



12th North East Workshop on Computational Intelligence Processing
Electronics and Communication Sciences Unit, ISI KOLKATA
and
North Eastern Regional Institute of Science and Technology
Nirjuli, Arunachal Pradesh
February 22 – February 25, 2012

Pulak Purkait, SRF,
Indian Statistical Institute
KOLKATA

Plan for this course

Wednesday Feb 22 – Motivation and Basics of Information Processing

Thursday Feb 23 – Some basic Algorithms and Applications

Thursday Feb 24 – Advance Issues and Advance Computation

Saturday Feb 25 – More cont.. + Discussion and closing

Speakers

Motivation and
Basics of
Computational
Intelligence
Processing



Pulak Purkait,
Soumitra Samanta,
Partha Pratim Mohanta,

Advance
Issues and
Advance
Computation



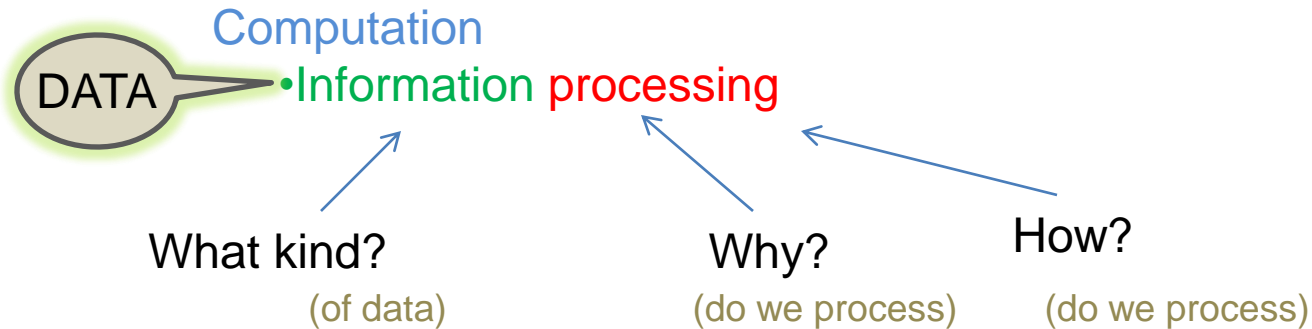
Swagatam Das,
Snehasis Mukherjee,
Yusuf Akhtar

Electronics and Communication Sciences Unit
Indian Statistical Institute, Kolkata

12th North East Workshop on Computational Intelligence Processing

Intelligence

- Ability to solve problems
- Examples of Intelligent Behaviors or Tasks
 1. Classification of texts based on content
 2. Heart disease diagnosis
 3. Chess playing
 4. Etc..



We will try to give answer of this three basic question in next two days.

Computation

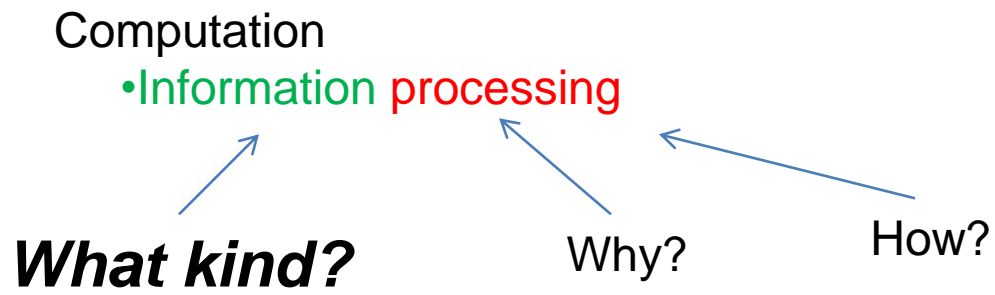
• Information processing

What kind?

Why?

How?





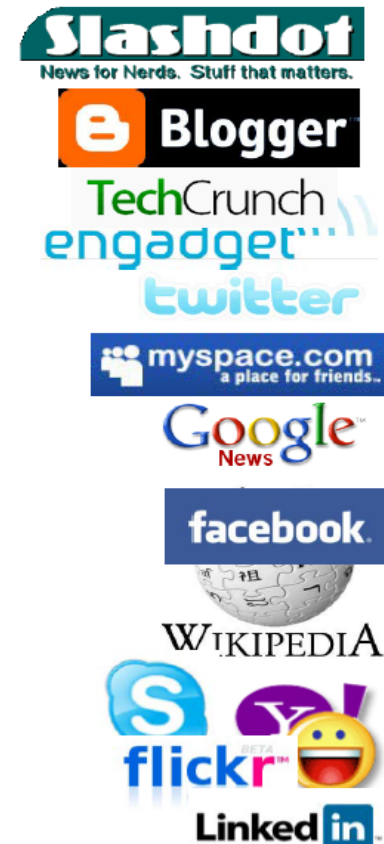
What kind? —> Information processing

Internet : Social Media

Social Media: Big change

- Web is no longer a static library that people passively browse
- Web is a place where people:
 - Consume and create content
 - Interact with other people:
 - Internet forums, Blogs, Social networks, Twitter, Wikis, Podcasts, Slide sharing, Bookmark sharing, Product reviews, Comments, ...
- Facebook traffic tops Google (for USA)
 - March 2010: FB > 7% of US traffic

