Processing of Information

Information processing :

- Human is able to do both logical and intuitive
- Computer is able to do the former, and far behind for latter

To do human like intuitive information processing, a computer should have:

- **Openness** (ability to cope up new– *feature of big data*)
- **Robustness** (stability, intolerability with distorted, incomplete, imprecise data -- *feature of big data*)
- **Real-time processing**

Fuzzy Logic is one of the approaches

Required characteristics to mimic human

In order to possess human like behavior, the system model should have at least the following characteristics

- Capabilities of handling **uncertainties** in different stages arising out of measurement errors and/or deficiency in input information
- **Flexibility** (adaptivity) to incorporate changes in its functionalities depending on the data characteristics
- Capability for **parallel and distributed** computing with close interaction among the processing elements
- **Robustness/ ruggedness** with respect to distortion, noise/ failure of components
- In addition, it is desirable that the computational paradigm be capable of **AR**, **efficient searching** and **handling non-numeric inputs**.

This has become particularly useful in addressing the current need of internet technology for mining voluminous, both in size and dimension, heterogeneous datasets.

Fuzzy Logic

Why is it needed: Situation 1

Suppose X is driving a car and X sees red light. X has to stop.

Decide when to press brake and how strongly

1. **Precise:** Find distance of red light and car (some accessories is needed) and current speed of car

   Apply rule: ***if the car is at a distance of d ft and moving at a speed of s ft/sec then press the brake with p poundal for t secs right now***

   Involves cost, labor, difficult

   Q. *Is it really needed? Such precision?* *NO*

   Reality: The car should stop before red light not hitting any car standing ahead of it

So, what do we do? : Situation 1

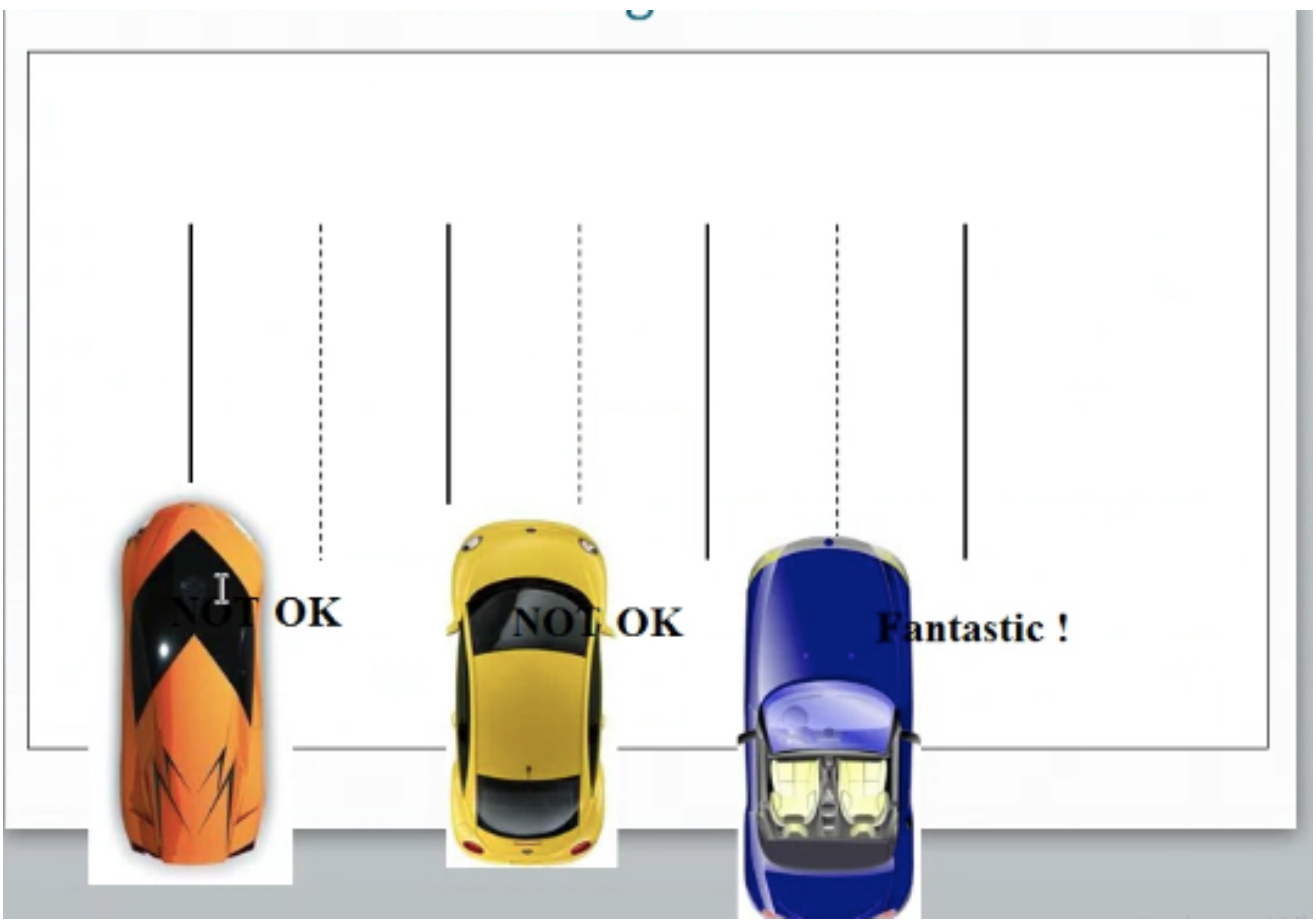
An approximate idea about the distance of the car and speed of the car is sufficient

2. The modified rule (Imprecise)

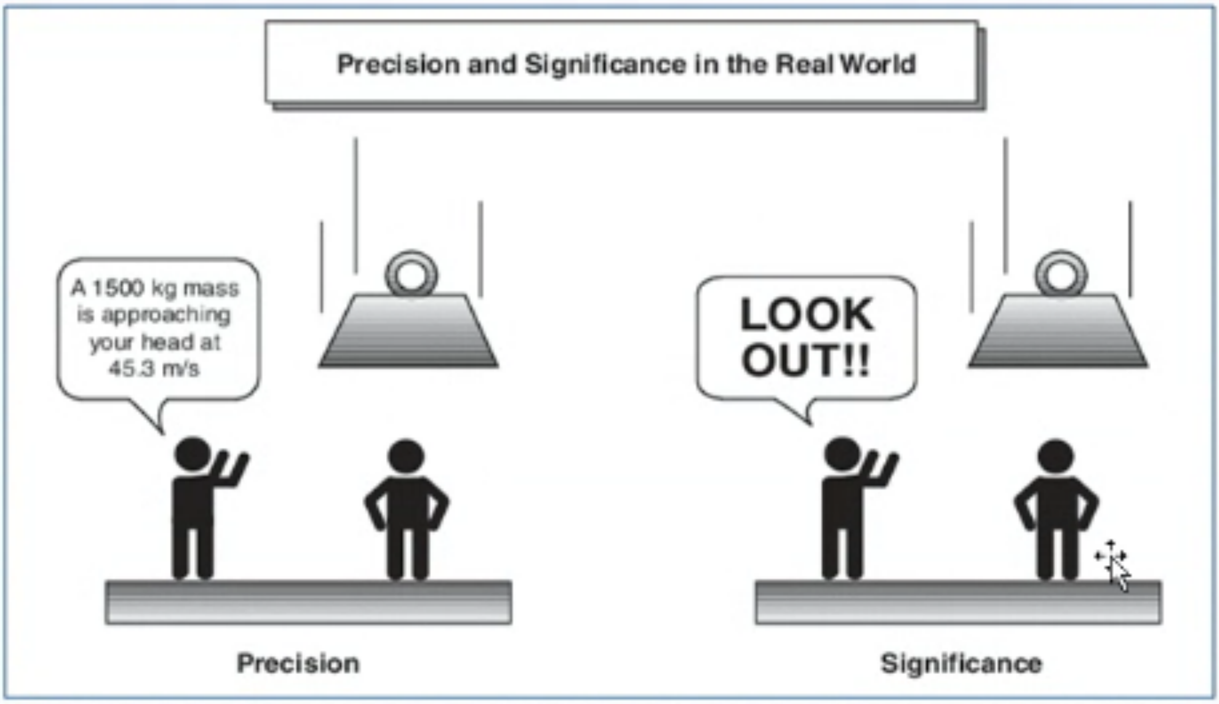
If the car is moving very fast and the Red Light is close then press the brake pretty hard

The rule is imprecise, generates approximate solution, less expensive, fast

FL Models vagueness, makes approximate decisions


NOT OK
NOT OK
Fantastic !

Situation 3

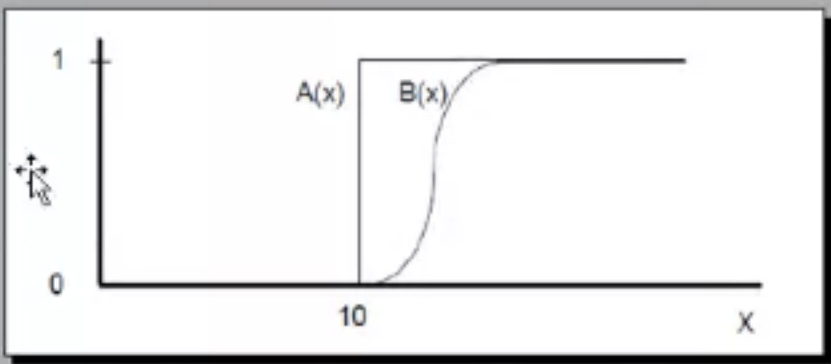

Precision and Significance in the Real World
A 1500 kg mass is approaching your head at 45.3 m/s
LOOK OUT!!
Precision
Significance


Buying bananas: yellow
Is it yellow?
Yes, μ=0.2
Is it yellow?
Yes, μ=0.3
Is it yellow?
Yes, μ=0.4
Is it yellow?
Yes, μ=0.5
Is it yellow?
Yes, μ=0.6
Is it yellow?
Yes, μ=0.7
Is it yellow?
Yes, μ=0.8
All are yellow--generalized set (fuzzy set)

Situation 5: Answer script Evaluation: We are "Soft" ! "Fuzzy" !

Fuzzy Sets

$A = \{ x \in \mathbf{X} \mid x > 10\}$	Boolean Set	$A: \mathbf{X} \to \{0, 1\}$
$B = \{ x \in \mathbf{X} \mid x >> 10\}$	Fuzzy Set	$B: \mathbf{X} \to [0, 1]$



Characteristic function of sets A(x) and B(x)

So, what do we realize?
Our world is imprecise !!!

- **Mathematical and Statistical techniques are often unsatisfactory.**
  - Experts make decisions with imprecise data in an uncertain world.
  - They work with knowledge that is rarely defined mathematically or algorithmically but use vague terminology with words.
- Fuzzy logic is able to use vagueness to obtain an answer. By considering different factors simultaneously, we get a better answer, one that is more suited to the situation.

Conventional Set and Fuzzy Set

Prof. Lotfi Zadeh, 1965

- A new way to represent vagueness in everyday life; generalization of conventional set (CS) theory
- CS contains objects that satisfy precise properties required for its membership

  FS -> imprecisely defined properties to varying degrees
- CS -> clearly defined boundaries

  FS -> meaning and significance are well understood but boundaries are not well defined

CS: set of integers >2 and < 10

FS: strong fever, high temperature, rich man, tall people ...

Fuzzy Set: Notation

A fuzzy set A of the Universe X is defined as a collection of ordered pairs:

A = { (μ_A(x), x), ∀x ∈ X}, where $\mu_A(x)$ ($0 \le \mu_A(x) \le 1$} gives the degree of belonging of the element x to the set A or, degree of possession of an imprecise property represented by A.

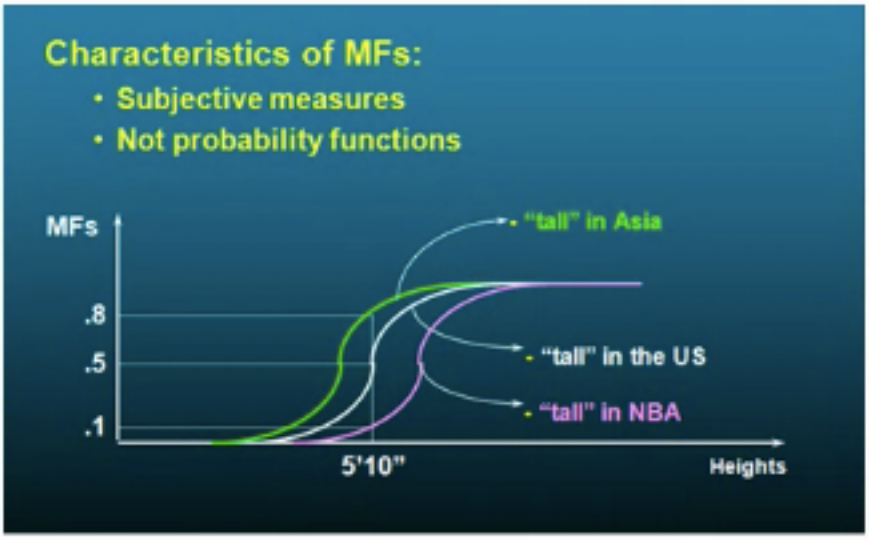
When the universe of discourse, X, is discrete and finite, then:

$$\underset{\sim}{A} = \left\{ \frac{\mu_{\underset{\sim}{A}}(x_1)}{x_1} + \frac{\mu_{\underset{\sim}{A}}(x_2)}{x_2} + \cdots \right\} = \left\{ \sum_i \frac{\mu_{\underset{\sim}{A}}(x_i)}{x_i} \right\}$$

When X, is continuous and infinite $\underset{\sim}{A} = \left\{ \int \frac{\mu_{\underset{\sim}{A}}(x)}{x} \right\}$

$\mu_A = 1$? $\mu_A = 0$?

