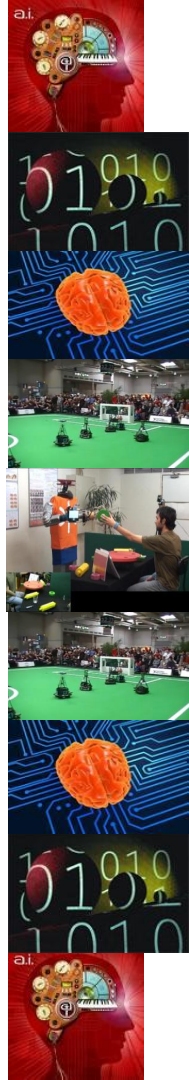
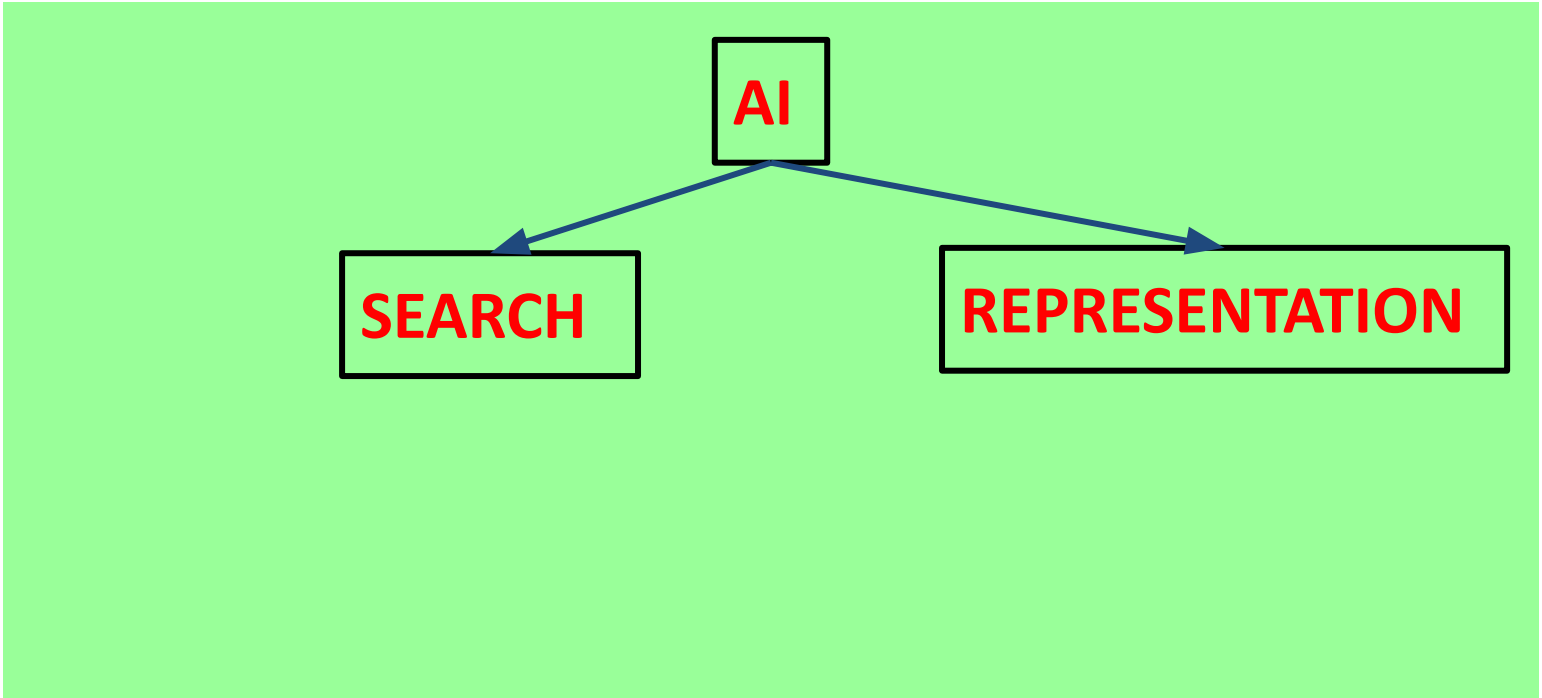
Introduction to Machine Learning

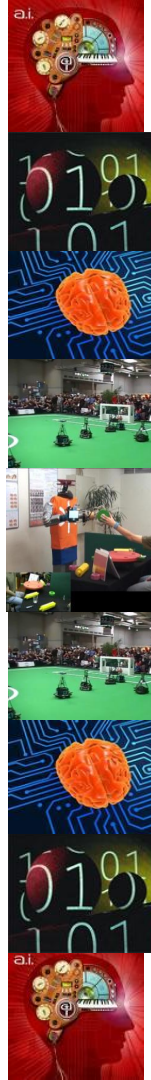
Lecture Outline

- Review of Search and Planning?
- Definition of Machine learning.
- Machine learning techniques