http://money.cnn.com/2010/03/16/technology/facebook_most_visited



What kind? → Information processing

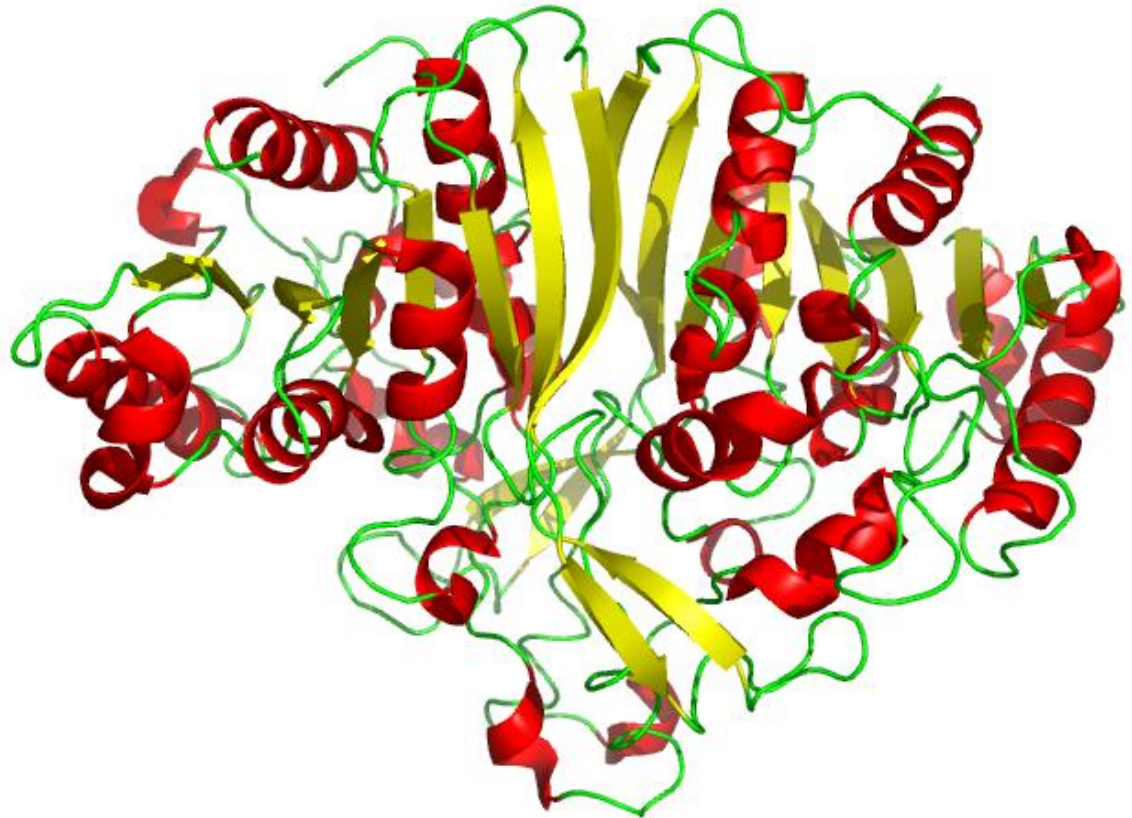
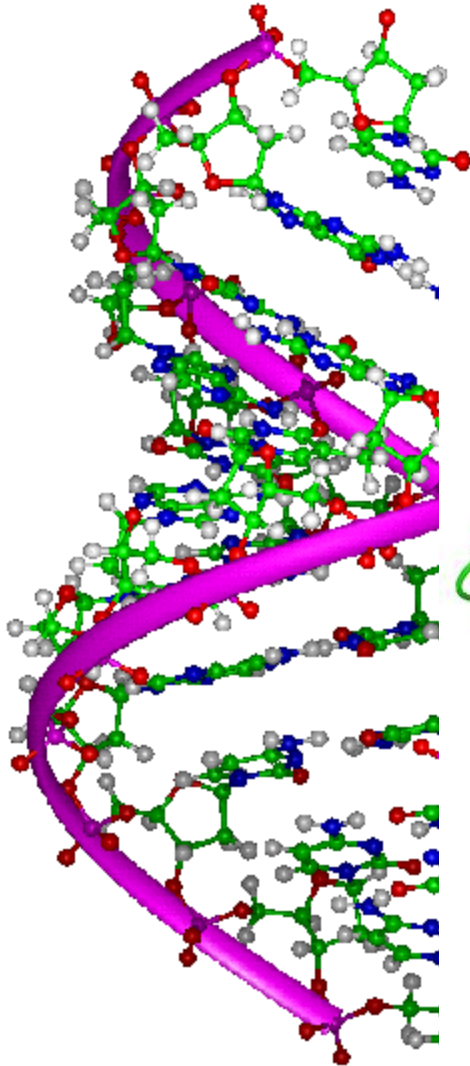
Computer Vision



What kind? —> Information processing

Bioinformatics :

The problem is as follows: given a sequence of amino acid residues, predict the placement of the main three dimensional sub-structures of the protein.

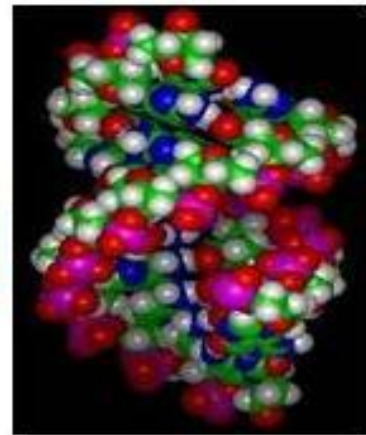
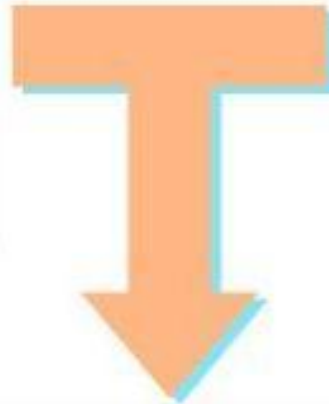


What kind? → Information processing

Bioinformatics :



Computer systems

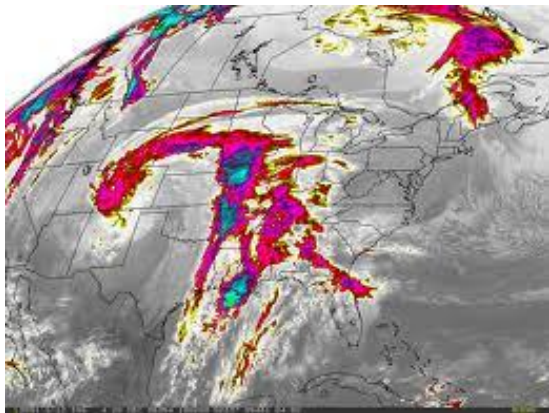


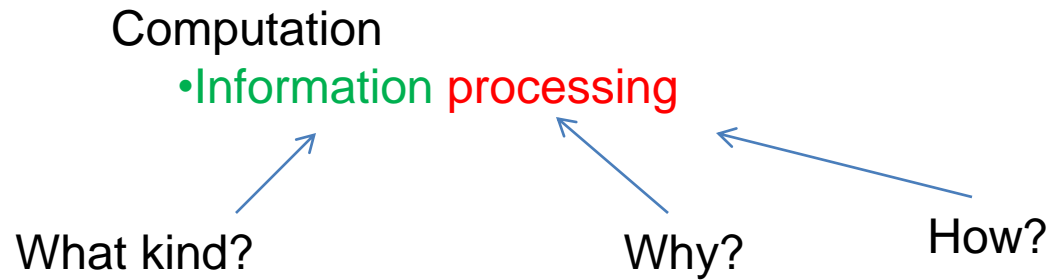
Biological systems

BIOINFORMATICS

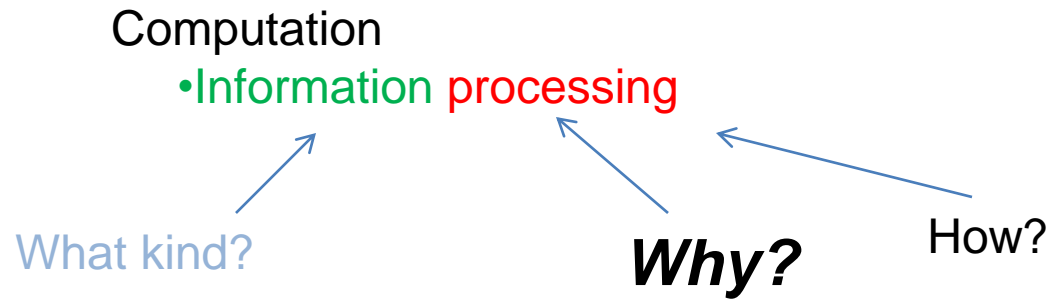
The world is driven by data : Data explains a lot

- Germany's climate research centre generates 10 petabytes per year
- Google processes 24 petabytes per day
- The Large Hadron Collider produces 60 gigabytes per minute (~12 DVDs)
- There are over 50m credit card transactions a day in the US alone.





- Internet : Social Media
- Computer Vision : Image and Video
- Bioinformatics
- Etc...



- Internet : Social Media
- Computer Vision : Image and Video
- Bioinformatics
- Etc...

Internet : Social Media

Consumer Brand Analytics

- What are people saying about our brand?