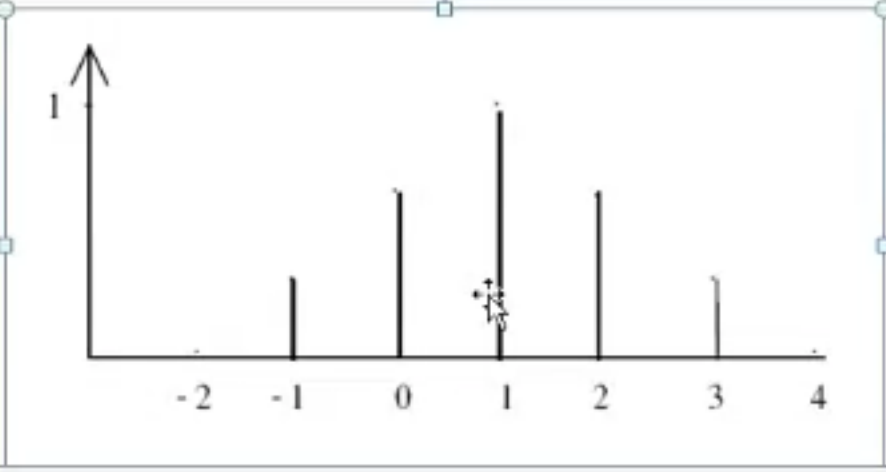
Fuzzy Set: Membership Function



Fuzzy Sets - Example

"numbers close to 1"

$$A = 0.0/-2 + 0.3/-1 + 0.6/0 + 1.0/1 + 0.6/2 + 0.3/3 + 0.0/4$$



Fuzzy Set: Membership Function

Grades of membership for fuzzy sets are subjectively assigned on the basis of context; no deterministic procedure to establish the validity; e.g., 5'4" – tall ??

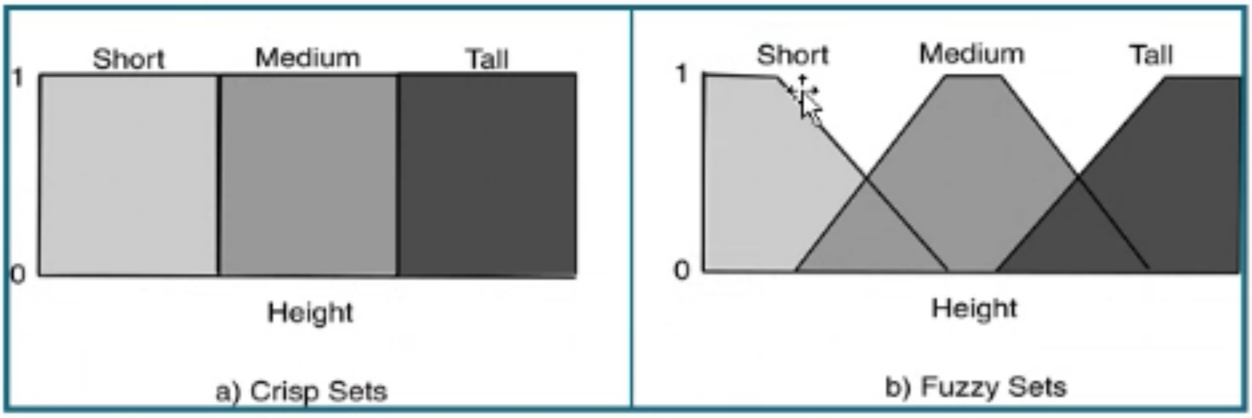
Flexibility of fuzzy set theory is associated with **the elasticity property** of the concept of its **membership function**

Higher the membership value lesser is the stretch needed to fit an object in that set.

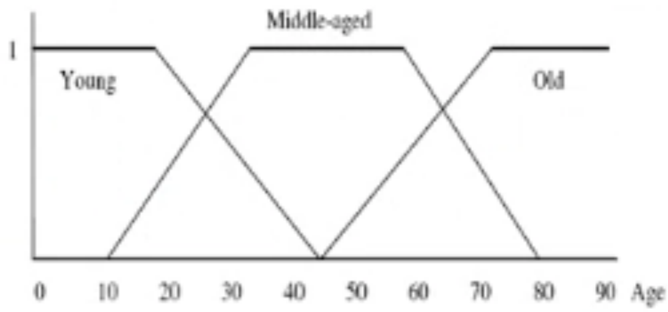
Imprecision

Words are used to capture imprecise notions, loose concepts or perceptions.

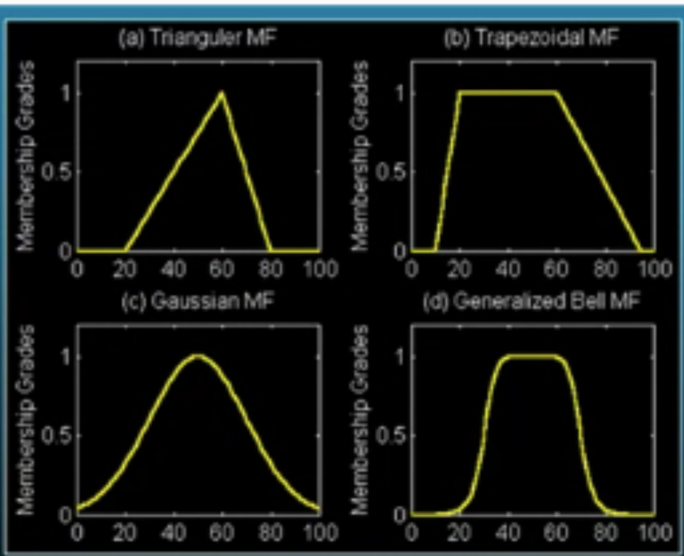
Fuzzy Sets



overlapping of the sets reflecting the real world more accurately than if we were using a traditional approach



Types of membership function



disp_mf.m

Types of membership function

- A *triangular* membership function is specified by three parameters {a, b, c}:

- **Triangle**(x; a, b, c) = 0 if $x \le a$;
  = $(x-a)/(b-a)$ if $a \le x \le b$;
  = $(x-b)/(c-b)$ if $b \le x \le c$;
  = 0 if $c \le x$.

Types of membership function

- **Trapezoid**(x; a, b, c, d) = 0 if $x \le a$;
  = $(x-a)/(b-a)$ if $a \le x \le b$;
  = 1 if $b \le x \le c$;
  = $(d-x)/(d-c)$ if $c \le x \le d$;
  = 0, if $d \le x$.

- **Sigmoid**(x; a, c) = $1/(1 + \exp[-a(x-c)])$ where *a* controls slope at the crossover point $x = c$.

Manipulation of Fuzzy sets: Fuzzy Logic

- FL: well defined reasoning system, elements being manipulated are fuzzy, rules of logic are well defined.

  It is a logic system which deals vague concept

- The task of translating human expressions in FL system is simple, since humans tend to communicate ideas and quantifications by using verbal rather than numeric descriptions (use linguistic variables e.g., *big, old, fast......)*

- To translate these values in FL system a *membership function is defined*. *fuzzy modifiers/ fuzzy quantifiers* are used to modify the function (e.g., very old, not very old, very very old, not at all old, too old, ...)

- FL: well defined reasoning system, elements being manipulated are fuzzy, rules of logic are well defined.

  It is a logic system which deals vague concept

- The task of translating human expressions in FL system is simple, since humans tend to communicate ideas and quantifications by using verbal rather than numeric descriptions (use linguistic variables e.g., *big, old, fast......)*

- To translate these values in FL system a *membership function is defined*. *fuzzy modifiers/ fuzzy quantifiers* are used to modify the function (e.g., very old, not very old, very very old, not at all old, too old, ...)
- Quantifiers: represent approximate quantification of FL

Fuzzy Logic

- A **fuzzy if-then rule** is of the form "If x is A then y is B" where A and B are linguistic values defined by fuzzy sets on universes of discourse *X* and *Y*, respectively.

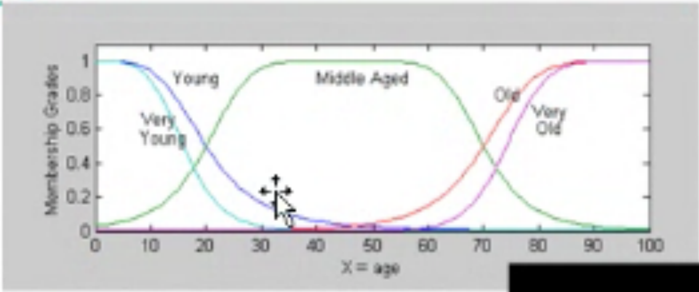
- "x is A" is called *antecedent* and "y is B" is called *consequent*.

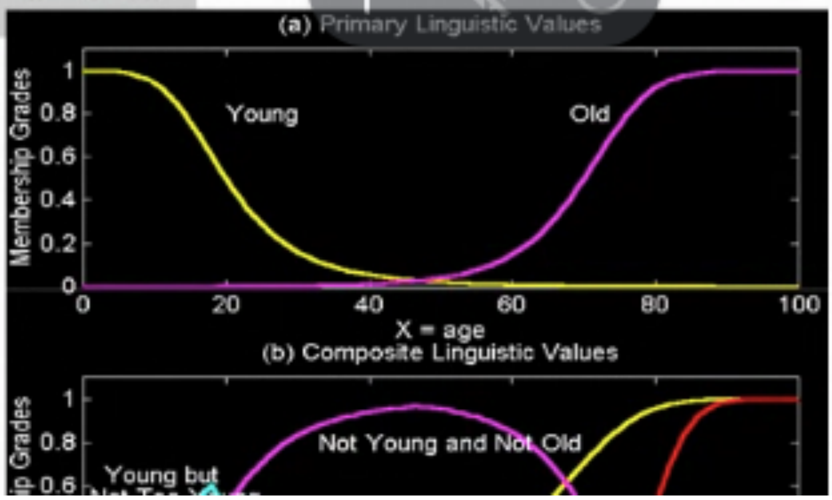
Example

- If pressure is high, then volume is small.
- If the road is slippery, then driving is dangerous.
- If the fruit is ripe, then it is soft.

where **high, small, ripe** are fuzzy sets.

Linguistic terms and hedges: Quantifiers, Modifiers





Fuzzy Logic then . . .

- is particularly good at handling uncertainty, vagueness and imprecision.
- especially useful where a problem can be described linguistically (using words).
- Devised to model human reasoning processes comprising

  vague predicate: large, beautiful, small

  partial truth: not very true, more or less false,

  linguistic quantifiers: most, almost, a few

Characteristics of Fuzzy set

- **Normality and height**: A FS A is normal if its μ attains 1
  supremum $\mu_A(x) = 1$
- **Support**(A) is set of all points x in X such that
  $\{(x|\ \mu_A(x) > 0\ \}$
- **Core**(A) is set of all points x in X such that
  $\{(x|\ \mu_A(x) = 1\ \}$
- **Crossover point** of a fuzzy set A is a point x in X such that
  $\{(x|\ \mu_A(x) = 0.5\ \}$
- **α-cut** of a fuzzy set A is set of all points x in X such that
  $\{(x|\ \mu_A(x) \geq \alpha\ \}$

Characteristics of Fuzzy set

- **Band width**: $|x2-x1|$ where x2 and x1 are crossover points.
- A is **Symmetric**: $\mu_A(c+x) = \mu_A(c-x)$ for all $x \in X$.
- **Cardinality**: scalar cardinality(A) = $\sum\mu_A(x)$
- **Concentration** and **Dilation**: If A is a linguistic value then $CON(A) = A^p$, $(p>1)$ and $DIL(A) = A^r$ (r is in 0 to 1) .

  Using these operations we can generate linguistic hedges e.g.,

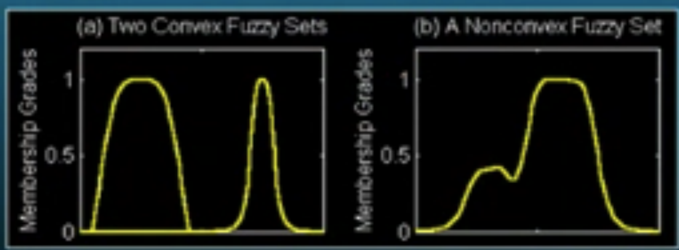
  - more or less old = DIL(old);
  - extremely old = CON(CON(CON(old))).