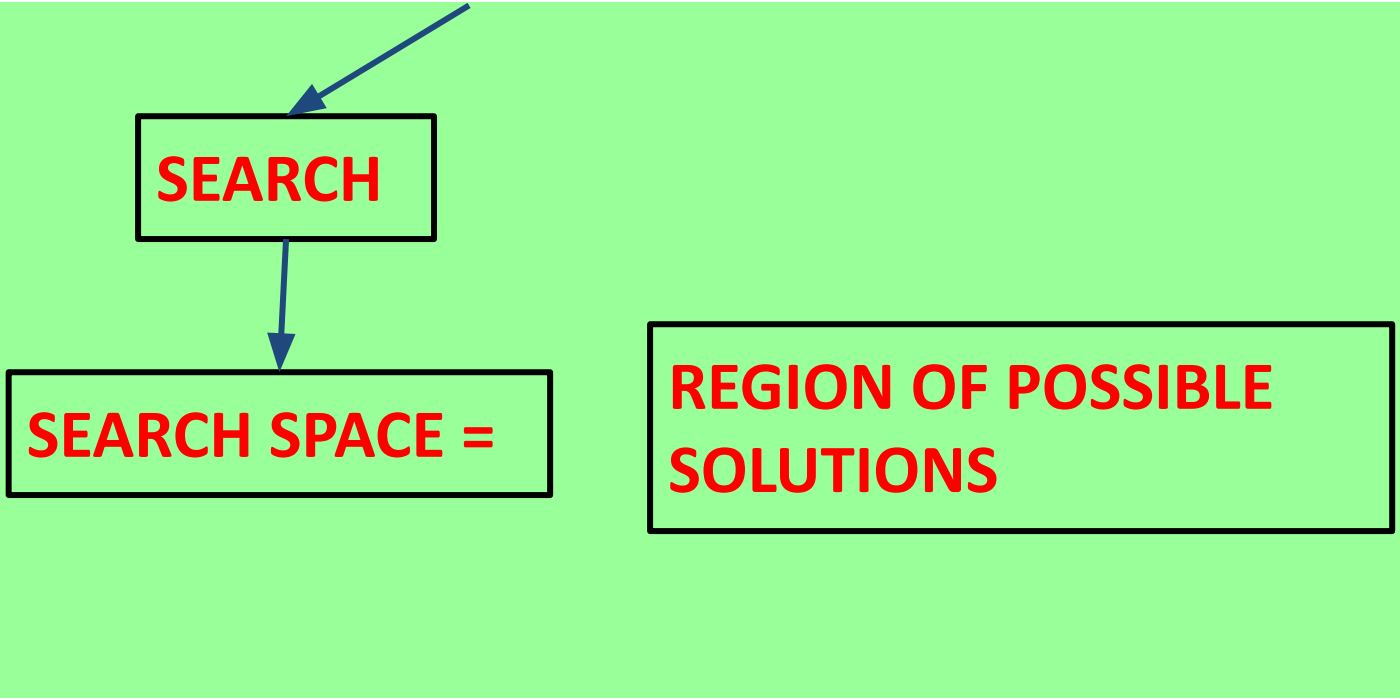


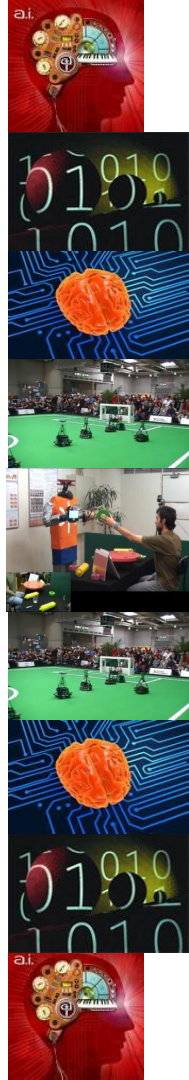
Review Search and Planning



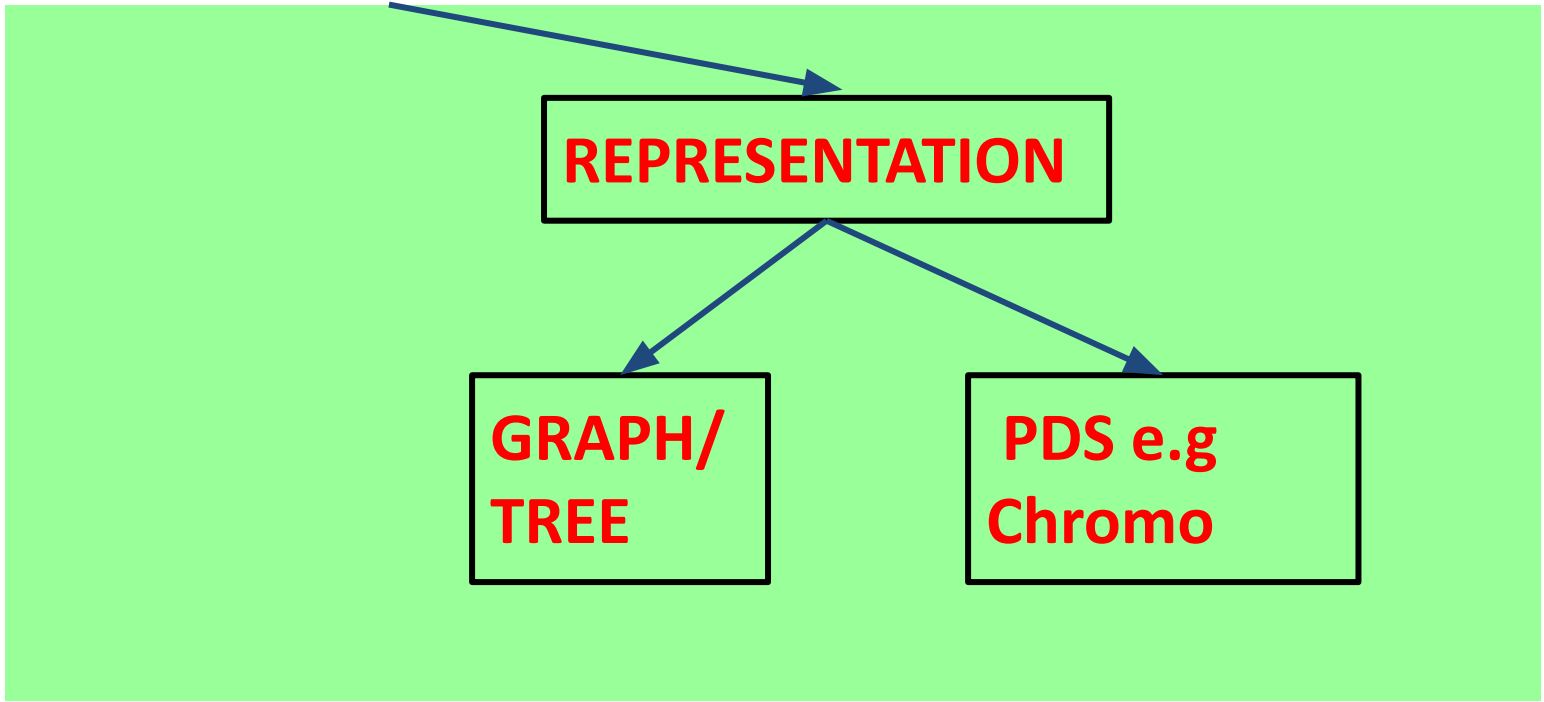


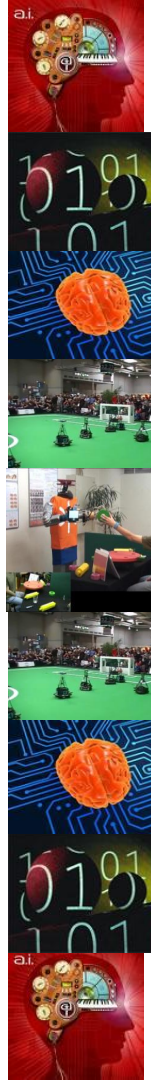
Review Search and Planning



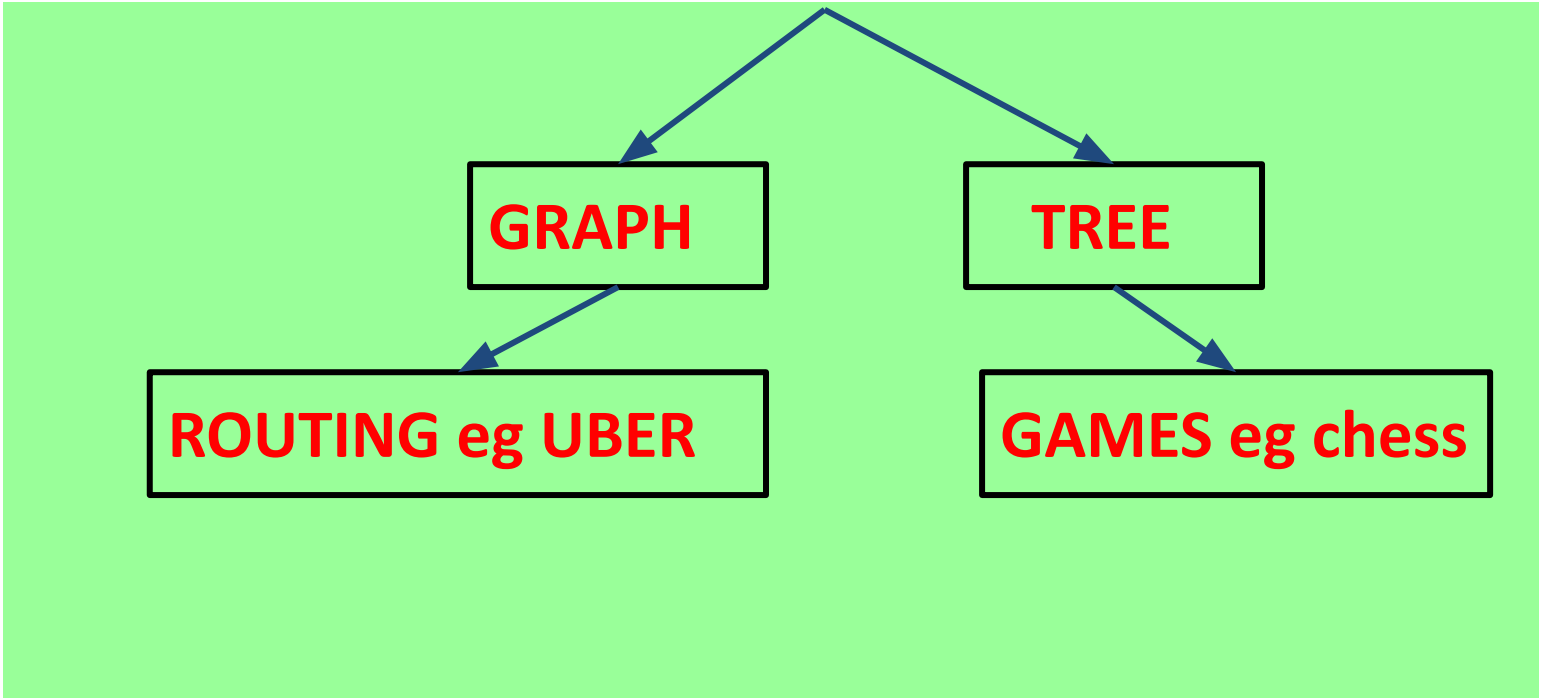


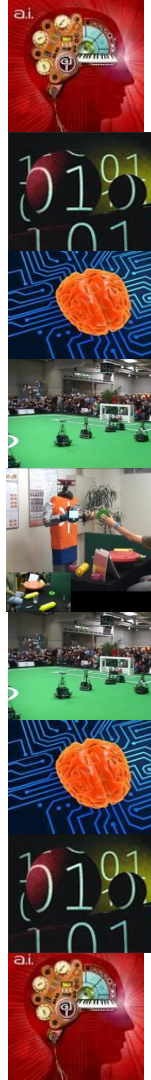
Review Search and Planning





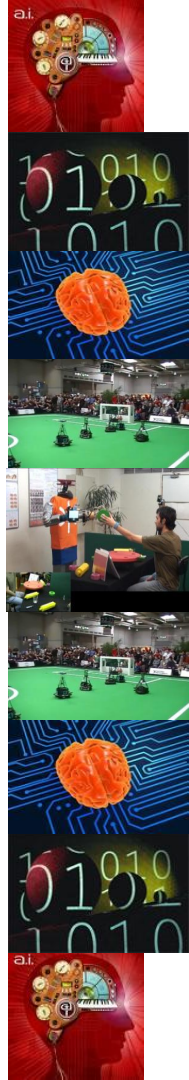
Review Search and Planning





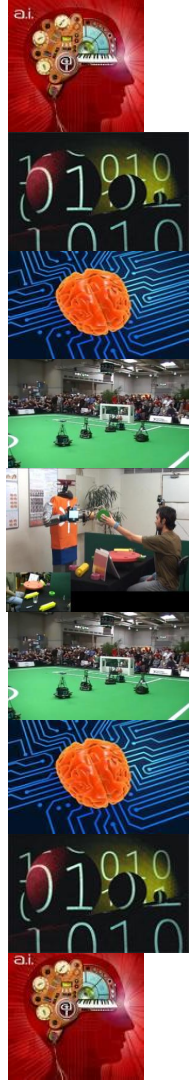
Uninformed Searches

- Definition (description)?
  - do not use additional information to guide the search process.
  - Ignore the cost of finding a solution.
- Examples.
  - Breadth first search, depth first search, DFS-ID
- Vulnerability
  - Computational costs ??



