COMP4211: Machine Learning

James Kwok

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

What is the Machine Learning?

Learning = Improving with experience at some task

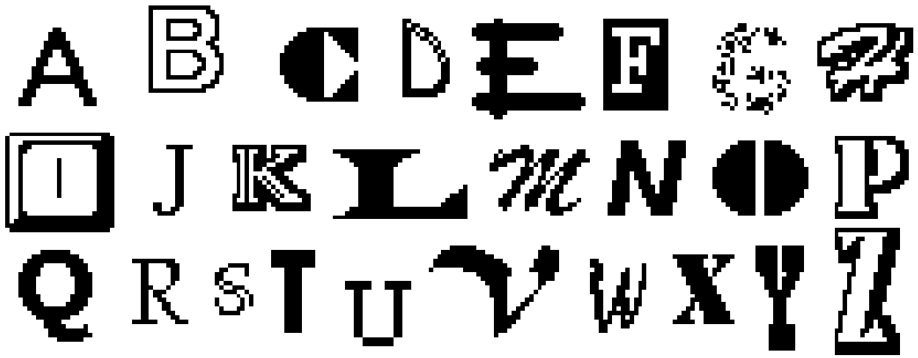
From [Simon 83]

- Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more effectively the next time

Machine Learning is Everywhere

Character recognition

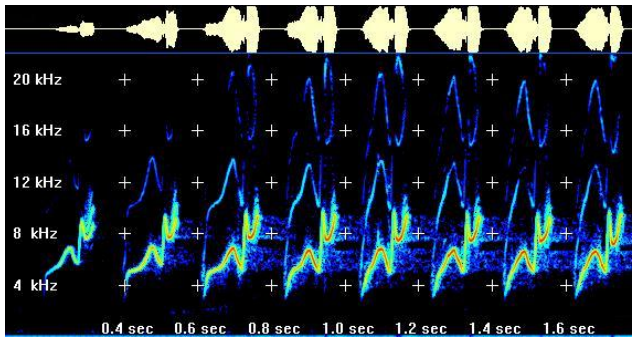
- raw data: image



- class: numerals, English (Chinese, etc.) characters

Speech recognition

- raw data: speech signal

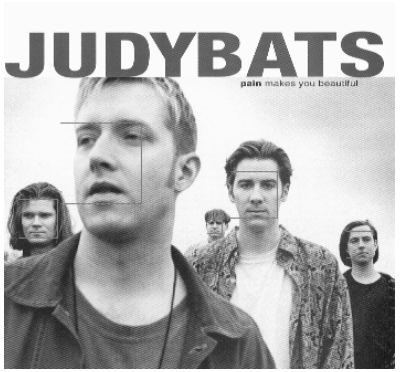


- class: spoken words

Examples...

Face detection

- raw data: image



- class: face/non-face

Fingerprint identification

- raw data: fingerprint image

- class: known/unknown person

Examples...

Document classification

- raw data: (web) document

EIA SAYS DISTILLATE, GAS STOCKS OFF IN WEEK

WASHINGTON, March 11 - Distillate fuel stocks held in primary storage fell by 8.8 mln barrels in the week ended March six to 119.6 mln barrels, the Energy Information Administration (EIA) said.

In its weekly petroleum status report, the Department of Energy agency said gasoline stocks were off 500,000 barrels in the week to 251.0 mln barrels and refinery crude oil stocks fell 1.2 mln barrels to 331.8 mln.

The EIA said residual fuel stocks fell 1.5 mln barrels to 36.4 mln barrels and crude oil stocks in the Strategic Petroleum Reserve (SPR) rose 200,000 barrels to 516.7 mln.

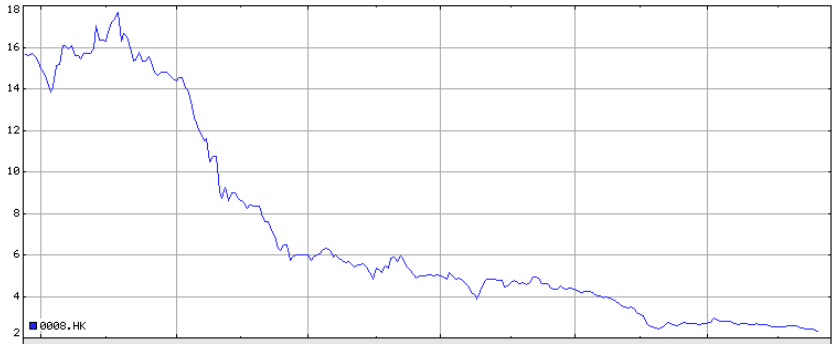
The total of all crude, refined product and SPR stocks fell 10.3 mln barrels to 1,564.8, it said.

- class: semantic categories (e.g. fuel, money, China)

Examples...

Financial engineering

- raw data: financial timeseries (e.g. stock prices)



- class: financially healthy / unhealthy company

Examples...

DARPA Grand Challenge

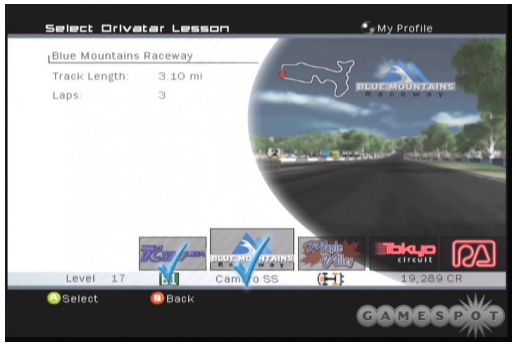
- a prize competition for driverless vehicles

2007 Urban Challenge

- urban area course; need to obey all traffic regulations

Examples...

Forza Motorsport



- drivatar: you teach it to drive like you do
- invite it into the passenger seat of a few different cars and let it observe how you drive

Bioinformatics

How Can This Be Solved?

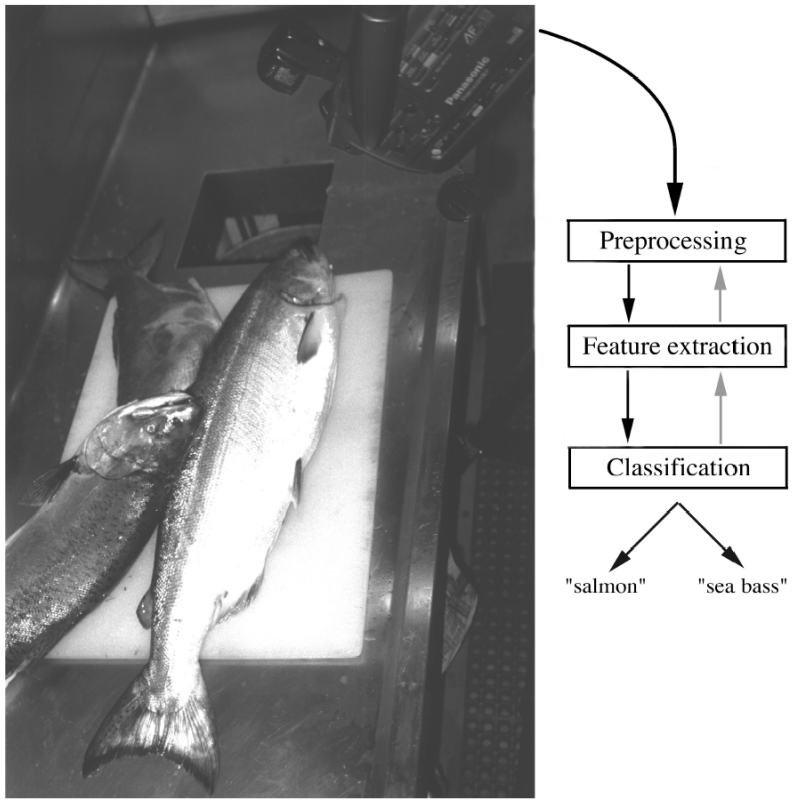
Is this an apple?

need training examples

apples

not apples

Example: Automatic Fish Sorting



Different types of fish differ in length, lightness, width, number and shape of fins, position of the mouth, etc

Also there are variations in lighting, position of the fish on the conveyer, etc

- noise
- needs pre-processing

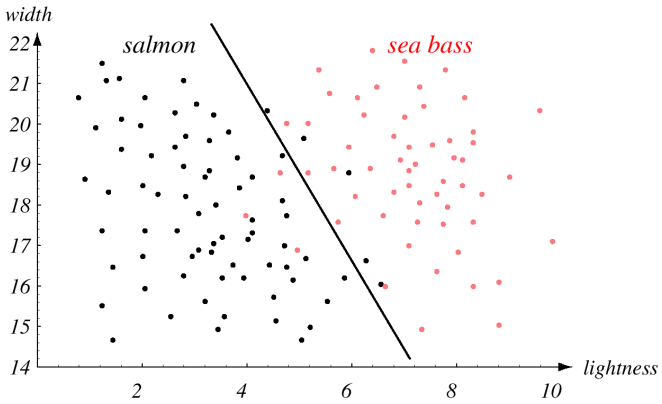
Pre-Processing

Example (remove the background)

Example (adjust for light level)

Feature Extraction

Consider each fish as a point in some feature space

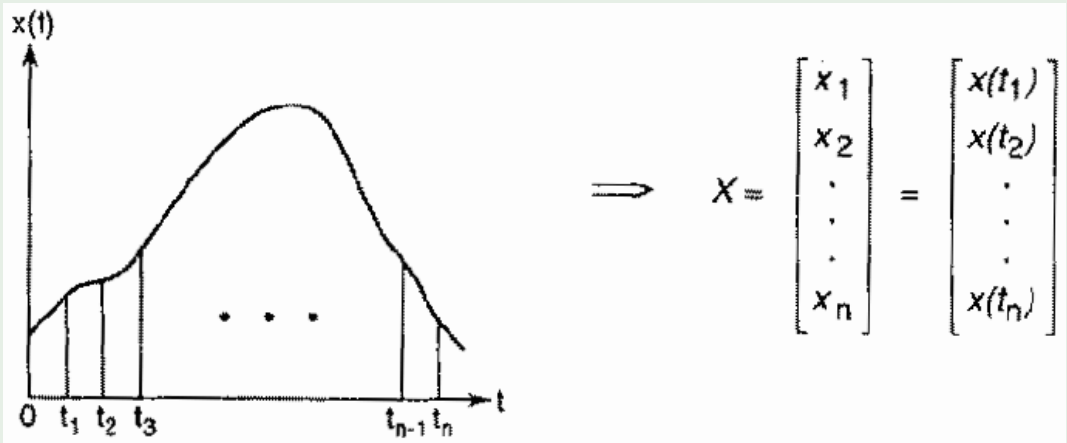


2-dimensional feature space:

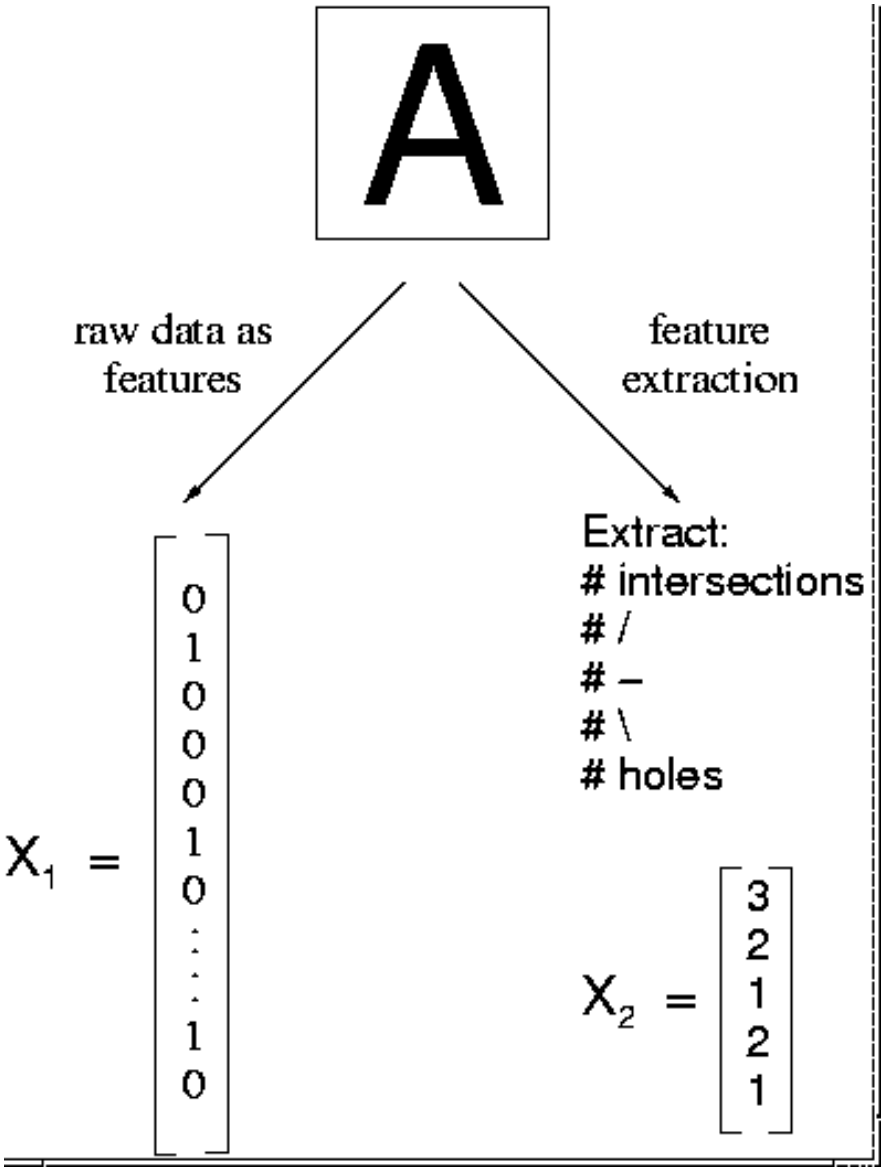
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- feature vector

Example (waveform)



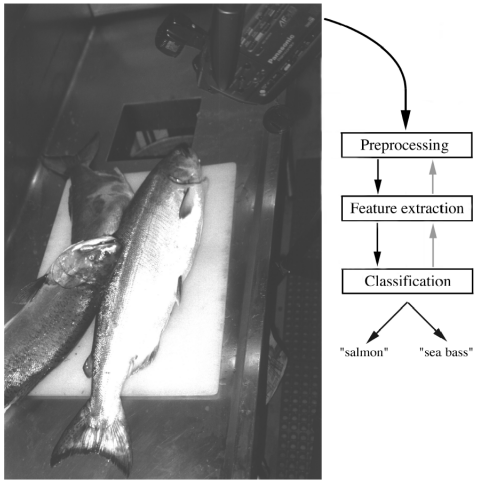
Example Features



Learning (Training)

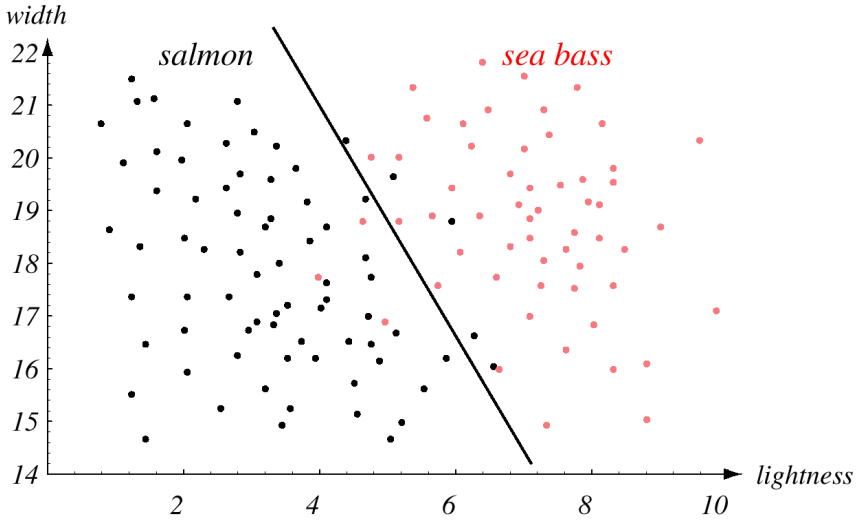
Training sample

- tell me which species a particular fish belong to (data collection)



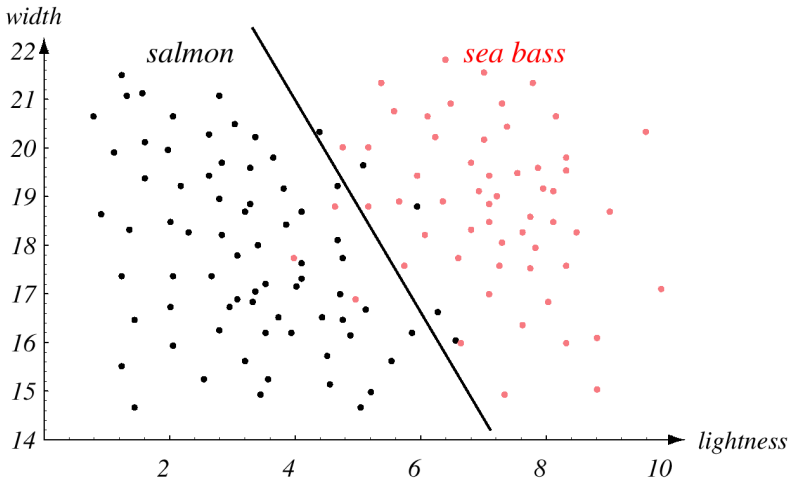
Partition the feature space into 2 regions, one for each type of fish

- decision boundary



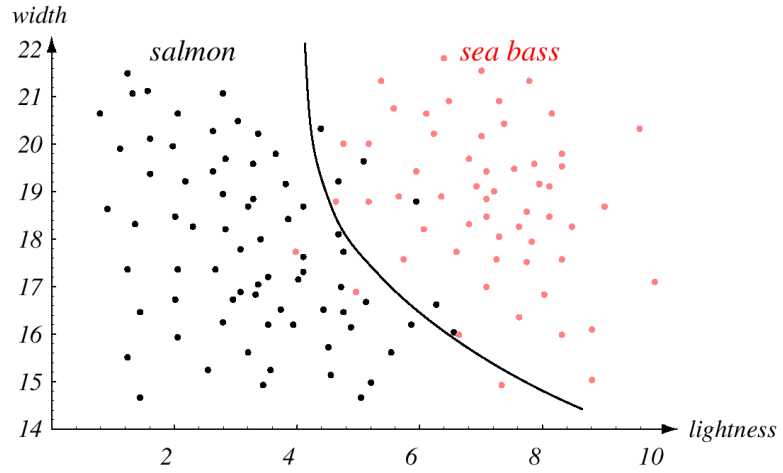
Classifier

Linear classifier



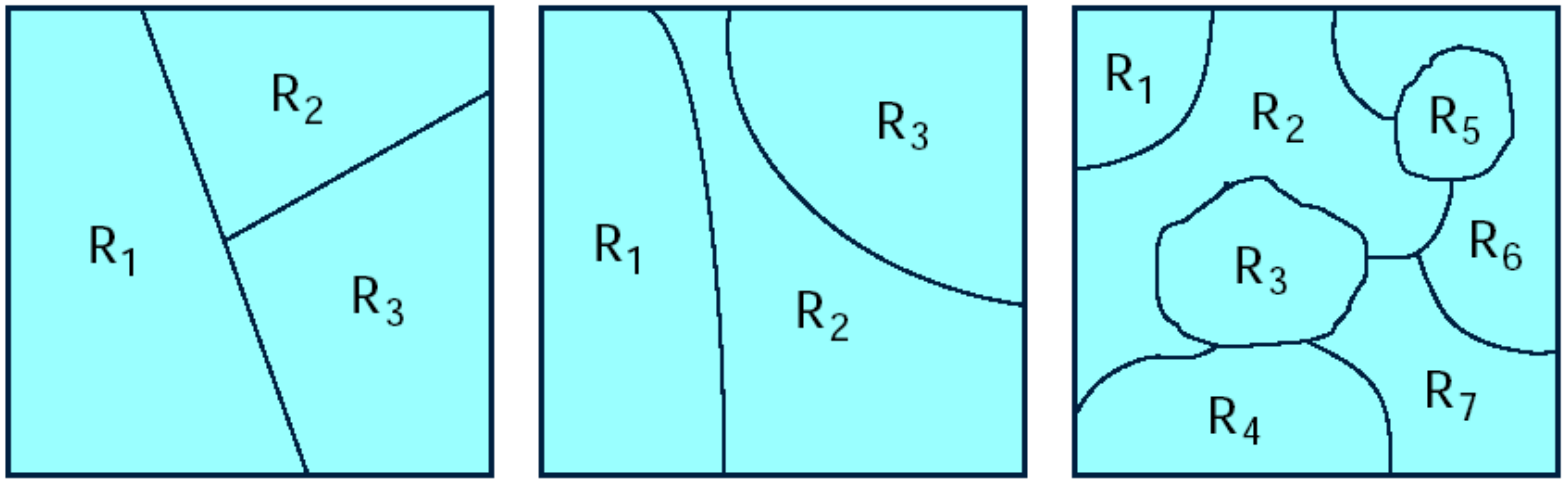
Can also use a more complicated decision boundary

- e.g. quadratic classifier



Classifier...