Convexity

A fuzzy set *A* is convex if for any λ in [0, 1],

$$\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \min(\mu_A(x_1), \mu_A(x_2))$$

Alternatively, *A* is convex is all its α-cuts are convex.

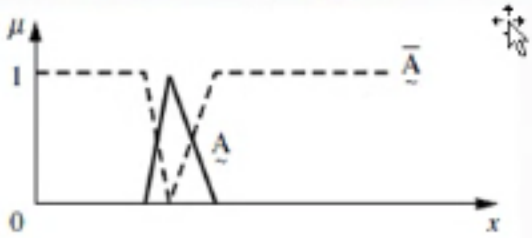


convexmf.m

Operations on Fuzzy Set: Complement

- **Fuzzy Complement:** The complement of A denoted by Ā or NOT :

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x)$$



Sugeno's Complement: c-s(a) = (1-a)/(1+sa) ; s>-1

Yager's Complement: c-w(a) = $(1-a^{w})^{1/w}$; w is +ve

Operations on Fuzzy Set: Intersection

- **Fuzzy Intersection:** The intersection of two fuzzy sets A and B , $A \cap B$ or *A AND B* is fuzzy set C whose membership function is specified by the

$$\mu_C(x) = \mu_{A\cap B} = \min(\mu_A(x), \mu_B(x)) = \mu_A(x) \wedge \mu_B(x)$$

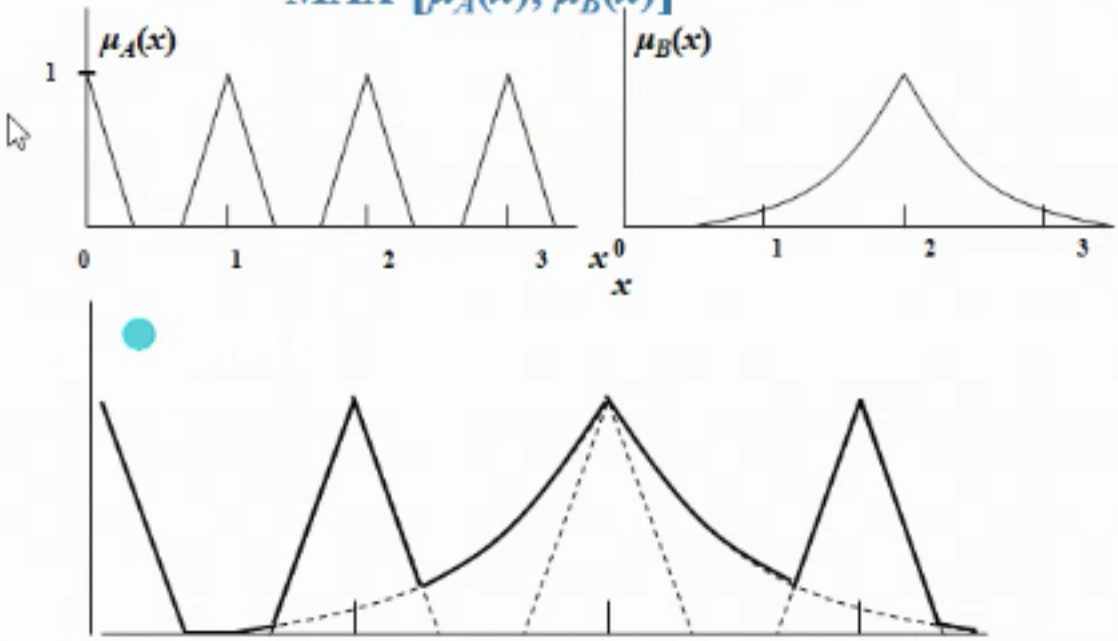
In general,

- $\mu_{A\cap B} = T(\mu_A(x), \mu_B(x))$ where T is T-norm operator. Some T-Norm operators:
  - Minimum: $\min(a,b)=a \wedge b$
  - Algebraic product: ab
  - Bounded product: $0 \vee (a+b-1)$

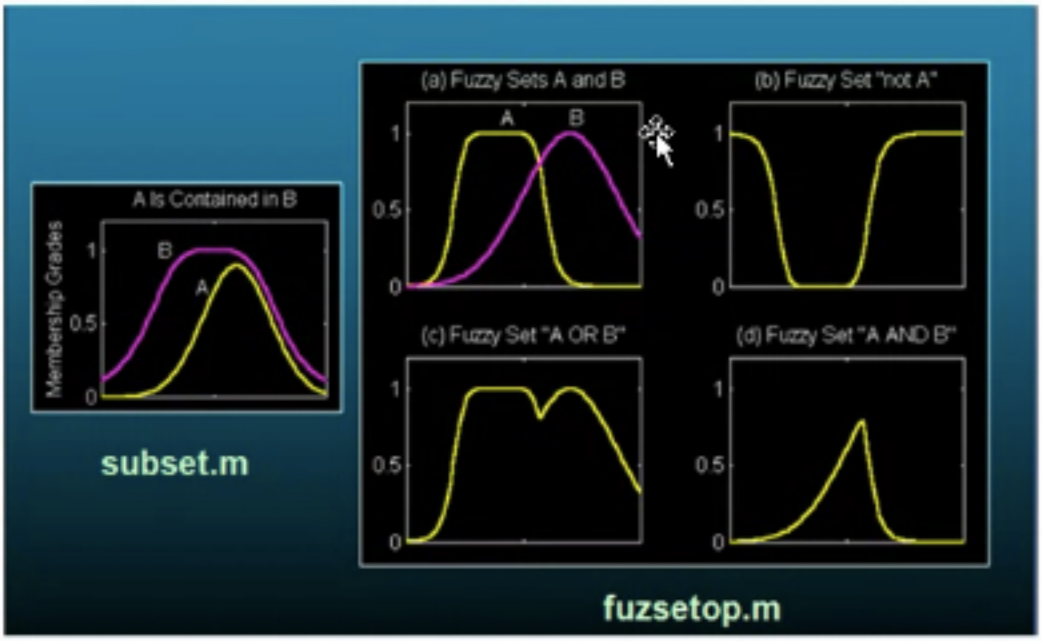
Example: Union

A OR *B* = *A*+*B* ={ *x* | (*x* is *near* an integer) OR (*x* is *close* to 2)}

$$= \text{MAX}\left[\mu_A(x), \mu_B(x)\right]$$



Operations on Fuzzy Set



Fuzzy Approach

Can you fill in?

Before FL era	In FL era
Math World	?
Hard data	?
Quantitative methods	?
Bivalent reasoning	?

Applications:
Consumer Electronics

From Siemens -Fuzzy washing machine

- **Automatic water level adjustment** - FL detects the type and amount of laundry in the tub allows as much water to enter the machine as is really needed for the loaded amount

  Less water will heat up quicker -> less energy consumption
- **Foam Detection** – Too much foam is compensated by an additional rinse cycle
- **Imbalance compensation** – In the event of imbalance, calculate the maximum possible speed, sets this speed and starts spinning

From Panasonic – Fuzzy washing machine

Matsushita – Fuzzy vacuum cleaner; Hitachi – Rice cooker

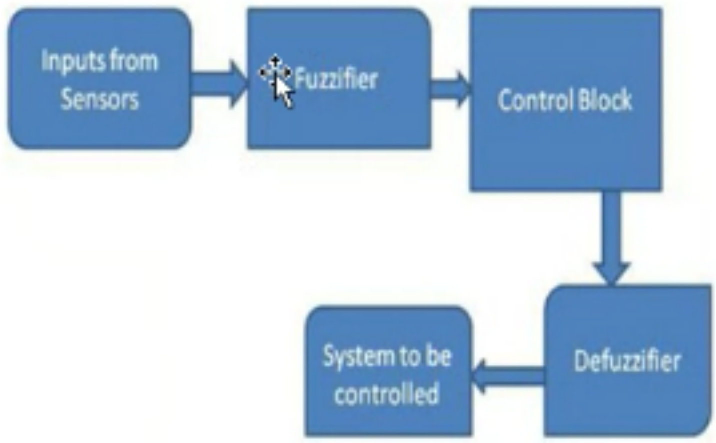
Sharp – Refrigerator; Canon, Minolta – camera- focus, exposure, zoom

Hitachi – Sendai subway system

Fuzzy Logic in Appliances: an example

Fuzzy Control System

A fuzzy control system consists of the following components:



A Fuzzy Logic Control System

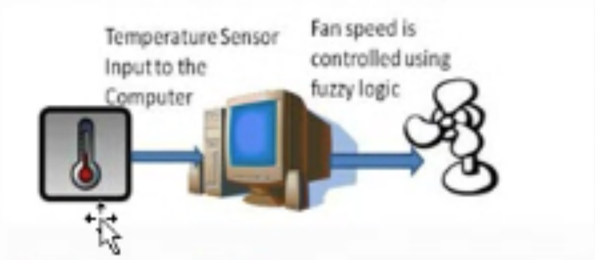
Fuzzy Control

- **Fuzzifier** transforms input (numeric) into linguistic
- **Controller** performs fuzzy logic operation of assigning outputs based on linguistic info; does AR similar to human;

  it consists of KB and IE; **KB consists of membership functions and fuzzy rules**, which are obtained from the knowledge of the system operations depending on environment
- **Defuzzifier** converts fuzzy output to the required output to control the system

Example: to control speed of the fan based on room temperature

- too hot: full speed
- bit hot: moderate
- too cold: lowest



- Fuzzifier: assigns linguistic variable
- If measured value is between: 30C to 40C- room is quite hot, 22C – 28C – moderate, 10C – 20C – cold
- Now, functioning of KB – it contains info. of these member functions as well as the rule base e.g.,

Clustering

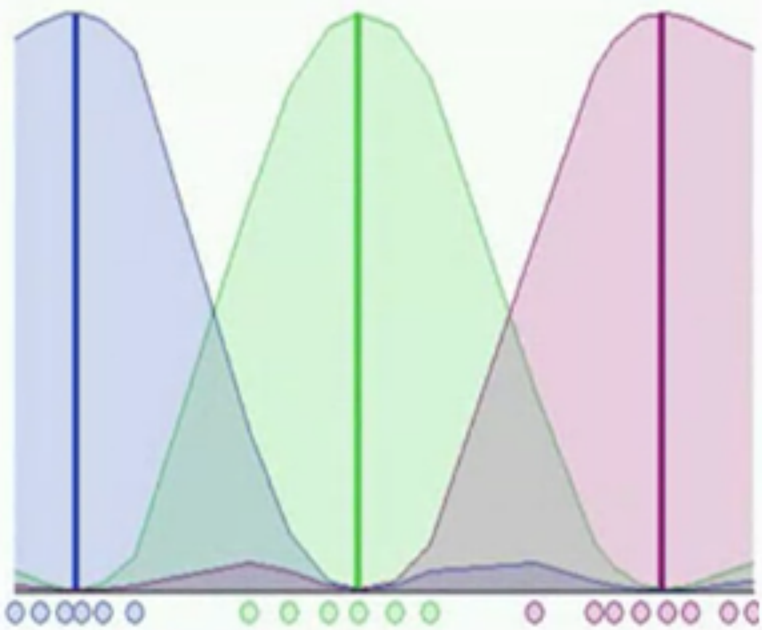
Classification

Fuzzy version

Fuzzy C-Means Clustering

Fixed number of clusters.

One centroid per cluster.



Clusters are fuzzy sets.

Membership degree of a point can be any number between 0 and 1.

Sum of all degrees for a point must add up to 1.

