Chapter- 6 : Machine Learning

- Machine learning is a branch of AI that uses algorithm to allow computer to evolve behaviors based on data collected from databases or gathered through sensors.
- Machine learning focuses on prediction based on known properties learned from the training data.
- The performance is usually evaluated learned with respect to reproduce known knowledge.

Why Machine Learning?

- Recent Progress in algorithms and theory
- Huge computational power is available
- Many tasks would benefit from adaptive systems:
  - Robot exploring Mars (or cleaning your house!)
  - Software agents (OS functions, web searching)
  - Speech, vision, language, …

Machine Learning Algorithms

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning (Environment provides feedback)
- Inductive learning
- Artificial Neural Network Learning
- Genetic Algorithm
- Bayesian Network..etc

Learning

- Learning is acquiring new or modifying existing knowledge, behaviors, skills and may involve synthesizing different types of information

Learning involves 3 factors:

Changes: Learning changes the learner. For machine learning the problem is determining the nature of these changes and how to best represent them.

Generalization: Learning leads to generalization. Performance must improve not only on the same task but on similar tasks

Improvement: Learning leads to improvements. Machine learning must address the possibility that changes may degrade performance and find ways to prevent it.

Learning Methods
- There are two different kinds of information processing which must be considered in a machine learning system

  - *Inductive learning* is concerned with determining general patterns, organizational schemes, rules, and laws from raw data, experience or examples.

  - *Deductive learning* is concerned with determination of specific facts using general rules or the determination of new general rules from old general rules.

    - A=B, B=C, then A =C

**Types of learning method**

1. **Rote Learning**
- It is a technique which focuses on memorization.
- Memorization - saving new knowledge to be retrieved when needed rather than calculated.
- It avoids understanding the inner complexities and inferences of the subject that is being learned.
- It works by taking problems that the performance element has solved and memorizing the problem and the solution.
- Only useful if it takes less time to retrieve the knowledge than it does to recomputed it

  Example: A.L. Samuels Checkers Player (1959-1967). It is a program that knows and follows the rules of checkers. It memorizes and recalls board positions it has encountered in previous games

2. **Learning by Analogy**
- Learning by analogy means acquiring new knowledge about an input entity by transferring it from a known similar entity.
- This technique transforms the solutions of problems in one domain to the solutions of the problems in another domain by discovering analogous states and operators in the two domains.

  Examples of Analogy Learning

Qa = 3   Qb = 9
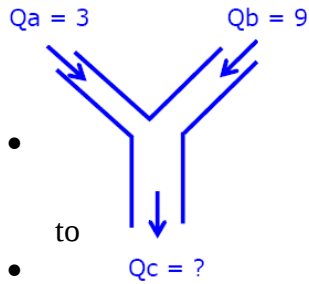
- to   Qc = ?
- **Hydraulic Problem**

I1   I2
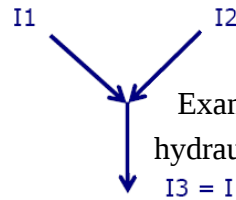
I3 = I1 + I2   Example: Infer by analogy the hydraulics laws that are similar to Kirchhoff's laws. Pressure drops like voltage drops

Kirchoff's First Law Hydrogen atom is like our solar system. The Sun has a greater mass than the Earth and attracts it, causing the Earth to revolve around the Sun. The nucleus also has a greater mass then the electron and attracts it. Therefore it is plausible that the electron also revolves around the nucleus

3. **Explanation Based Learning (EBL)**
- Humans appear to learn quite a lot from one example.
- Human learning is accomplished by examining particular situations and relating them to the background knowledge in the form of known general principles.
- This kind of learning is called "Explanation Based Learning (EBL)".
  - Tea cup example

4. **Learning by Example (Inductive learning)**
- **Learning by example** is a general learning strategy where a concept is learned by drawing inductive inferences from a set of fact.
- AI systems that learn by example can be viewed as searching a concept space by means of a decision tree.
- The best known approach to constructing a decision tree is called **ID3 (Iterative Dichotomizer 3** developed by J. Ross Quinlan in 1975)

