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,

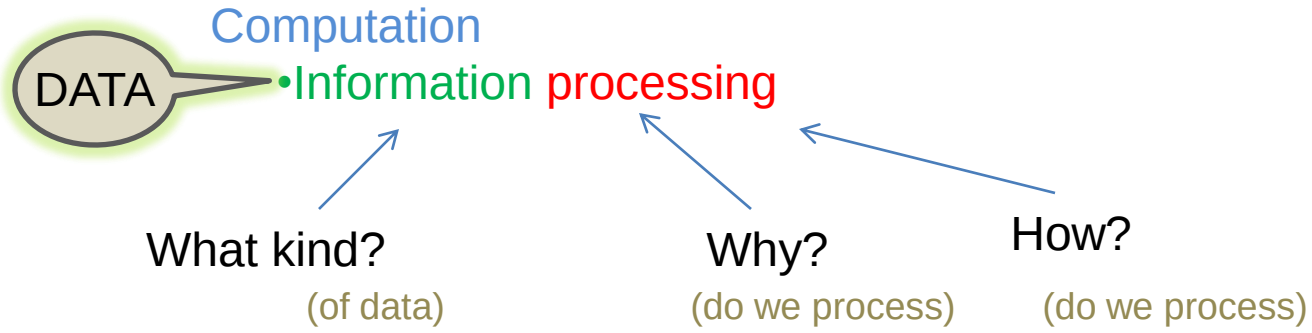
Advance Issues and Advance Computation

Swagatam Das,
Snehasis Mukherjee,
Yusuf Akhtar

Electronics and Communication Sciences Unit
Indian Statistical Institute, Kolkata

Intelligence

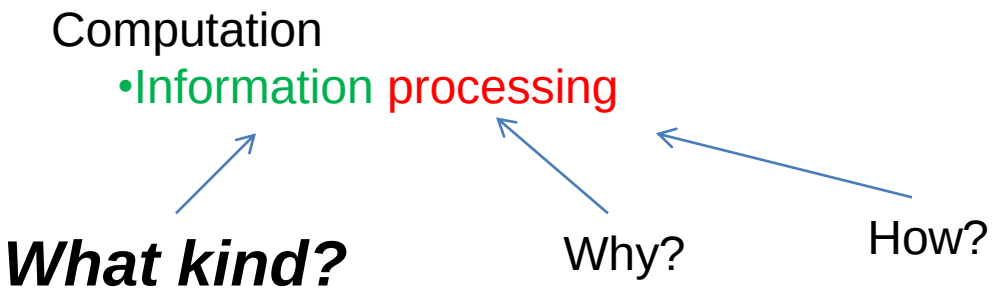
- Ability to solve problems
- Examples of Intelligent Behaviors or Tasks
  1. Classification of texts based on content
  2. Heart disease diagnosis
  3. Chess playing
  4. Etc..



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

Computation

- Information processing


Computation
•Information processing
What kind?
Why?
How?