Notation

- U_{ij} : the degree of membership of x_i in the cluster j
- x_i : the ith pattern of d-dimensional measured data
- c_j : the d-dimensional center of the cluster

Fuzzy C-Means Clustering

C-Means	$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \left(\sum_{k,\ x_k \in G_i} d(\mathbf{x}_k - \mathbf{c}_i) \right)$
Fuzzy C-Means (FCM)	summing over all data points fuzziness exponent $J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j}^{n} u_{ij}^{m} d_{ij}^{2}$ membership degree

Fuzzy C-Means Clustering

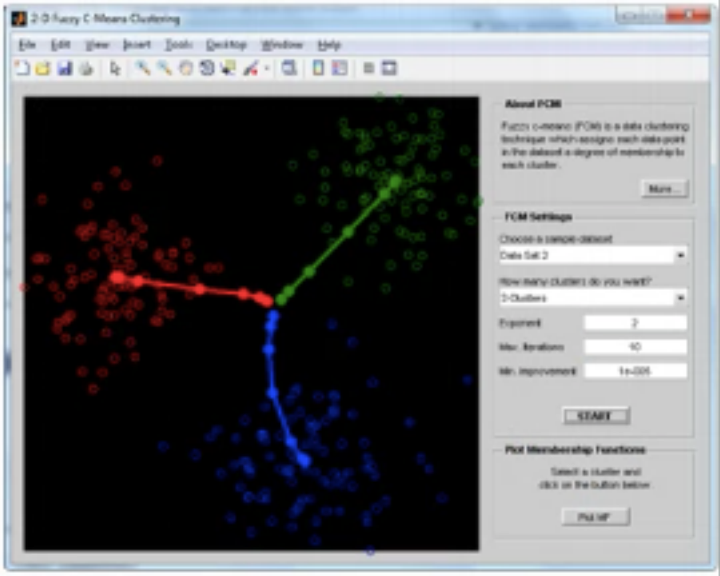
pick c centroids at random

assign membership degrees according to:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}}$$

move each centroid to the following position:

$$\mathbf{c}_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} \mathbf{x}_j}{\sum_{j=1}^{n} u_{ij}^{m}}$$



Algorithm

$W = \{x_1, x_2, \ldots, x_n\}$: A set of n labelled patterns

u_{ij} for all $i = 1,2, \ldots C$ and $j = 1,2, \ldots n$

BEGIN

Input y, a unknown sample

Set $K, 1 \le K \le n$

Initialize $i = 1$

DO UNTIL (K - nearest neighbors found)

Compute distance from y to x_i

Find K - nearest neighbors

END DO UNTIL

For $i = 1$ to C

Compute the membership of y to i - th class

$$u_i(y) = \frac{\sum_{j=1}^{K} u_{ij}\left(1/\|y - x_j\|^{2/(m-1)}\right)}{\sum_{j=1}^{K}\left(1/\|y - x_j\|^{2/(m-1)}\right)} \qquad (m > 1)$$

END FOR

Classify y to the class with the maximum $u_i(y)$

END

Toolboxes and Libraries for FL

Fuzzy Logic Toolbox for MATLAB:
http://www.mathworks.com/products/fuzzylogic/index.html

Fuzzy Logic package for Java (jFuzzyLogic)
http://jfuzzylogic.sourceforge.net/html/index.html

Fuzzy Logic libraries for C++ (JFuzzyQt)

http://sourceforge.net/projects/jfuzzyqt/

Application Areas

- Agricultural Machinery & Produce
- Bio-inspired Systems
- Condition Monitoring
- Data Mining
- Decision Support
- Fault Diagnosis
- Industrial Electronics
- Intelligent Information Retrieval
- Manufacturing Systems
- Multi-objective Optimisation
- Power and Energy
- Process Optimisation
- Robotics
- System Identification & Modelling
- Telecommunications
- Virtual Reality
- Autonomous Reasoning
- Biomedical Engineering
- Consumer Electronics
- Data Visualisation
- Engineering Design Optimisation
- Human-Machine Interface
- Intelligent Agents
- Internet Tools
- Motion Control and Power Electronics
- Nano and Micro-systems
- Process and System Control
- Reactive Distributed AI
- Signal or Image Processing
- Systems Integration
- Time Series Prediction
- Vision or Pattern Recognition

Fuzzy ≠ Probability

- Example
  - #1 Probability of having poison is 0.8
  - #2 Membership of having poison is 0.8
  - **Which one will you choose to survive?**

Overview

- What is Learning?
- Machine Learning
- Types of Learning

  Supervised, Unsupervised, Semi-supervised, Reinforcement
- Issues
- Applications

What is Learning?

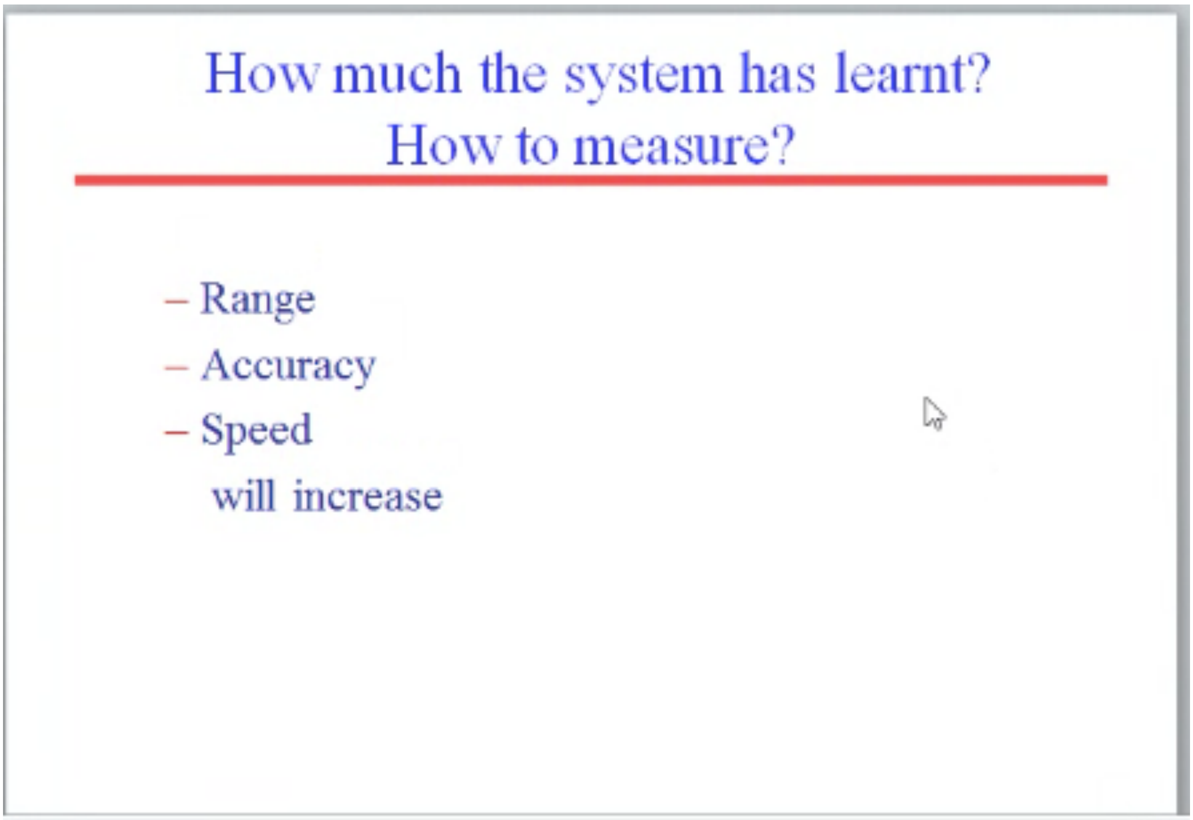
"The acquisition of knowledge or skills in (something) through study, experience, or being taught."

- One of the distinctive attributes of intelligent behaviour.
- Does not necessarily involve consciousness
- Finds statistical regularities or other relevant patterns in the data.

Defining a Learning Problem

- A program learns from experience *E* with respect to task *T* and performance measure *P*, if its performance at task T, as measured by P, improves with experience E.

- Example:
  - Task: Solving a sum
  - Performance: % of correct answers
  - Experience: Practice examples, exercises


How much the system has learnt?
How to measure?
– Range
– Accuracy
– Speed
will increase

Machine Learning…….

- Machine learning is about designing algorithms that allow a computer to learn from experience and data.

Machine learning is used in cases where:

- There is an intuition that a certain rule exists
- But, we do not know it or cannot express it mathematically

So, we learn the rules from data

Example

For the academic year 2017, the student selection committee is given the task to select students for admission to their institution.

Students filled up application forms and submitted them to the Institution.

Example

The selection committee selected some candidates based on their marks in Mathematics, Physics and Chemistry; and their financial status.

- Similar procedure was followed for the academic years 2018 & 2019.

- Then an analysis was made on the performance of the students admitted in 2017, 2018 & 2019.

Example

Now, in 2020, the selection committee again got the task and thought of an automated process.

- BUT HOW?

They did not have a specific rule to define the selection criteria

Example

MACHINE LEARNING
IS THE ANSWER

The committee gave the marks of previous year students' application as well as the data whether they were selected or not to the computer, trained it and tested it-
LEARNING

Example

Given a student **x** defined by his name, address, marks in the subjects and his financial status the machine learns the

target function

f(x)=y

where *y* is the output whether the candidate will be selected or not

Marks of (Maths+ Physics+ Chemistry) >200 ⟶ Selected

Marks of (Maths+ Physics+ Chemistry) >175 & Financially poor ⟶ Selected

Example

Given a student **x** defined by his name, address, marks in the subjects and his financial status the machine learns the

target function

$f(x)=y$

where y is the output whether the candidate will be selected or not

Marks of (Maths+ Physics+ Chemistry) >200 → Selected

Marks of (Maths+ Physics+ Chemistry) >175 & Financially poor → Selected

Marks of (Maths+ Physics+ Chemistry) =178 Financially well —Selected?→ Not Selected

Another Example

A bank wants to know whether to assign loan to a person or not

ID	Age	Has_Job	Own_House	Credit_Rating	Class
1	young	false	false	fair	No
2	young	false	false	good	No
3	young	true	false	good	Yes
4	young	true	true	fair	Yes
5	young	false	false	fair	No
6	middle	false	false	fair	No
7	middle	false	false	good	No
8	middle	true	true	good	Yes
9	middle	false	true	excellent	Yes
10	middle	false	true	excellent	Yes
11	old	false	true	excellent	Yes
12	old	false	true	good	Yes
13	old	true	false	good	Yes
14	old	true	false	excellent	Yes
15	old	false	false	fair	No

Bank has these data of previous years

Another Example

A bank wants to know whether to assign loan to a person or not

Learn a classification model from the data
Use the model to classify future loan applications into
Yes (approved) and
No (not approved)
What is the class for following case/instance?

Another Example

A bank wants to know whether to assign loan to a person or not

Learn a classification model from the data
Use the model to classify future loan applications into
Yes (approved) and
No (not approved)
What is the class for following case/instance?

Age	Has_Job	Own_house	Credit-Rating	Class
young	false	false	good	?

Another Example

Data: Loan application data
Task: Predict whether a loan should be approved or not.
Performance measure: accuracy.

No learning: classify all future applications (test data) to the majority class (i.e., Yes):

ID	Age	Has_Job	Own_House	Credit_Rating	Class
1	young	false	false	fair	No
2	young	false	false	good	No
3	young	true	false	good	Yes
4	young	true	true	fair	Yes
5	young	false	false	fair	No
6	middle	false	false	fair	No
7	middle	false	false	good	No
8	middle	true	true	good	Yes
9	middle	false	true	excellent	Yes
10	middle	false	true	excellent	Yes
11	old	false	true	excellent	Yes
12	old	false	true	good	Yes
13	old	true	false	good	Yes
14	old	true	false	excellent	Yes
15	old	false	false	fair	No

Another Example

Data: Loan application data
Task: Predict whether a loan should be approved or not.
Performance measure: accuracy.

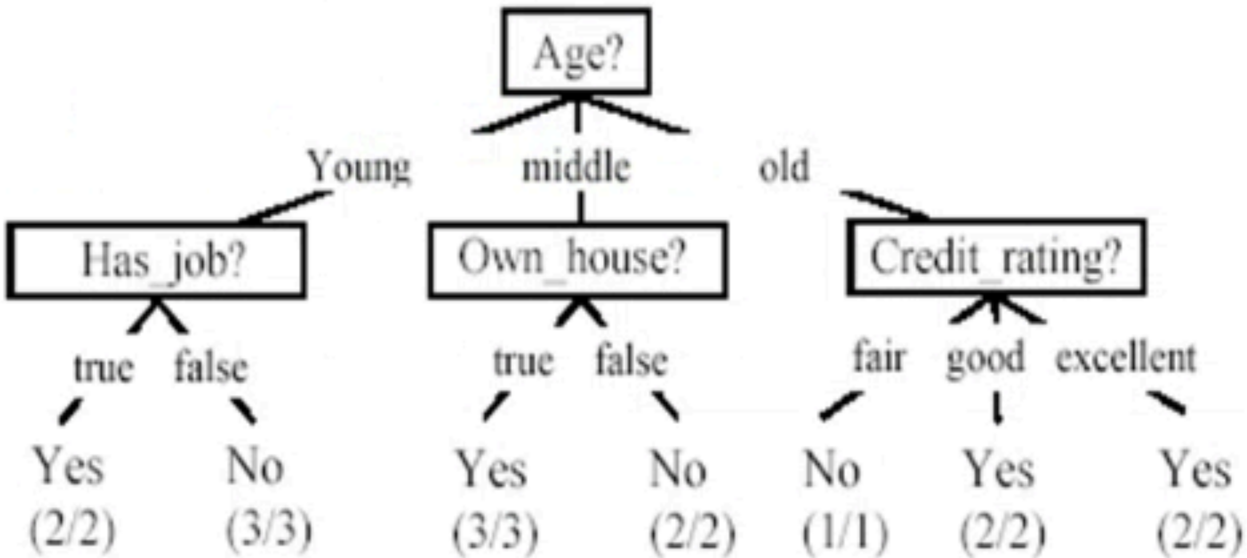
No learning: classify all future applications (test data) to the majority class (i.e., Yes):
Accuracy = 9/15 = 60%.
We can do better than 60% with learning.