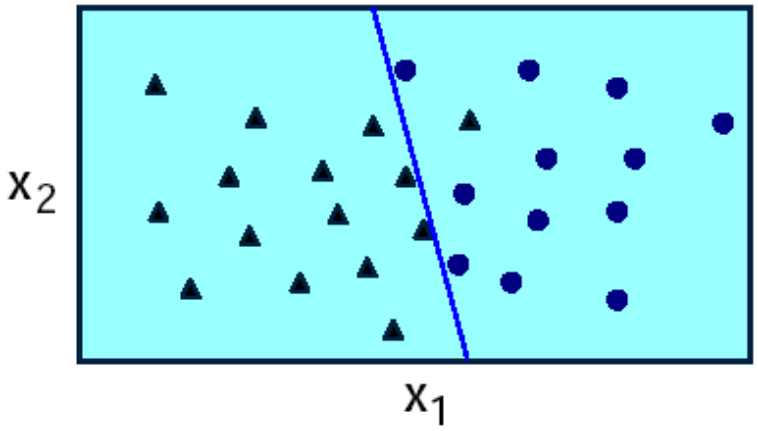
- can be even more complicated



How to learn the boundary (parameters)?

How to Handle New Images?

Use the trained classifier to classify the new image



How to measure the classifier performance?

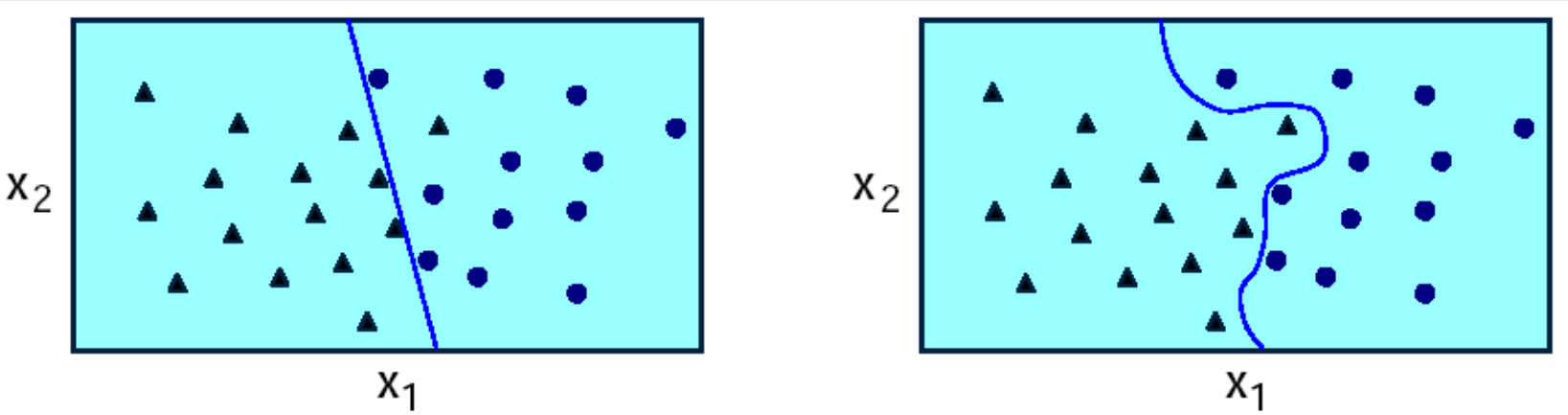
- classification error rate
  - % patterns that are assigned to the wrong category
- other aspects may be important too
  - e.g. computational complexity
  - e.g. user-friendliness

How to Handle New Images?...

Will the classifier work for this unseen fish?

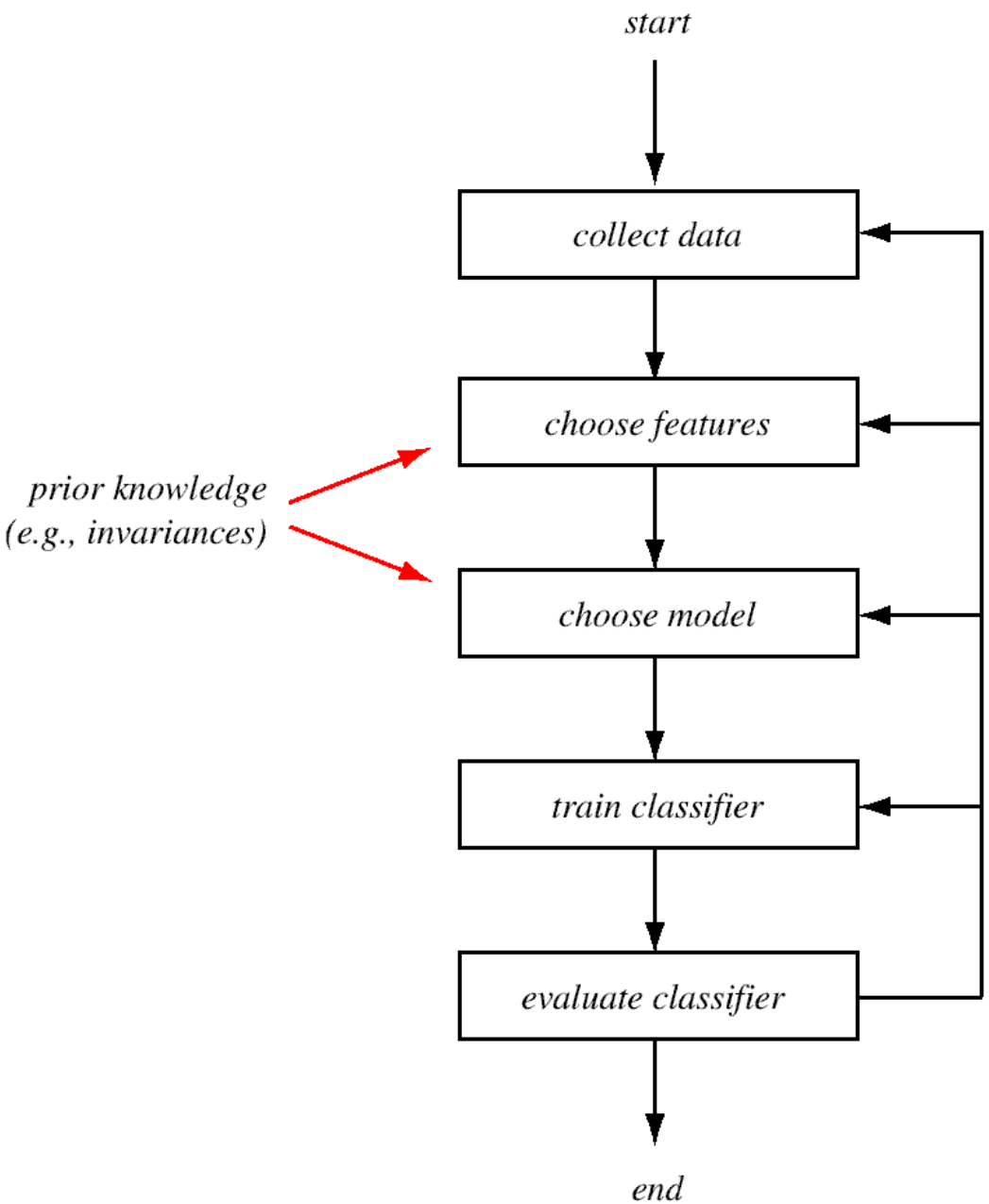
- issue of generalization

Example



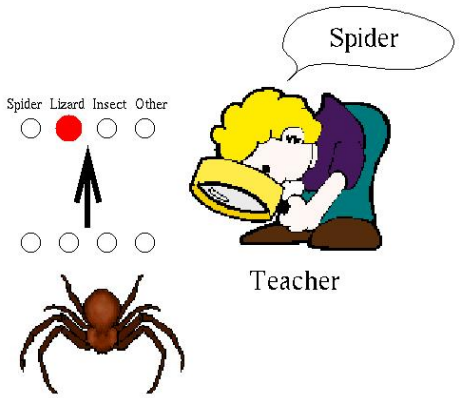
Which model is better?

Classifier Design



Learning Paradigm: Supervised Learning

The learner is provided with a set of inputs together with the corresponding desired outputs

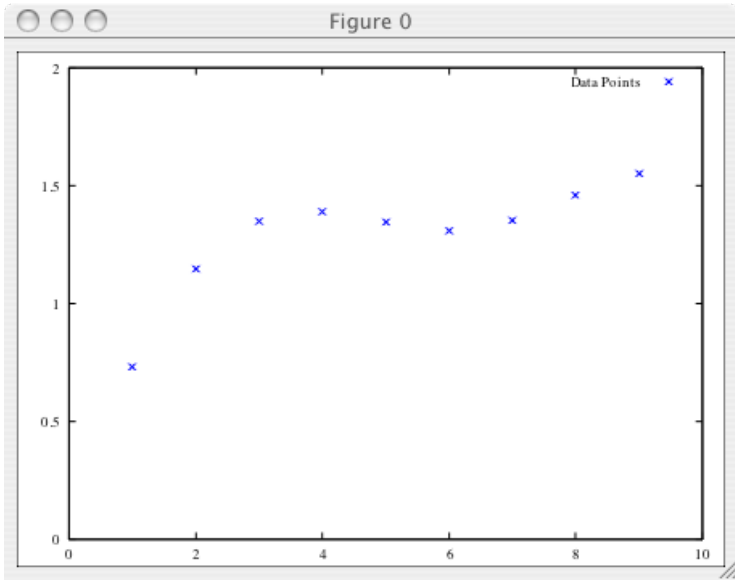


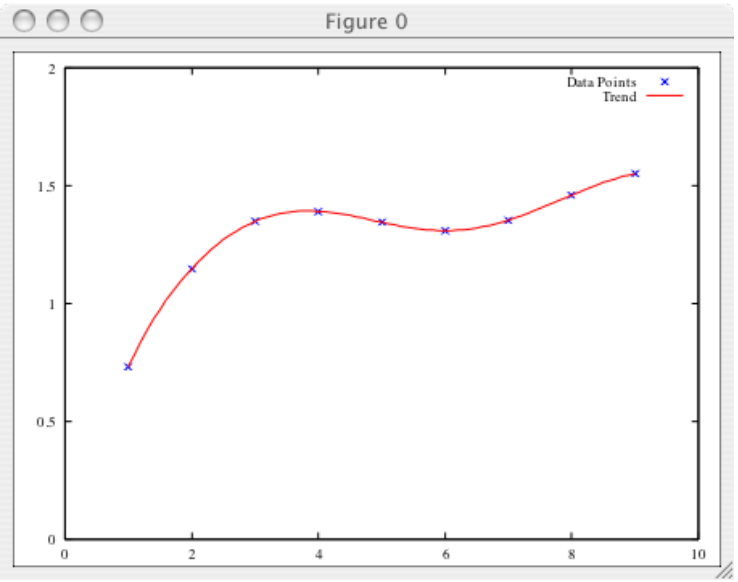
- has a "teacher"