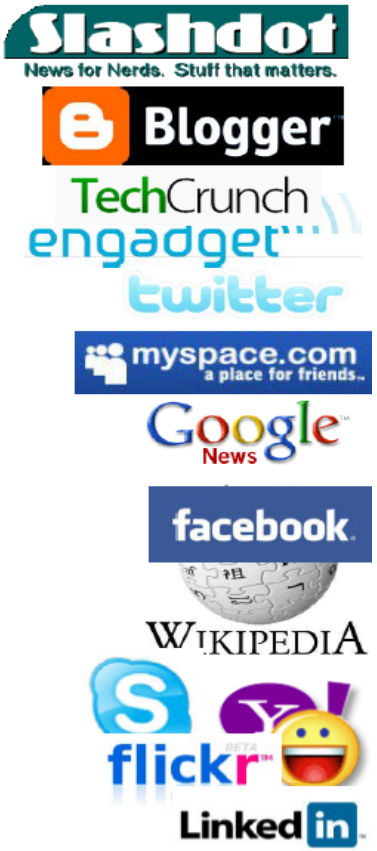
What kind? ⟶ Information processing

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

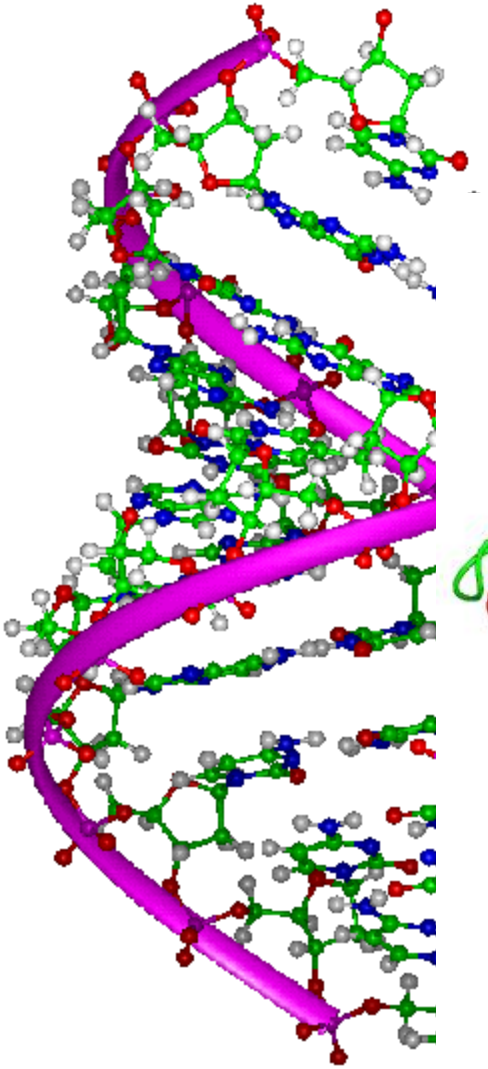


What kind? → Information processing

Computer Vision

What kind? ⟶ Information processing

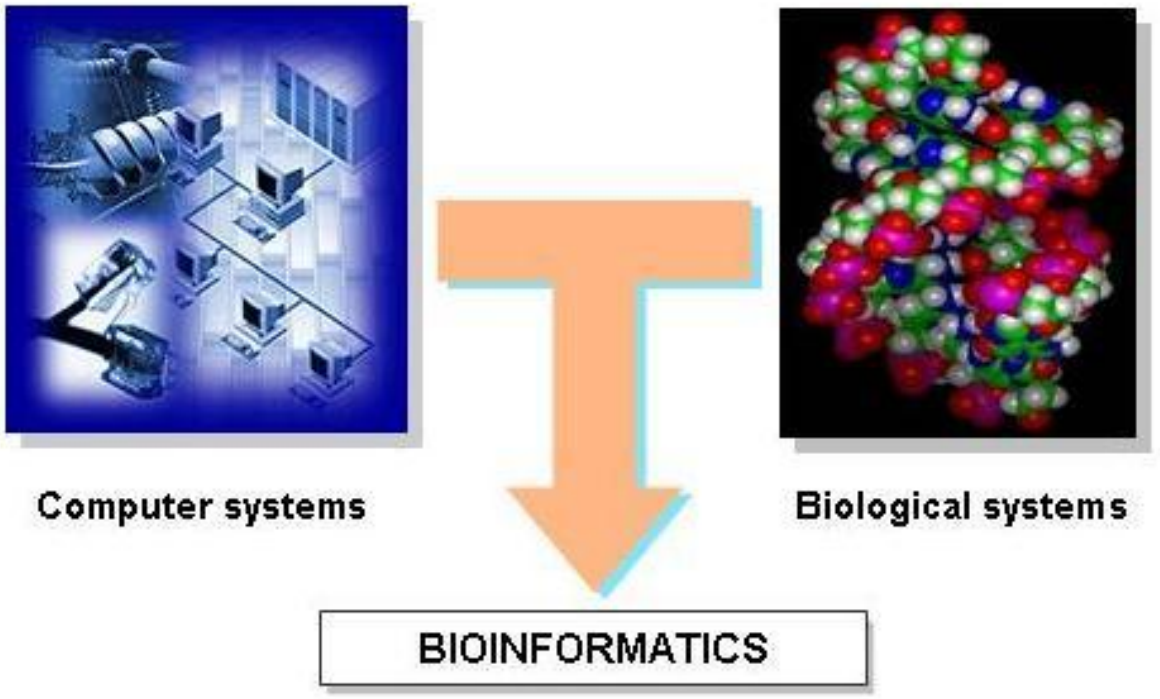
Bioinformatics :



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.

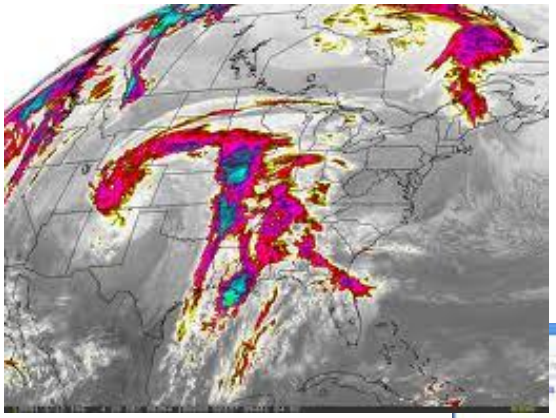
What kind? ⟶ Information processing

Bioinformatics :



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.



Computation

•Information processing

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

Computation

•Information processing

What kind?

Why?

How?

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

Information processing ← Why?

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

Information processing ← Why?

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?