ID	Age	Has_Job	Own_House	Credit_Rating	Class
1	young	false	false	fair	No
2	young	false	false	good	No
3	young	true	false	good	Yes
4	young	true	true	fair	Yes
5	young	false	false	fair	No
6	middle	false	false	fair	No
7	middle	false	false	good	No
8	middle	true	true	good	Yes
9	middle	false	true	excellent	Yes
10	middle	false	true	excellent	Yes
11	old	false	true	excellent	Yes
12	old	false	true	good	Yes
13	old	true	false	good	Yes
14	old	true	false	excellent	Yes
15	old	false	false	fair	No

Another Example

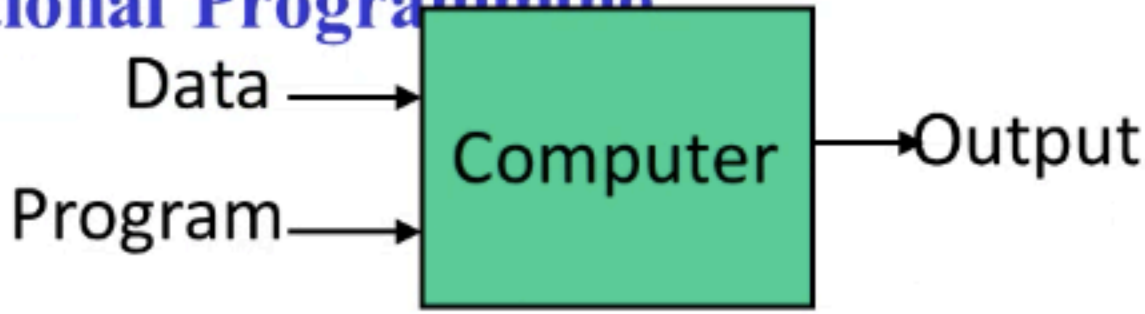
A decision tree approach

ID	Age	Has_Job	Own_House	Credit_Rating	Class
1	young	false	false	fair	No
2	young	false	false	good	No
3	young	true	false	good	Yes
4	young	true	true	fair	Yes
5	young	false	false	fair	No
6	middle	false	false	fair	No
7	middle	false	false	good	No
8	middle	true	true	good	Yes
9	middle	false	true	excellent	Yes
10	middle	false	true	excellent	Yes
11	old	false	true	excellent	Yes
12	old	false	true	good	Yes
13	old	true	false	good	Yes
14	old	true	false	excellent	Yes
15	old	false	false	fair	No

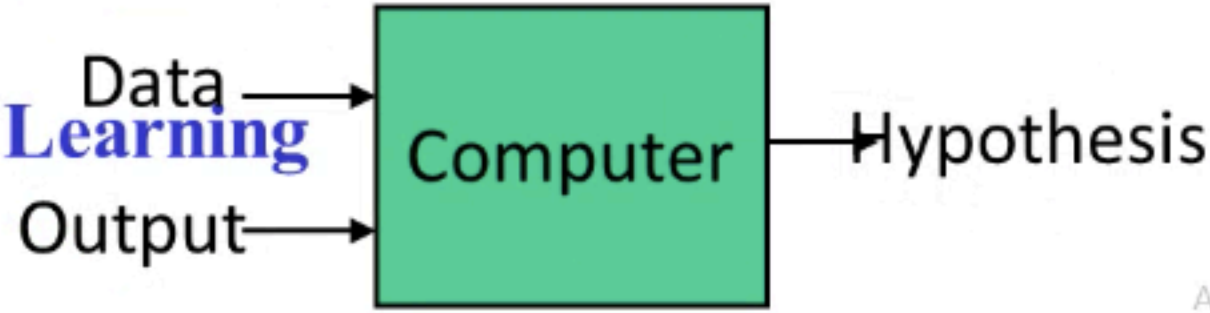


Traditional Programming vs. Machine Learning

Traditional Programming



Machine Learning

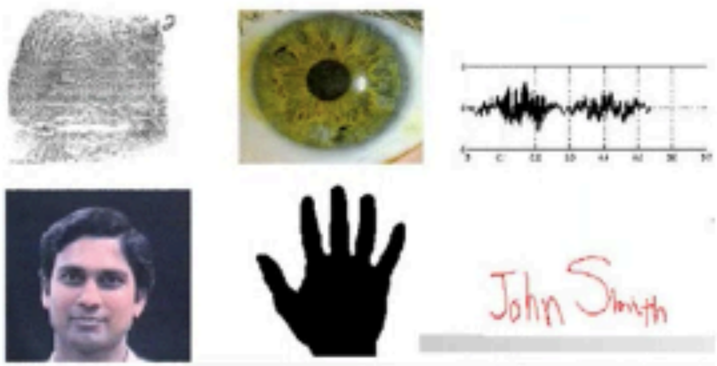


What is a Pattern?

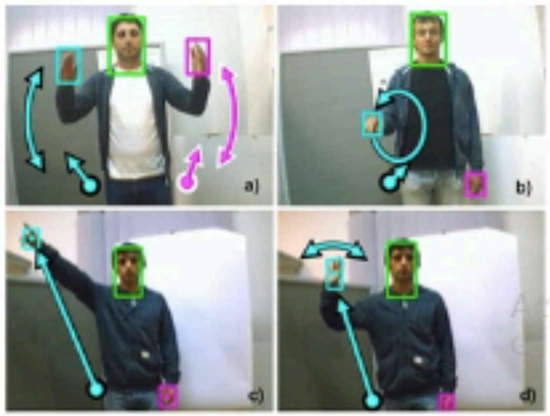
- An object or event.
- Represented by a vector **x** of values corresponding to various features.

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

biometric patterns



hand gesture patterns



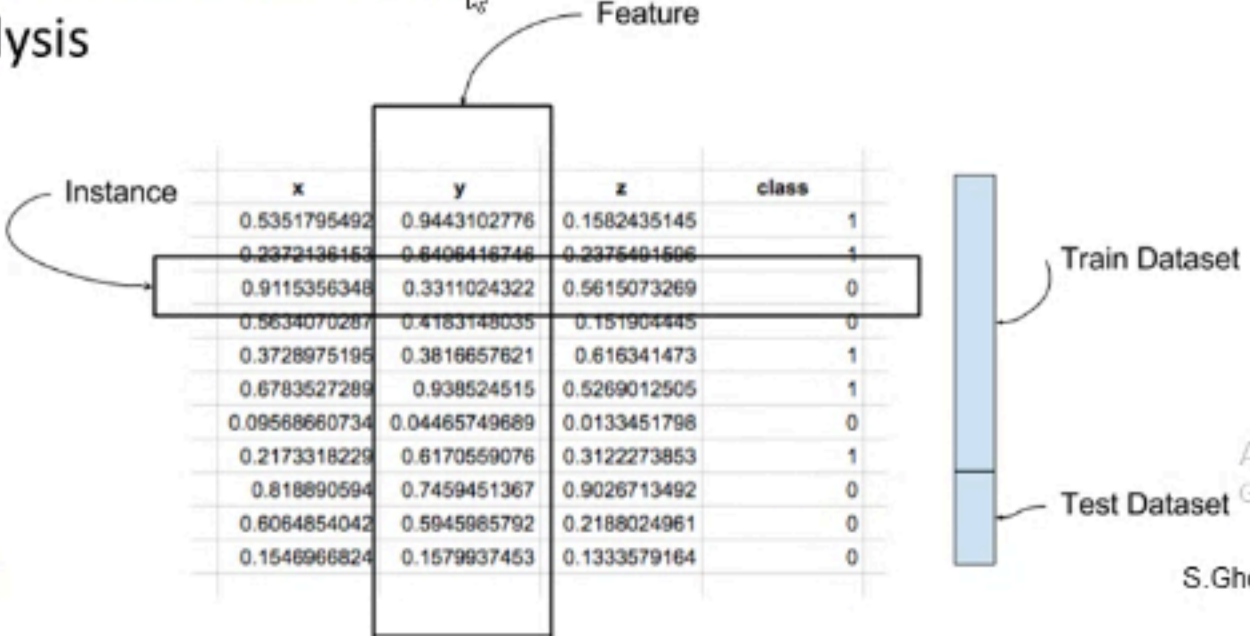
What is a Feature?

- A pattern is defined by a set of features/ attributes/ …

Person	Ht (x1)	Wt (x2)	Class (y)
1	6”	89 kg	American/1
2	5’4”	75	Indian/0
3			
10001	5.5	79	??

f: x→y
Y = f(x), form of f is unknown
Known:
(x1,x2), y
f(x1,x2)

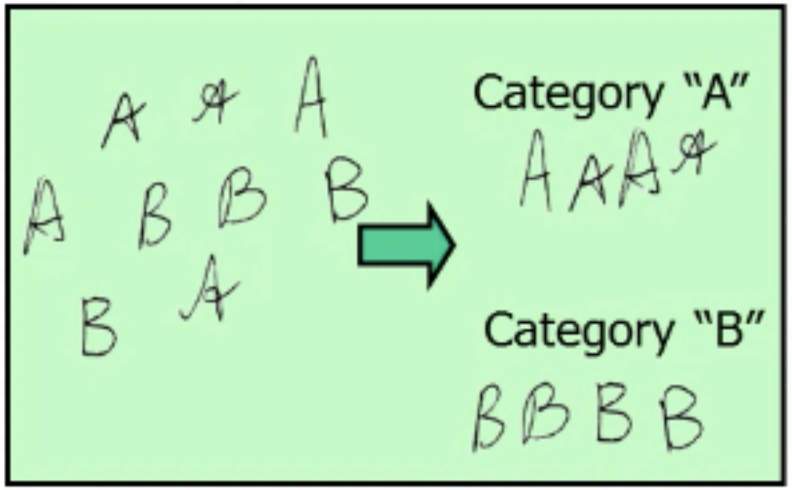
- A measurable property of the object that can be used for analysis



x	y	z	class
0.5351795492	0.9443102776	0.1582435145	1
0.2372136153	0.6406416746	0.2375401506	1
0.9115356348	0.3311024322	0.5615073269	0
0.5634070287	0.4183148035	0.151904445	0
0.3728975195	0.3816657621	0.616341473	1
0.6783527289	0.938524515	0.5269012505	1
0.09568660734	0.04465749689	0.0133451798	0
0.2173318229	0.6170559076	0.3122273853	1
0.818890594	0.7459451367	0.9026713492	0
0.6064854042	0.5945965792	0.2188024961	0
0.1546966824	0.1579937453	0.1333579164	0

Pattern Recognition
Classification

- Assign an object or an event (pattern) to one of several known categories (or classes).

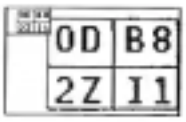


Data Variability

- Intra-class variability

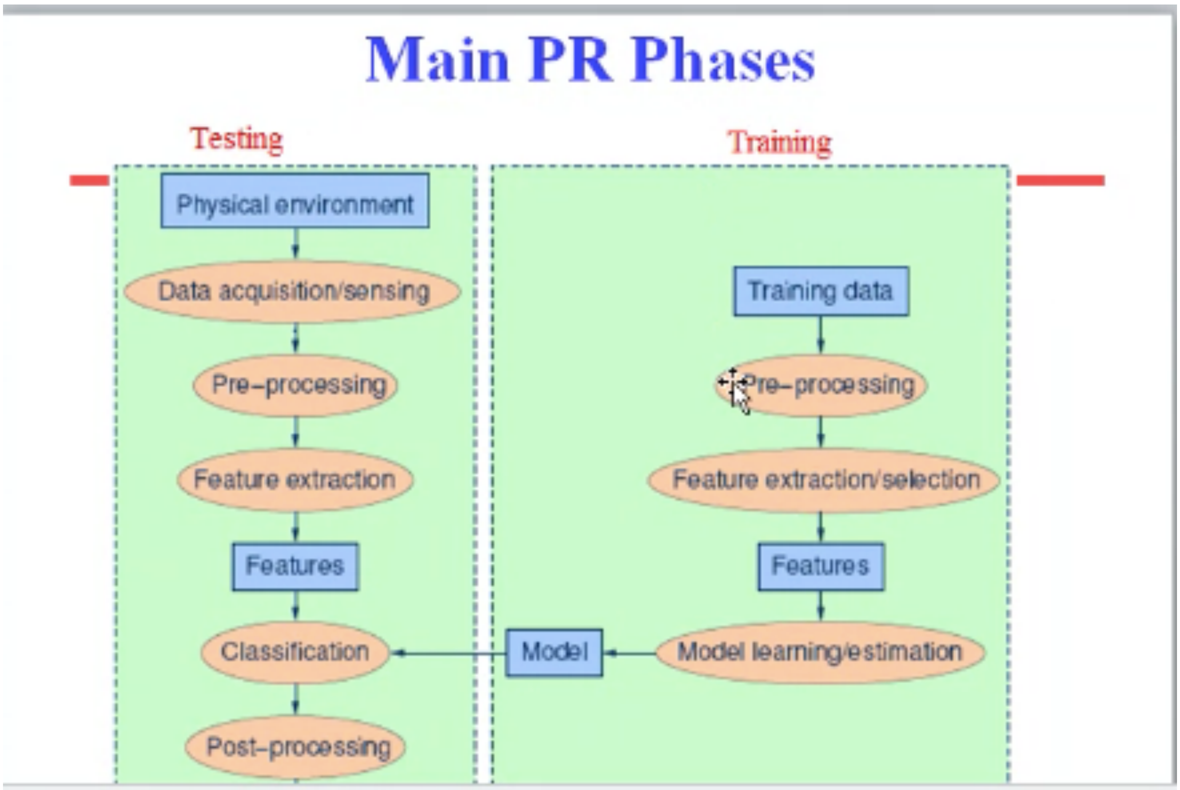
The letter "T" in different typefaces

- Inter-class variability



Letters/Numbers that look similar

- We typically deal with this issue by collecting a large number of examples and a "good" set of features.