Marketing Communications

- Significant spending on marketing, advertising:
Companies trying to position their products
- Brand analytics helps to determine whether such campaigns are effective

Product reviews

- Automatically mine product reviews for information on product features, new requests, ... Easy to use, Comfortable chair, Light weight, Sturdy, Good price

Root of the Network

Articles/Blogs are enough to get all sort of Information



The Netflix problem

- Netflix database
 - About half a million users
 - About 18,000 movies
- People rate movies
- Sparsely sampled entries



Recommend DVD according to users flavor

Information processing ← Why?

Internet : Social Media

amazon.com®

ebay



Mashup Patterns: Designs and Examples for the Modern Enterprise (Paperback)

by Michael Dornt (Author)

List Price: \$39.99

Price: **\$34.37** & this item ships for FREE with Super Saver Shipping. [Details](#)

You Save: \$5.62 (14%)

Pre-order Price Guaranteed. [Learn more.](#)

This title has not yet been released.

You may pre-order it now and we will deliver it to you when it arrives. Ships from and sold by Amazon.com. Gift-wrap available.

[Low](#) from \$34.37

[See larger image](#)

[Publisher: learn how customers can search inside this issue.](#)

Please tell the publisher:

I'd like to read this book on Kindle

Don't have a Kindle? [Get yours here.](#)

Special Offers and Product Promotions

- Pre-order Price Guarantee! Order now and if the Amazon.com price decreases between your order time and the end of the day of the release date, you'll receive the lowest price. [Here's how](#) (restrictions apply)

Customers Who Bought Related Items Also Bought

Page 1 of 13



Predicting Collective Intelligence: Building Smarter Systems by Toby Segaran

★★★★★ (48) \$27.04



Don't Make Me Think: A Common Sense Approach to Web User Interface Design by Steve Krug

★★★★★ (459) \$26.40



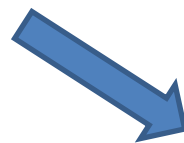
High Performance Web Sites: Essential Knowledge for Front-End Web Developers by Steve Souders

★★★★★ (31) \$20.44



Designing Web Navigation: Optimizing the User Experience by James Kalbach

★★★★★ (18) \$26.49



Information processing ← Why?



Internet : Social Media



People You May Know [See All](#)



Swarnali Banerjee

28 mutual friends

[Add Friend](#)



Nikhil Bellarykar

41 mutual friends

[Add Friend](#)



Apurba Das

7 mutual friends

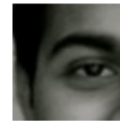
[Add Friend](#)



Anindita Chakrabarty

17 mutual friends

[Add Friend](#)



Gourab Saha

3 mutual friends

[Add Friend](#)



Shampa Bhattacharjee

4 mutual friends

[Add Friend](#)



Riddhipratim Basu

17 mutual friends

[Add Friend](#)

Information processing ← Why?

Internet : Social Media



More ▾ 16 of 2,461 < > ⚙

Feb 13 (7 days ago) ☆ ↶ ▾

amiya bhowmick

Add to circles

Show details

Ads – Why these ads?

These ads are based on emails from your mailbox. Visit Google's [Ads Preferences Manager](#) to learn more, block specific advertisers, or opt out of personalized ads.

Rs.1.45 Crore Pension
কম টাকা Invest আর
কোটিপতি হয়ে যান
[Investment.PolicyBazaar.com](#)

Wedding Invitations/Cards
Find Wedding Cards Providers
in your City? Call 4444 4444
[weddings.getit.in](#)

ASR Instruments
SIR Test Equipment
& SIR Test Services



Information processing ← Why?



Like Dislike + Add to Share

359,924

241 likes, 30 dislikes

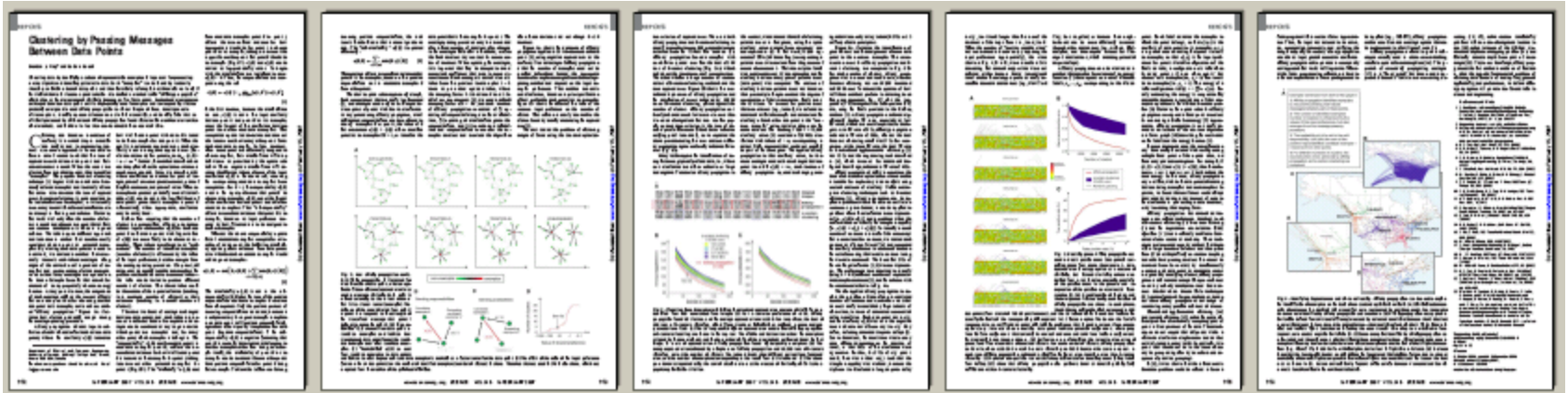
Uploaded by [nafisa125](#) on 16, 2007

[Rabindra Sangeet](#) (Prem Porjai)

Artist: Bonna & saadi

-  **Sraboner Dharar Moton**
by saikatsanju
163,154 views
3:42
-  **A demo on how to use FEMALE CONDOM**
by carrotcommunications
1,096,835 views
2:07
-  **Jokhon porbe na mor payer chinho ei bate**
by dhrubadey
257,905 views
6:32
-  **Shreya Ghosal Ekla Chalo Re.mov**
by sunilaanand
340,298 views
5:06
-  **বন্ধুত্ব তে- শাহানা বাজাপতি**
by nurulghani
44,513 views
3:47
-  **Tomar Khola Hawa ~ Shahana Bajpaie**
by taniamoni
83,540 views
5:00
-  **Tagor song_ জন্মি বড় বড় মূসুর বাঘন ছিল...**
by nazmul911
103,726 views
3:04

Text Document analysis : Topic Model



How can we develop algorithms that will

- Produce a summary of the document?
- Find similar documents?
- Predict document layouts that are suitable for different readers?

Information processing ← Why?

