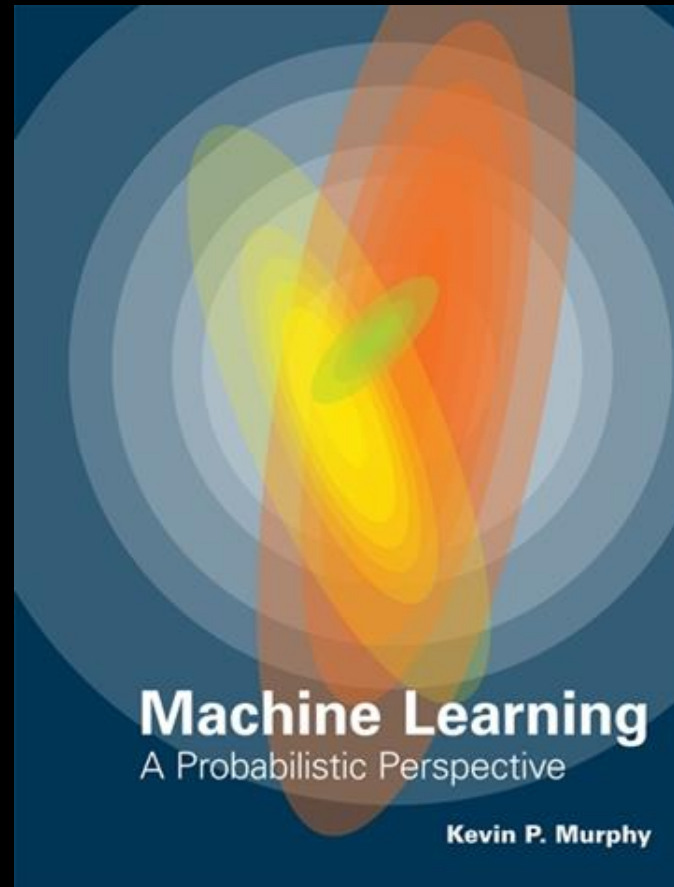
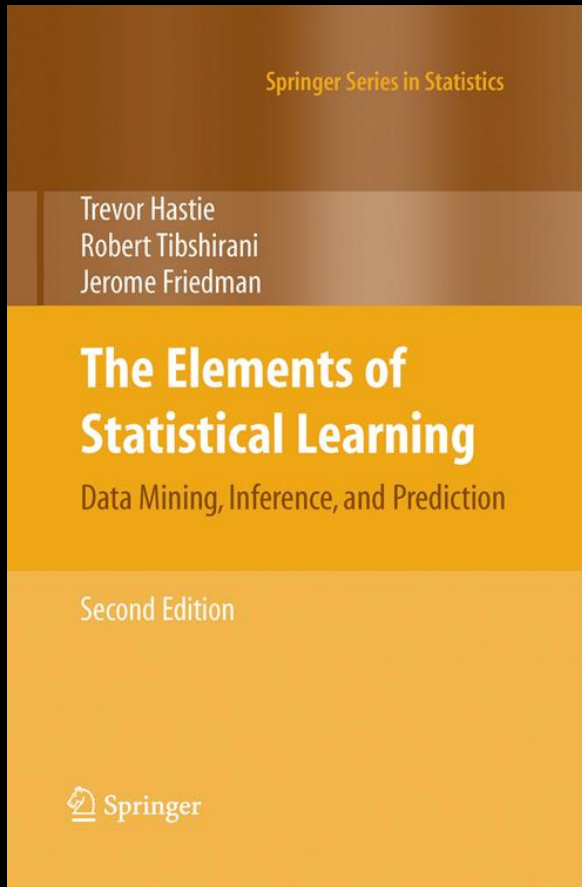