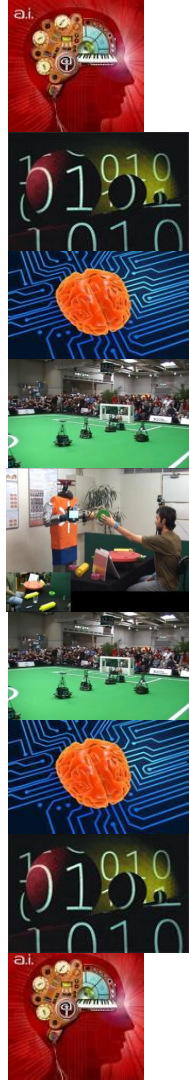
Informed Searches

- Definition (description)?
  - Use additional information to guide the search process.
  - Heuristics function used.
- Examples.
  - Best first search, A*search.
- Vulnerability
  - Although quicker than uninformed -local optima.



Metaheuristics- Single Point

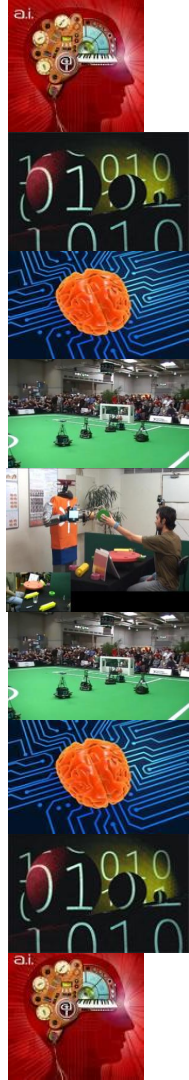
- Search single point
  - Tabu search
  - Simulated Annealing
- Key functionality
  - Tabu
    - taboo list
  - Simulated Annealing
    - temperature parameter
  - slow
- Real-word application
  - Search , routing, packing, scheduling



Metaheuristic - Multi-Point

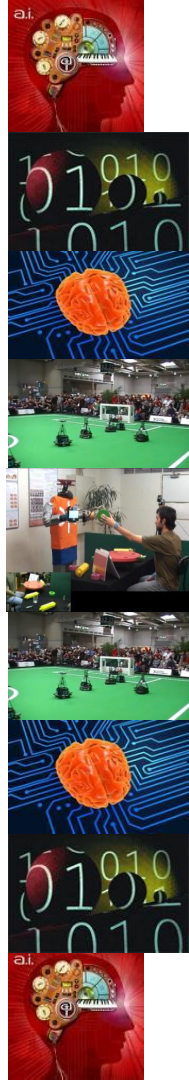
- Population - Evolutionary Computation
  - Evolutionary Algorithms
  - Swarm Intelligence.

- Principles
  - EA- Darwin's principle of natural selection.
    - Population , survival of the fittest.
  - SI- Swarming principle .
    - Population , collective advantage



Evolutionary Algorithms

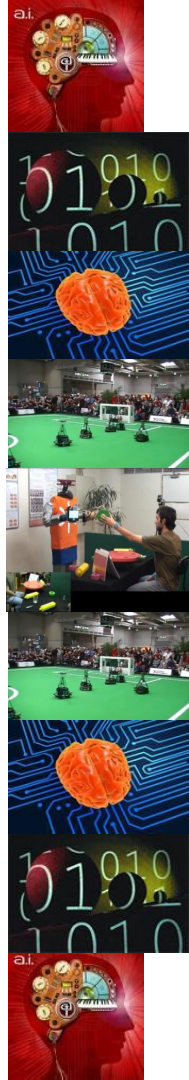
- Genetic Algorithm
  - Representation - fixed length chromosomes , genes
  - Search space
    - Solution space
- Genetic Programming
  - Representation - syntax trees
  - Search space
    - Program space
- Grammatical Evolution
  - Representation - variable length chromosomes
  - Search space
    - Program space



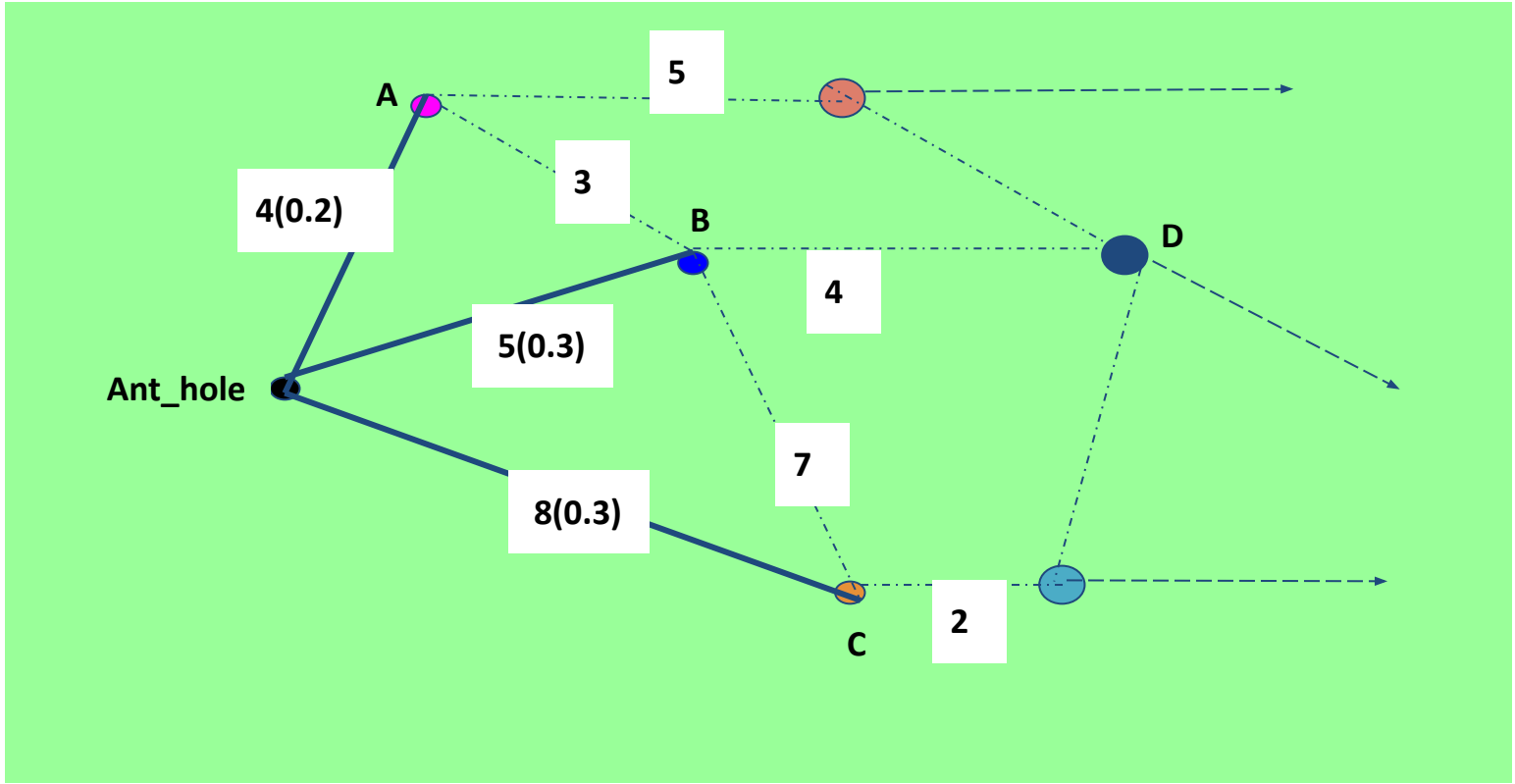
Swarm Intelligence

- Particle Swarm Optimization
  - Representation
    - Particles with velocity and direction
  - Search space
    - Solution/program space

- Ant Colony Optimization
  - Representation
    - Artificial Ants, (pheromone & heuristic values)
  - Search space
    - Solution/program