ID3

- ID3 is an algorithm used in decision tree learning to generate a decision tree.
- Decision tree consists of decision nodes and leaf nodes connected by arcs.
- ID3 builds the tree from the top down
- Entropy or Information gain is used to select the most useful attribute for classification
- **Entropy(H) = - $\sum (p_i)\log_2(p_i)$**
  - Entropy is the basis of Information Theory
  - Entropy is a measure of randomness, hence the smaller the entropy the greater the information content

ID3 algorithm

- Create a root node for tree
- If all examples are positive, then create a positive node and stop
- If all examples are negative, then create a negative node and stop
- Otherwise
  - Calculate entropy, information gain to select root node and branch node.(nodes with highest information gain or minimum entropy is selected as root node)
  - Partition the examples into subset
  - Repeat until all examples are classified

Eg: Refer class note

Learning Framework

There are four major components in a learning system:
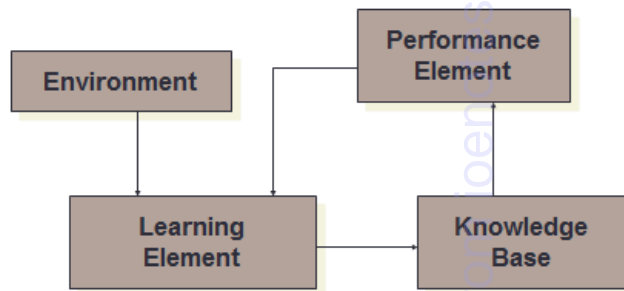
Environment

- The environment refers the nature and quality of information given to the learning element.
- The nature of information depends on its level (the degree of generality with respect to the performance element)
  - High level information is abstract, it deals with a broad class of problems
  - Low level information is detailed; it deals with a single problem.
- The quality of information involves
  - noise free
  - reliable
  - ordered

Learning Elements

- Acquire new knowledge through learning elements. Learning may be of

- Rote learning
- Learning by examples
- Learning by analogy
- Explanation based learning…etc
- The learning elements should have access to all internal actions of the performance element.

The Knowledge Base

The knowledge base should be

1. Expressive
    - Knowledge should be represented in understandable way
2. Modifiable
    - it must be easy to change the data in the knowledge base
3. Extendibility
    - The knowledge base must contain meta-knowledge (knowledge on how the data base is structured) so the system can change its structure

The Performance Element

- The performance element analyzes how complex the learning is and how learning is being performed?
- Complexity depends upon type of task. For learning, the simplest task is classification based on a single rule while the most complex task requires the application of multiple rules in sequence.
- The learning elements should have access to all internal actions of the performance elements.
- Transparency, the learning element should have access to all the internal actions of the performance element.

Genetic Algorithm

- A genetic algorithm maintains a population of candidate solutions for the problem at hand, and makes it evolve by iteratively applying a set of stochastic operators.
- It is a variation of stochastic beam search.
- Inspired by **biological evolution process**
- Uses concepts of "**Natural Selection**" i.e. "**Survival of the fittest"** and "**Genetic Inheritance**"
- Particularly well suited for hard problems where little is known about the underlying search space
- Widely used in business, science and engineering
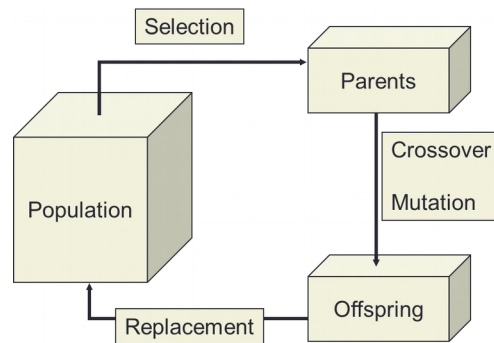
*Genetic process in Nature*

Stochastic operators

- **Selection** replicates the most successful solutions found in a population at a rate proportional to their relative quality
- **Crossover** decomposes two distinct solutions and then randomly mixes their parts to form novel solutions
- **Mutation** randomly produces a candidate solution.