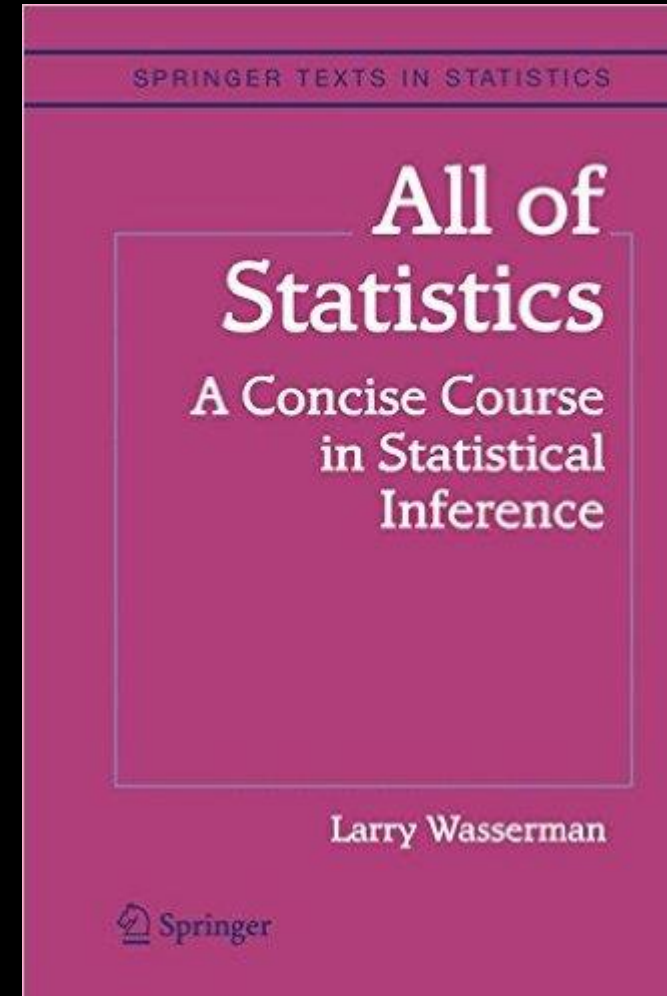
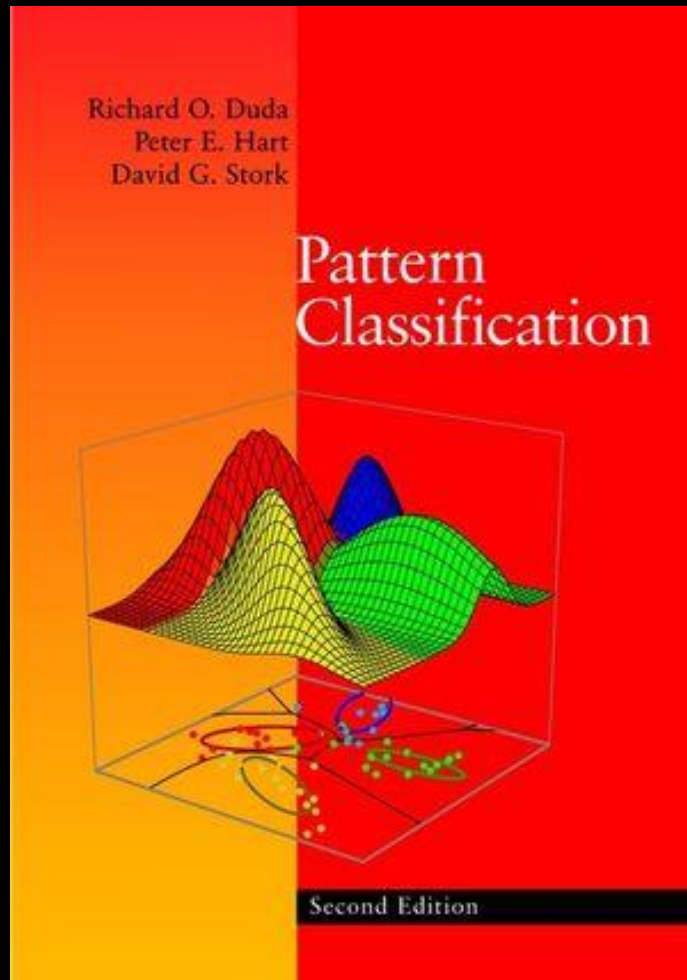
Statistical Machine Learning, Larry Wasserman Carnegie  
Mellon PhD level