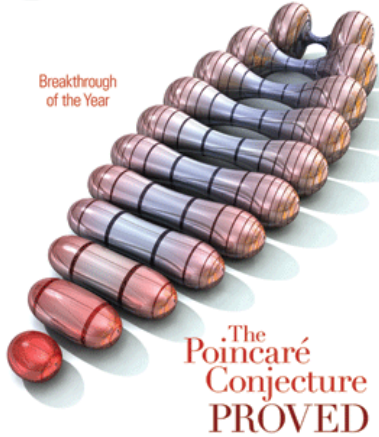


Example: Science Magazine Papers

Science

22 December 2006 | \$19

Breakthrough
of the Year



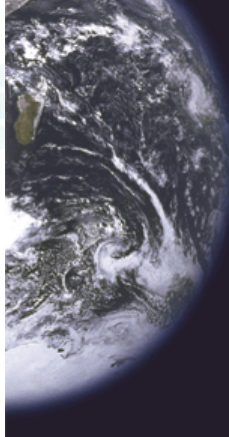
The Poincaré
Conjecture
PROVED

AAAS

Science

12 December 2003

Vol. 302 No. 5652
Pages 1845-2016 \$10

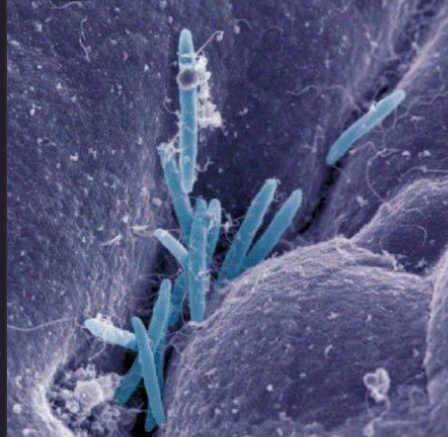


TRAGEDY
OF THE
COMMONS?

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Science

21 January 2011 | \$19



Downloaded from www.sciencemag.org

Science

7 October 2009 | \$19



Ardipithecus ramidus

AAAS

Science

1 February 2008 | \$19

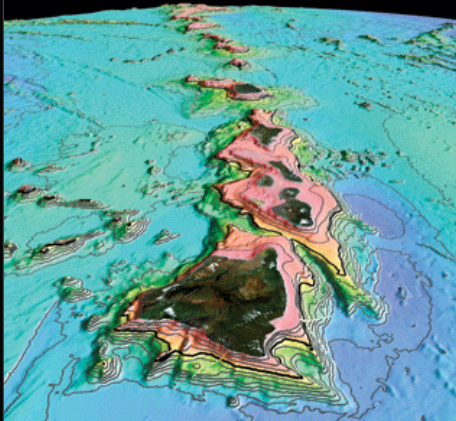


Gordon
Research Conferences

AAAS

Science

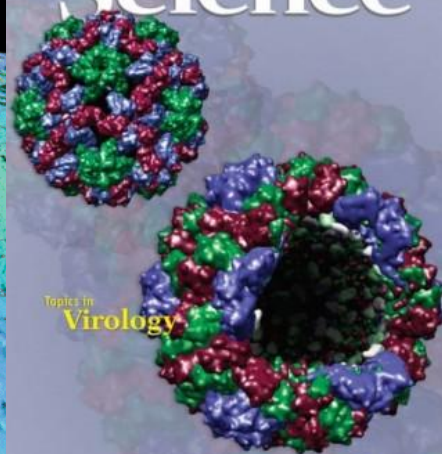
4 December 2009 | \$19



AAAS

Science

12 May 2006 | \$19

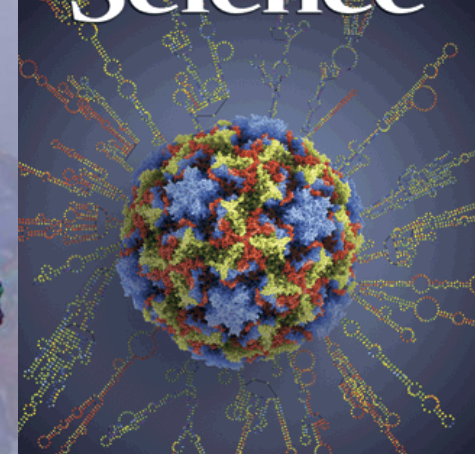


Topics in
Virology

AAAS

Science

3 April 2009 | \$19



AAAS

Example: Science Magazine Papers

- Dataset with approx. 12K papers from Science Magazine
- Selected concepts from model with K=200

Topic

universe	0.0439	drug	0.0672	cells	0.0675	sequence	0.0818	years	0.156
galaxies	0.0375	patients	0.0493	stem	0.0478	sequences	0.0493	million	0.0556
clusters	0.0279	drugs	0.0444	human	0.0421	genome	0.033	ago	0.045
matter	0.0233	clinical	0.0346	cell	0.0309	dna	0.0257	time	0.0317
galaxy	0.0232	treatment	0.028	gene	0.025	sequencing	0.0172	age	0.0243
cluster	0.0214	trials	0.0277	tissue	0.0185	map	0.0123	year	0.024
cosmic	0.0137	therapy	0.0213	cloning	0.0169	genes	0.0122	record	0.0238
dark	0.0131	trial	0.0164	transfer	0.0155	chromosome	0.0119	early	0.0233
light	0.0109	disease	0.0157	blood	0.0113	regions	0.0119	billion	0.0177
density	0.01	medical	0.00997	embryos	0.0111	human	0.0111	history	0.0148
bacteria	0.0983	male	0.0558	theory	0.0811	immune	0.0909	stars	0.0524
bacterial	0.0561	females	0.0541	physics	0.0782	response	0.0375	star	0.0458
resistance	0.0431	female	0.0529	physicists	0.0146	system	0.0358	astrophys	0.0237
coli	0.0381	males	0.0477	einstein	0.0142	responses	0.0322	mass	0.021
strains	0.025	sex	0.0339	university	0.013	antigen	0.0263	disk	0.0173
microbiol	0.0214	reproductive	0.0172	gravity	0.013	antigens	0.0184	black	0.0161
microbial	0.0196	offspring	0.0168	black	0.0127	immunity	0.0176	gas	0.0149
strain	0.0165	sexual	0.0166	theories	0.01	immunology	0.0145	stellar	0.0127
salmonella	0.0163	reproduction	0.0143	aps	0.00987	antibody	0.014	astron	0.0125
resistant	0.0145	eggs	0.0138	matter	0.00954	autoimmune	0.0128	hole	0.00824

Information processing ← Why?