Main PR Phases
Testing
Training
Physical environment
Data acquisition/sensing
Pre–processing
Feature extraction
Features
Classification
Post–processing
Training data
Pre–processing
Feature extraction/selection
Features
Model learning/estimation
Model

Learning

- Supervised Learning
  - ✓ *Number of classes and class labels of some data patterns are known for supervision (Known as training samples)*
- Unsupervised Learning
  - ✓ *Cluster labels are not known*
  - ✓ *Number of clusters may not be known*
- Semi-supervised Learning
  - ✓ *Class labels of only a very small data patterns are known for supervision*
- Reinforcement Learning
  - ✓ *Indirect indication of correctness*

Supervised Learning

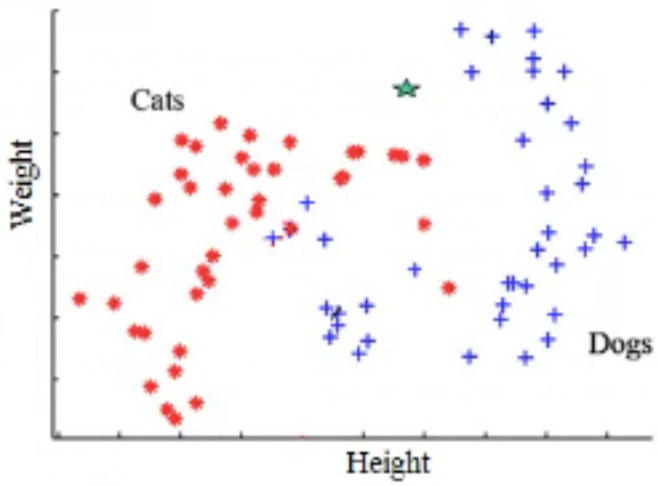
- There is a learner and a supervisor(teacher).
- Under the complete supervision of the teacher, the learner acquires the knowledge about the data.

Class: Dog

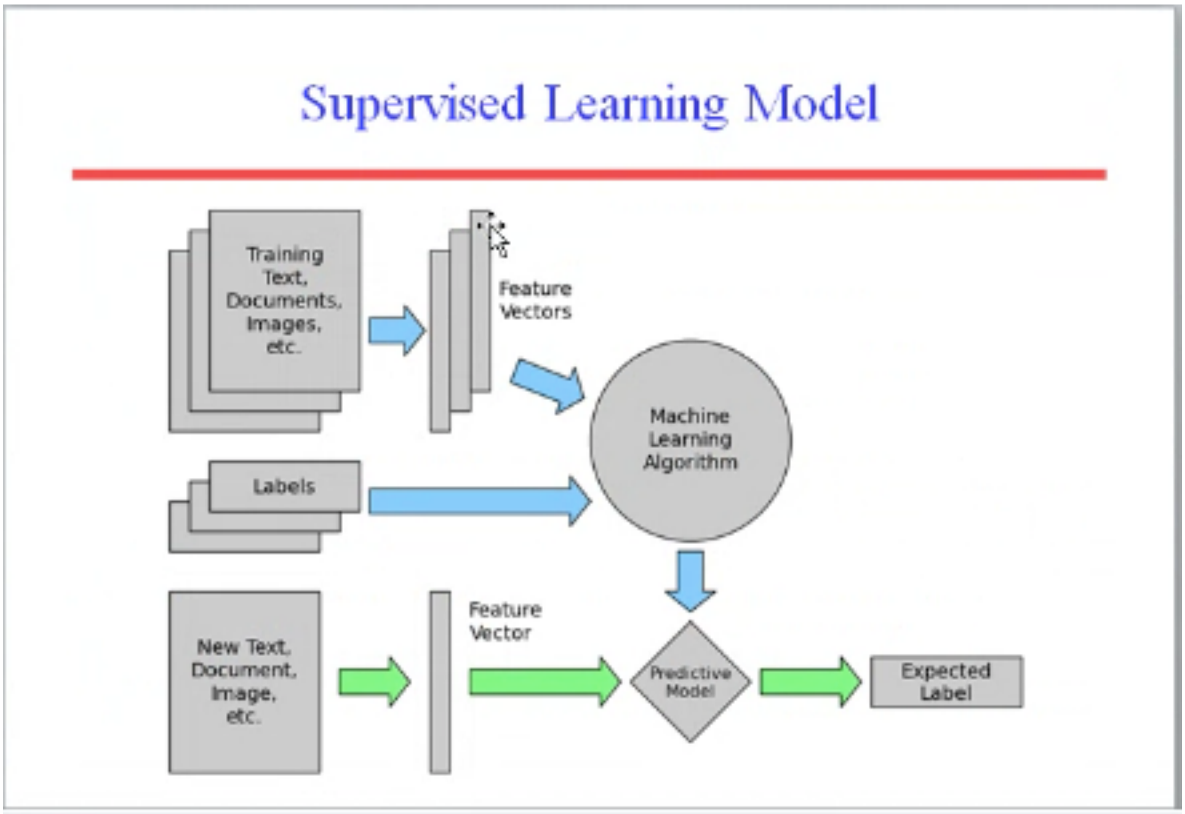
Training examples

Class: Cat

Supervised Learning



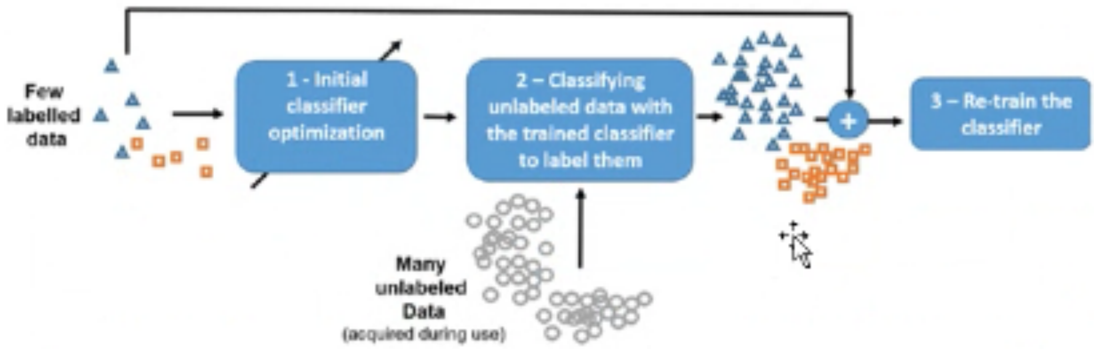
Classify the unknown test sample on the basis of the available training data


Supervised Learning Model
Training
Text,
Documents,
Images,
etc.
Feature
Vectors
Machine
Learning
Algorithm
Labels
New Text,
Document,
Image,
etc.
Feature
Vector
Predictive
Model
Expected
Label

Unsupervised Learning

- Unsupervised learning draws inferences from datasets consisting of input data without labeled responses.
- Therefore this type of learning has no teacher
- It forms a model to extract the underlying structure of the given data

Semi Supervised Learning



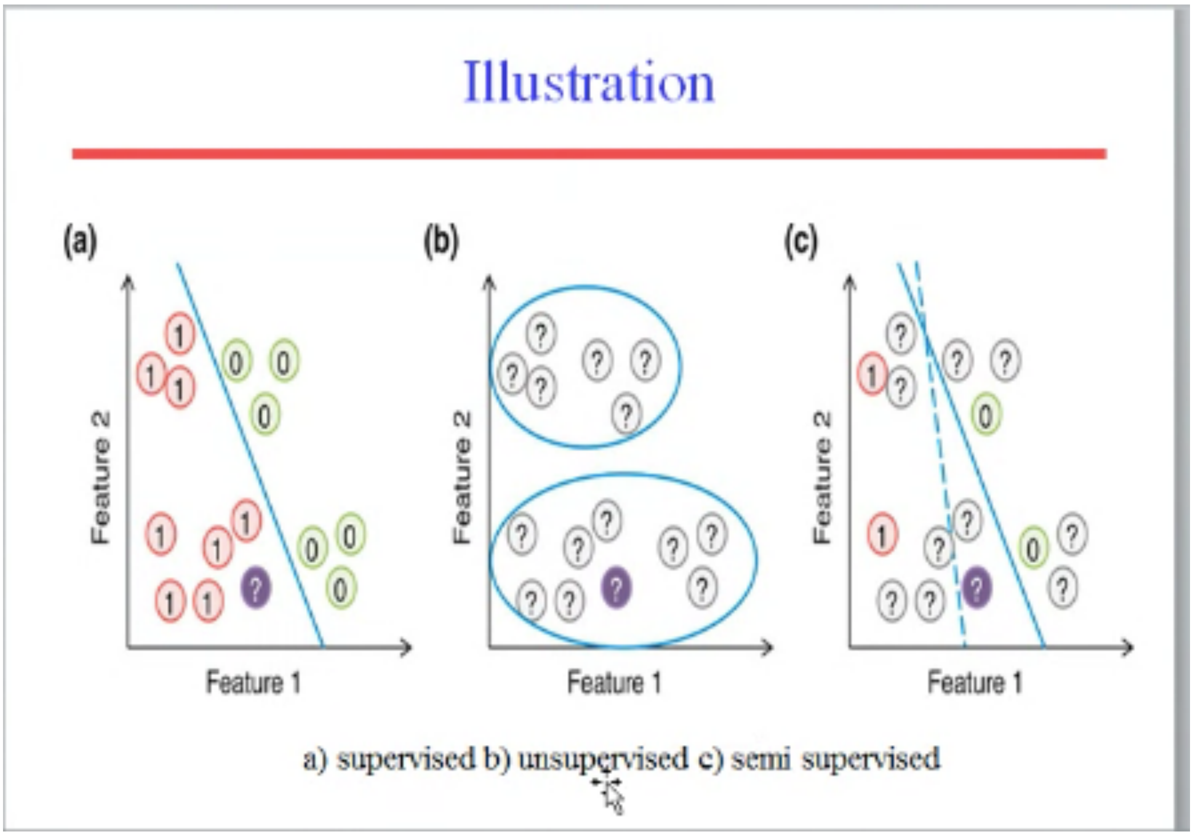
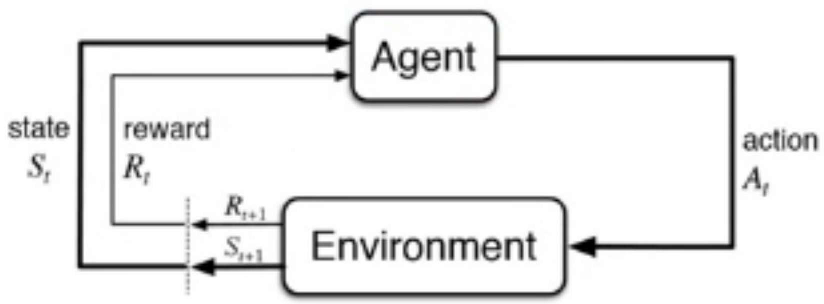


Illustration
(a)
(b)
(c)
Feature 2
Feature 1
Feature 2
Feature 1
Feature 2
Feature 1
a) supervised b) unsupervised c) semi supervised

Reinforcement Learning



- More general than supervised / unsupervised
- Learn from interaction with environment to achieve a goal
- Taking suitable action to maximize (in terms of reward) in a particular situation

Classification: Applications

- Assign object/event to one of a given finite set of categories.
  - Protein sequence analysis
  - Medical diagnosis (given symptoms diagnose the diseases)
  - Credit card applications or transactions
  - Fraud detection in e-commerce
  - Worm detection in network packets
  - Spam filtering in email
  - Recommended articles in a newspaper
  - Recommended books, movies, music, or jokes
  - Financial investments
  - Spoken words
  - Handwritten characters (identify the character)

Measuring Performance

- Classification accuracy
- Solution correctness
- Solution quality (efficiency)
- Speed of performance

0	0	0	0.4	0.2
0	1	1	0.8	0.99
1	0	1	0.9	0.8
1	1	0	0.1	0.001

Learning is making useful changes in our minds

Marvin Minsky

LEARN ----UNLEARN----RELEARN

- Primary task of all biological neural systems is to control various functions (mainly behavioral).
- Human being can do it almost instantaneously and without much effort. e.g., recognizing a scene or music immediately.
- **Artificial Neural Network** (ANN) or Neural Network (NN) models try to **simulate** the **biological neural network** with electronic circuitry.
- Also known as **Connectionists Model/ Parallel Distributed Processing** (PDP).

Purpose : To achieve human like performance (particularly in pattern recognition & image processing).