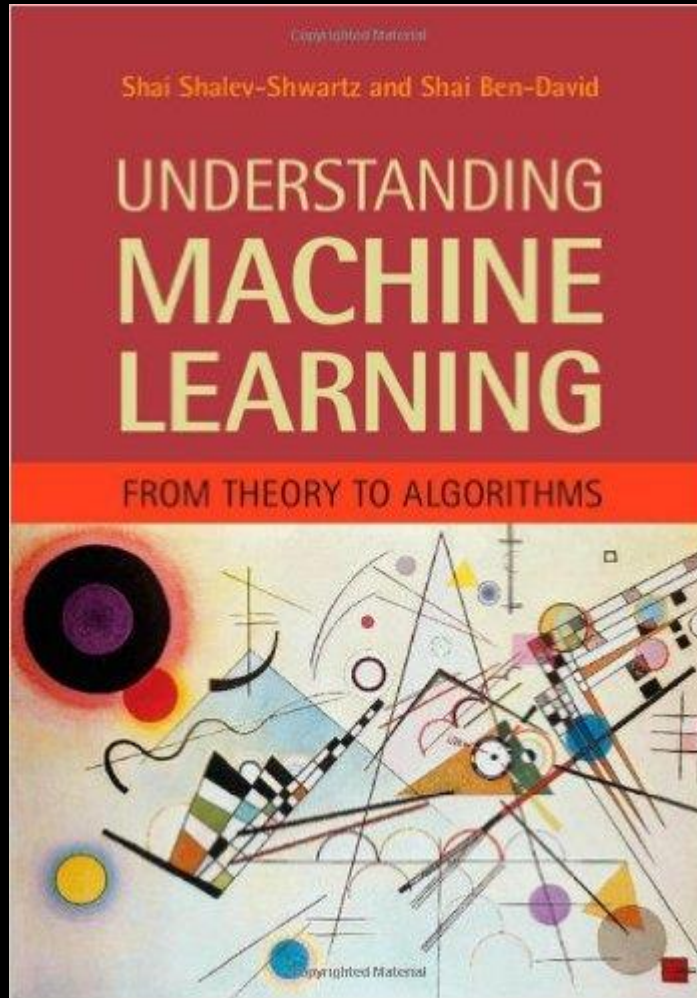
## The Strategy



## The Strategy



## The Strategy





## The Strategy



## The Strategy

11.  $\int_0^{\infty} e^{-x} x^n dx = \infty$  and hence

$$\int_0^{\infty} x^{n-1} \left\{ \phi(0) - \frac{x}{1} \phi(1) + \frac{x^2}{2} \phi(2) - \dots \right\} dx = \frac{1}{n-1} \phi(-n)$$

Sol.  $\int_0^{\infty} e^{-x} x^n dx = e^{-x} \{ x^n + nx^{n-1} + n(n-1)x^{n-2} + \dots \}$   
 where  $x=0 = \infty$  by IV 10 Cor.

$$f(0) \int_0^{\infty} e^{-x} x^{n-1} dx = \frac{1}{n} f(0)$$

$$\frac{f'(0)}{1} \int_0^{\infty} e^{-x} x^{n-1} dx = \frac{1}{n} \frac{f'(0)}{1}$$

$$\frac{f''(0)}{2!} \int_0^{\infty} e^{-x} x^{n-1} dx = \frac{1}{n^2} \frac{f''(0)}{2!}$$

and so on.

Adding up all the results we have

$$\int_0^{\infty} x^{n-1} \left\{ f(0) - \frac{x}{1} f(1) + \frac{x^2}{2!} f(2) - \dots \right\} dx = \frac{1}{n-1} f\left(\frac{1}{n}\right)$$

Let  $f(x^n) = \phi(x)$  then  $f\left(\frac{1}{n}\right) = \phi(-n)$ .

Cor 1.  $\int_0^{\infty} x^{n-1} \left\{ \phi(0) - x \phi(1) + x^2 \phi(2) - \dots \right\} dx = \frac{\pi \phi(-n)}{\sin \pi n}$