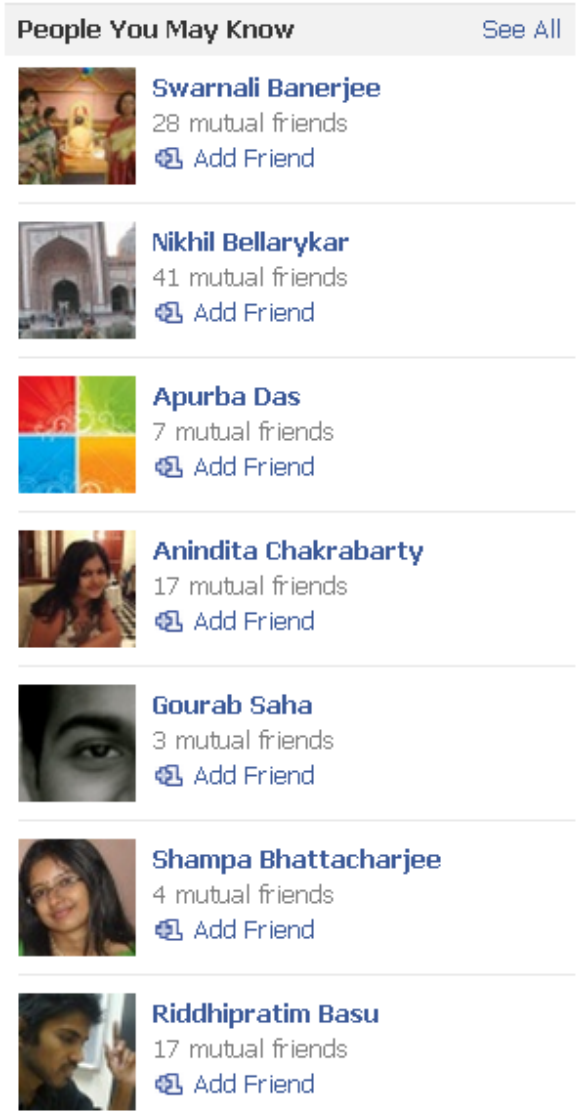
Internet : Social Media

amazon.com

ebay

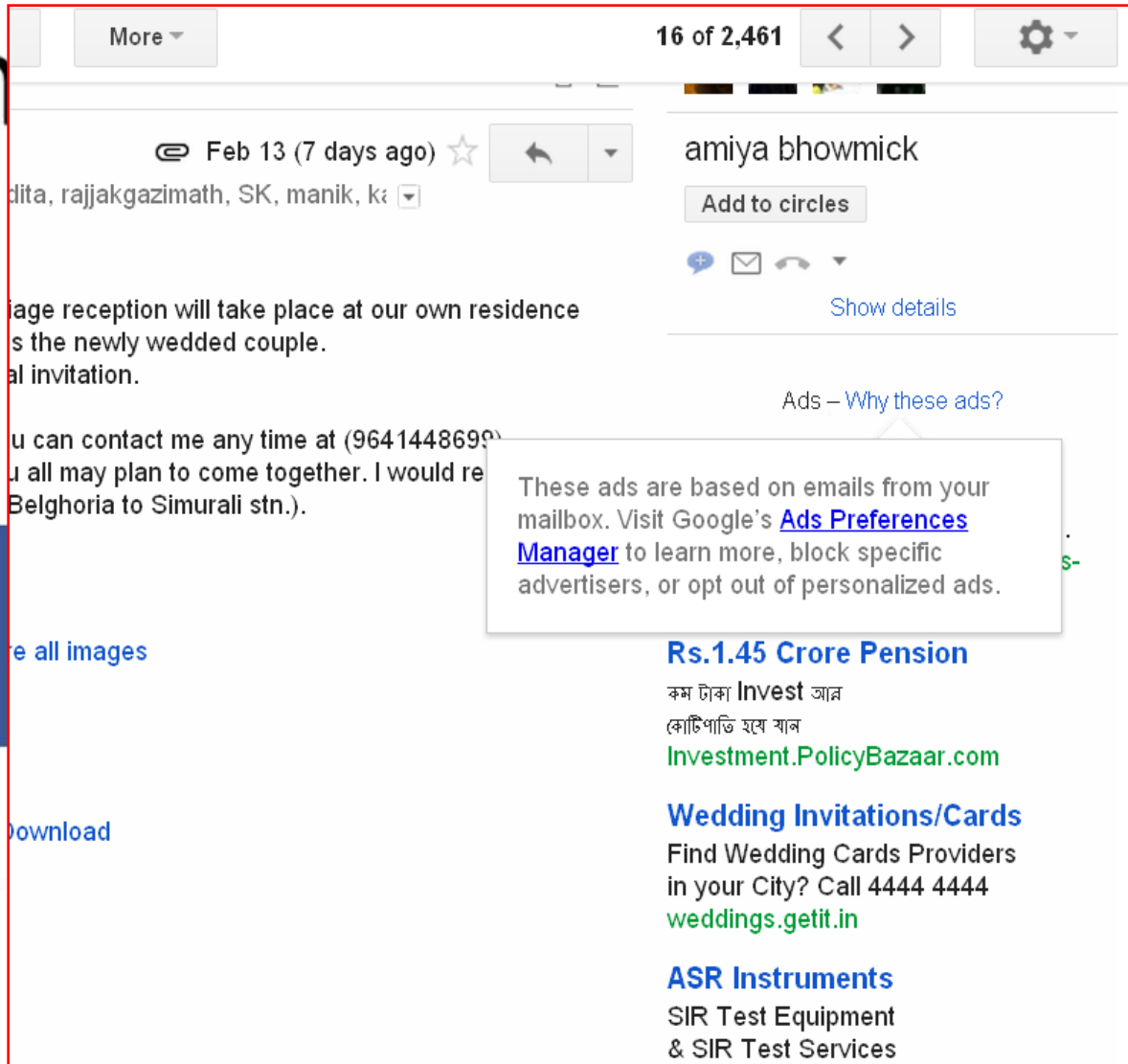
Information processing ← Why?

Internet : Social Media



Information processing ⟵ Why?

Internet : Social Media



YouTube

Information processing ⟵ Why?

Sraboner Dharar Moton
by saikatsanju
163,154 views

A demo on how to use FEMALE CONDOM
by carrotcommunications
1,096,835 views

Jokhon porbe na mor payer chinho ei bate
by dhrubadey
257,905 views

Shreya Ghosal Ekla Chalo Re.mov
by sunilaanand
340,298 views

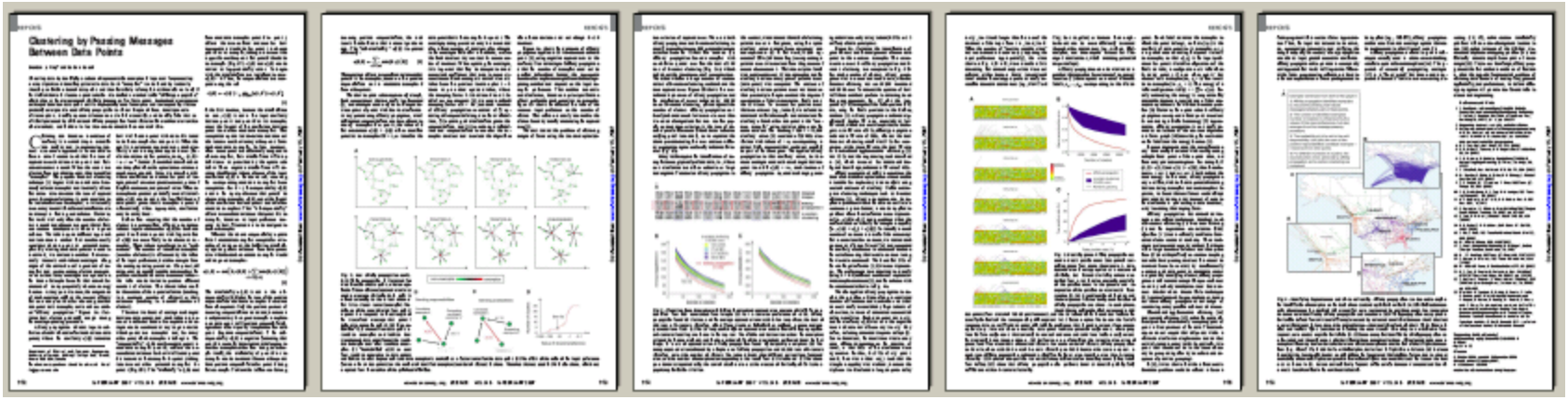
মকাভরে ঐ- মাহানা বাজপেয়ি
by nurulghani
44,513 views