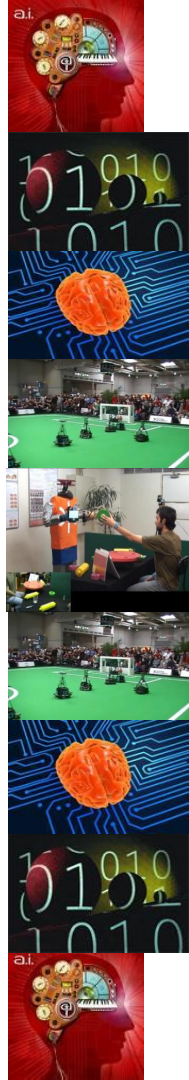
Ant Colony Optimization



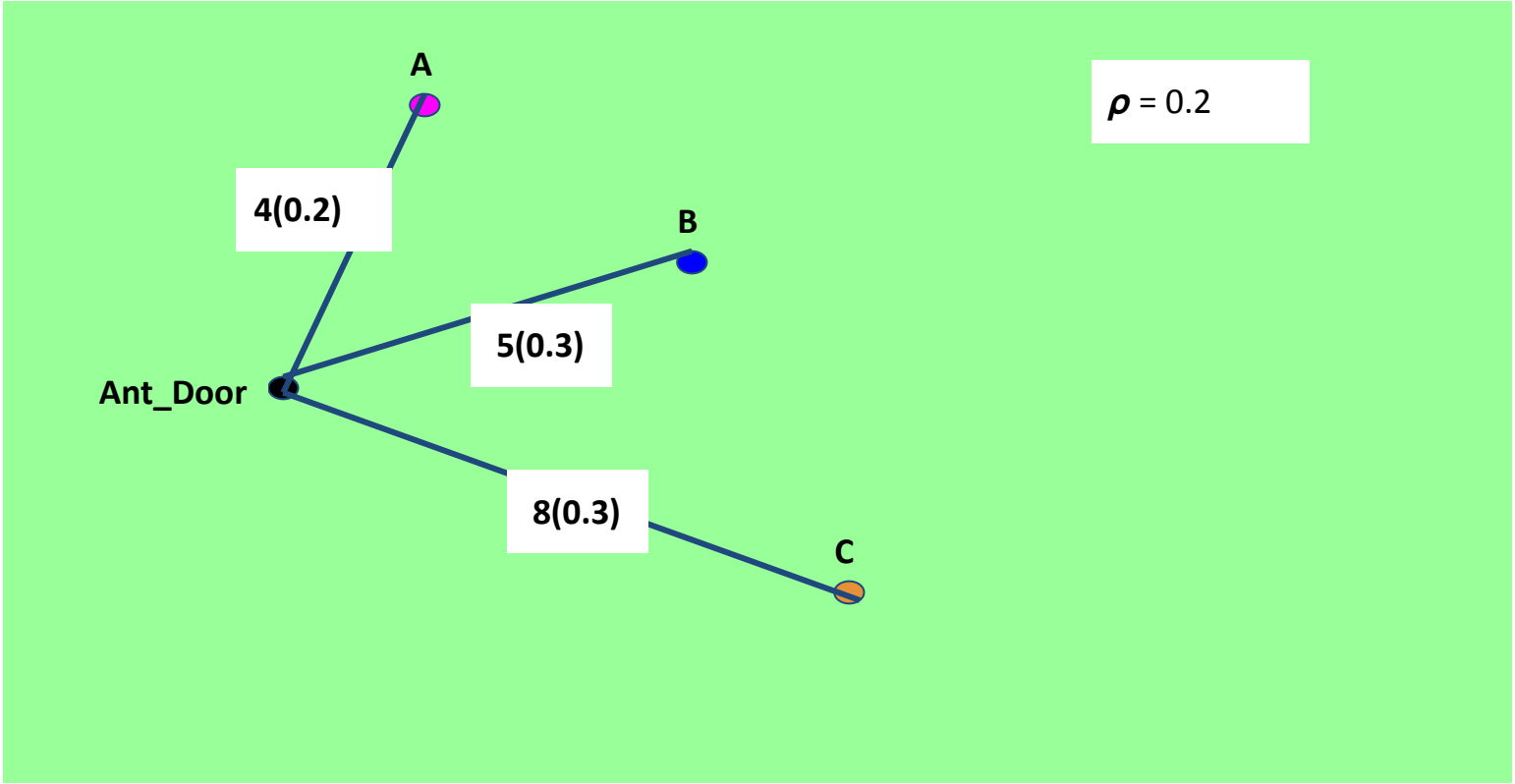
ACO Equation Updates

$$p(c_{ij}|s^p) = \frac{\tau_{ij}^{\alpha} * \eta_{ij}^{\beta}}{\sum_{c_{il} \in N(s^p)} \tau_{il}^{\alpha} * \eta_{il}^{\beta}}, \forall c_{il} \in N(s^p) \quad (1)$$

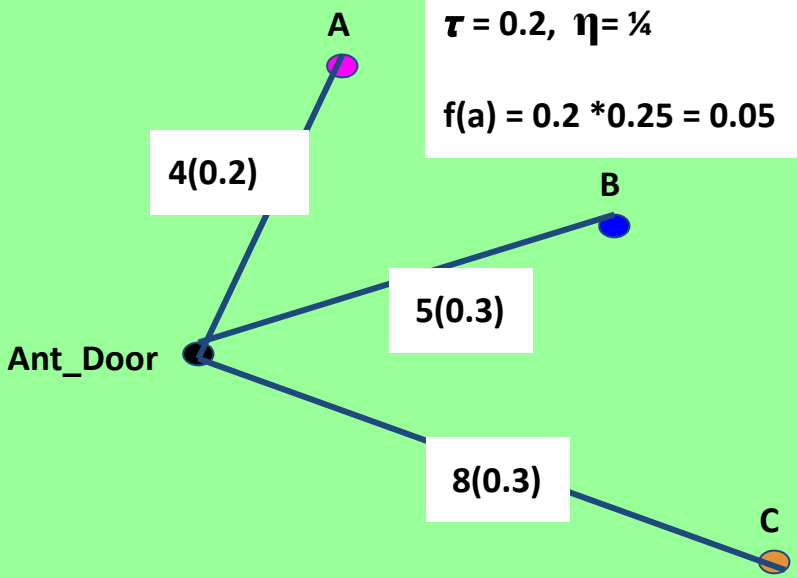
$$\tau_{ij} = (1 - \rho)\tau_{ij} + \rho \sum_{s|c_{ij}} F(s) \quad (2)$$