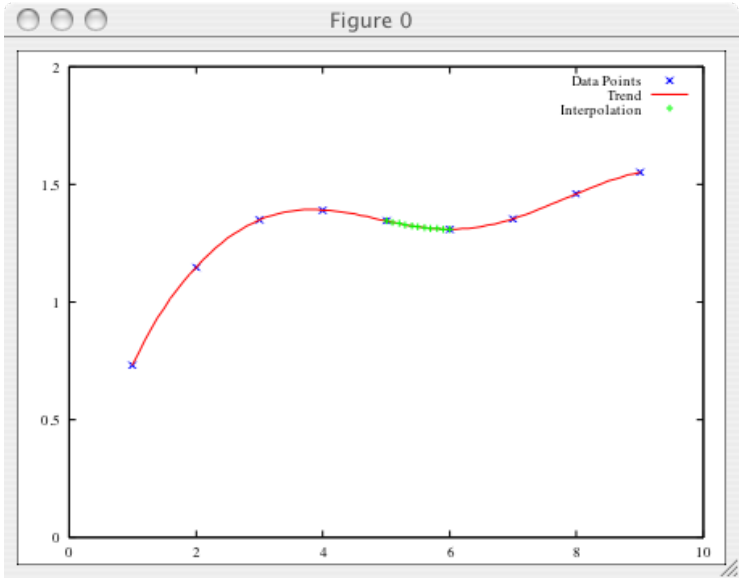
Example

- teaching kids to recognize different animals
- graded examinations with correct answers provided

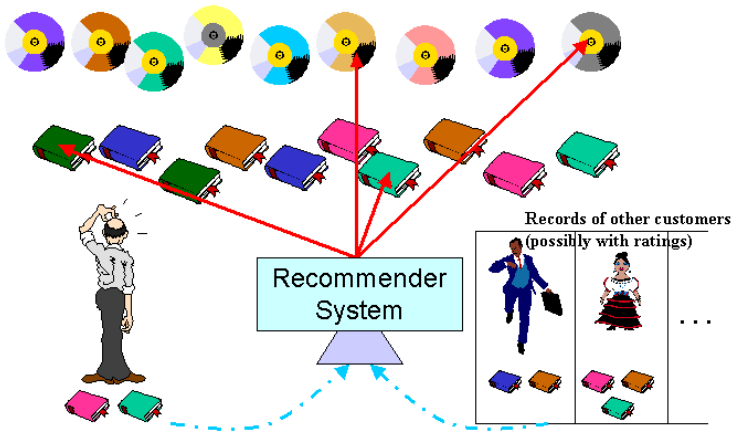
Example: Regression

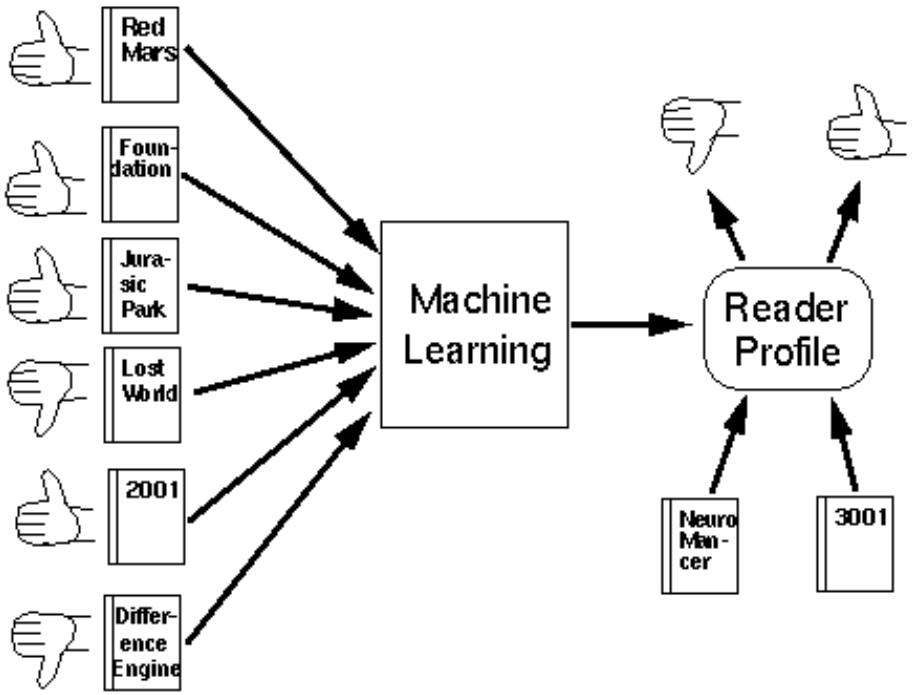




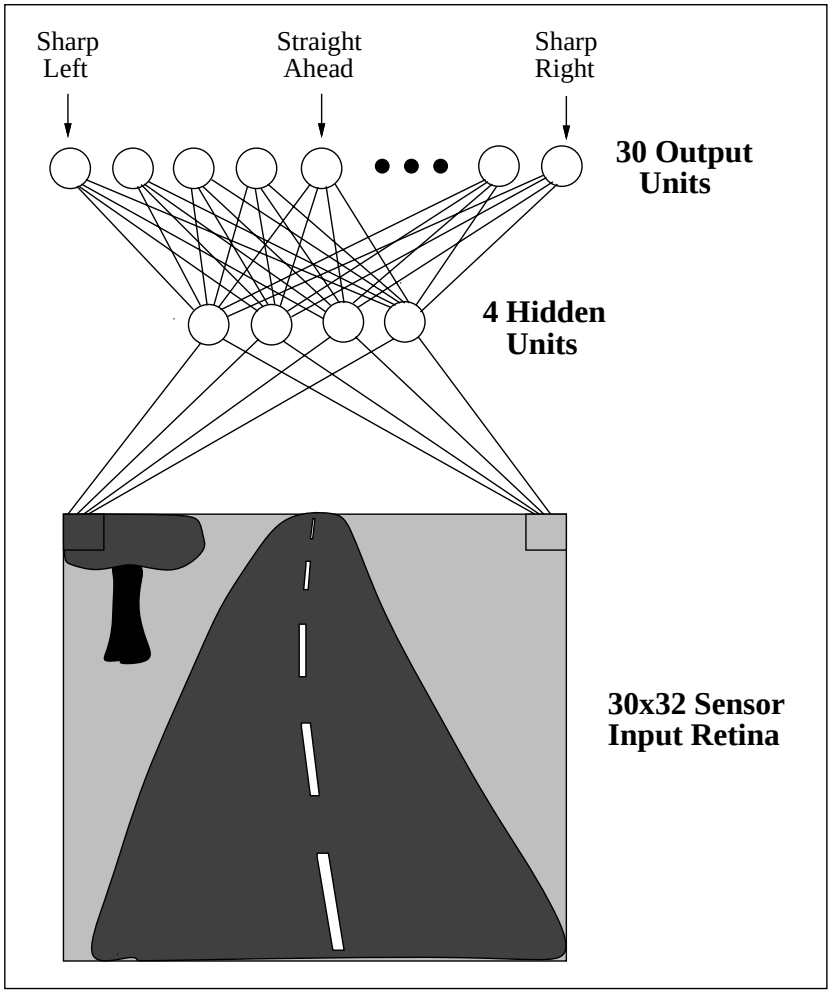


Example: Recommender System





Example: Control

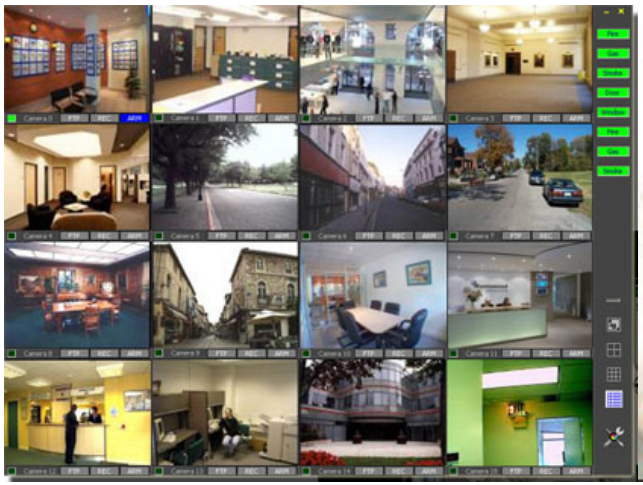


Examples

- text classification

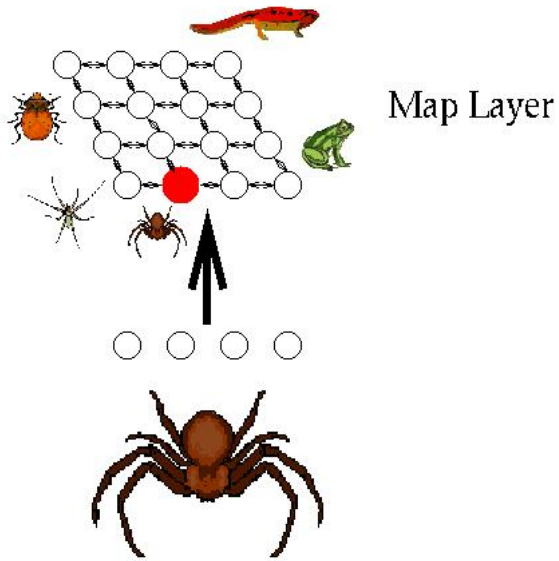
Google™ YAHOO!®

- face detection (video surveillance, digital camera)



Learning Paradigm: Unsupervised Learning

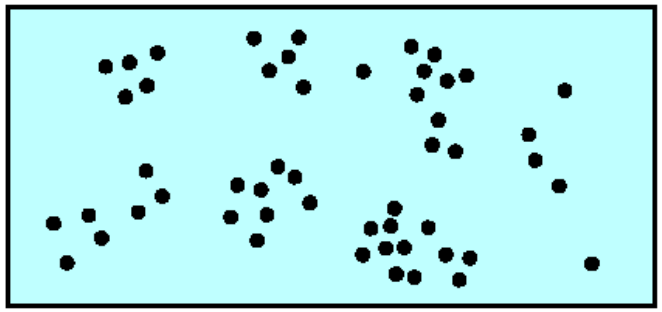
Training examples as input patterns, with no associated output