Tomar Khola Hawa ~~ Shahana Bajpaie
by taniamoni
83,540 views

Tagor song_ আমি বড় বড় মুখর বাদল দিনে...
by nazmul911
103,726 views

Information processing ← Why?

Text Document analysis : Topic Model

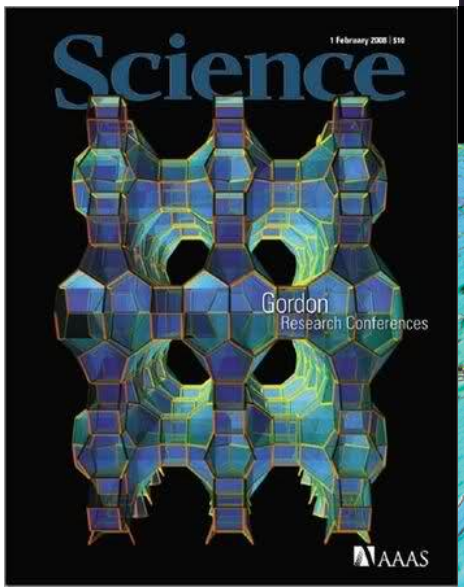


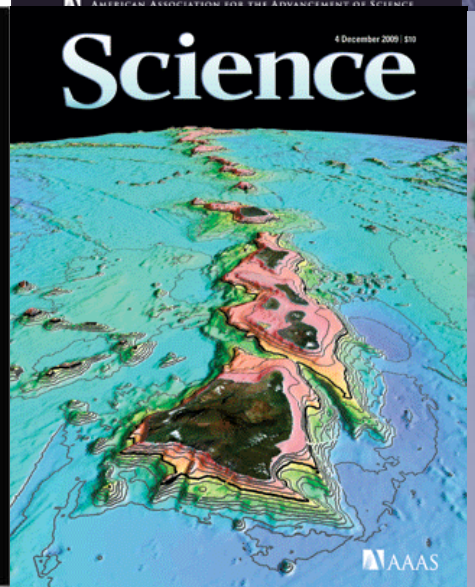
How can we develop algorithms that will

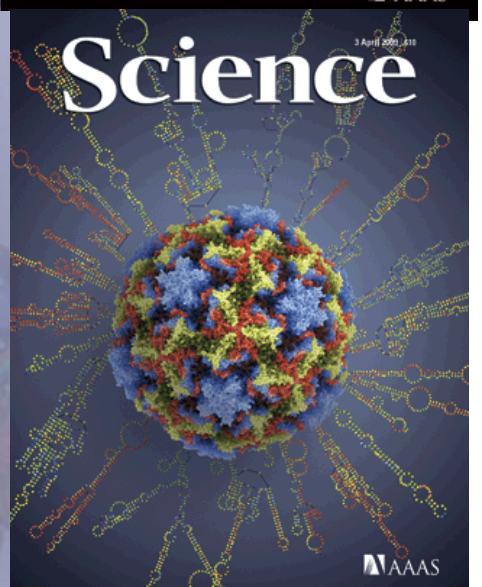
- Produce a summary of the document?
- Find similar documents?
- Predict document layouts that are suitable for different readers?

Information processing ← Why?

Example: Science Magazine Papers







Information processing ← Why?