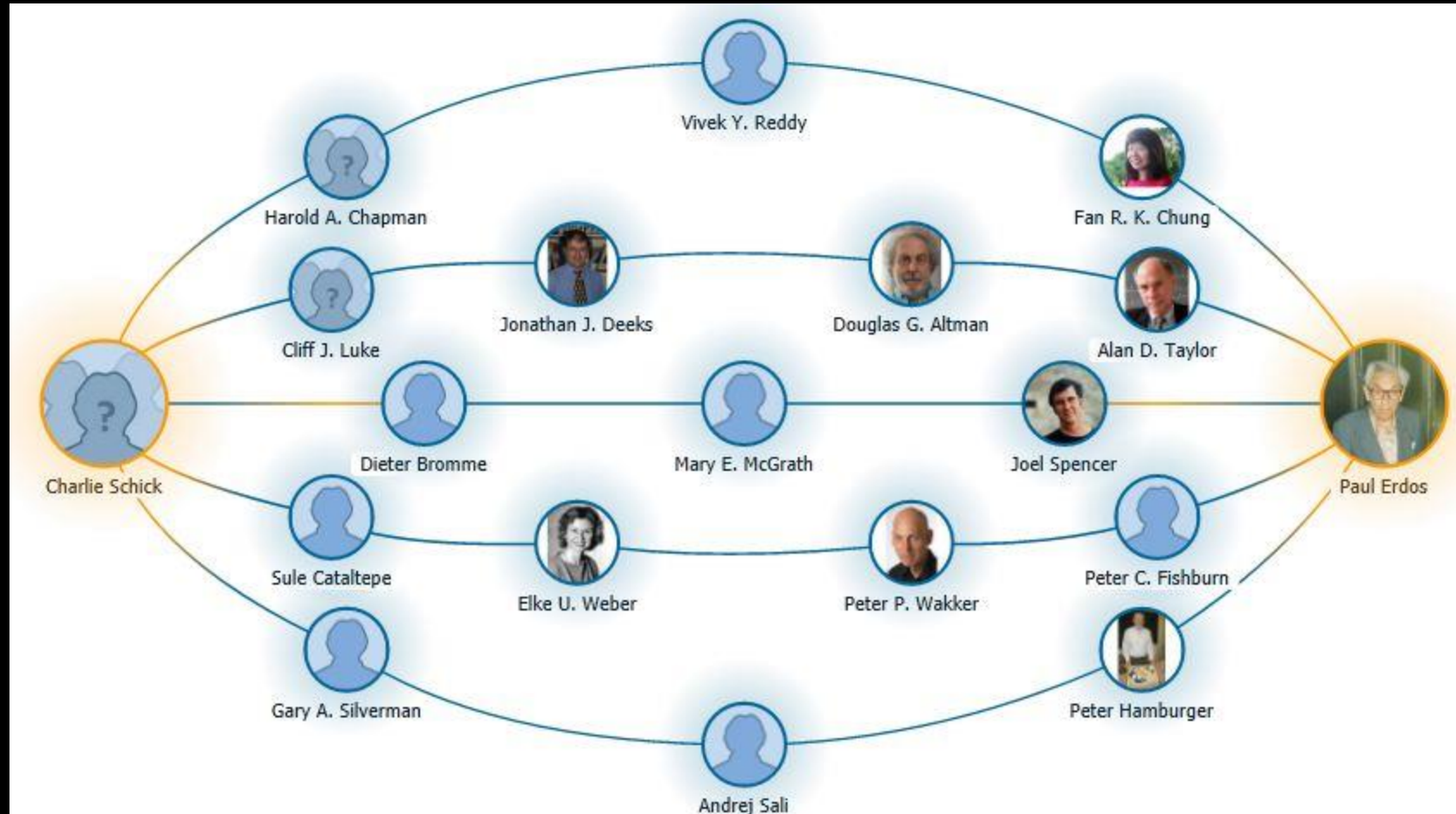
Cor 2.  $\int_0^{\infty} x^{n-1} \left\{ \phi(0) - \frac{x^2}{2} \phi(2) + \frac{x^4}{4!} \phi(4) - \dots \right\} dx = \frac{1}{n-1} \phi(-n) \times \cos \frac{\pi n}{2}$

Cor 3.  $\int_0^{\infty} \left\{ \phi(0) - \frac{x}{1} \phi(1) + \frac{x^2}{2!} \phi(2) - \dots \right\} \cos nx dx$   
 $= \phi(-1) - n^2 \phi(-3) + n^4 \phi(-5) - \dots$

Cor 4.  $\int_0^{\infty} \left\{ \phi(0) - x^2 \phi(2) + x^4 \phi(4) - \dots \right\} \cos nx dx$   
 $= \frac{\pi}{2} \left\{ \phi(-1) - \frac{\pi}{4} \phi(-2) + \frac{\pi^2}{2!} \phi(-3) - \frac{n^2}{4} \phi(-5) + \dots \right\}$