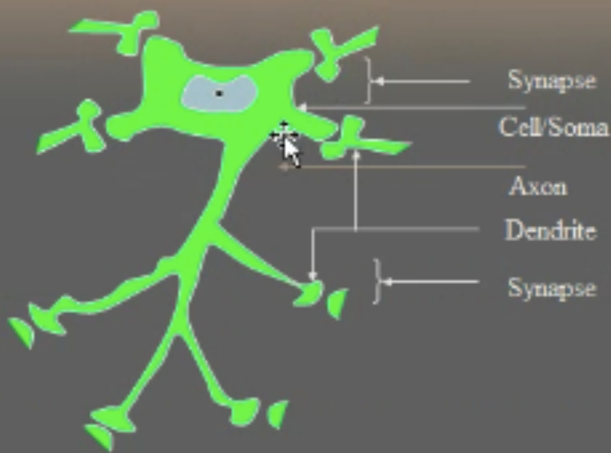
Definition

Definition : Massively parallel interconnected network of simple processing elements which are intended to interact with the objects of the real world in the same way as biological systems do.

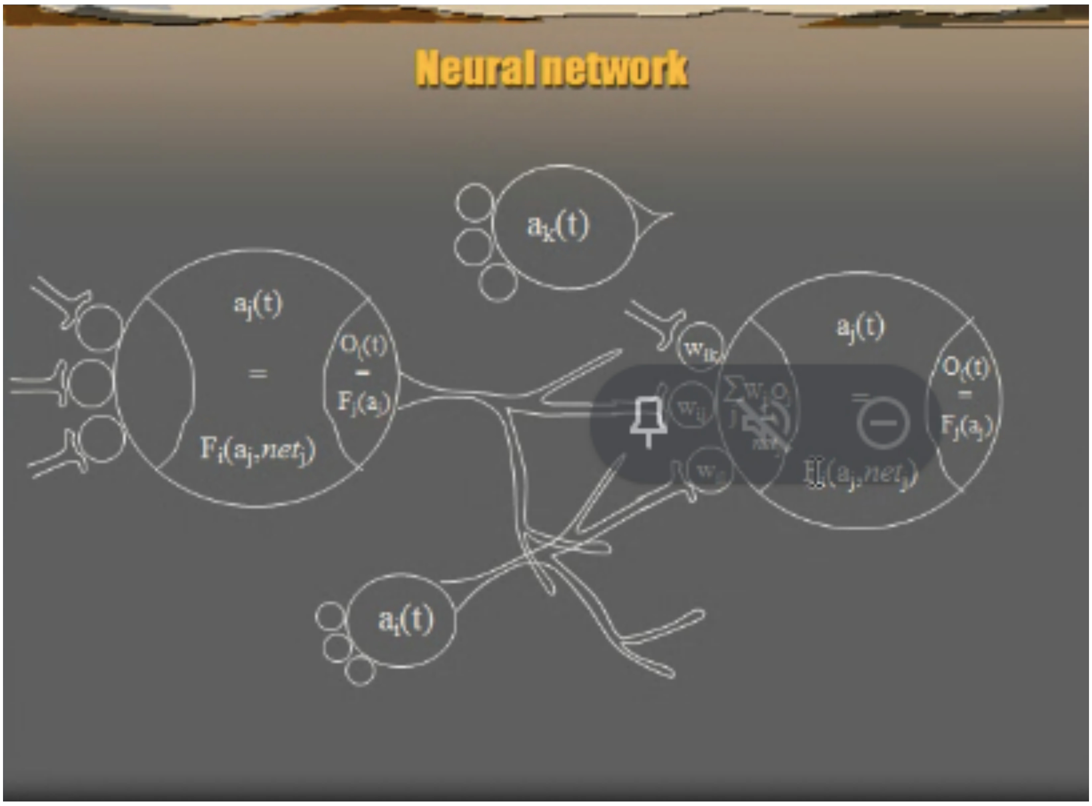
- **NN** models are extreme simplifications of human neural systems.

- Computational elements (neurons/nodes/ processors) are **analogous** to that of the fundamental constituents (**neurons**) of the biological nervous system.

Similarity between BNN and ANN



- Gets input via synaptic connection
- Accumulated input is transformed to a single output
- Output is transmitted through axon
- If input > 0, the neuron fires
- Total output ⇐ firing rate



Neural network
ak(t)
aj(t)
=
Fj(aj,netj)
Oj(t)
=
Fj(aj)
wjk
wji
ai(t)

General framework of neural networks

- **Processing units**
  - Receives input from connected neurons, compute an output value and sends it to other connected neurons.
  - Three types of units - *input, output, hidden.*
- **Output value** - $o_i(t)=f(I_i(t))$
  - Total input for i^th^ neuron is I_i.
  - f is a *threshold* or *squashing* function.
- **Unidirectional connections** (w_{ij})
  - $w_{ij} < 0 \rightarrow$ unit u_j inhibits unit u_i.
  - $w_{ij} = 0 \rightarrow$ unit u_j has no direct effect on unit u_i.
  - $w_{ij} > 0 \rightarrow$ unit u_j excites unit u_i.

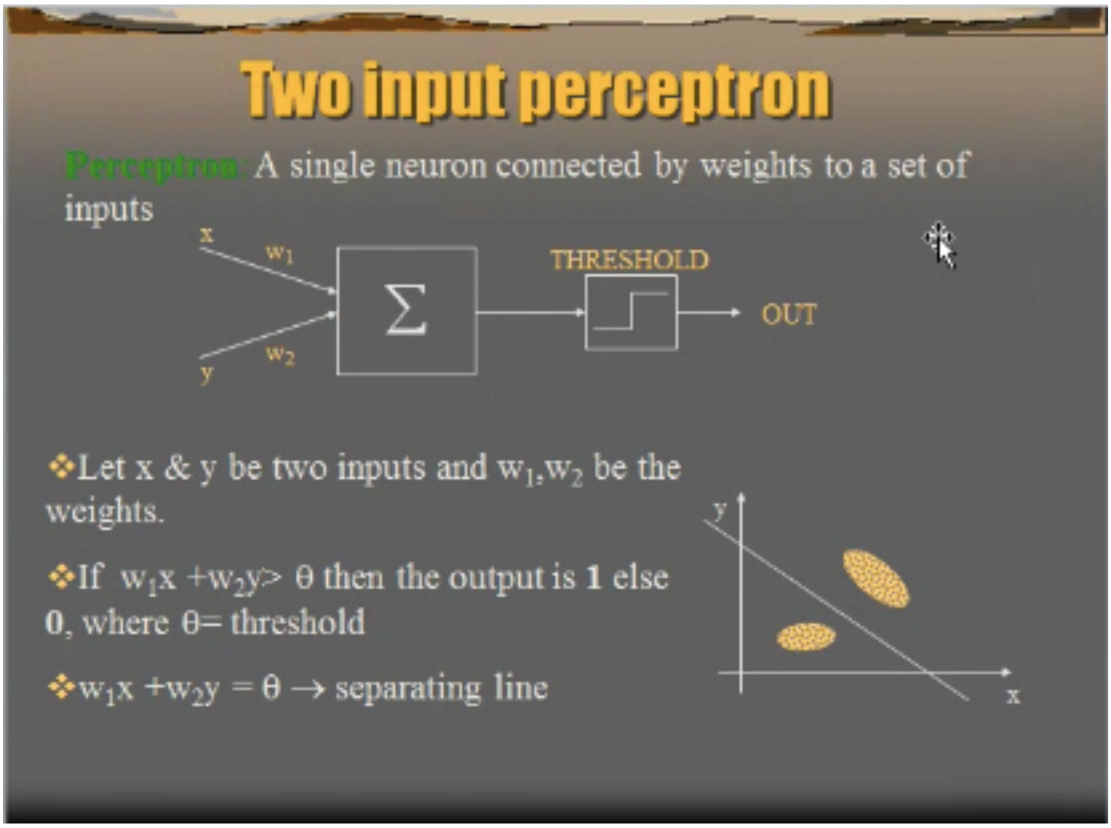
Major advantages

- adaptivity - adjusting the connection strengths to new data/information,
- speed - due to massively parallel architecture,
- robustness - to missing, confusing, ill-defined/noisy data,
- ruggedness - to failure of components,
- optimality - as regards error rates in performance.

Popularly used NN models

Some common feature are there; but **differ in finer details**.

- Multi-layer perceptron (hetero associator/supervised classifier)
- Hopfield's model of associative memory (auto associator/CAM)
- Kohonen's model of self-organizing neural network (regularity detector/ unsupervised classifier)
- Radial basis function network (supervised)
- Adaptive resonance theory (regularity detector)
- Cellular neural network
- Neo-cognitron


Two input perceptron
Perceptron: A single neuron connected by weights to a set of inputs
x
w1
y
w2
Σ
THRESHOLD
OUT
❖Let x & y be two inputs and w1,w2 be the weights.
❖If w1x +w2y> θ then the output is 1 else 0, where θ= threshold
❖w1x +w2y = θ → separating line
y
x

Learning rule

Learning: Present a set of input patterns, adjust the weights until the desired output occurs for each of them.

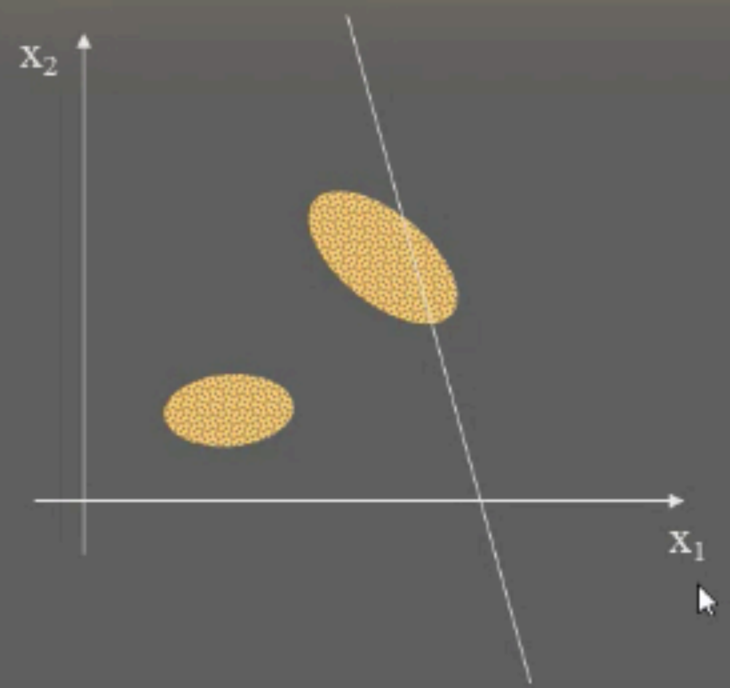
$w_i(t+1) = w_i(t) + \Delta_i;$

$\Delta_i = \eta\ \delta\ x_i;$

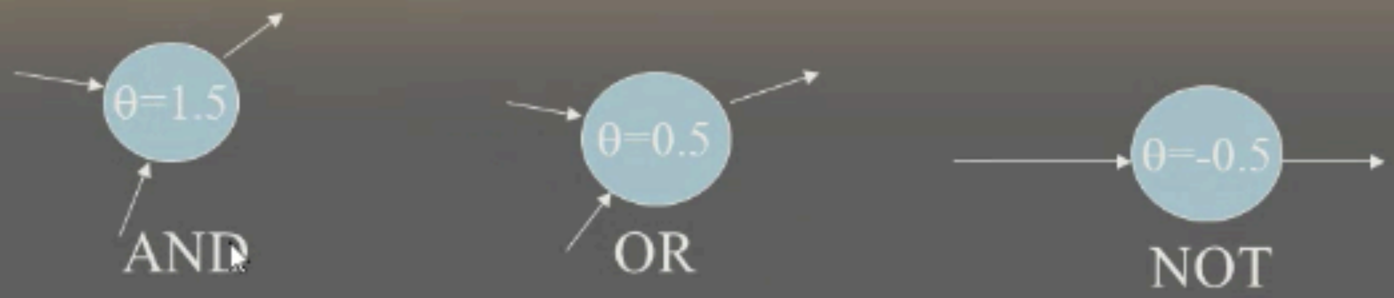
$\delta = T - A$ (i.e., target – actual).

- If the sets of patterns are linearly separable, the single layer perceptron algorithm is guaranteed to find a separating hyperplane in a finite number of steps.