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

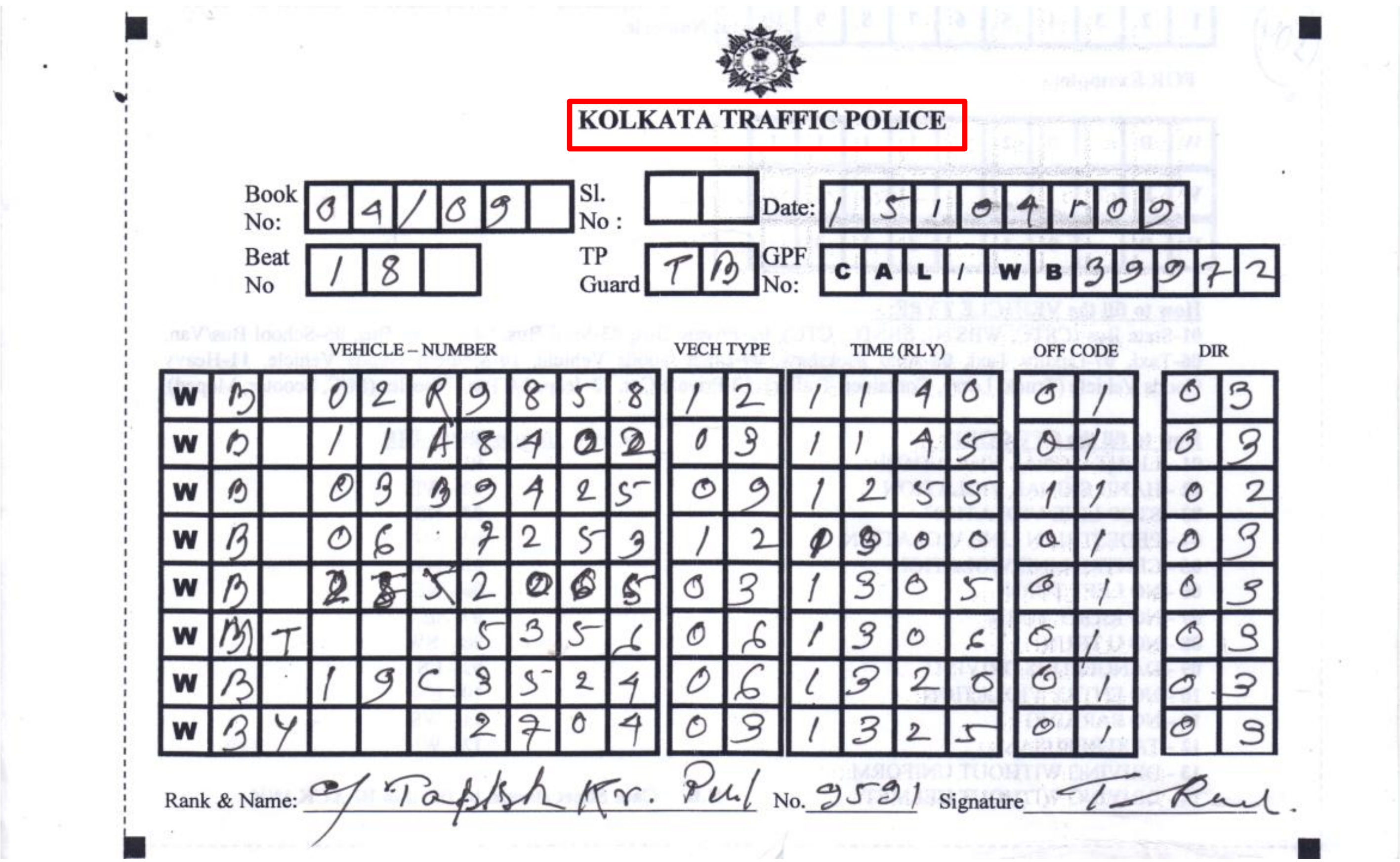
A **facial recognition system** is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source

Query Image

Image Database

Information processing ← Why?

Image Document : Optical Character(Text / Handwritten) Recognition



KOLKATA TRAFFIC POLICE

Book No: 0 4 / 0 9 | Sl. No : | Date: 1 5 / 0 4 / 0 9

Beat No: 1 8 | TP Guard: T B | GPF No: C A L / W B 3 9 9 7 2

		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 9	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 9 0 0	0 1	0 3
W	B	2 5 S 2 0 6 5	0 3	1 3 0 5	0 1	0 3
W	M T	5 3 5 6	0 6	1 3 0 6	0 1	0 3
W	B	1 9 C 3 5 2 4	0 6	1 3 2 0	0 1	0 3
W	3 Y	2 7 0 4	0 9	1 3 2 5	0 1	0 3

Rank & Name: S/ Tapash Kr. Paul No. 9591 Signature T. Paul

Information processing ← Why?

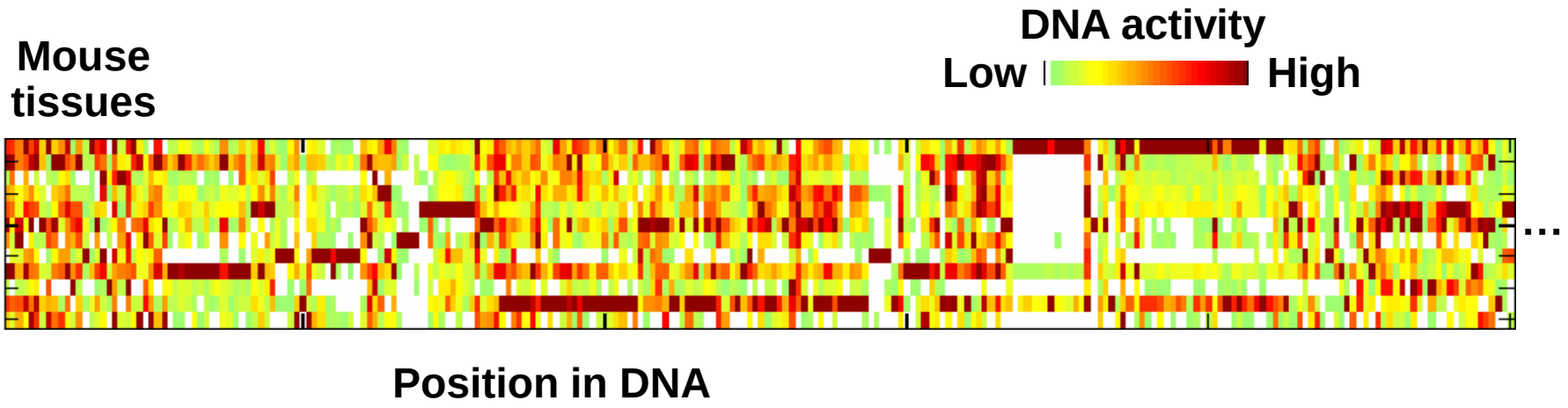
Computer vision : video

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



How can we develop algorithms that will

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

Computation

•Information processing

•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…

Computation

•Information processing

What kind?

Why?

How?