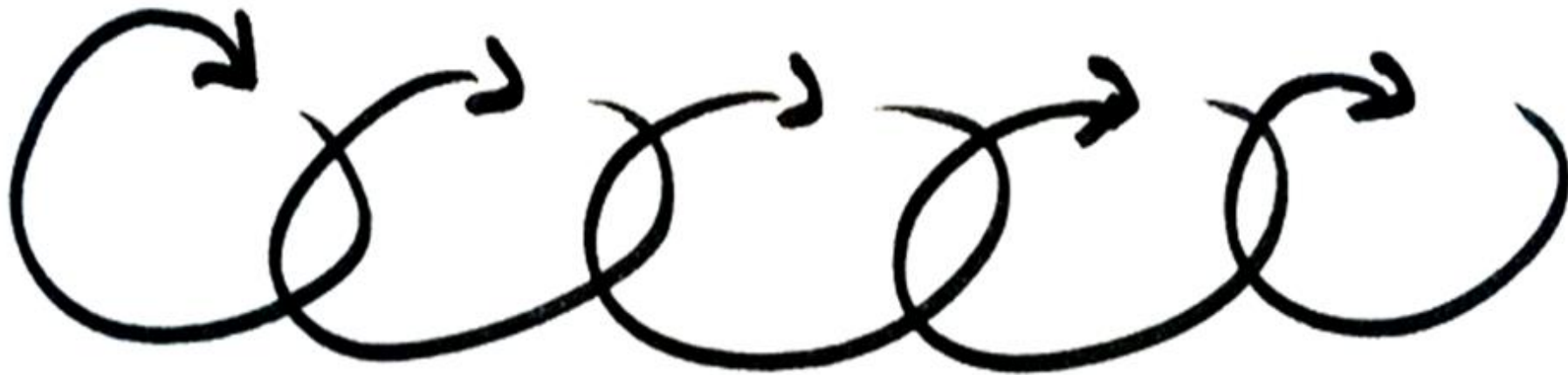
## The Strategy



## The Strategy



## The Strategy





## The Strategy



## IV. The Specialist

- Deep Learning
- Reinforcement Learning
- Computer Vision
- Natural Language Processing

You want to have a special ability on one or two subject, so you can get a job



## Neural Networks Hugo Larochelle, Montreal Univ PhD level

### ARTIFICIAL NEURON

**Topics:** connection weights, bias, activation function

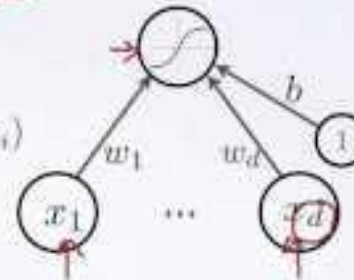
- Neuron pre-activation (or input activation):

$$\underline{a(\mathbf{x})} = \underline{b} + \sum_i \underline{w_i x_i} = \underline{b} + \underline{\mathbf{w}^\top \mathbf{x}}$$

- Neuron (output) activation:

$$h(\mathbf{x}) = g(a(\mathbf{x})) = g(b + \sum_i w_i x_i)$$

- $\mathbf{w}$  are the connection weights
- $b$  is the neuron bias
- $g(\cdot)$  is called the activation function





**Information State**

An **information state** (a.k.a. **Markov state**) contains all useful information from the history.

**Definition**

A state  $S_t$  is **Markov** if and only if

$$\mathbb{P}[S_{t+1} \mid S_t] = \mathbb{P}[S_{t+1} \mid S_0, \dots, S_t]$$

- "The future is independent of the past given the present"

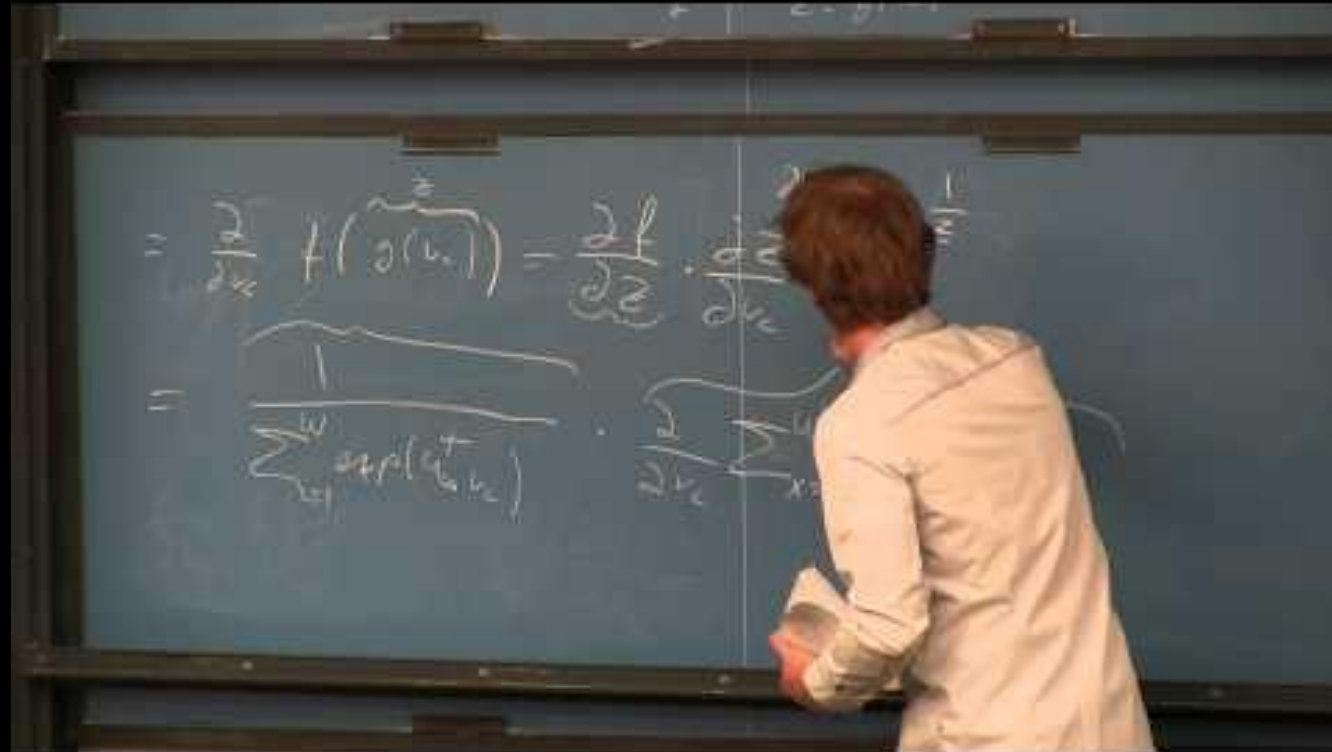
$$H_{t:t} \rightarrow S_t \rightarrow H_{t+1:t}$$