Ant Colony Optimization



Ant Colony Optimization



$\tau = 0.2,\ \eta = \frac{1}{4}$

f(a) = 0.2 *0.25 = 0.05

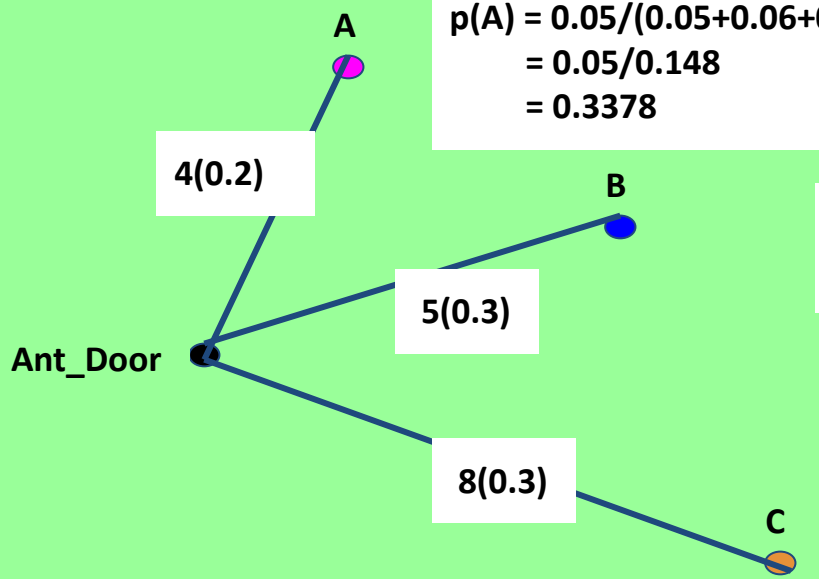
$\tau = 0.3,\ \eta = \frac{1}{8}$

f(b) = 0.3 *0.2 = 0.06

$\tau = 0.3,\ \eta = \frac{1}{8}$

f(c) = 0.3 *0.125 = 0.038

Ant Colony Optimization

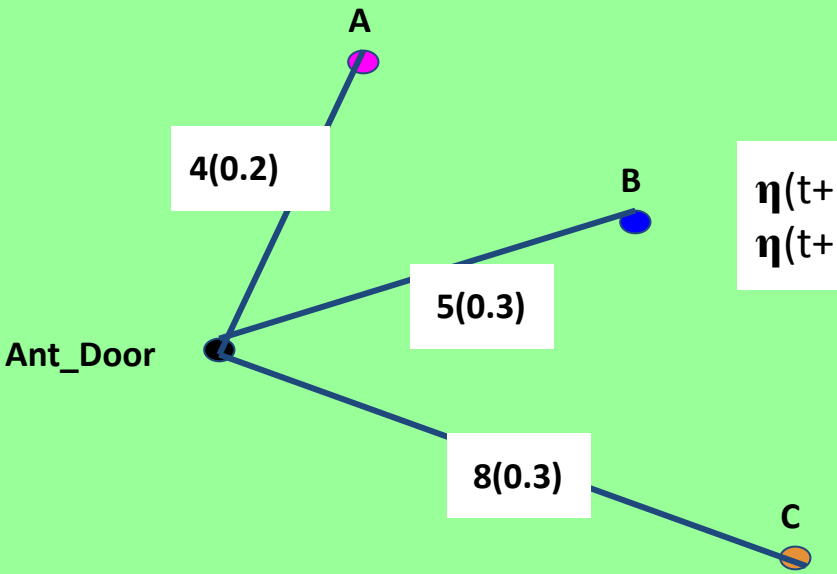


p(A) = 0.05/(0.05+0.06+0.038)
= 0.05/0.148
= 0.3378

p(b) = 0.06/0.148
=0.4054

p(c) = 0.038/0.148
= 0.2568

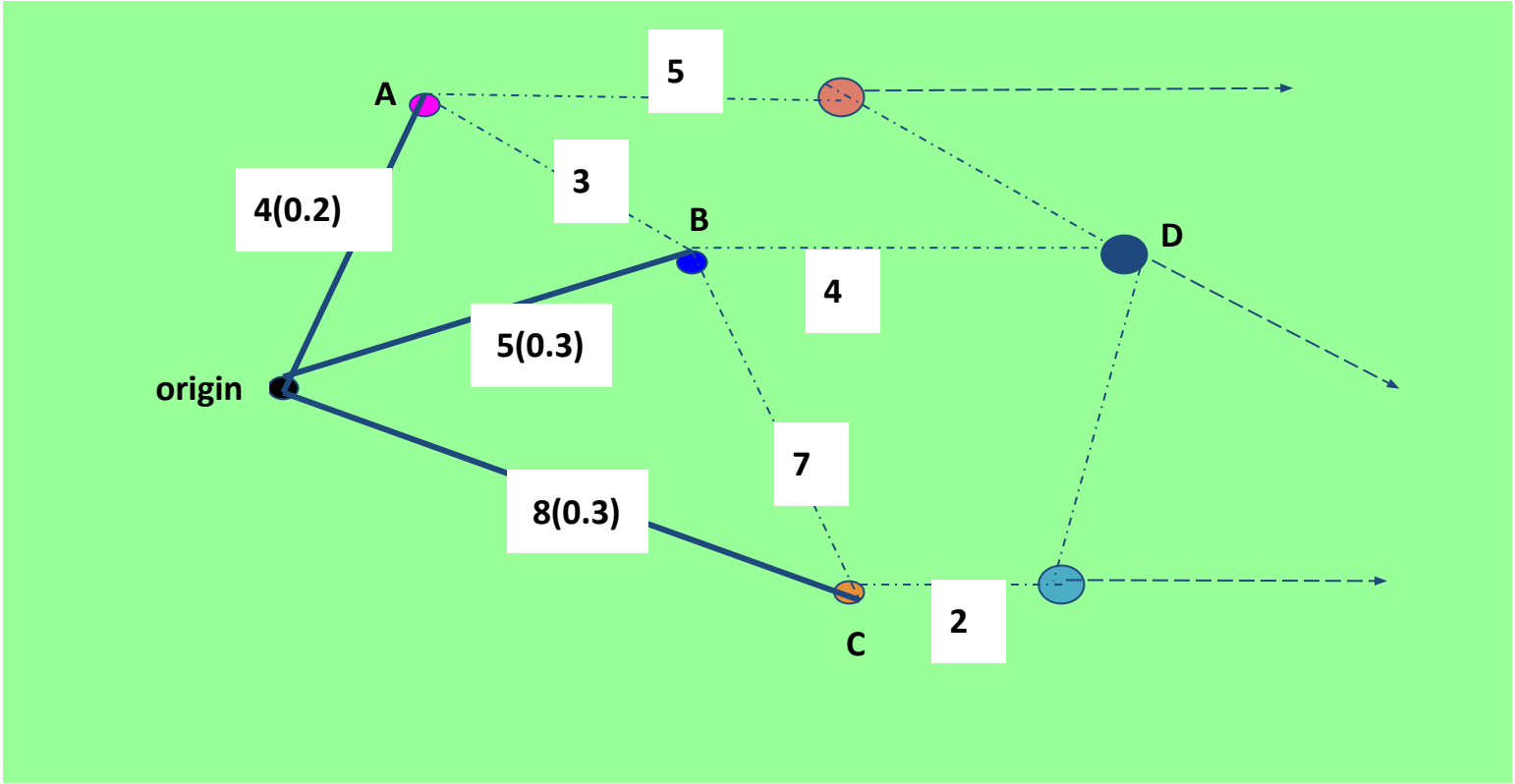
ACO- Update



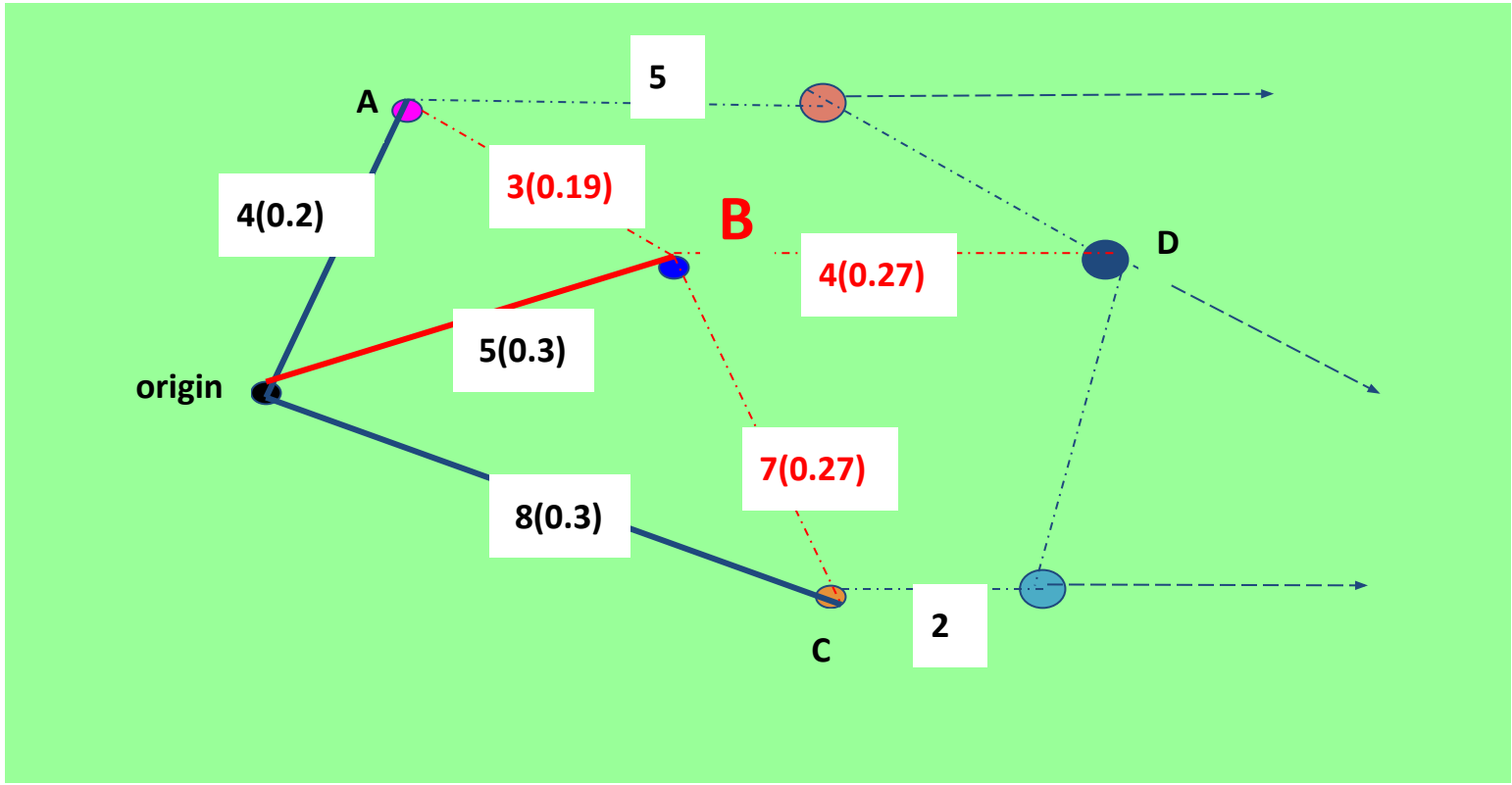
$\eta(t+1) = (1-\rho)*\eta + \rho*\sum F(s)$

$\eta(t+1) = (1-0.2)*\ 0.3 + 0.2*0.148 = 0.27$

ACO @ t=0

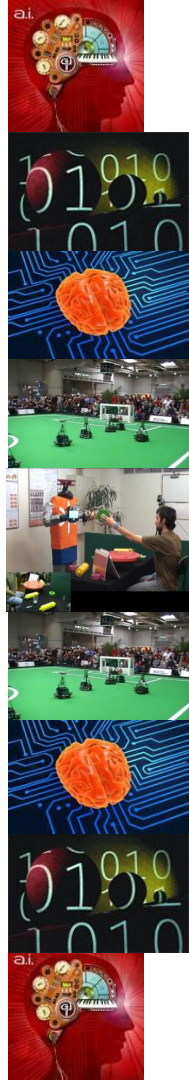


ACO @ t+1

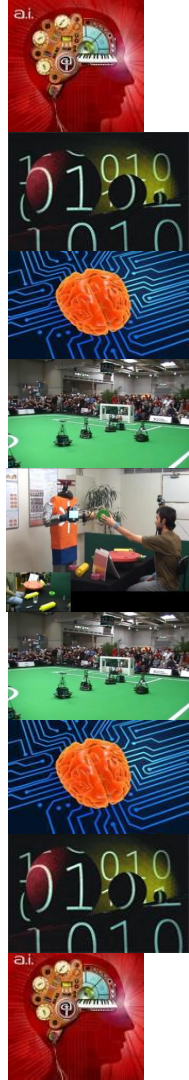


Adaptation & Hybridisation

- Adaptive Algorithms.
  - Make changes during the search.
- Hybridisation.
  - Two or more algorithms are combined.