Image Document : summary of an image

Annotate images



SKY WATER TREE
MOUNTAIN PEOPLE



SCOTLAND WATER
FLOWER HILLS TREE



SKY WATER BUILDING
PEOPLE WATER



FISH WATER OCEAN
TREE CORAL



PEOPLE MARKET PATTERN
TEXTILE DISPLAY

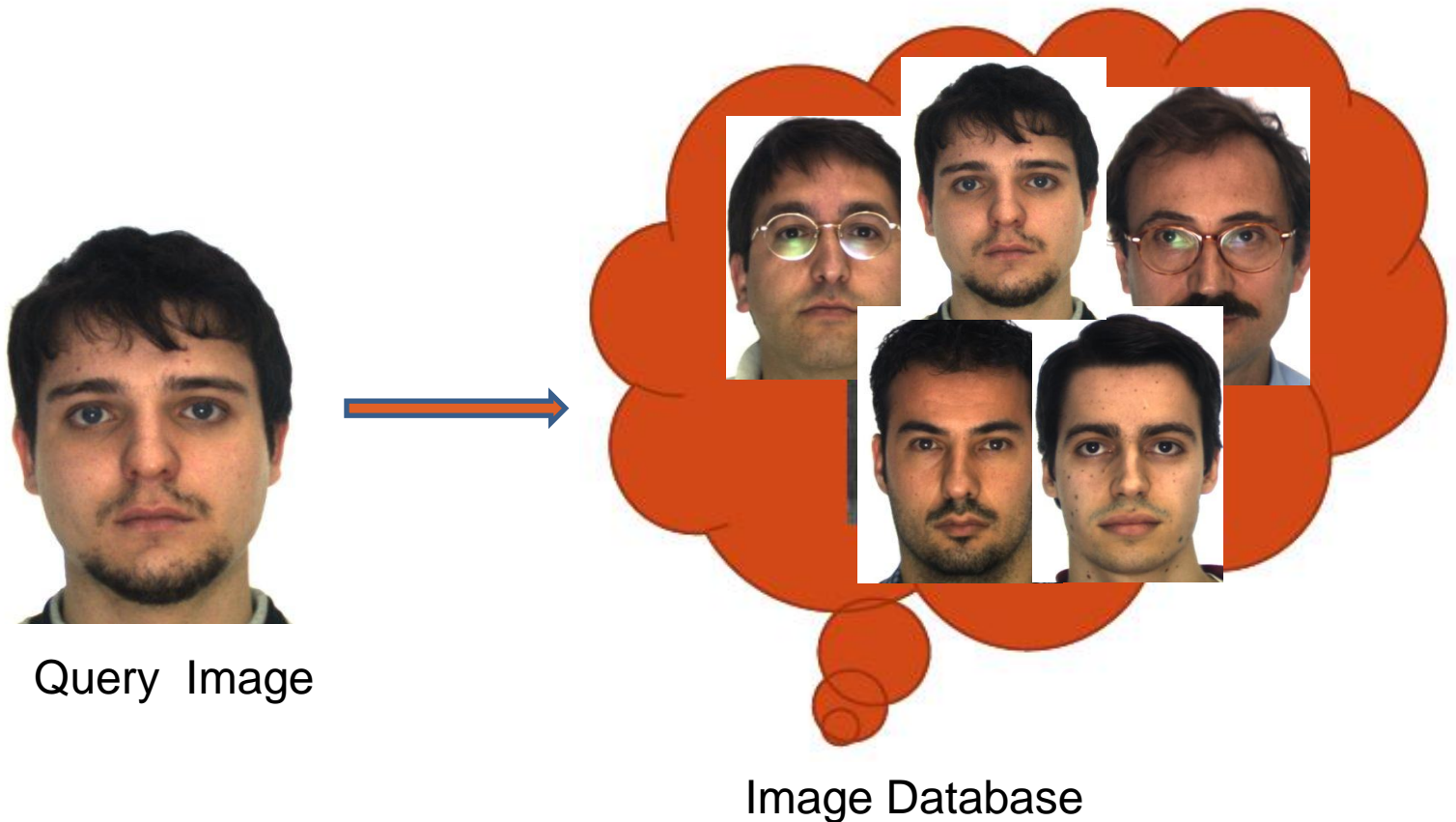


BIRDS NEST TREE
BRANCH LEAVES

Information processing ← Why?


Image Document

A **facial recognition system** is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source



Information processing ← Why?

Image Document : Optical Character(Text / Handwritten) Recognition



KOLKATA TRAFFIC POLICE

Book No: Sl. No: Date:

Beat No: TP Guard: GPF No:

		VEHICLE - NUMBER						VECH TYPE	TIME (RLY)			OFF CODE	DIR						
W	B	0	2	R	9	8	5	8	1	2	1	1	4	0	0	1	0	3	
W	B	1	1	A	8	4	0	2	0	3	1	1	4	5	0	1	0	3	
W	B	0	3	B	9	4	2	5	0	9	1	2	5	5	1	1	0	2	
W	B	0	6		7	2	5	3	1	2	0	0	0	0	0	1	0	3	
W	B	2	5	5	2	0	0	5	0	3	1	3	0	5	0	1	0	3	
W	B	T				5	3	5	2	0	6	1	3	0	2	0	1	0	3
W	B	1	9	C	8	5	2	4	0	6	1	3	2	0	0	1	0	3	
W	B	4				2	7	0	4	0	3	1	3	2	5	0	1	0	3

Rank & Name: of Tapsh Kr. Paul No. 9591 Signature T. Paul

Information processing ← Why?

Computer vision : video

How can we develop algorithms that will

- Track objects
- Recognize objects
- Segment objects
- Denoise the video
- Determine the state (eg, gait) of each object

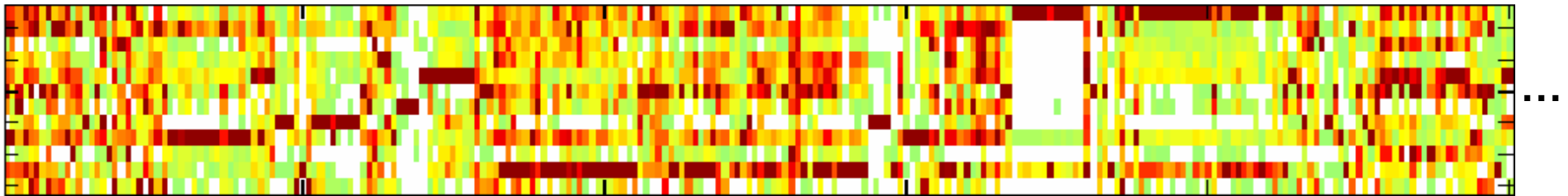


Information processing ← Why?

Bioinformatics

Mouse
tissues

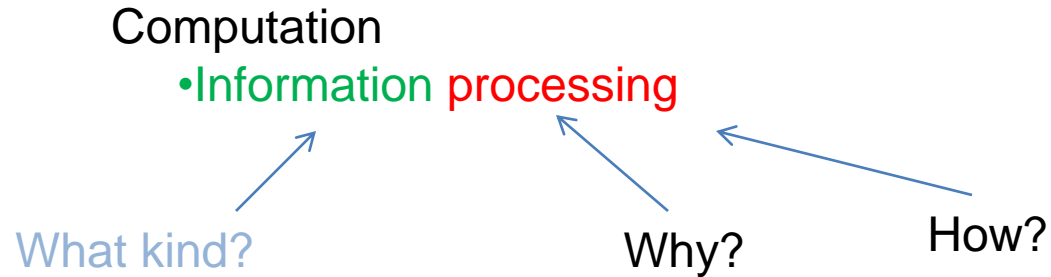
DNA activity
Low |  High



Position in DNA

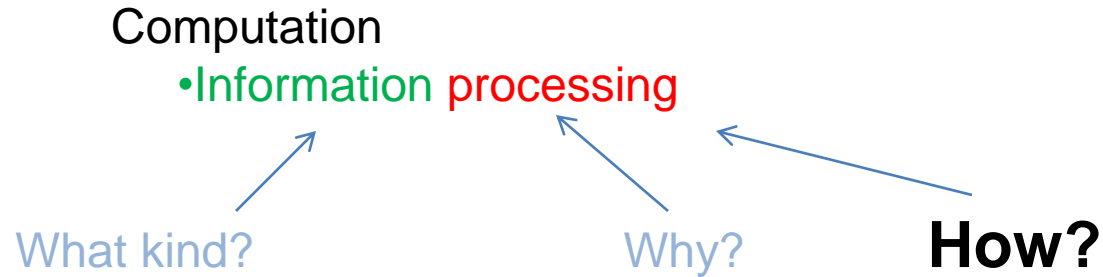
How can we develop algorithms that will

- Find start sites and stop sites of genes, by looking for common patterns of activity?