Comparison between Genetic Algorithm and Nature

Genetic Algorithm	Nature
Optimization problem	Environment
Feasible solutions	Individuals living in that environment
Solution Quality (fitness function)	Individual's degree of adaption to its surrounding environment
A set of feasible solutions	A population of organisms
Stochastic operators	Selection, Crossover and mutation in nature's evolutionary process
Iteratively applying a set of stochastic operators on a set of feasible solutions	Evolution of populations to suit their environment

Genetic Algorithm

- GA starts with k randomly generated states (population)
- A state is represented a string over a finite alphabet (often a string of 0s and 1s)
- Evaluation function (fitness function) defines fitness value of each states.
- Produce the next generation of states by selection, crossover, and mutation.
- The primary advantage of GA comes from crossover operation.

Crossover point randomly chosen
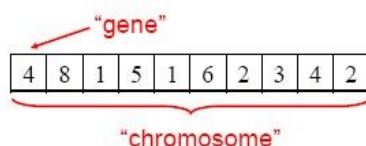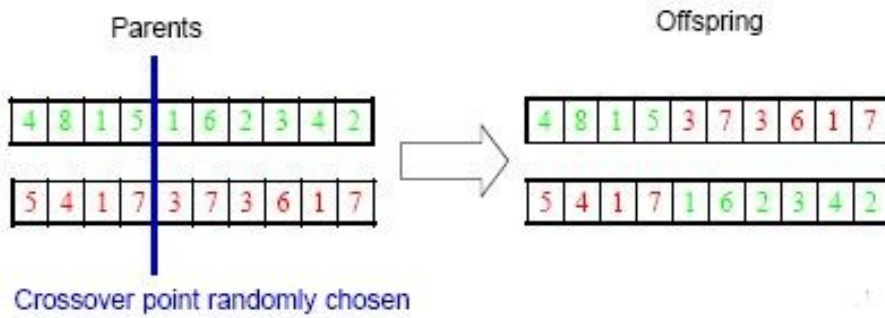
Algorithm

- produce an initial population of individuals
- evaluate the fitness of all individuals
- while (solution not found)
    - o select fitter individuals for reproduction
    - o recombine between individuals
    - o mutate individuals
    - o evaluate the fitness of the modified individuals
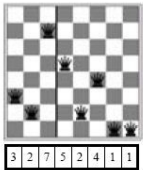   - o generate a new population
- End while

GA flowchart
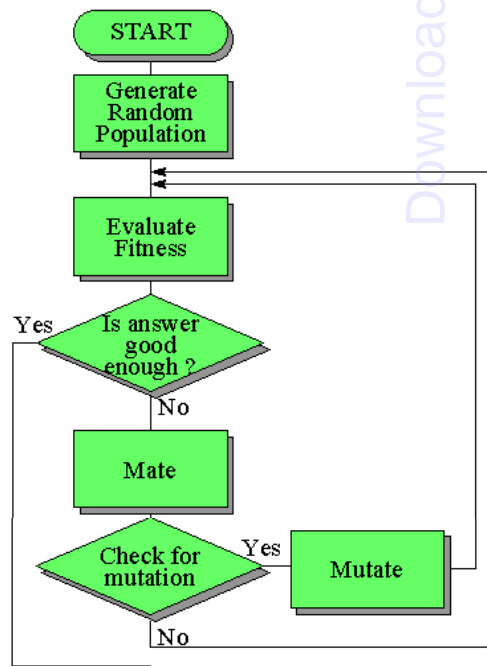
Disadvantage

- GA is better if the problem does not have any mathematical model for the solution.
- GA is less efficient in terms of speed of convergence.
- GA has tendency to get stuck in local maxima rather than global maxima.