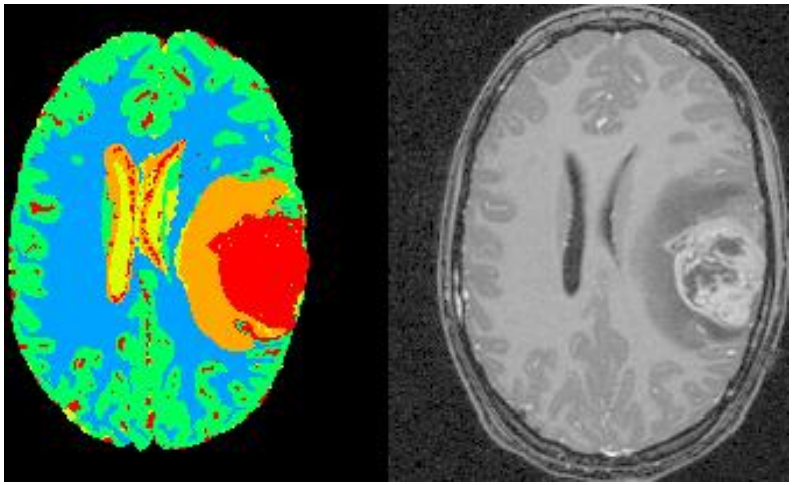
- no teacher
- similarity measure exists to detect groupings / clusterings

Uses of Unsupervised Learning

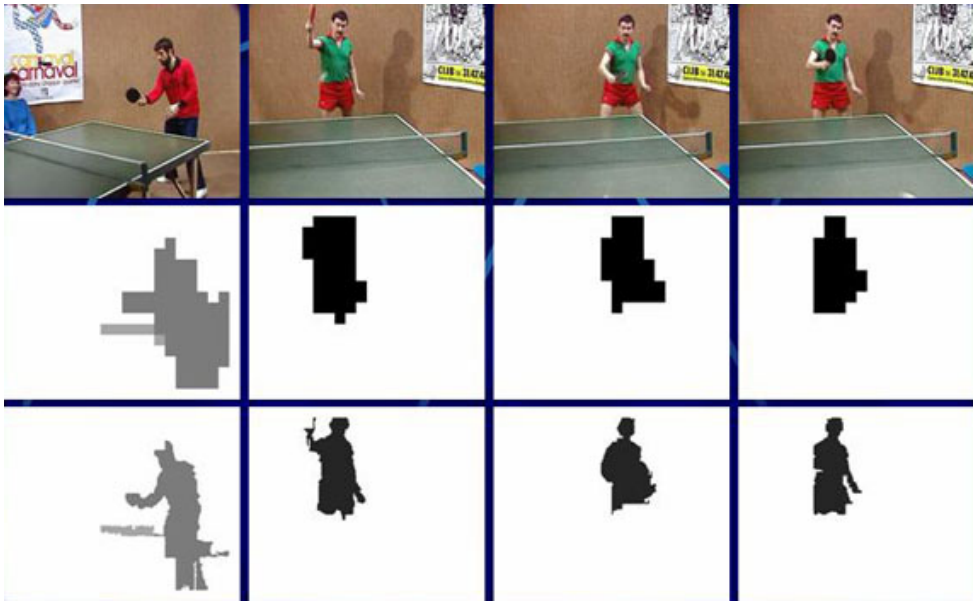
- find clusters



  - in the early stages of an investigation, it may be helpful to perform exploratory data analysis to gain some insight into the nature or structure of the data



Examples

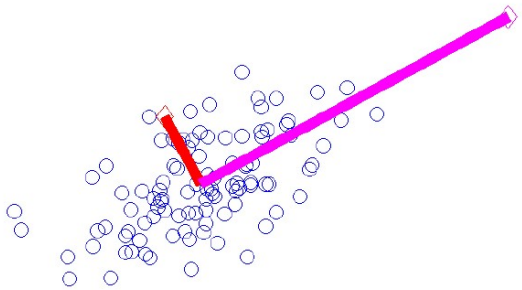


Star Trek IV 0.024 Star Trek II 0.023 Star Trek VI 0.023 Star Trek III 0.021 The Fifth Element 0.018	Dr. Strangelove 0.029 A Clockwork Orange 0.020 Delicatessen 0.018 Cinema Paradiso 0.018 Brazil 0.017717
The Rock 0.553 Eraser 0.232 Independence Day (ID4) 0.089 Mission: Impossible 0.077 Trainspotting 0.021	The Piano 0.288 The Remains of the Day 0.077 In the Name of the Father 0.067 Forrest Gump 0.052 Shadowlands 0.047

Uses of Unsupervised Learning...

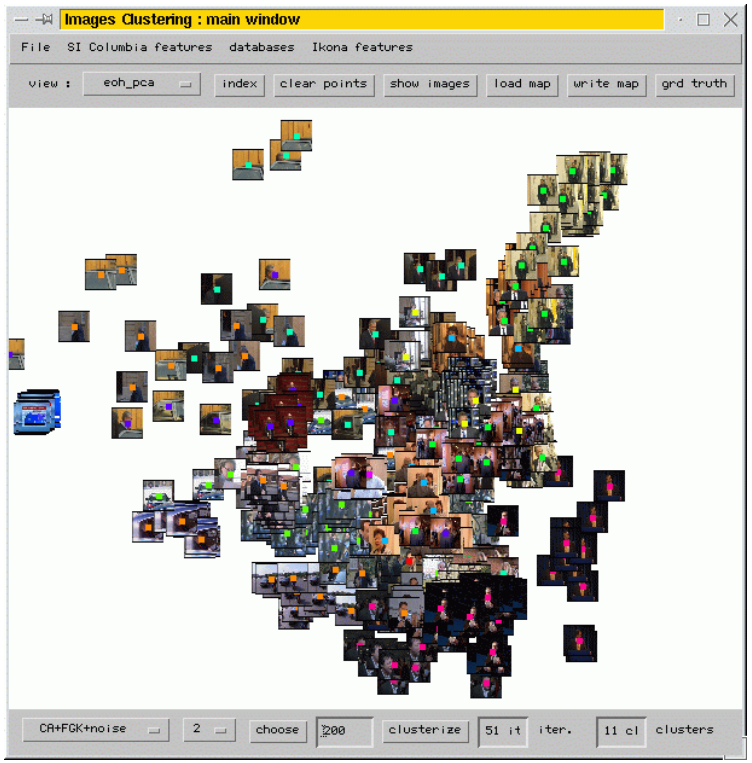
- find features or preprocess existing features for the subsequent pattern classification problem (supervised learning)

principal component analysis (PCA)



Example (eigenface)

Example: Content-based Image Retrieval



- visualize image locations and distances

Uses of Unsupervised Learning...

- find the least likely observations from a dataset (outlier detection)

Example (network intrusion detection)

Detect whether someone is trying to hack the network or doing anything else unusual on the network

File Logs Settings Help

Application	Protocol	Local Address	Remote Address	State	Creation Time	Rx [Bytes]	Type	Tx [Bytes]
SVCHOST.EXE	TCP	all:135	----------	Listening	27/Feb/2007 09:41:27	0	0	0
SVCHOST.EXE	TCP	all:3389	----------	Listening	27/Feb/2007 09:41:33	0	0	0
SVCHOST.EXE	TCP	all:1025	----------	Listening	27/Feb/2007 09:41:29	0	0	0
SYSTEM	TCP	192.168.123.128:139	----------	Listening	27/Feb/2007 09:41:30	0	0	0
PERSFW.EXE	TCP	all:44334	----------	Listening	27/Feb/2007 09:41:43	0	0	0
PERSFW.EXE	TCP	all:44334	localhost:1986	Connected In	27/Feb/2007 16:45:42	2931	<attack>	152516
SVCHOST.EXE	TCP	all:5000	----------	Listening	27/Feb/2007 09:42:15	0	0	0
LSASS.EXE	TCP	all:27155	----------	Listening	27/Feb/2007 09:42:17	0	0	0
PUTTY.EXE	TCP	all:1898	uststu1.ust.hk:22	Connected Out	27/Feb/2007 15:54:54	445204	<normal>	141398
CCAPP.EXE	TCP	localhost:1035	----------	Listening	27/Feb/2007 09:42:23	0	0	0
SSHCLIENT.EXE	TCP	all:1639	lcpu2.cse.ust.hk:22	Connected Out	27/Feb/2007 11:43:48	135198	<normal>	3764
ICQ.EXE	TCP	all:21470	----------	Listening	27/Feb/2007 14:41:02	0	0	0
ICQ.EXE	TCP	all:1759	64.12.24.205:5190	Connected Out	27/Feb/2007 14:41:04	59285	<normal>	12191
MSNMSGR.EXE	TCP	all:1722	by2msg1104003.ph...	Connected Out	27/Feb/2007 14:38:33	176880	<normal>	129012
PFWADMIN.EXE	TCP	all:1986	localhost:44334	Connected Out	27/Feb/2007 16:45:42	278215	<normal>	2931
LSASS.EXE	UDP	all:500	----------	Listening	27/Feb/2007 09:41:40	0	0	0
SYSTEM	UDP	192.168.123.128:138	----------	Listening	27/Feb/2007 09:41:30	205	0	40995
LSASS.EXE	UDP	localhost:1033	----------	Listening	27/Feb/2007 09:42:23	0	0	2
SYSTEM	UDP	192.168.123.128:137	----------	Listening	27/Feb/2007 09:41:30	1088	0	17982
PERSFW.EXE	UDP	all:44334	----------	Listening	27/Feb/2007 09:41:43	0	0	0
IEXPLORE.EXE	UDP	localhost:1600	----------	Listening	27/Feb/2007 11:17:52	1173	0	1173
IEXPLORE.EXE	UDP	localhost:1697	----------	Listening	27/Feb/2007 14:09:48	15	0	15

Outlier Detection

Example (database cleaning)

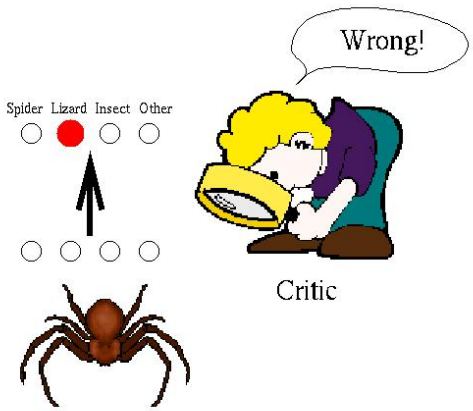
want to find out whether someone stored bogus information in a database (typos, etc.), mislabelled digits, ugly digits, bad photographs in an electronic album