- Internet : Social Media
- Computer Vision : Image and Video
- Bioinformatics
- Etc...

- Internet :
 1. Media Analysis
 2. Brand Management
 3. Customer recommendations
 4. Inherent topic
 5. Etc...
- Computer Vision :
 1. Surveillance
 2. Entertainment
 3. Robotics
 4. Etc..
- Bioinformatics
 1. DNA analysis
 2. Protein Structure analysis
- Etc...



- Internet : Social Media
- Computer Vision : Image and Video
- Bioinformatics
- Etc...

- Internet :
 1. Media Analysis
 2. Brand Management
 3. Customer recommendations
- Computer Vision :
 1. Surveillance
 2. Entertainment
 3. Robotics
 4. etc
- Bioinformatics
 1. DNA analysis
 2. Protein Structure analysis
- Etc...

Information processing ← How?

Machine Learning & Pattern Recognition

Information

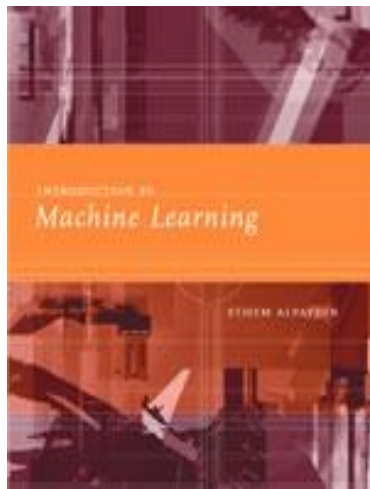


Desired Goal

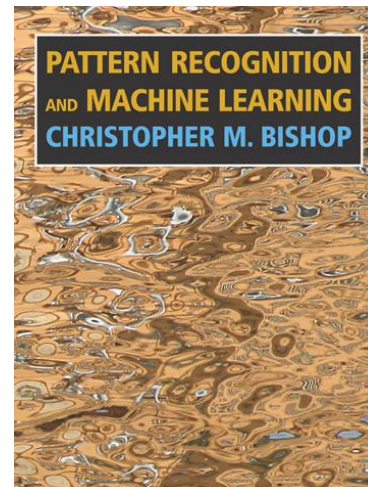
An introduction to

Machine Learning & Pattern Recognition

*Pulak Purkait, SRF,
Indian Statistical Institute
KOLKATA*



ETHEM ALPAYDIN
© The MIT Press, 2004



CHRISTOPHER M. BISHOP
© Springer (2006)

Definition of Machine Learning

- self-configuring data structures that allow a computer to do things that would be called “intelligent” if a human did it
- “making computers behave like they do in the movies”



I Robot



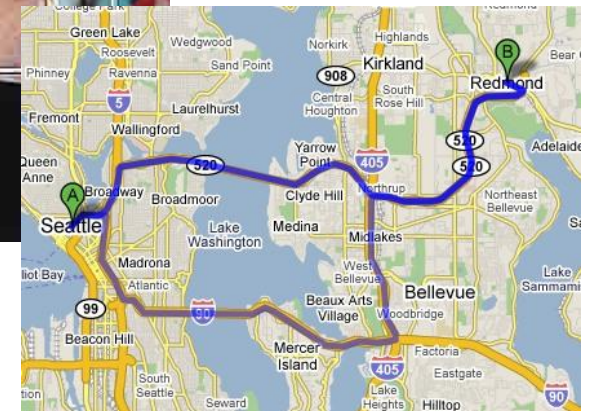
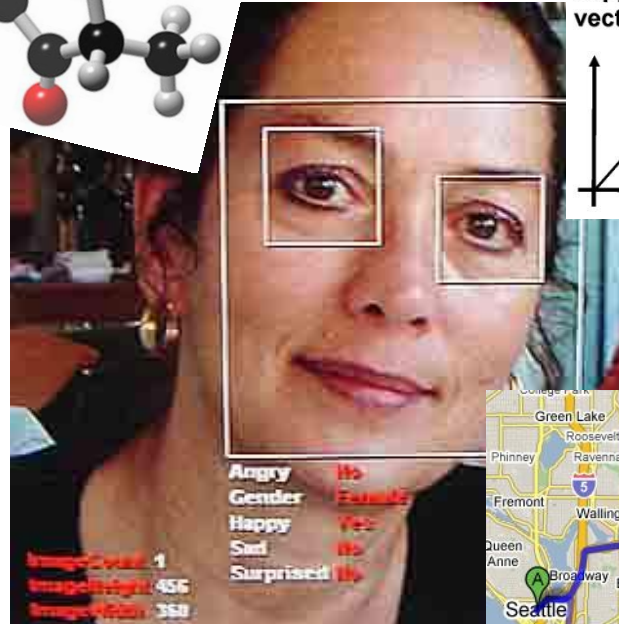
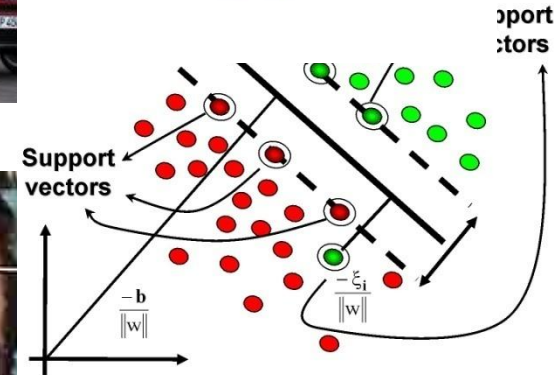
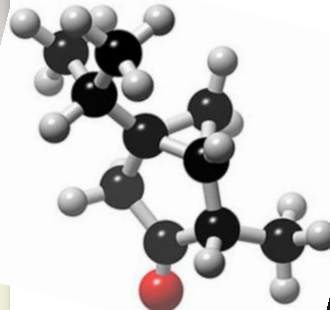
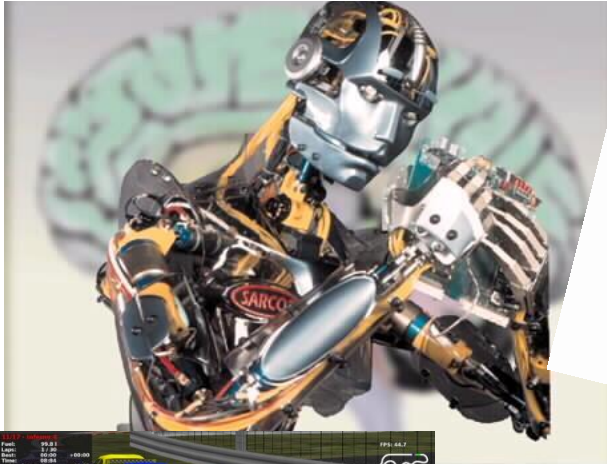
Wall-E



Transformer-3



Why "Learn" ?



Application of Machine Learning is in Everywhere

Learning from Data

How can we extract knowledge from data to help humans take decisions?

How can we automate decisions from data?

How can we adapt systems dynamically to enable better user experiences?