•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

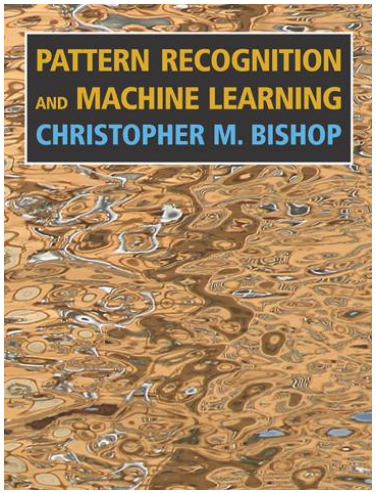
Information → Desired Goal

An introduction to

Machine Learning & Pattern Recognition

Pulak Purkait, SRF,
Indian Statistical Institute
KOLKATA

ETHEM ALPAYDIN

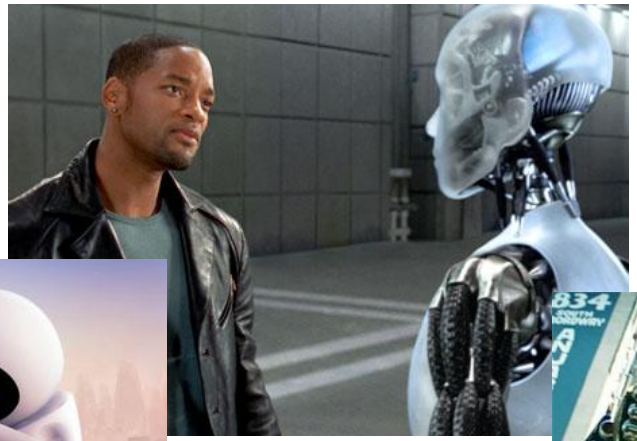


CHRISTOPHER M. BISHOP

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”

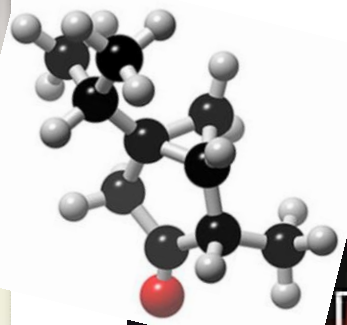


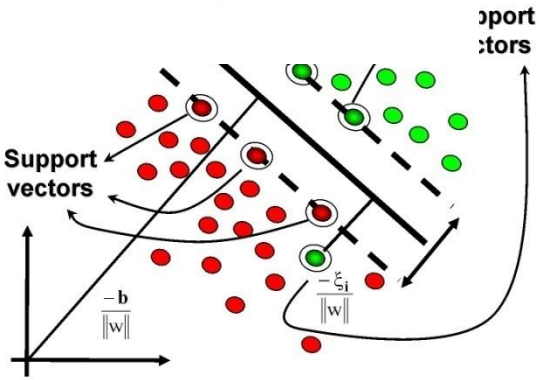
I Robot

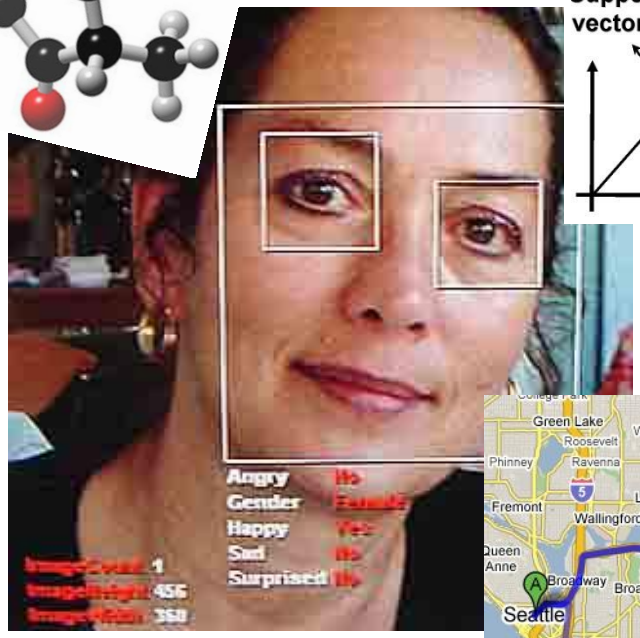
Wall-E

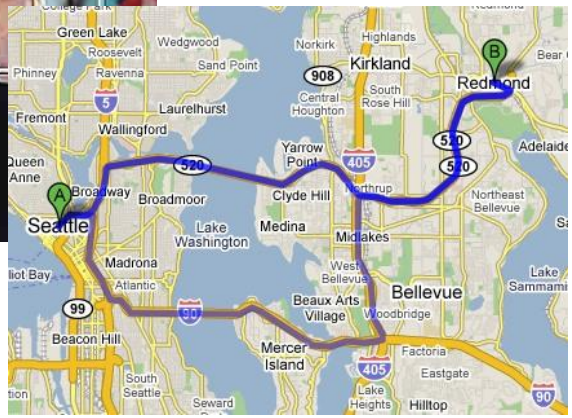
Transformer-3

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?