- Once the state is known, the history may be thrown away
- i.e. The state is a sufficient statistic of the future
- The environment state  $S_t^*$  is Markov
- The history  $H_t$  is Markov

## Deep Learning Vision Andrei Karpathy, Stanford PhD level



## Deep Learning NLP Richard Socher, Stanford PhD level





## The Strategy

### Image Recognition And Drones



Capturing image data





## The Strategy






## The Strategy

# True startup story

- Startup builds exchange for ads on webpages
- Clients bid on opportunities, market takes a cut
- System gets popular
- Stuff works better if ads and pages are matched
  - Programmer adds a few IF ... THEN ... ELSE clauses (system improves)
  - Programmer adds even more clauses (system sort-of improves, ruleset is a mess)
  - Programmer discovers decision trees (lots of rules, but they work better)
  - Programmer discovers boosting (combining many trees, works even better)
- Startup is bought ... (machine learning system is replaced entirely)

## The Strategy



Cornell University  
Library

arXiv.org
 

Logged in as **nsanders** | [Help](#) | [My Account](#) | [Logout](#)

Start >> **Add Files** >> Process >> Metadata >> Preview

### Submission files

Your submission to the archive must be in one of the following formats (listed in order of preference):

- [\(La\)TeX](#), [AMS\(La\)TeX](#), [PDFLaTeX](#)
- [DOCX \(Word 2007\)](#)
- [PDF](#)
- [PostScript](#)
- [HTML with JPEG/PNG/GIF images](#)

!

If your submission is (La)TeX, then you must submit the source (plus necessary macros and figures), not derivative dvi, Postscript, or PDF (see [Why TeX?](#)). For more information on formats and other submission details see [Submission Help](#).

Add files
 

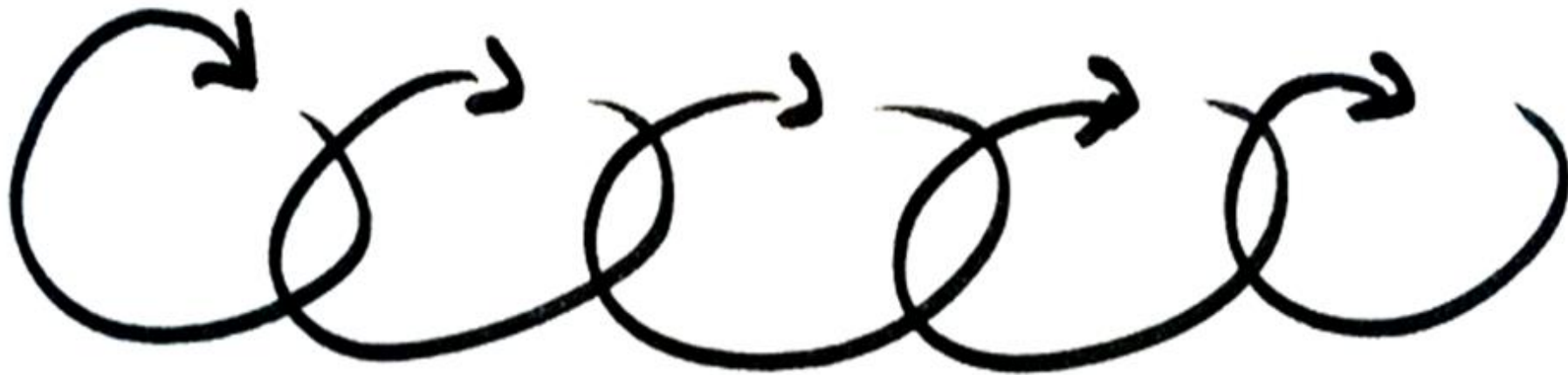
Choose...

Upload file

Continue: Process Files

(Please upload article files to process.)