Write code to explicitly
do the above tasks



Write code to make the computer
learn how to do the tasks



Pattern Recognition (What ?)

Duda Hart, 1973 : “A field concerned with machine recognition of meaningful regularities”.

Bezdek, 1981: “Pattern recognition is the search for structure in data”

Pattern Recognition (What ?)

Is it the same object?



- The task is **Pattern Recognition**
- The way we do it called **Machine Learning**



Basic Task of ML/PR

- Association
- Supervised Learning
 - Classification
 - Regression
- Unsupervised Learning
- Reinforcement Learning

Basic Task of ML/PR

- Association
- Supervised Learning
 - Classification
 - Regression
- Unsupervised Learning
- Reinforcement Learning



Learning Associations

- Basket analysis:

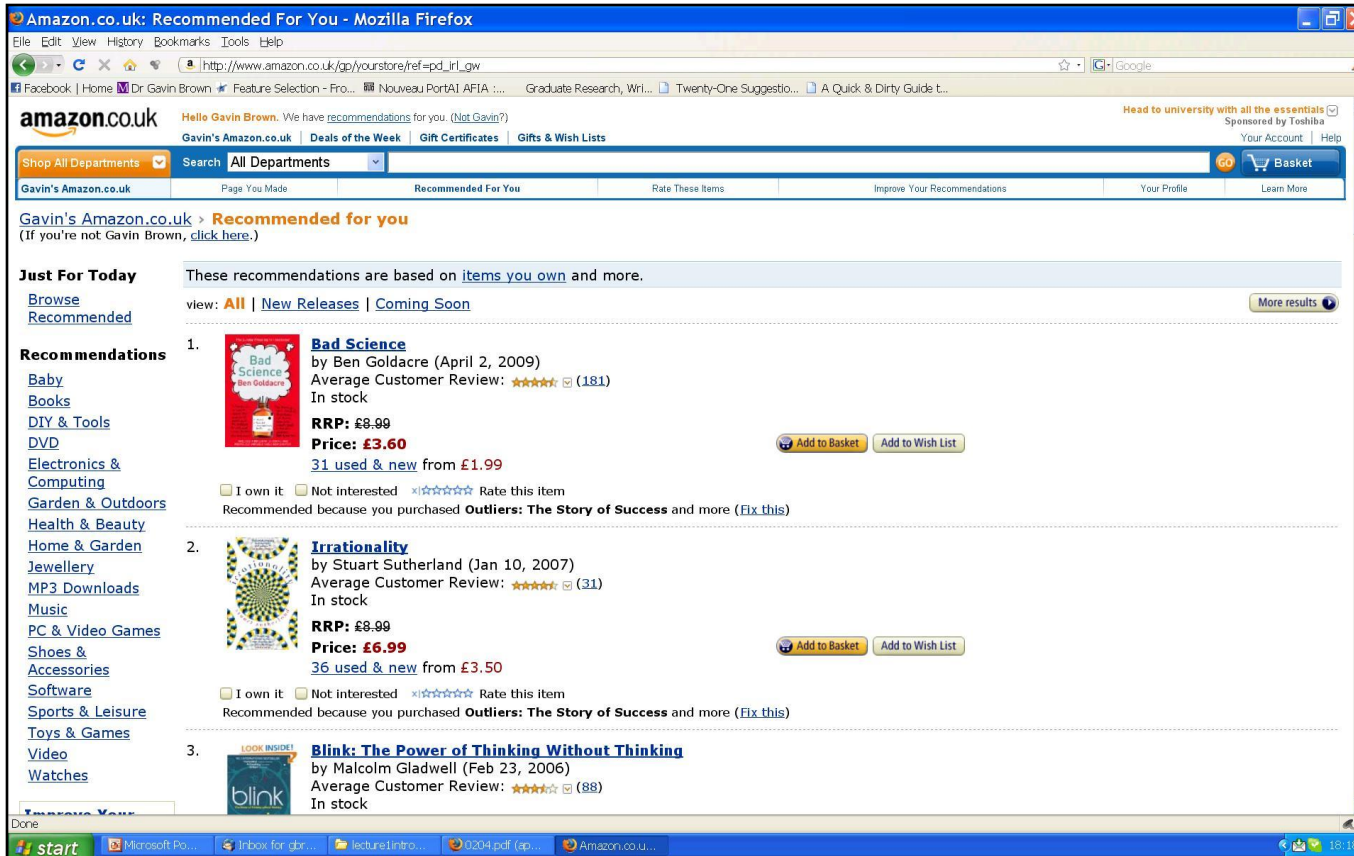
$P(Y | X)$ probability that somebody who buys X also buys Y where X and Y are products/services.

Example: $P(\text{chips} | \text{beer}) = 0.7$

- Matrix Completion

	t_1	t_2	.	.	t_m
u_1	0.1			0.7	
u_2					0.3
.				0.3	
.	0.4				
u_n		0.2			0.6

- Using machine learning to recommend books.



ALGORITHMS

Collaborative Filtering
Nearest Neighbour
Clustering

Classification

Features

- Predict unseen data from seen data

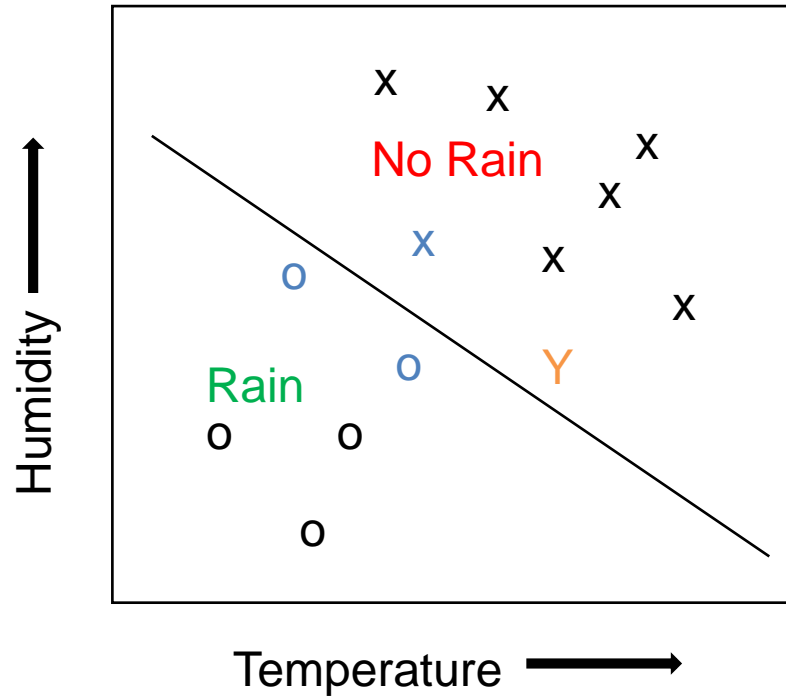
Temperature	Humidity	Rain	
55	10%	no	Class I
40	40%	yes	Class II
95	20%	no	Class I
75	45%	yes	Class II

- Will it rain at 70 degrees and 30% humidity?