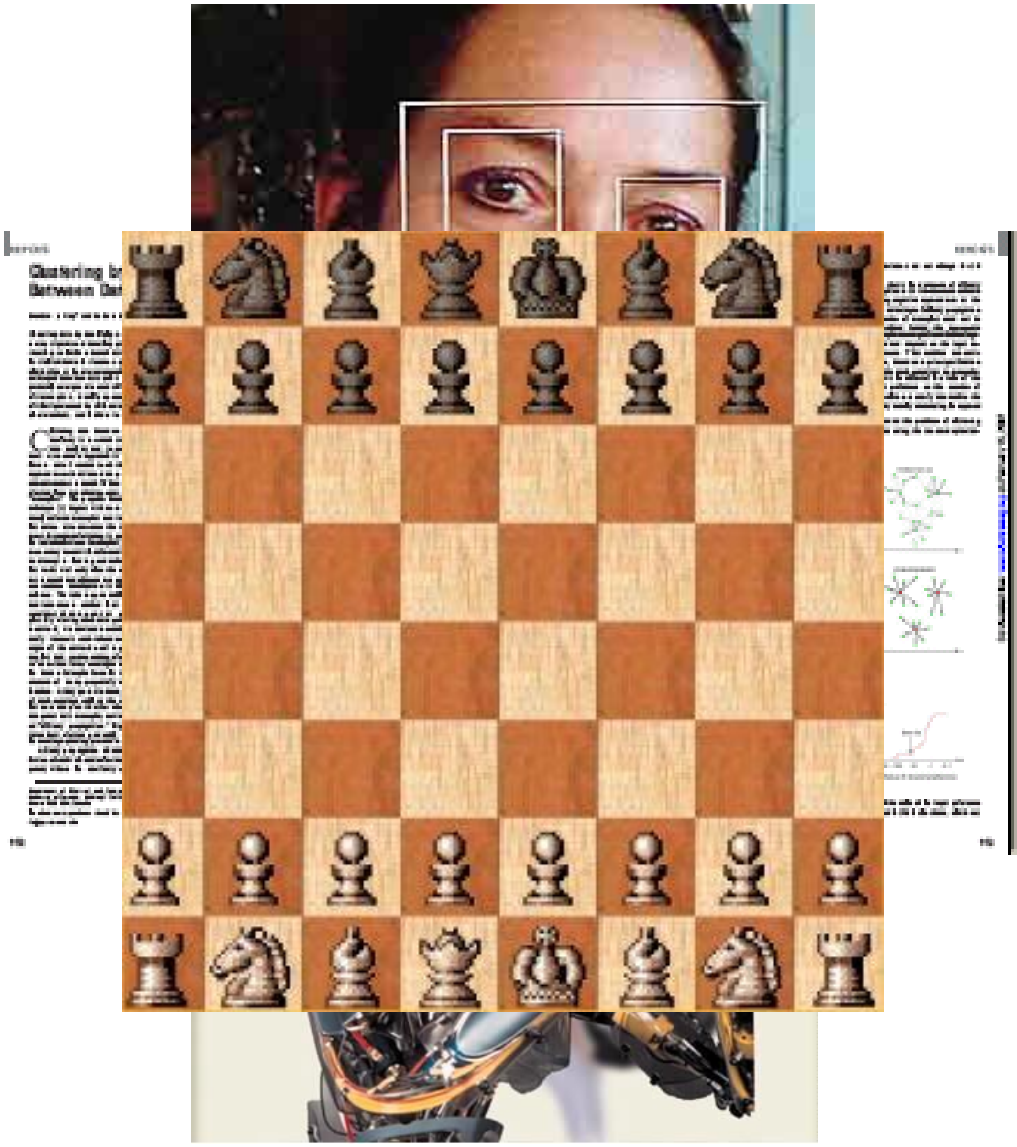
- The task is Pattern Recognition
- The way we do it called Machine Learning

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:

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

  Example: P (chips | 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

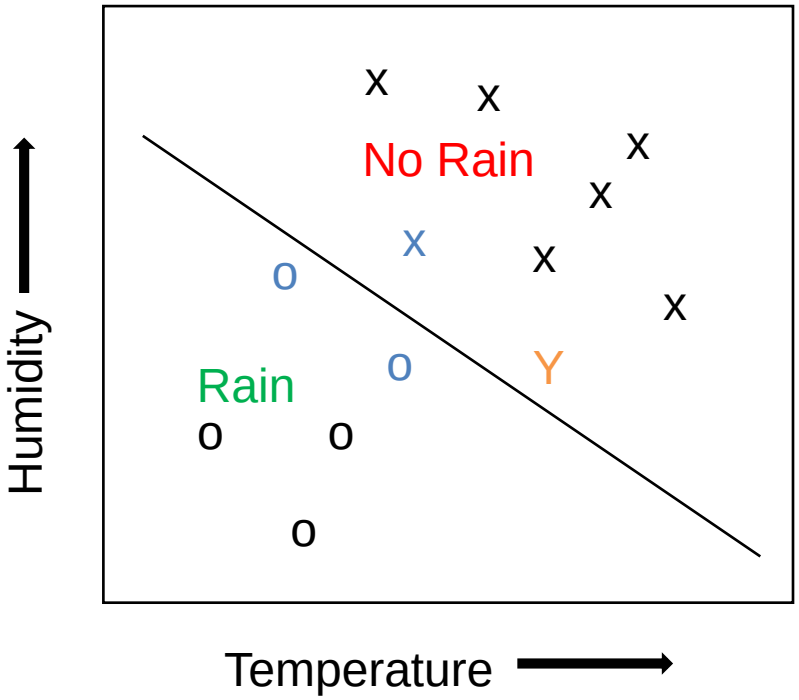
Features

- 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

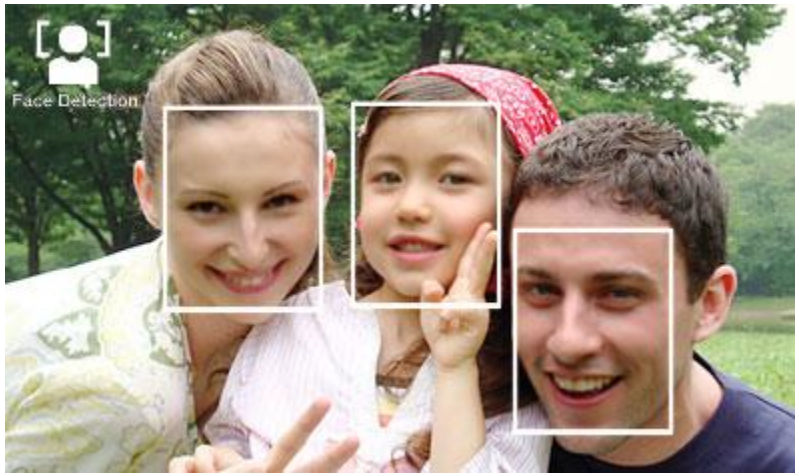
- Will it rain at 70 degrees and 30% humidity?