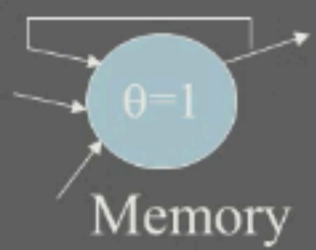
Change of weights



Boolean functions

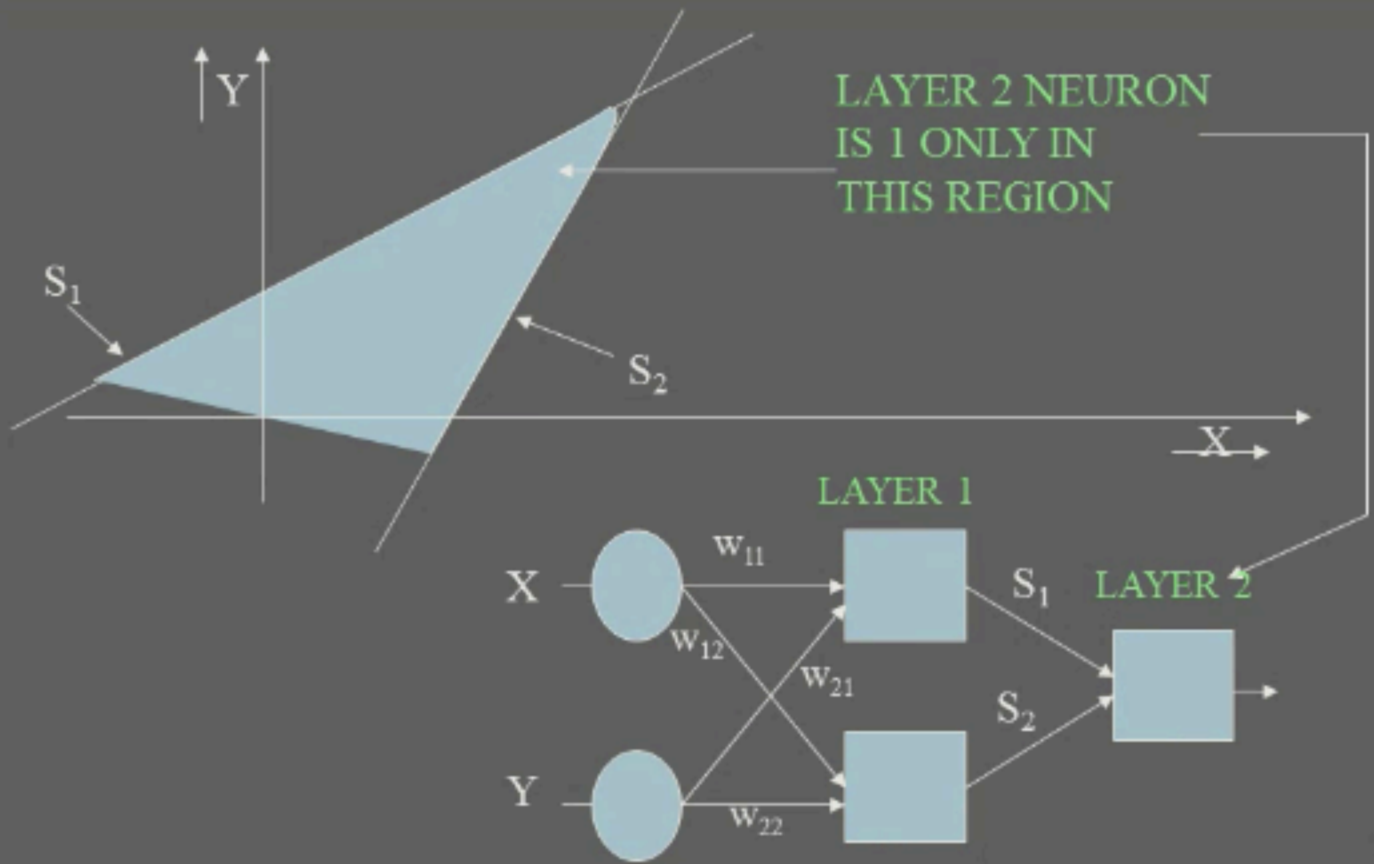


How to design other gates (NOR, NAND) ?



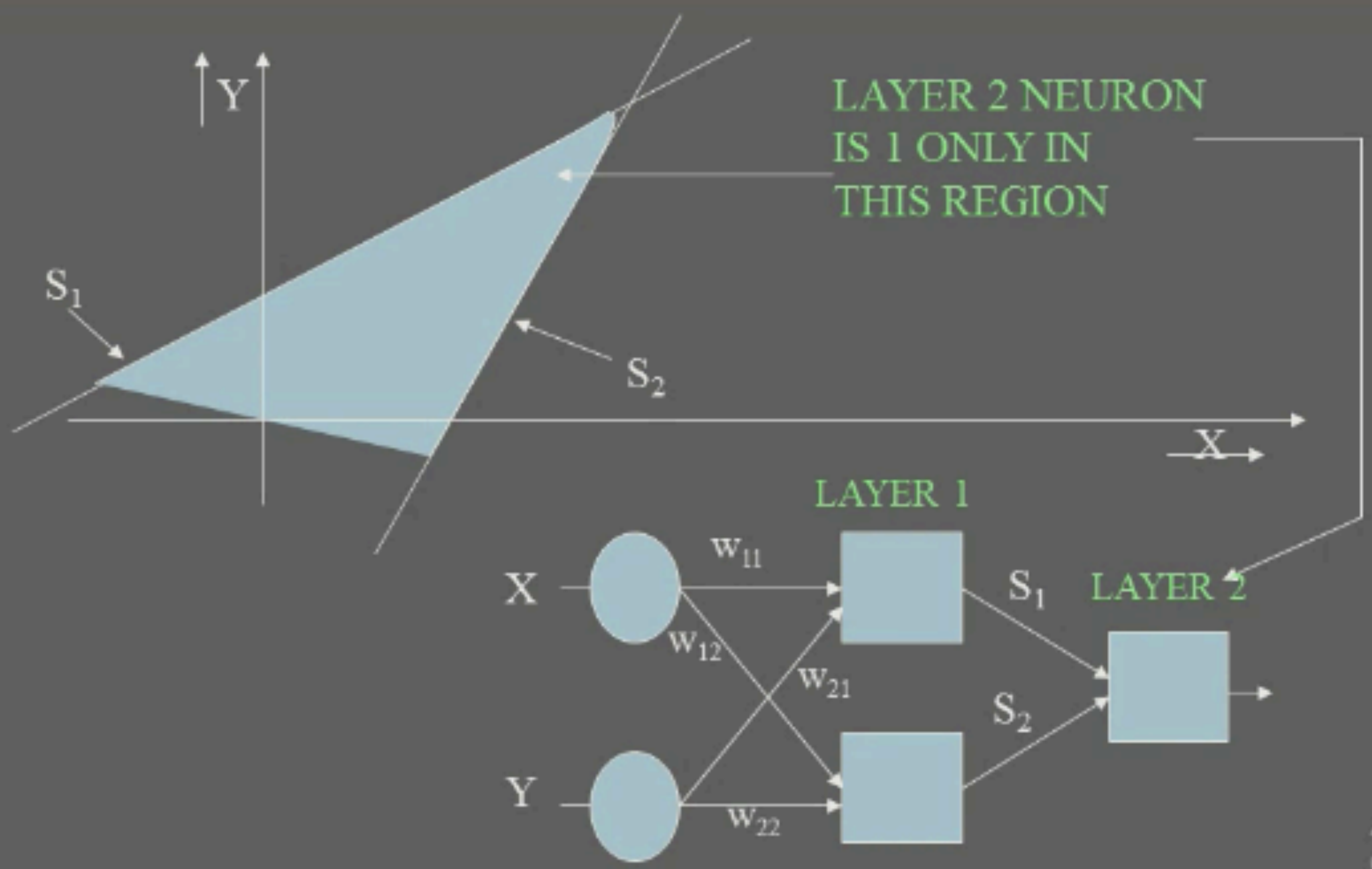
Cascading layers

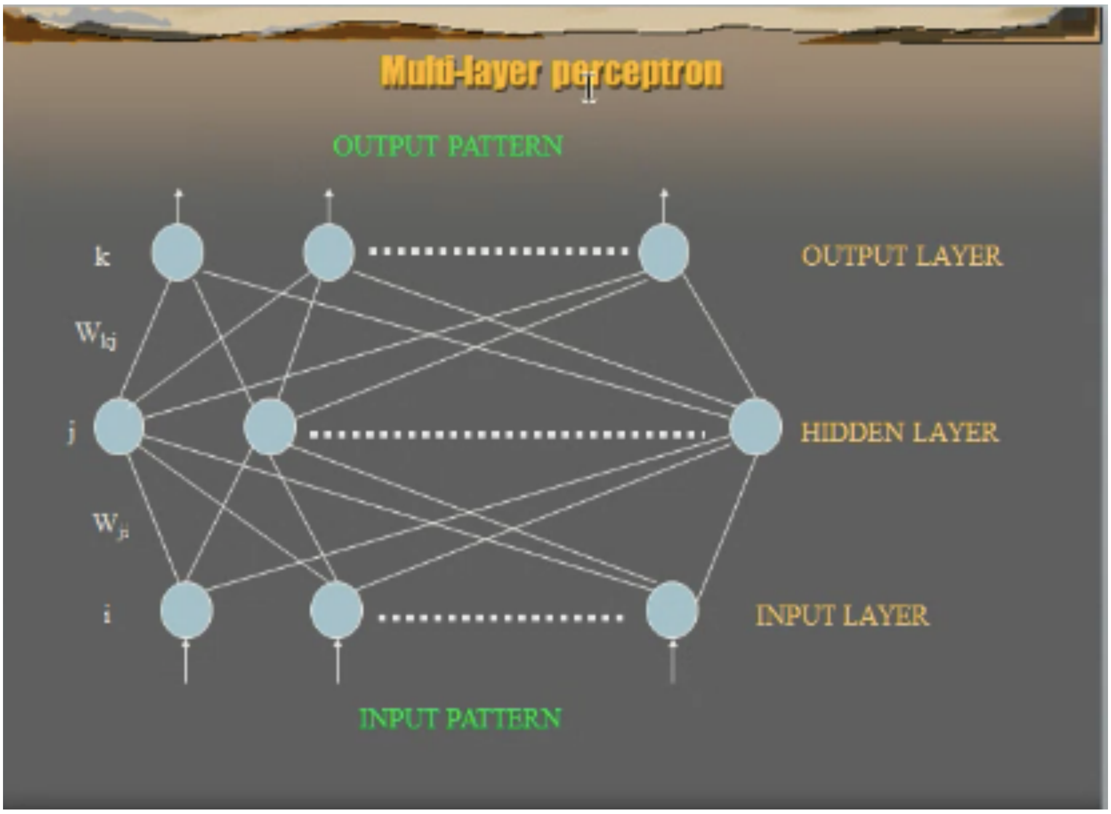
Two layers : Generates convex decision regions



Cascading layers

Two layers : Generates convex decision regions





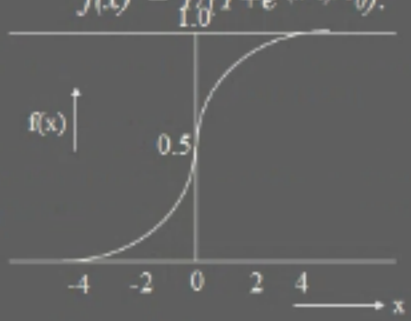
Multi-layer perceptron
OUTPUT PATTERN
k
OUTPUT LAYER
W kj
j
HIDDEN LAYER
W ji
i
INPUT LAYER
INPUT PATTERN

- The output of a node i is obtained as

$$o_i = f(I_i), \quad f \text{ is the activation function.}$$

- Mostly the activation function is sigmoidal/squashing, with the form (smooth, non-linear, differentiable & saturating),

$$f(x) = 1/(1+e^{-(x-\theta)/\theta_0}).$$



- Initially very small random values are assigned to the links/weights.

Parameter updating

- For learning (*training*) we present the input pattern $X=\{x_i\}$, and ask the net to adjust its set of weights/biases in the connecting links such that the desired output $T=\{t_i\}$ is obtained at the output layer.
- Then another pair of X and T is presented for learning.
- Learning tries to find a simple set of weights and biases that will be able to discriminate among all the input/output pairs presented to it.
- The output $\{o_i\}$ will not be the same as the target $\{t_i\}$.

  Error is,

$$E = \frac{1}{2}\sum_i (t_i - o_i)^2$$

- For learning the correct set of weights error is E is reduced as rapidly as possible.
- Use gradient descent technique.

If $o_j = \dfrac{1}{1+e^{-\left(\sum_i w_{ji} o_i - \theta_j\right)}}$ then $f'(I_j) = \dfrac{\partial o_j}{\partial I_j} = o_j(1-o_j)$

and thus we get

$$\Delta w_{ji} = \begin{cases} \eta\left(-\dfrac{\partial E}{\partial o_j}\right) o_j(1-o_j) o_i & \rightarrow \text{output layer} \\ \eta\left(\sum_k \delta_k w_{kj}\right) o_j(1-o_j) o_i & \rightarrow \text{hidden layer} \end{cases}$$

- A large value of η corresponds to rapid learning but might result in oscillations.
- A momentum term of $\alpha \Delta w_{ji}(t)$ can be added to increase the learning rate without oscillation.

$$\Delta w_{ji}(t+1) = \eta \delta_j o_i + \alpha \Delta w_{ji}(t)$$

- The second term is used to specify that the change in w_{ji} at $(t+1)^{th}$ instant should be somewhat similar to the change undertaken at instant t.

Designing Optimum Architecture

- Design of an optimum neural network for a given problem is still not formally specified.
- Pruning/growing algorithm are used for optimizing the architectures of neural networks.
- In growing, people start with a small architecture and gradually add neurons/weights/layer to get an optimum architecture.
- In pruning, people start with a big architecture and gradually delete neurons/weights/layer to get an optimum architecture.
- Genetic algorithms are also used to design optimum architectures.