An example

Example: 8-queens

- Fitness Function: number of nonattacking pairs of queens (28 is the value for the solution)
- Represent 8-queens state as an 8 digit string in which each digit represents position of queen

3 2 7 5 2 4 1 1

16

Example: 8-queens (Selection)

Notice 3 2 7 5 2 4 1 1 is selected twice while 3 2 5 4 3 2 1 3 is not selected at all
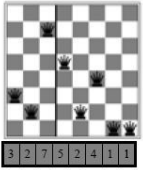
19

Fuzzy Learning

- In 1965 **Lotfi Zadeh**, published his famous paper "fuzzy sets". Zadeh extended the work on possibility theory into a formal system of mathematical logic and introduced a new concept for applying natural language terms. This new logic for representing and manipulating fuzzy terms was called fuzzy logic.

Example: 8-queens (Mutation)

Mutation corresponds to randomly selecting a queen and randomly moving it in its column

21

- **Traditional Logic:** Traditional Boolean logic uses sharp distinctions. For instance Tom with height 181cm is tall. If we draw a line at 180 cm, David with height 179cm is small. Is David really small?

- **Fuzzy logic**: It is a form of knowledge representation suitable for notions that can't be defined precisely but which depend upon their contexts. A way to represent variation or imprecision in logic
- **Fuzzy** means "not clear, distinct or precise or blurred". It is a concept of partial truth, where truth value may range between completely true or completely false.
- In contrast with traditional logic theory where binary sets have two valued logic (True/False), fuzzy logic variables may have value that ranges in degree from 0 to 1.
- Fuzzy logic is a form of multi-valued logic.
- Fuzzy logic reflects how people think. It attempts to model our sense of our decision making and our common sense.
- Example: Temperature, Height, Speed, Distance, Beauty
  - Motor is running really hot
  - Tom is a very tall guy

Crisp (Traditional) Variables

Crisp variables represent precise quantities. It denotes sharp distinctions.

- X = 3.1415
- A Î {0,1}
- Men Î {Tall, short}
- Speed Î { slow, fast}

Range of logical values in Boolean and fuzzy logic
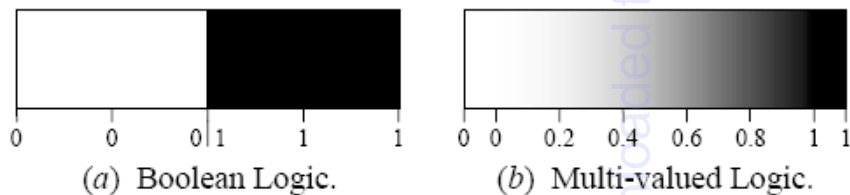
Crisp and fuzzy sets example

Name	Height, cm	Degree of Membership	
		Crisp	Fuzzy
Chris	208	1	1.00
Mark	205	1	1.00
John	198	1	0.98
Tom	181	1	0.82
David	179	0	0.78
Mike	172	0	0.24
Bob	167	0	0.15
Steven	158	0	0.06
Bill	155	0	0.01
Peter	152	0	0.00
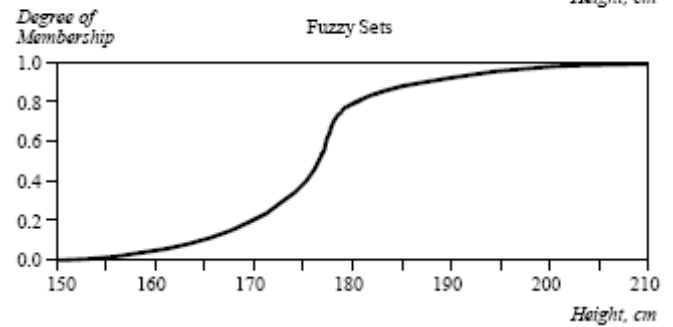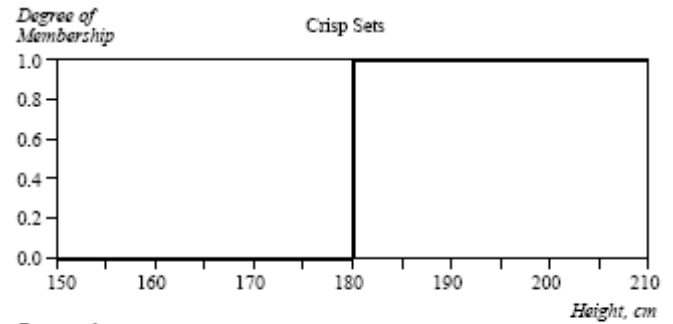
- In fuzzy theory, fuzzy set A of universe X is defined by function μ_A (x) called membership function of set A.