Introduction Machine Learning

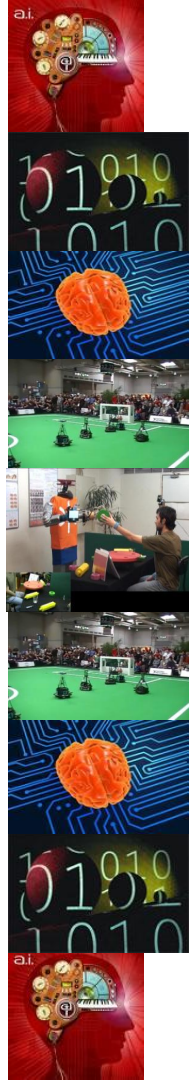


Machine Learning Outline

Define Machine Learning

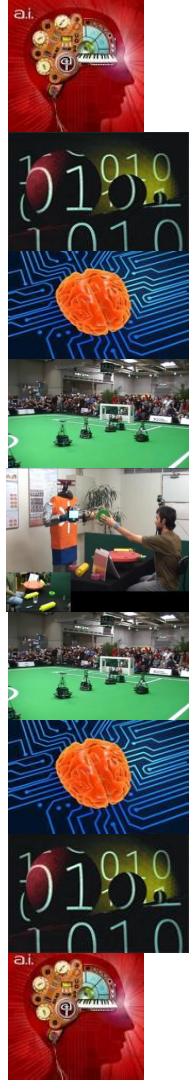
Machine Learning Data.

Overview of Machine Learning Techniques.



Introduction

- What is machine learning?
  - Extract patterns from data.
  - Generalize.
- Types of machine learning
  - Supervised learning
  - Unsupervised learning
  - Reinforcement learning

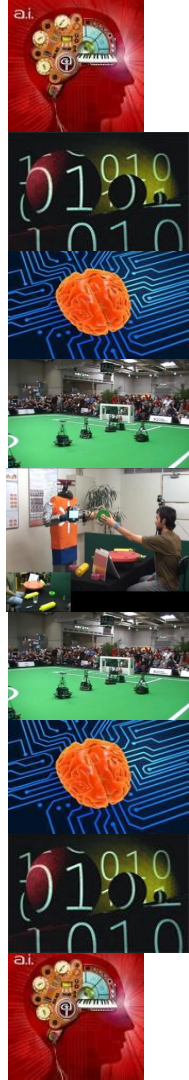


Machine Learning

The desire is to get machines to learn through examples.

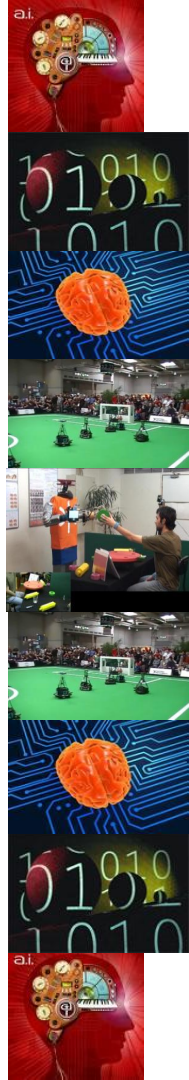
This is how humans learn.

Given enough examples a machine should be able to learn on its own.



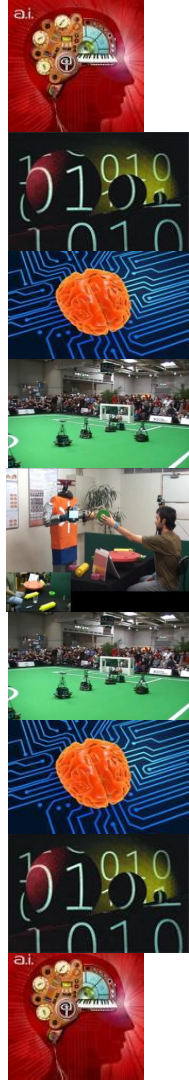
Data Formulation

- Data for machine learning comes from many sources.
- For machine learning we have a universe of objects.
- An object has a number of variables that describe its **properties**.
- In ML variables are often referred to as **attributes**.



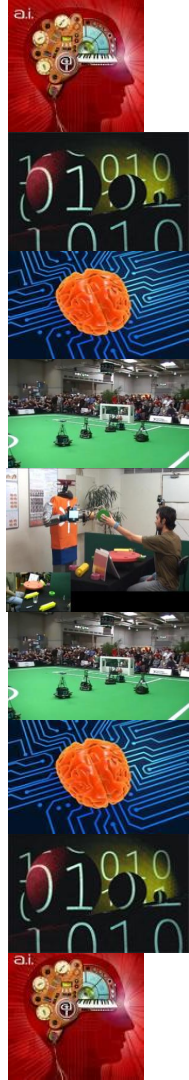
Data Formulation

- The set of attribute values corresponding to each of the objects is called an **instance**.
- The complete set of data available to us for an application is called a **dataset.**
- The dataset is **labelled** if there is one attribute given special significance and the aim is to predict its value.



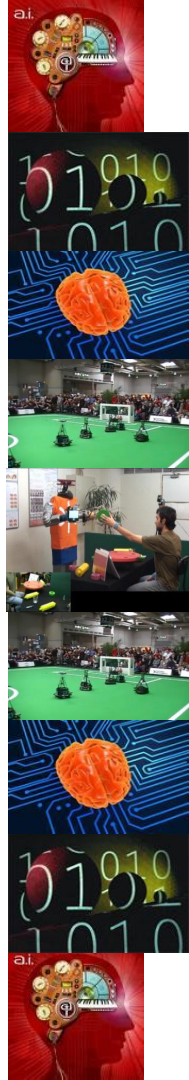
Data Formulation

- The significant variable is commonly referred to as the **class**.
- When there is no such significant attribute we call the dataset **unlabelled**.



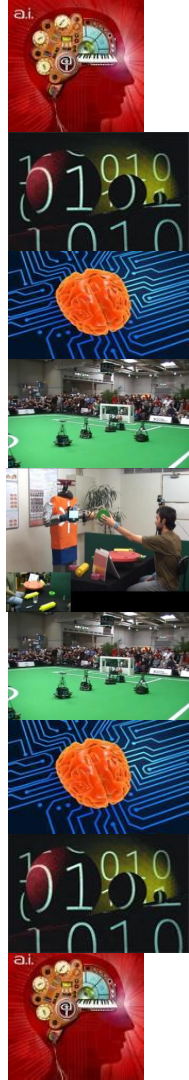
Types of Variables

- **Nominal** variables are used to put objects into categories, e.g. the name or colour of an object.
- **Binary** variable is a special case of a nominal variable that takes only two possible values: true or false, 1 or 0 etc.
- **Ordinal** variables are similar to nominal variables, except that they have values that can be arranged in a meaningful order, e.g. small, medium, large.



Types of Variables

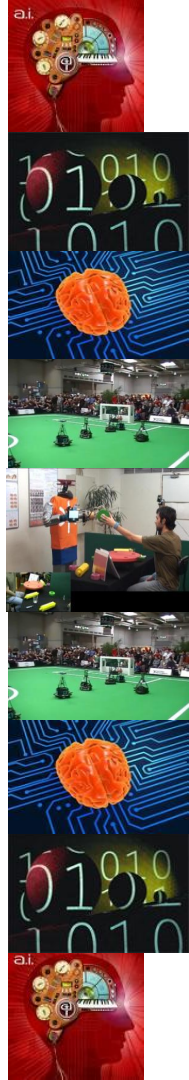
- **Integer** variables are ones that take values that are genuine integers, for example ‘number of children’.
- **Interval-scaled** variables are variables that take numerical values which are measured at equal intervals from a zero point or origin e.g temperature.
- **Ratio-scaled** variables are similar to interval-scaled variables except that the zero point does reflect the absence of the measured characteristic.



Classification of Variables

- **Categorical** corresponding to nominal, binary and ordinal variables.

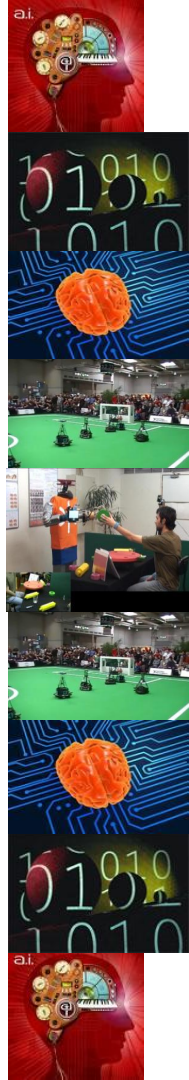
- **Continuous** corresponding to integer, interval-scaled and ratio-scaled variables.



Pre-processing

- The hardest task may be to get the data into a standard form in which it can be analysed.

- In real-world datasets erroneous values can be recorded for a variety of reasons, including measurement errors, subjective judgements and malfunctioning or misuse of automatic recording equipment.