Learn a Machine
using Training Data

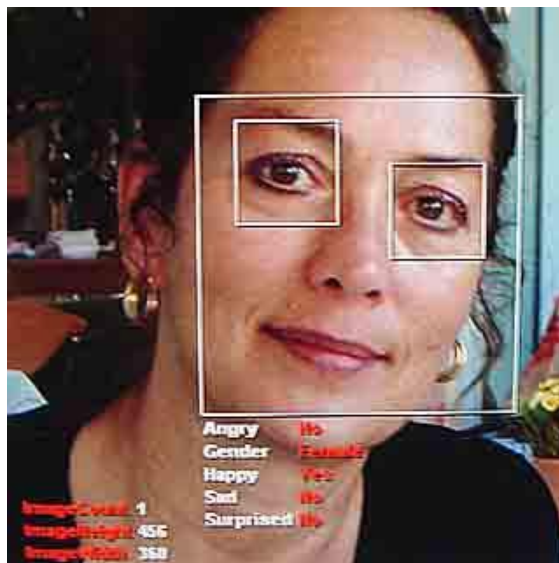
Class ?

Basic Strategy of Classification



Draw a boundary between the classes

- Using machine learning to identify faces and expressions.



ALGORITHMS

Decision Trees
Adaboost

Classification: Some other Applications

- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
 - Use of a dictionary or the syntax of the language.
 - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Medical diagnosis: From symptoms to illnesses
- ...

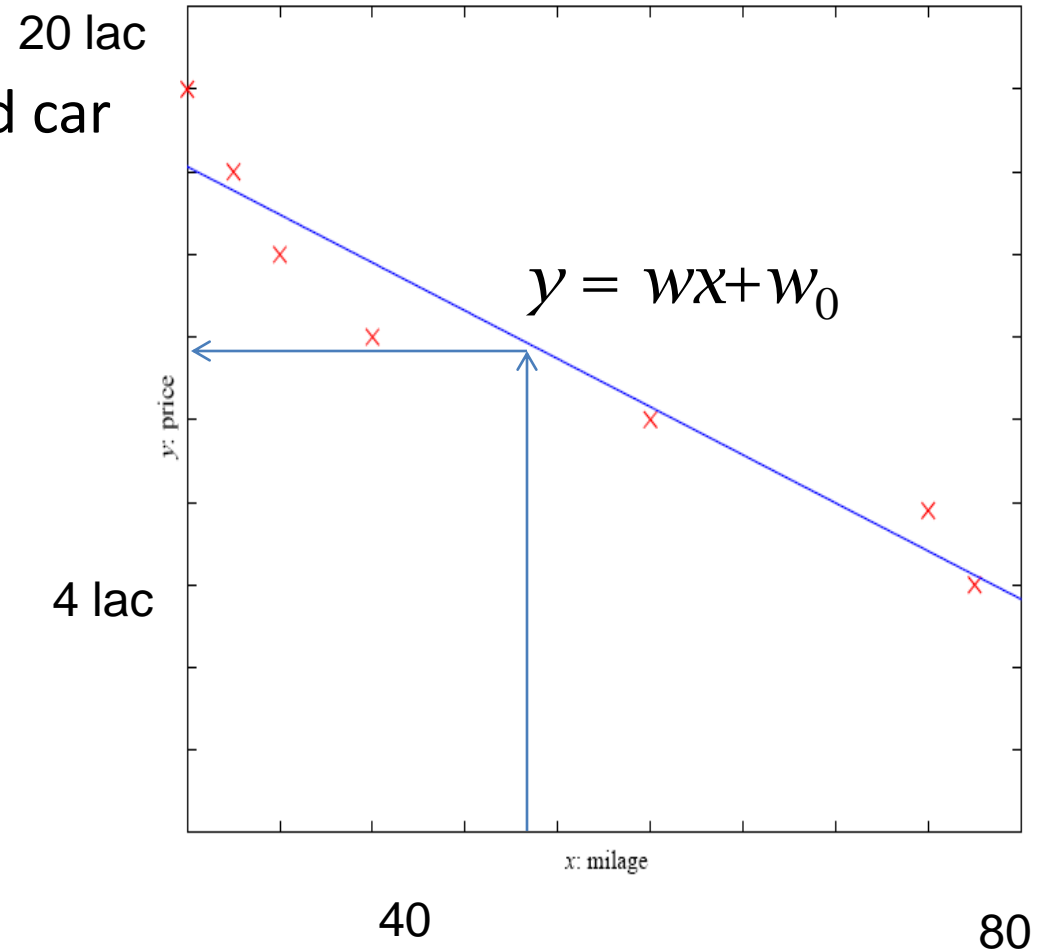
Regression

- Example: Price of a used car
- x : car attributes
- y : price

$$y = g(x | \vartheta)$$

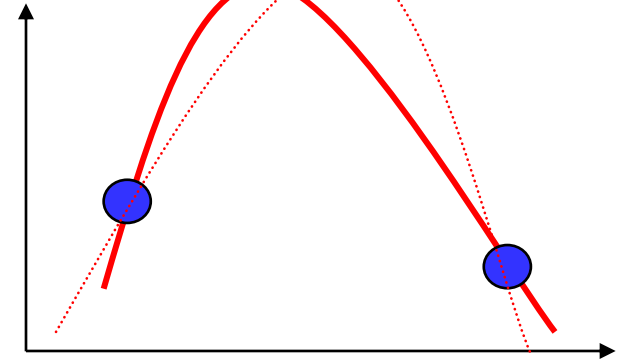
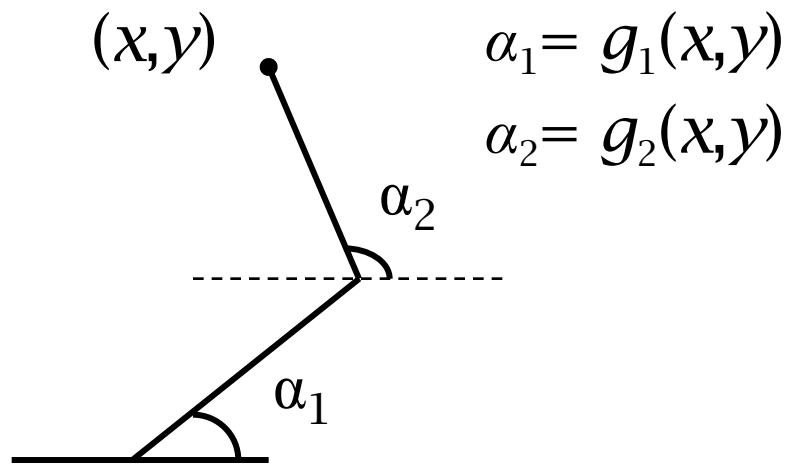
$g()$ model,

ϑ parameters



Regression Applications

- Navigating a car: Angle of the steering wheel
- Kinematics of a robot arm



- Response surface design

Supervised Learning: Uses

- **Prediction of future cases:** Use the rule to predict the output for future inputs
- **Knowledge extraction:** The rule is easy to understand
- **Compression:** The rule is simpler than the data it explains
- **Outlier detection:** Exceptions that are not covered by the rule, e.g., fraud

Unsupervised Learning

- Learning “what normally happens”
- No output
- Clustering: Grouping similar instances
- Example applications
 - Customer segmentation in Customer relationship Management (CRM)
 - Image compression: Color quantization
 - Bioinformatics: Learning motifs

Reinforcement Learning

- Learning a policy: A sequence of outputs
- No supervised output but delayed reward
- Credit assignment problem
- Game playing
- Robot in a maze
- Multiple agents, partial observability, ...



Bshop.flv