## The Strategy



## IV. The Researcher

- Track down interesting paper
- Replicate some research
- Read the journals
- Get a PhD?

You want to have a special ability on one or two subject, so you can get a job

## The Strategy

Arxiv Sanity Preserver

Built by @karpathy to accelerate research.

Serving last 25909 papers from cs.[CV|CL|LG|NE]/stat.ML

User:

Pass:

[Login or Create](#)

Fork me on GitHub

New to arxiv-sanity? Check out the [introduction video](#).

X



most recent

top recent

recommended

library

Only show v1

Showing most recent Arxiv papers:

**Product Graph-based Higher Order Contextual Similarities for Inexact Subgraph Matching**

Anjan Dutta, Josep Lladós, Horst Bunke, Umapada Pal

2/1/2017 cs.CV

1702.00391v1 [pdf](#)

[show similar papers](#) | [review](#)



Many algorithms formulate graph matching as an optimization of an objective function of pairwise quantification of nodes and edges of two graphs to be matched. Pairwise measurements usually consider local attributes but disregard contextual information involved in graph structures. We address this issue by proposing contextual similarities between pairs of nodes. This is done by considering the tensor product graph (TPG) of two graphs to be matched, where each node is an ordered pair of nodes of the operand graphs. Contextual similarities between a pair of nodes are computed by accumulating weighted walks (normalized pairwise similarities) terminating at the corresponding paired node in TPG. Once the contextual similarities are obtained, we formulate subgraph matching as a node and edge selection problem in TPG. We use contextual similarities to construct an objective function and optimize it with a linear programming approach. Since random walk formulation through TPG takes into account higher order information, it is not a surprise that we obtain more reliable similarities and better discrimination among the nodes and edges. Experimental results shown on synthetic as well as real benchmarks illustrate that higher order contextual similarities add discriminating power and allow one to find approximate solutions to the subgraph matching problem.

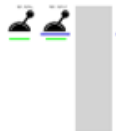



## The Strategy

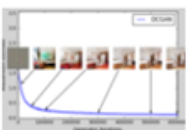
Receive the best of GitXiv right in your inbox.
 
[Get Newsletter](#)
✕

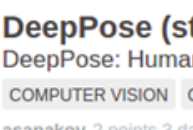
Collaborative Open Computer Science

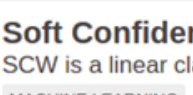
View: [Top](#) [New](#) [Best](#) [Daily](#)

- 

**Learning to reinforcement learn**  
 DeepMind's Meta-RL A3C algorithm  
 DEEP REINFORCEMENT LEARNING (DRL)  
 Christos Iraklis Tsatsoulis 3 points 2 days ago 0 Comments
- 

**Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning**  
 CONVOLUTIONAL NEURAL NETWORKS (CNN) DEEP LEARNING (DL)  
 Christos Iraklis Tsatsoulis 1 point 2 days ago 0 Comments
- 

**Wasserstein GAN**  
 An alternative to traditional GAN training  
 ADVERSARIAL NETWORKS GENERATIVE  
 Christos Iraklis Tsatsoulis 4 points 2 days ago 0 Comments
- 

**DeepPose (stg-1) on TensorFlow**  
 DeepPose: Human Pose Estimation via Deep Neural Networks implemented on TensorFlow  
 COMPUTER VISION CONVOLUTIONAL NEURAL NETWORKS (CNN) DEEP LEARNING (DL)  
 asanakoy 2 points 3 days ago 0 Comments
- 

**Soft Confidence-Weighted Learning**  
 SCW is a linear classifier with fast learning, low memory usage, and high accuracy.

## The Strategy

[Home](#)
[Login](#)

# NIPS 2016

Monday December 05 -- Saturday December 10, 2016

Centre Convencions Internacional Barcelona, Barcelona SPAIN

[2016 Pricing »](#)
[Registration Closed »](#)

[Dates](#)
[Calls ▾](#)
[Program Books ▾](#)
[Schedule ▾](#)
[Barcelona ▾](#)
[Committees](#)

[View Earlier Meetings »](#)
[View our Sponsors »](#)
[2016 Pre-proceedings »](#)
[NIPS 2017 Long Beach »](#)

### Videos

Videos from the Tutorials and Conference are now linked into the schedule. Also PDF's of papers are linked to posters. Follow Facebook/Twitter (below) for further video announcements.

[View Full Schedule »](#)

### Awards

We would like to congratulate the winners of the best paper, best student paper, best demonstration and best reviewers awards.

[View Awards »](#)

### Tutorials Mon Dec 5th

The tutorial times and rooms have not been set yet. View the list of tutorials using the button below.