Example (fraud detection)

credit cards, telephone bills, medical records

Learning Paradigm: Reinforcement Learning

Training examples as input-output pairs, with evaluative output only



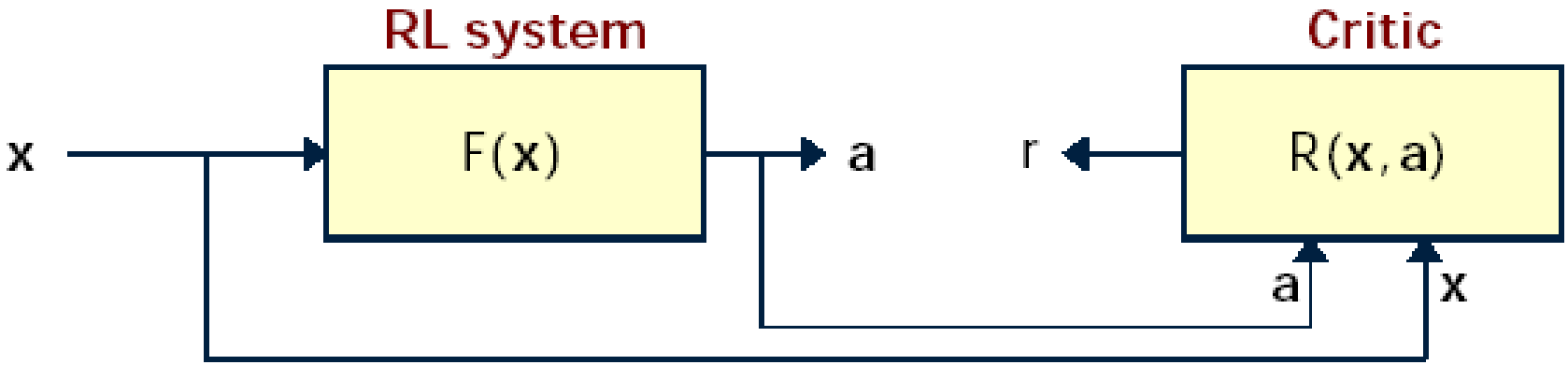
- try to increase the reinforcement it receives

Example

- graded examinations with only overall scores but no correct answers

Reinforcement Learning

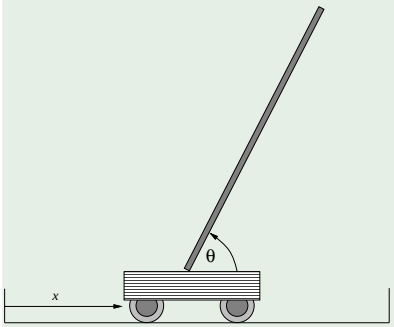
Learning from interacting with an environment to achieve a goal



Learning a mapping from states to actions to maximize long-term reward

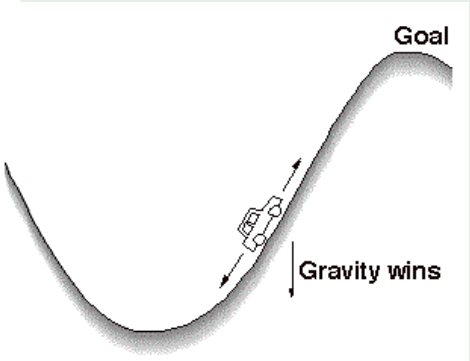
- demo

Example (Pole balancing)



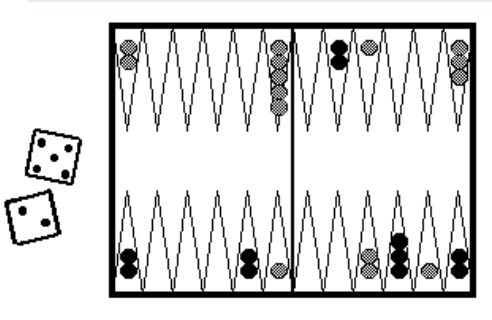
- goal: balance the pole as long as possible
- states: dynamic states of cart-pole system
- actions: push left, push right
- rewards: always 0 unless pole falls or cart hits end of track, in which case -1

Example (Mountain car)



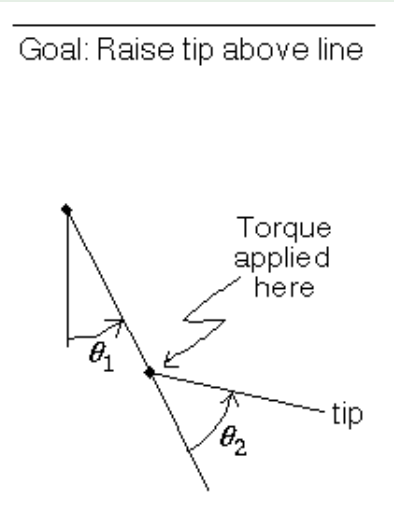
- goal: minimize time to the "goal"
- states: car's position and velocity
- actions: forward, reverse, none
- rewards: always -1 until car reaches the goal

Example (Backgammon)



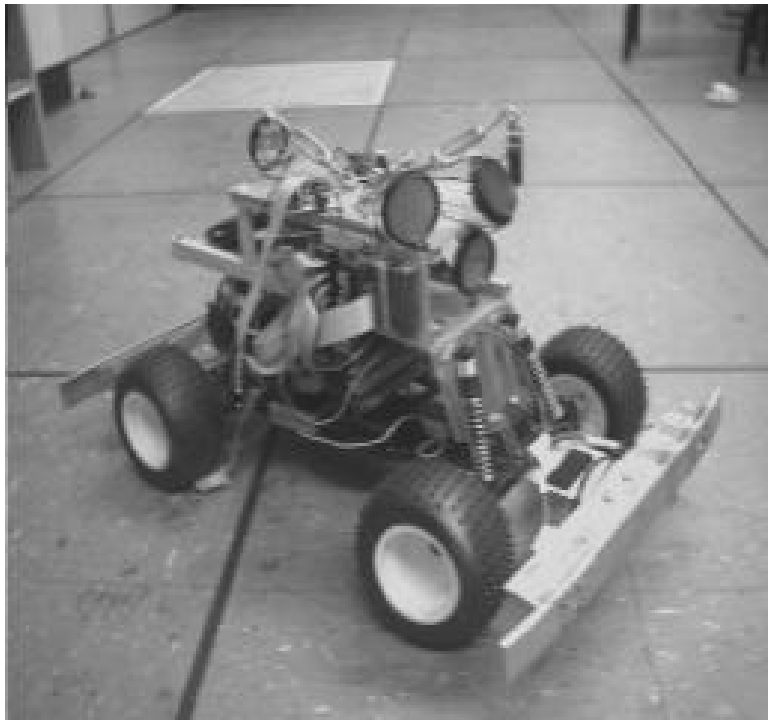
- goal: win
- states: configurations of the playing board
- actions: moves
- rewards: win: $+1$, lose: -1, else: 0

Example (Acrobat)



- goal: minimize time to "goal"
- state variables: 2 joint angles, 2 angular velocities
- rewards: -1 per time sweep

Example: Robotics

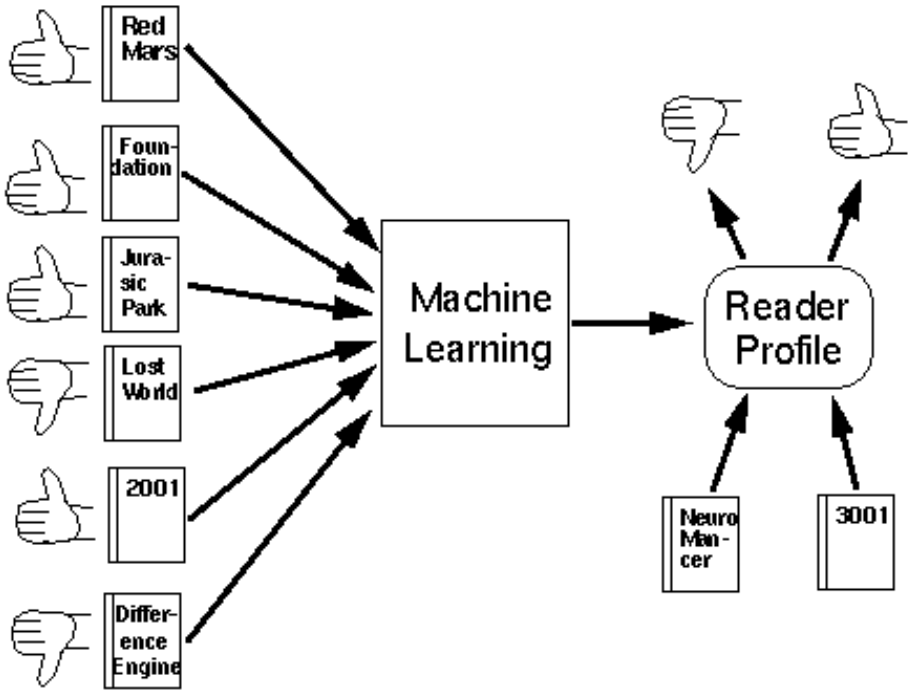


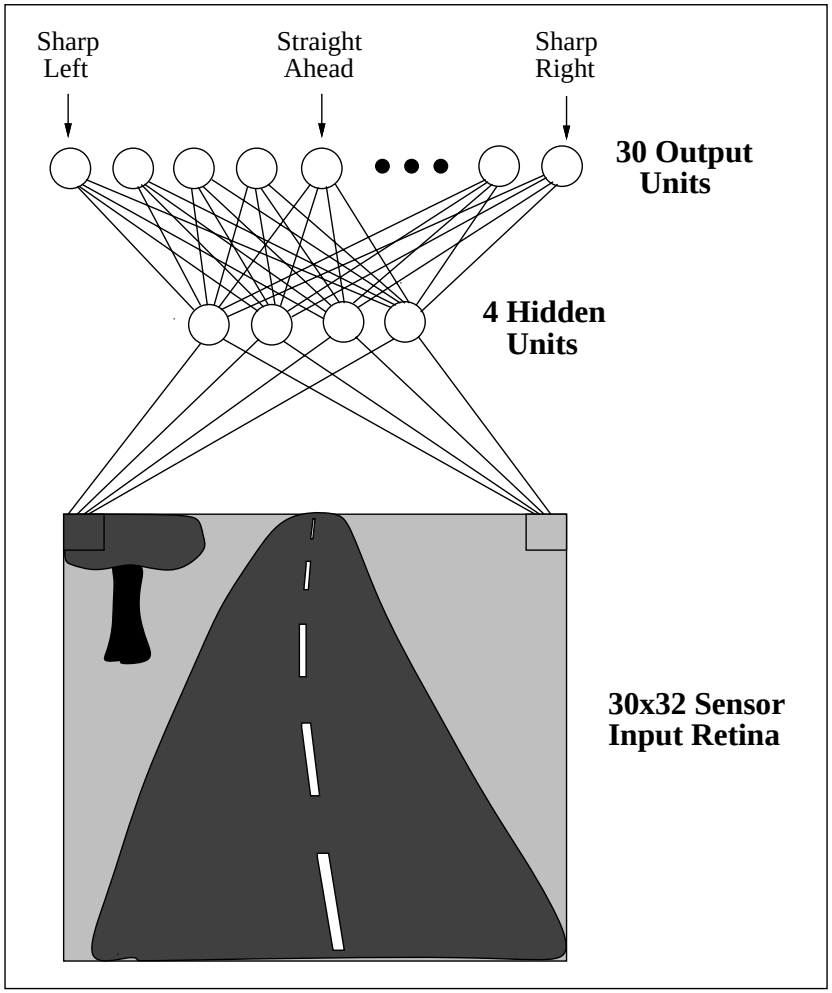
- input: sensory information and reinforcement signal
- output: avoid negative reward and try to have positive reward (obstacle avoidance, wall following, etc.)