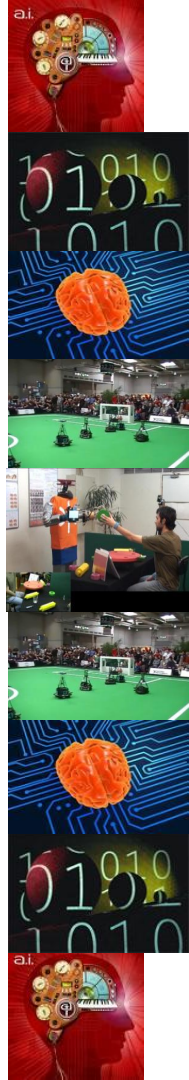


Pre-processing

Missing Values

- Discard Instances - delete all instances where there is at least one missing value and use the remainder.
  (advantage ?, disadvantage ?)

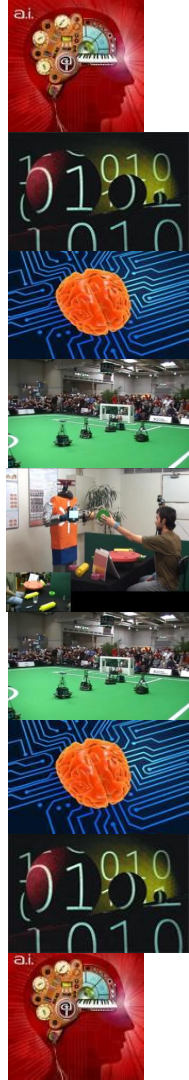
- Replace by -
  - most frequent.
  - average value.



Pre-processing

Reducing the Number of Attributes

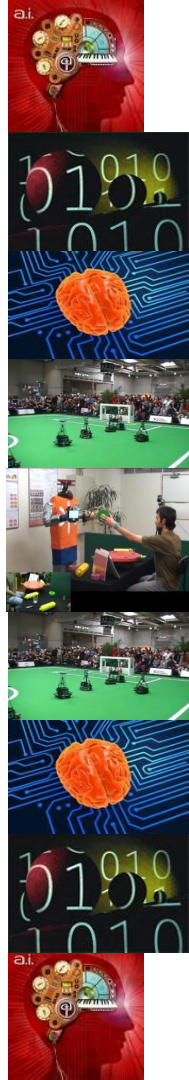
- For some datasets there can be substantially more attributes than there are instances.
- The term **feature reduction** or **dimension reduction** is generally used for this process.
- Eg Dob and age



Normalisation

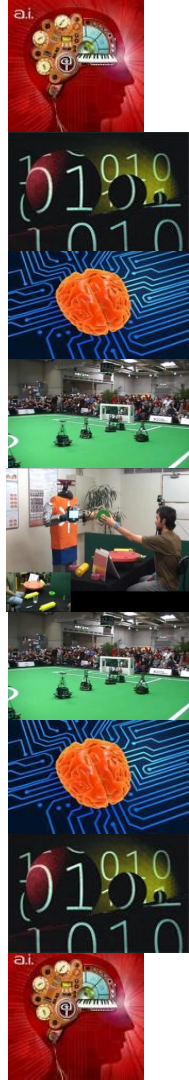
Min-max scaling = $$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- Is the process of transforming data values in a dataset to the same scale.
- Normalisation is the process of scaling data to within a certain range usually 0-1.
- Why normalise - when a feature exhibits high numeric variation in its values.



Supervised Learning

- Machine learning using labelled data is known as **supervised learning**.
- If the designated attribute is categorical, the task is called **classification**.
- If the designated attribute is numerical, the task is called the task is called **regression**.



Supervised Learning

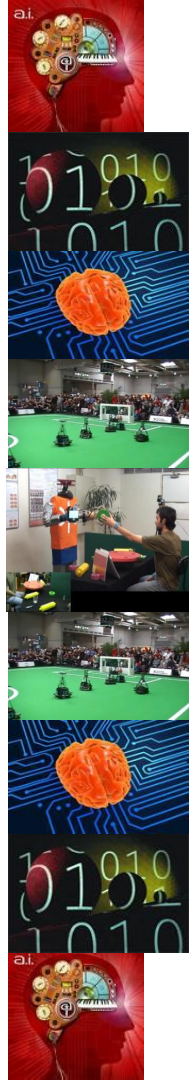
Data is divided into

- -Training
- -Validation
- -Testing

Split - (60%, 10% ,30%) OR (70% 30%)

Importance of appropriate data

- -K-fold cross-validation

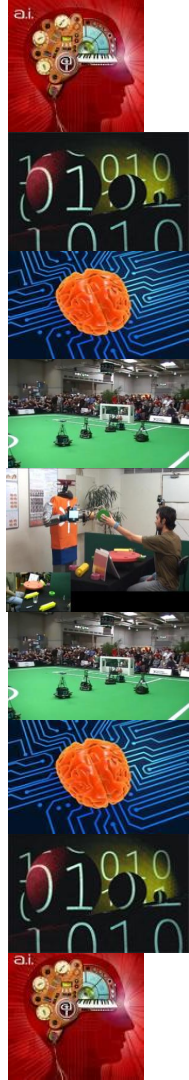


Unsupervised Learning

In unsupervised learning the data provided is not labelled, e.g. the attributes are provided in an instance but not the class.

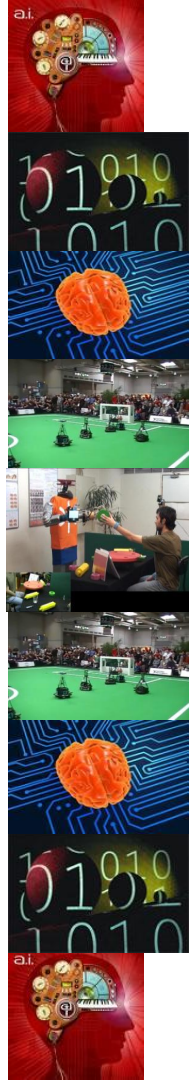
- Grouping algorithms are used.

- Clustering is commonly used.

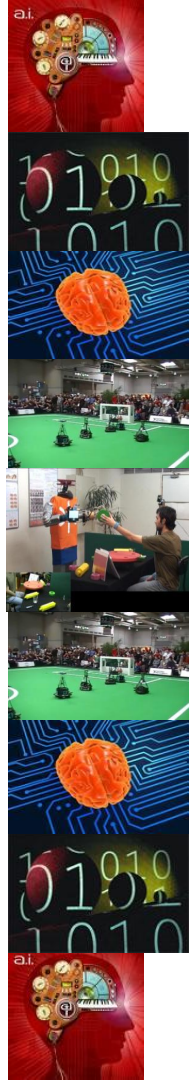


Reinforcement Learning

- Learning takes place as the problem is solved.
- Involves a set of agents solving a problem over a number of time steps in a particular environment.
- Actions taken by agents in the environment.
- Assessment and reward/punishment.
- Effect of the reward in the next time step.

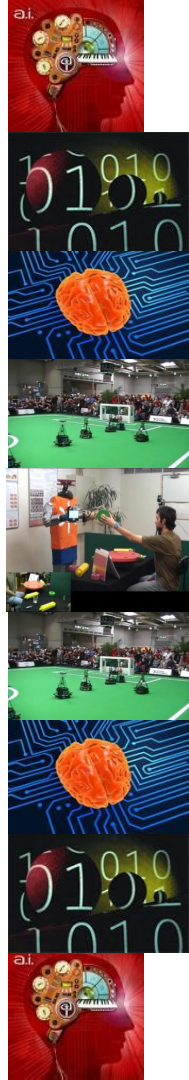


Machine Learning Techniques



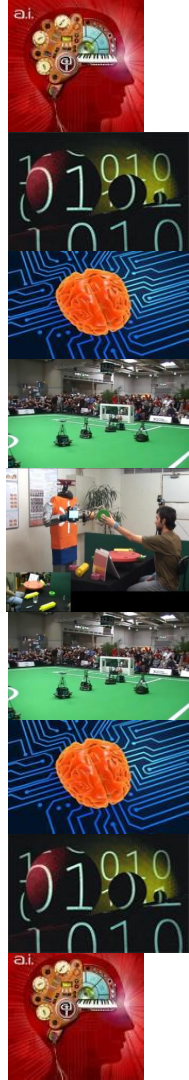
ML Techniques

- K-Nearest neighbour.
- Decision Trees.
- Regression.
- Support Vector Machines
- Artificial Neural Networks.



Nearest Neighbour

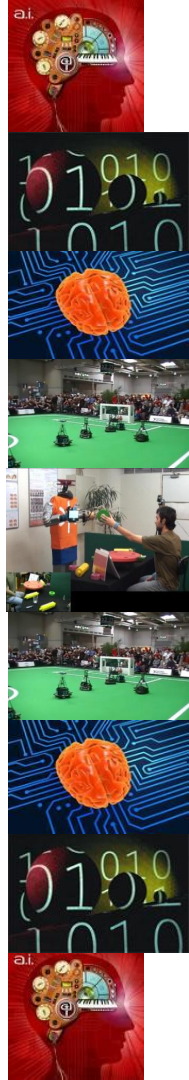
- Assigns the output for a new instance depending on the output of its neighbours.
- K-Nearest neighbour.
- Determining the nearest neighbours.
- Majority voting.



Nearest Neighbour

a	b	c	d	e	f	Class
yes	no	no	6.4	8.3	low	negative
yes	yes	yes	18.2	4.7	high	positive
yes	no	no	6.6	8.0	low	????