[View Tutorials »](#)

### Workshops Dec 9-10

Fifty workshops will take place over Friday and Saturday December 9th and 10th.

[View Workshops »](#)

### Invited Speakers Dec 5th-8th

Yann LeCun (Facebook & NYU), Susan Holmes (Stanford), Kyle Cranmer (NYU), Saket Navlakha (Salk Institute), Drew Purves (Deep Mind), Marc Raibert (Boston Dynamics), Irina Rish (IBM)

[View Workshops »](#)

### Sponsorship

Sponsorship of the NIPS Conference contributes to our success every year. Become a [sponsor »](#)

[View our Sponsors »](#)

### Symposia Thu Dec 8th

- Deep Learning Symposium
- Machine Learning and the Law
- Program Learning with Recurrent Neural Nets

(included with "Conference Sessions" and/or "Workshops") [View Symposia »](#)

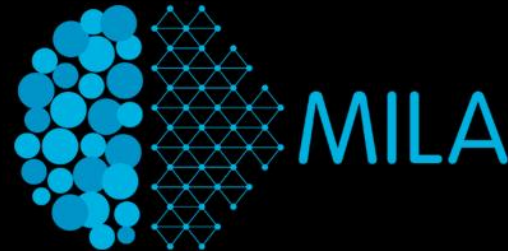
### Demonstrations Dec 6 - 7

Tue and Wed evening each have 10 demos showcasing novel technology in the following areas: interactive models, learning from demonstration, realtime visualization of learning models, hardware technology, biologically-inspired learning models, and robotics. [View Demos »](#)

## The Strategy

[HOME](#)[MILA](#)[RESEARCH](#)[TRAINING AND OUTREACH](#)[FRANÇAIS](#)

Montreal  
Institute for  
Learning  
Algorithms



*Dedicated to understanding the fundamentals of learning and intelligence*

### OPEN PROFESSOR POSITION IN MACHINE LEARNING AT UDEM

[Tenure-Track Professor Position in the Field of Applied Machine Learning](#) – Université de Montréal

The Department of Computer Science and Operations Research at Université de Montréal is seeking applications for a full-time tenure-track professor position at all ranks in areas related to machine learning and its applications (e.g., natural language, medicine, perception, recommendation systems, data mining).

Application deadline is January 15 2017, and the expected starting date is June 1st 2017. For inquiries, contact Houari Sahraoui, Chair, at the address in the official posting.


### FACULTY



## The Strategy

Carnegie Mellon University
☒ Search Only Machine Learning Department

SCHOOL OF COMPUTER SCIENCE | CONTACT US



**ML**  
MACHINE LEARNING  
DEPARTMENT

**Prospective Students**

**PhD Programs in ML**

Masters Programs in ML

Undergraduate Programs in ML

**Research**

**Teaching**

**People**

**Honors and Awards**

**News**

**Seminars and Talks**

---

Information for Current MLD Students

Contact Us

[\[MACHINE LEARNING DEPARTMENT\]](#) > [Prospective Students](#) > [PhD Programs in ML](#)

**PhD Program in Machine Learning**

The Ph.D. Program in Machine Learning is for students who are interested in research in Machine Learning and Computational Statistics. The program is operated jointly by faculty in the School of Computer Science and Department of Statistics.

**Joint PhD Program in Statistics and Machine Learning**

This PhD program differs from the Machine Learning PhD program in that it places significantly more emphasis on preparation in statistical theory and methodology. Similarly, this program differs from the Statistics PhD program in its emphasis on machine learning and computer science. The Joint Ph.D. Program in Machine Learning and Statistics is a new program aimed at preparing students for academic careers in both CS and Statistics departments at top universities.

**Joint PhD Program in Machine Learning and Public Policy**

The Joint Ph.D. Program in Machine Learning and Public Policy is a new program operated jointly by faculty in Machine Learning and the Heinz College (Schools of Public Policy, Information Systems, and Management). Students will gain the skills necessary to develop new state-of-the-art machine learning technologies and apply these successfully to real-world policy domains.

**Joint PhD Program in Neural Computation and Machine Learning**

This Joint PhD program trains students in the application of Machine Learning to Neuroscience by combining core elements of the ML PhD program and the Program in Neural Computation (PNC) offered by the Center for the Neural Basis of Cognition (CNBC).

Machine Learning Department | 5000 Forbes Avenue, Gates Hillman Center 8th Floor, Pittsburgh, PA 15213

[Carnegie Mellon Home](#)

# PACMANN AI

Thank you, Question?

Adityo Sanjaya  
PACMANN AI Head of Research  
Email: [adityosanjaya3.14@gmail.com](mailto:adityosanjaya3.14@gmail.com)  
+62 55777490099