Variants and Extensions

Multiclass classification

Binary classification





Multiclass Classification

Reducing multiclass classification to binary classification

1. one-vs-rest
   - binary classification: "C_i" vs "not C_i"
   - select the class that is most certain
2. one-vs-one
   - binary classification: C_i vs C_j
   - aggregate the results by voting

Variants and Extensions...

Multilabel classification

- an instance can have multiple labels

Example

- tags: building, city, and hkust
- Flickr (as of 2010): $>$ 20 millions unique tags

Example

- text categorization
- gene functions analysis in bioinformatics

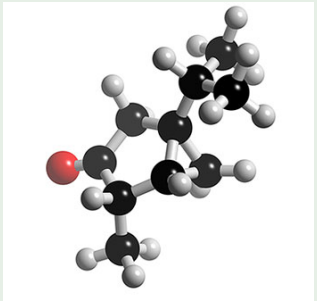
Variants and Extensions...

Traditional supervised learning problem:

- each pattern has a label
- not always the case

Example (drug activity prediction)

- predict whether a drug molecule can bind to the targets (enzymes or cell-surface receptors)



- each drug molecule can have multiple low-energy shapes or conformations
- according to the chemical theories, if there is at least one conformation binds to the target, the drug is called active
- chemical data cannot tell us the activity of each conformation

Multi-Instance Learning...

- each bag (molecule) consists of multiple instances (conformations)

- different bags may have different numbers of instances

MI assumption:

- a bag is positive when at least one of its instances is positive
- a bag is negative when all its instances are negative
- labels are associated with bags, not with the instances

Multi-Instance Learning...

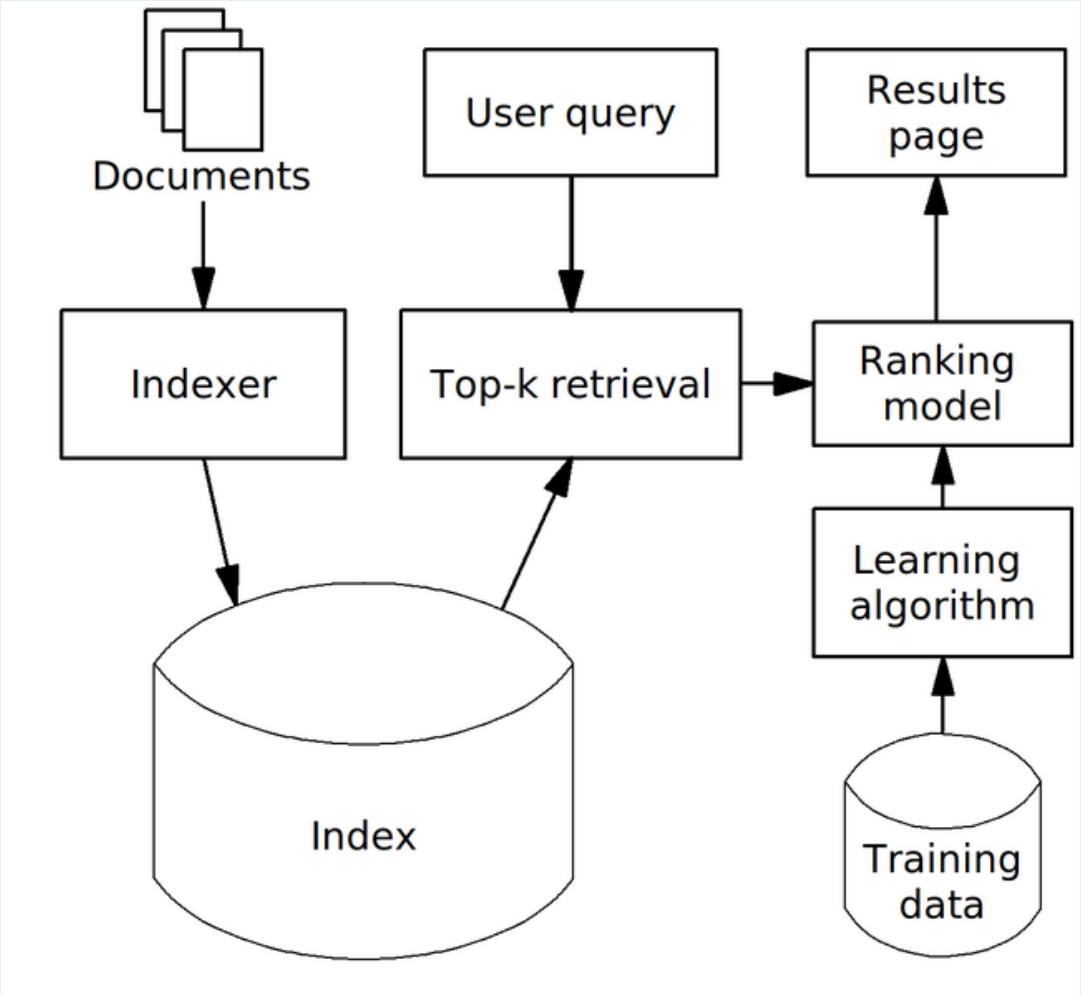
Example (Content-Based Image Retrieval (CBIR))

Task: Classify/retrieve images based on content

- each image (bag) is composed of several segments (instances)
- an image is labeled positive when at least one of its segments is positive; and negatively labeled otherwise

Ranking

Example



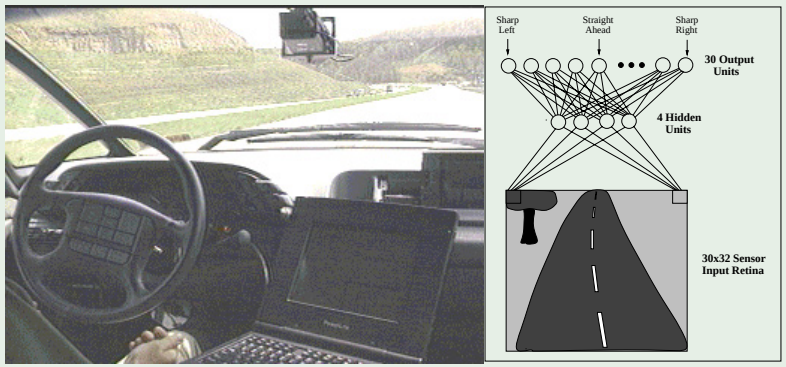
Variants and Extensions...

Multitask learning

Example (digit recognition)

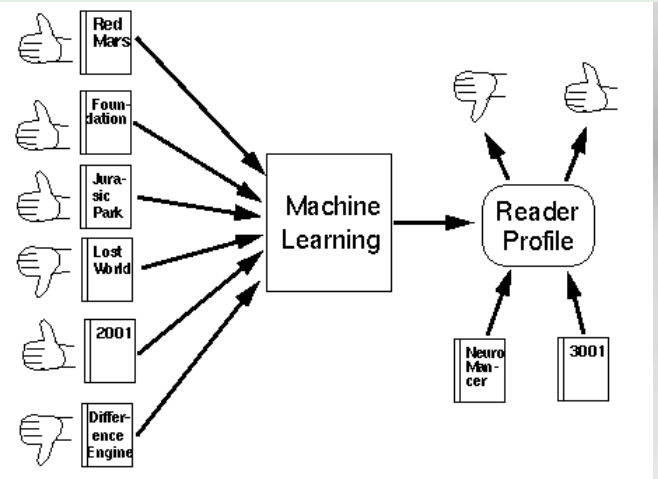
Example (autonomous vehicle)

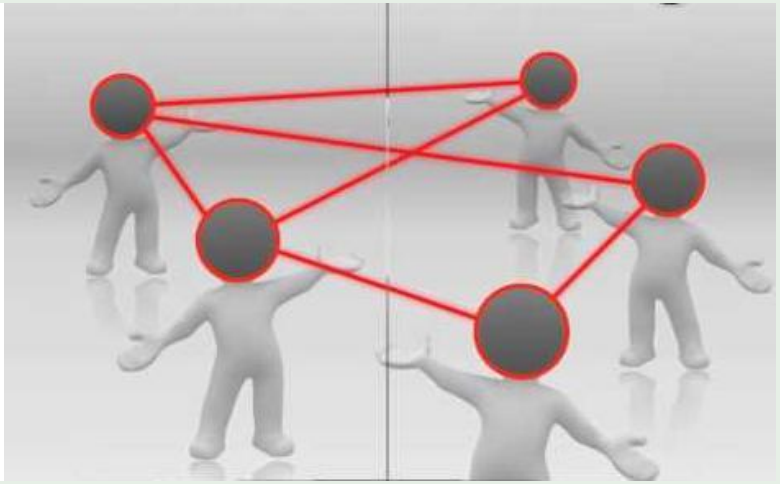
ALVINN (Autonomous Land Vehicle In a Neural Network)



Multitask Learning...

Example



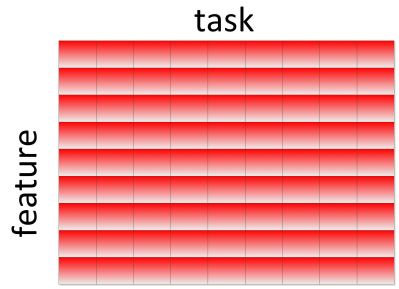


- a learning task can often be divided into several related subtasks
  - they share information!
- learning them together is more advantageous than learning each one independently, especially when there are insufficient training samples

Multitask Learning...