$$\mu_A \text{ (x) : } X \to [0,1] ,$$

Where, μ_A (x) = 1, if x is totally in A

= 0, if x is not in A

$0 < \mu_A$ (x) < 1, if x is partially in A.

- For any element x of X, membership function μ_A (x) equals the degree to which x is an element of set A. This degree ranges from 0 to 1, represents degree of membership, also called membership value of element x in set A.

μ_A: $X \to [0,1]$ the membership function of A.

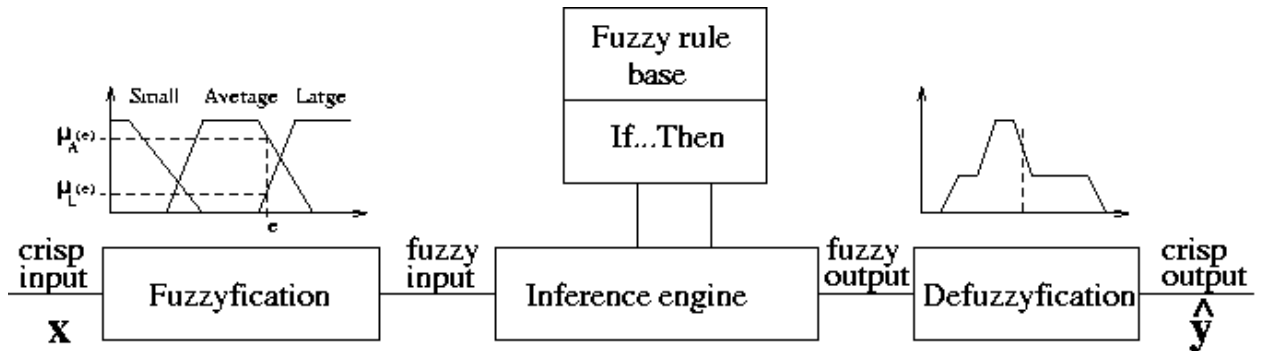μ_A (x) ∈ [0,1] is the degree of membership x in A.

- A fuzzy variable is often denoted by its membership function.

Fuzzy Inferences
- Two approaches of fuzzy inference are
  - Mamdani Inference
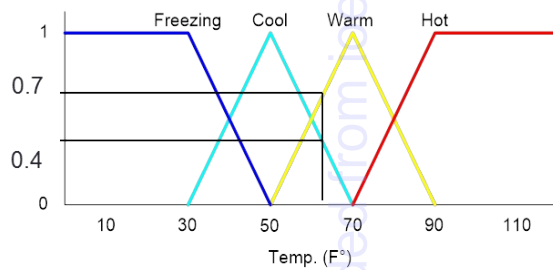
- Sugeno fuzzy inference

Mamdani inference applied in four stages:
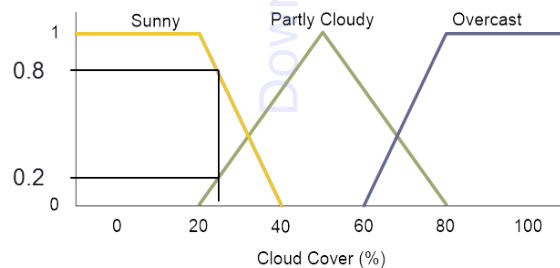
i. *Fuzzyfication of input variables:*

- Determines an input's membership in overlapping sets.
- Fuzzy Control combines the use of fuzzy linguistic variables with fuzzy logic

- 65 F° ⇒ Cool = 0.4, Warm= 0.7

- 25% Cover ⇒Sunny = 0.8, Cloudy = 0.2

ii. *Rule Evaluation*

- The second step is to take the fuzzified inputs, (such as $m_{(x=A1)}$ = 0.5, $m_{(x=A2)}$ = 0.2, $m_{(y=B1)}$ = 0.1 and $m_{(y=B2)}$ = 0.7), and apply them to the antecedents of the fuzzy rules.