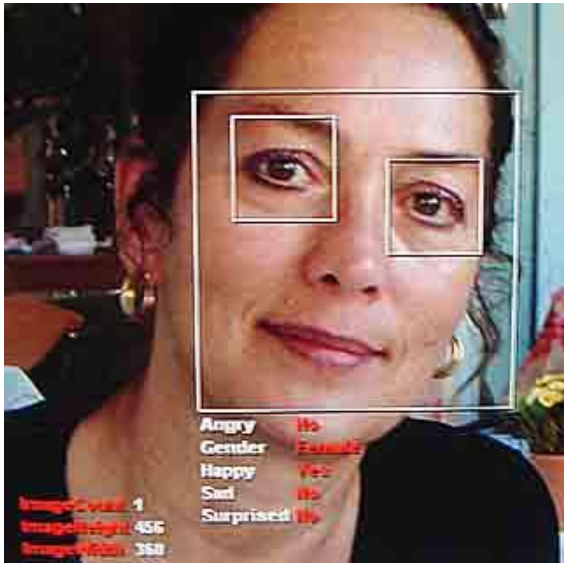
Basic Strategy of Classification



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

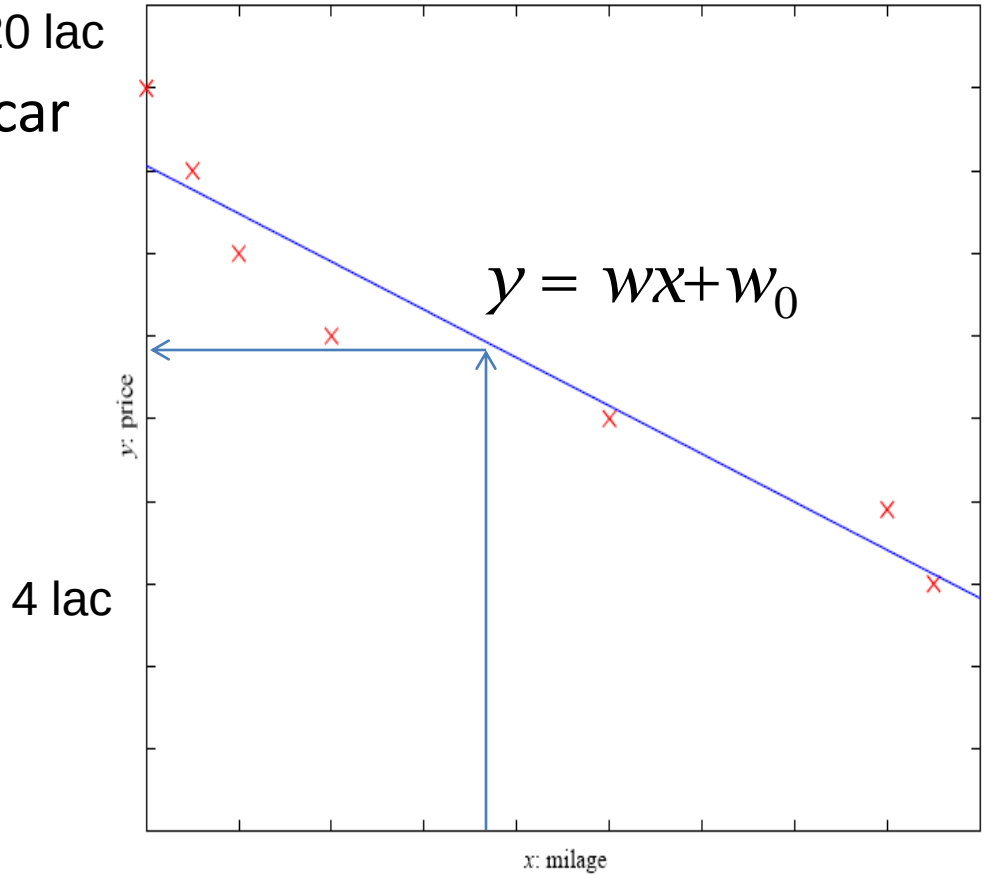
- Example: Price of a used car
- x : car attributes

  y : price

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

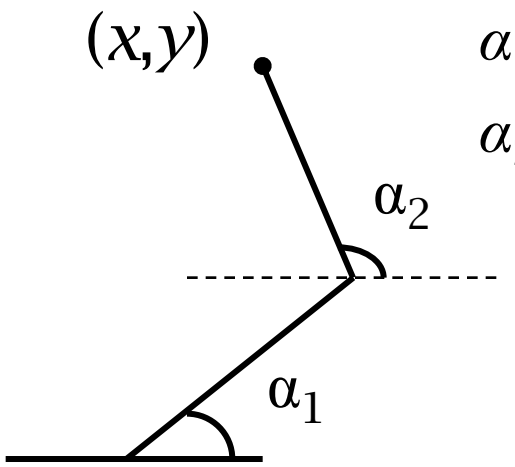
  $g(\ )$ model,

  ϑ parameters

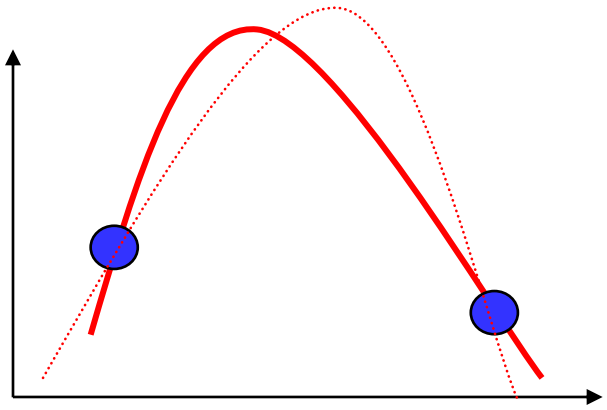


Regression Applications

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



$$\alpha_1 = g_1(x,y)$$
$$\alpha_2 = g_2(x,y)$$



- 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, ...

Bishop.flv