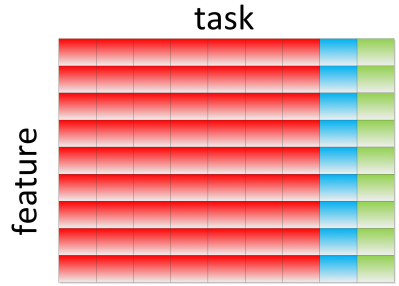
Learning with highly correlated tasks

- all the task weights form one cluster



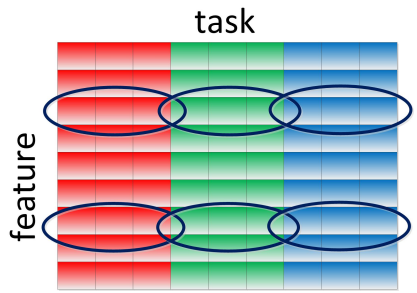
Learning with outlier tasks

- all the task weights form one cluster with a few outliers



Learning with clustered tasks

- all the task weights form several clusters



Variants and Extensions...

Transfer learning

- transfer learning in one context to another context that shared similar characteristics

Example

- depends on the proportion to which the learning task and the transfer task are similar

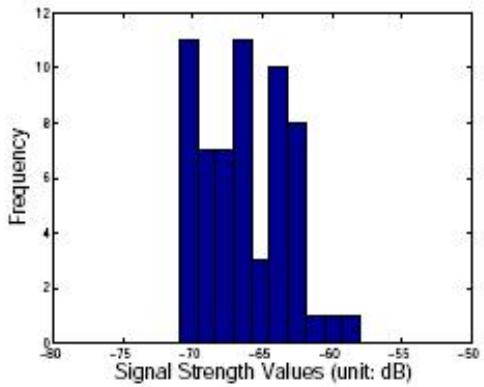
Example: Location Estimation

Locating users in a wireless environment

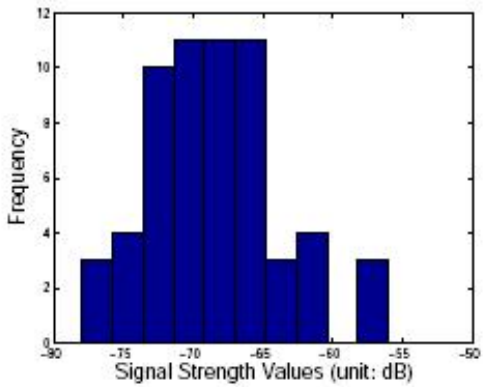
- a mobile device can be localized using signals received from various transmitters (access points)

In a complex indoor WiFi environment, the environment can be very dynamic

- caused by unpredictable movements of people, radio interference and signal propagation
- distribution of RSS values in different time periods may be significantly different



(a) Time Period 1



(b) Time Period 2

Multi-view learning

- instances may have disjoint multiple representations

Example (document classification)

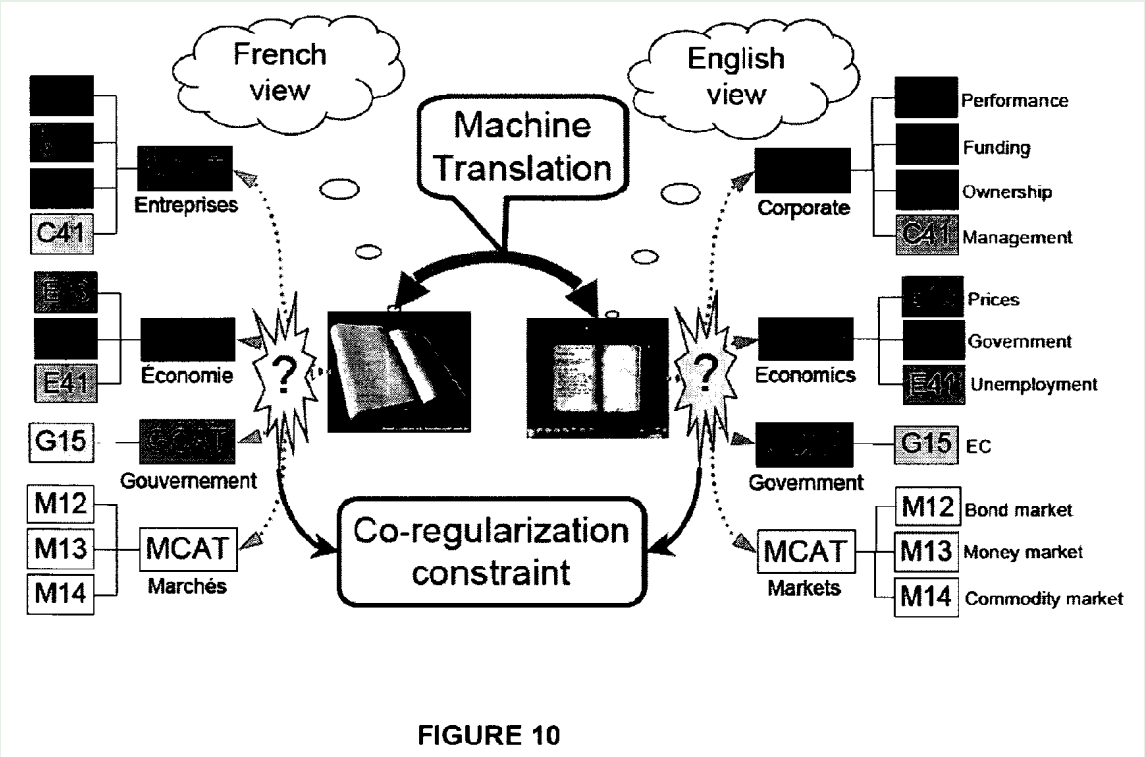
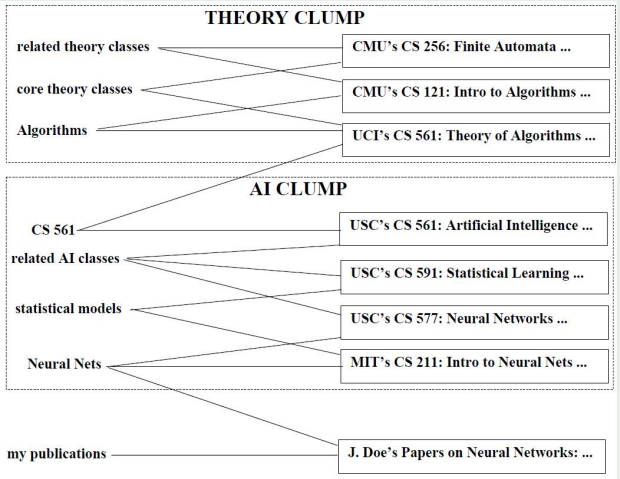


FIGURE 10

- both views need to agree

Multi-View Learning...

Example (webpage classification)



views

1. words in the hyperlinks pointing to the webpages
2. words in the webpages

Variants and Extensions...

Supervised learning

- learn from solved examples in a book
- In-class closed book exam

Active learning

- Only unsolved problems
- Can ask an expert a few questions

Example (Is this a chinese?)

chinese

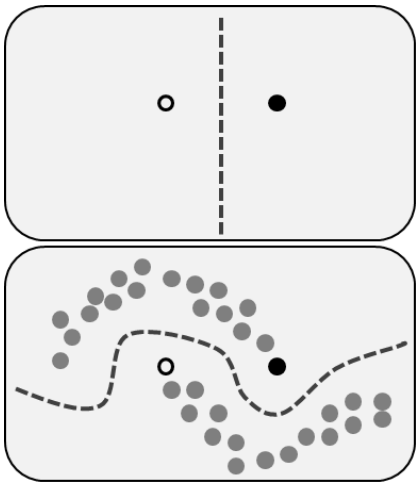
non-chinese

chinese?

Variants and Extensions...

Semi-supervised learning

- uses both labeled and unlabeled data for training
- typically a small amount of labeled data with a large amount of unlabeled data
  - labeled data: expensive and tedious human effort
  - unlabeled data: inexpensive

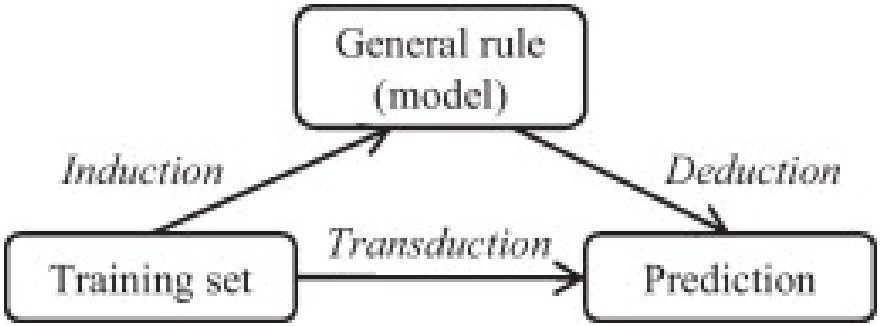


smooth on data manifold

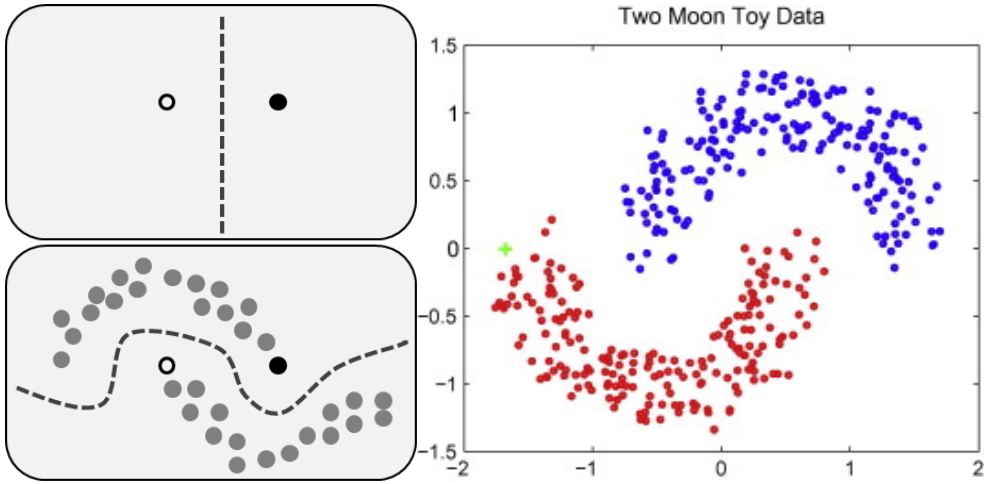
- book example + back of the book questions

Variants and Extensions...

Transductive learning



- produce label only for the available unavailable data



- Take home exam

Variants and Extensions...

Discriminative clustering

- adopt a cost function originally used for classification as a clustering criterion

Example (maximum margin clustering)

- maximum the margin between clusters