11

- If a given fuzzy rule has multiple antecedents, the fuzzy operator (AND or OR) is used to obtain a single number that represents the result of the antecedent evaluation.
- This number (the truth value) is then applied to the consequent membership function.

iii. *Aggregation of the rule outputs:*
- Aggregation is the process of unification of the outputs of all rules.
- Take the membership functions of all rule consequents previously scaled and combine them into a single fuzzy set.
- Determine outputs based on inputs and rules.
- The input of the aggregation process is the list of clipped or scaled consequent membership functions, and the output is one fuzzy set for each output variable.

iv. *Defuzzification:*
- Fuzziness helps us to evaluate the rules, but the final output of a fuzzy system has to be a crisp number.
- The input for the defuzzification process is the aggregate output fuzzy set and the output is a single number
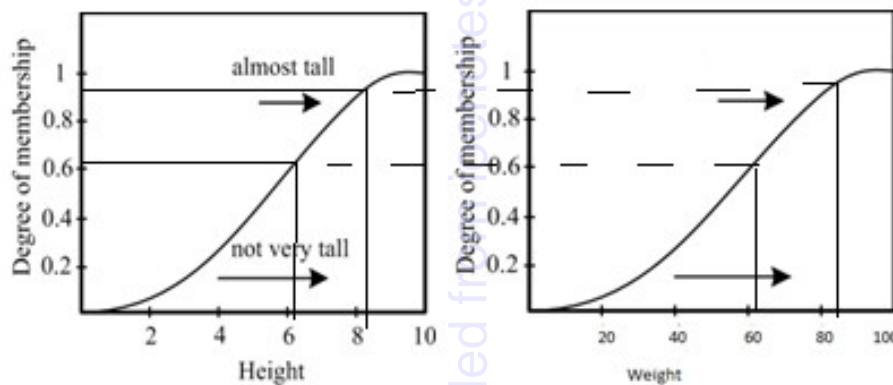
Drawbacks

- Fuzzy logic deals with imprecision, and vagueness, but not uncertainty
- Requires tuning of membership functions
- Fuzzy Logic control may not scale well to large or complex problems

Boltzmann machine

- A **Boltzmann machine** is a type of stochastic recurrent neural network.

- A Boltzmann machine, like a Hopfield network, is a network of units with an "energy" defined for the network.

- Boltzmann machines can be seen as the stochastic, generative counterpart of Hopfield nets.

- First examples of a neural network capable of learning internal representations, and are able to solve difficult problems.

- Have not proven useful for practical problems in machine learning or inference.

- Theoretically exciting due to Hebbian nature of their training algorithm, as well as their parallelism and the resemblance of their dynamics to simple physical processes.

- If the connectivity is constrained, the learning can be made efficient enough to be useful for practical problems.

- The global energy, E, in a Boltzmann machine is identical in form to that of a Hopfield network:

$$E = -(\sum_{i<j} w_{ij}\, s_i\, s_j + \sum_i \theta_i\, s_i)$$

Where:

- w_{ij} is the connection strength between unit j and unit i.

- s_i is the state, $s_i \in \{0, 1\}$, of unit i.

- θ_i is the bias of unit i in the global energy function. ($-\theta_i$ is the activation threshold for the unit.)

The connections in a Boltzmann machine have two restrictions:

- $w_{ii} = 0 \qquad \forall i$. (No unit has a connection with itself.)

- $w_{ij} = w_{ji} \qquad \forall i, j$. (All connections are symmetric.)

Often the weights are represented in matrix form with a symmetric matrix W, with zeros along the diagonal.