- We can predict its classification using that of the first instance, i.e. as ‘negative’.
- It is usual to base the classification on those of the **k** (3 or 5) nearest neighbours not just one.

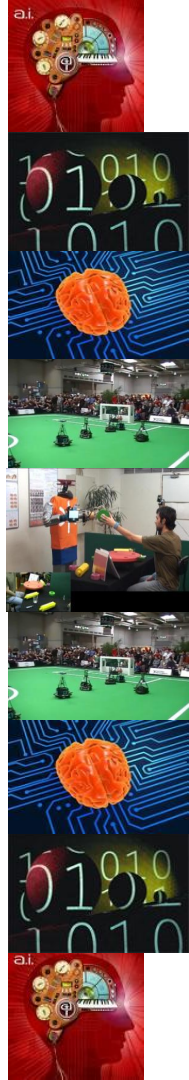


K-Nearest Neighbour

Distance Metrics

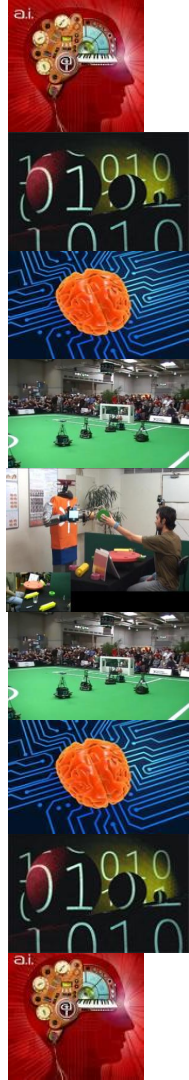
1. Euclidean distance- length of a line segment between 2 pts.

2. Manhattan distance e.g (12,9) and (4,2)
   distance = (12-4) + (9-2) = 15.

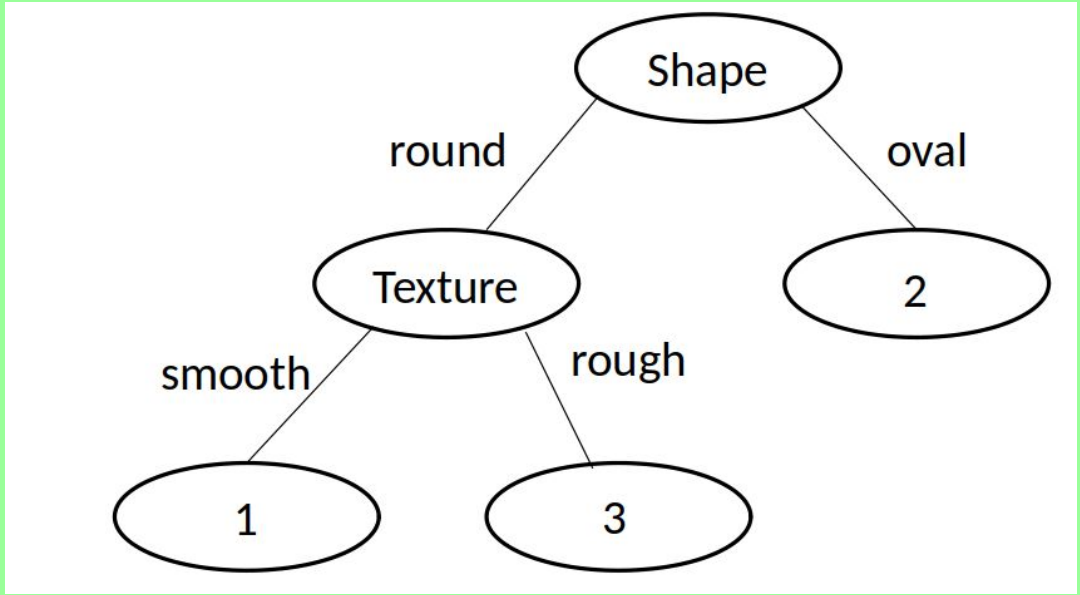


Decision Tree

- The decision tree is the classifier.
- The internal nodes of the tree represent problem attributes.
- The leaf nodes represent classes
- Performs binary and multiclass classification
- Derived by induction algorithms such as ID3 and C4.5

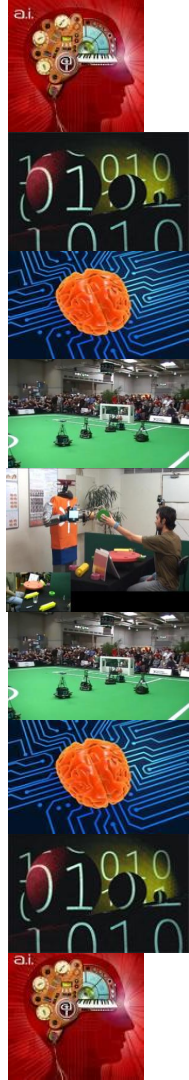


Decision Tree



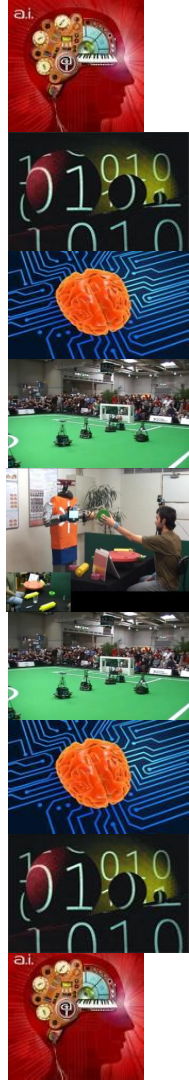
Regression

- Finds patterns in data.
- Finds the relationship between a set of independent variables and a dependent variable.
- Linear vs. logistic regression.
- Evaluation - Root mean square error.

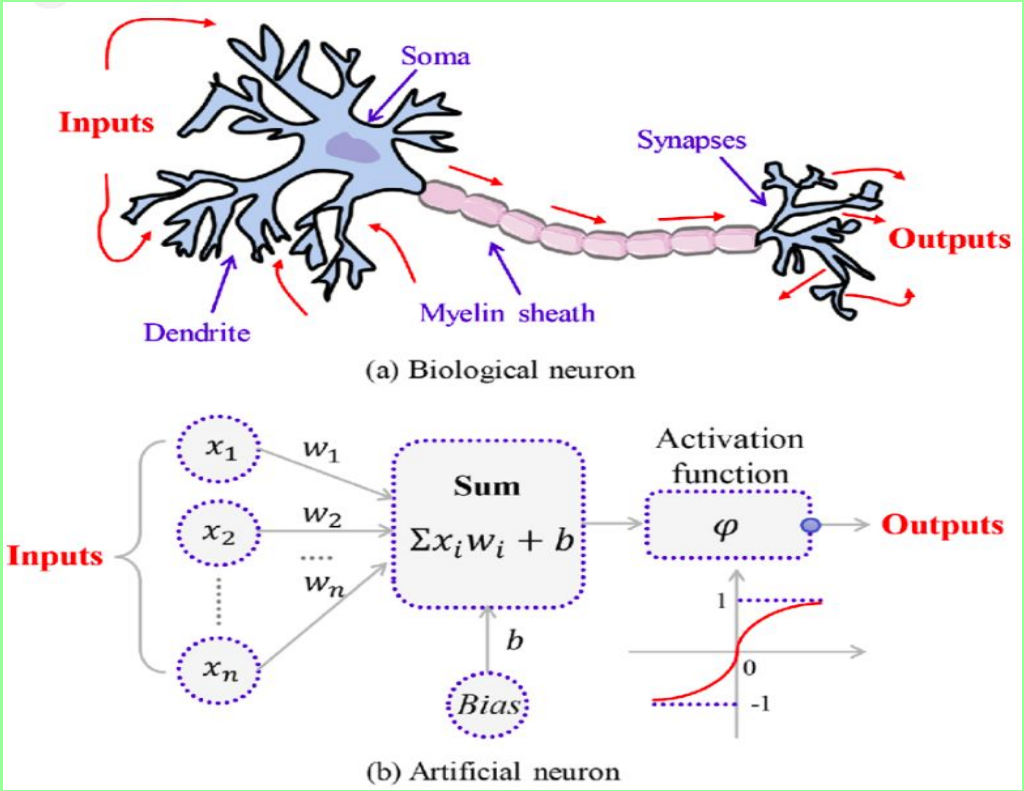


Support Vector Machines

- Build classification and regression models from data
- Linear vs. nonlinear classification
- Unsupervised learning

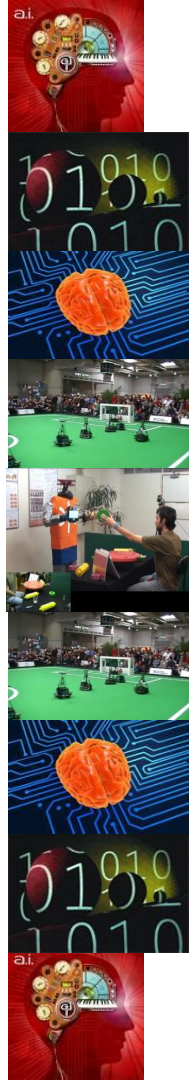


Biological Neuron vs Artificial Neuron



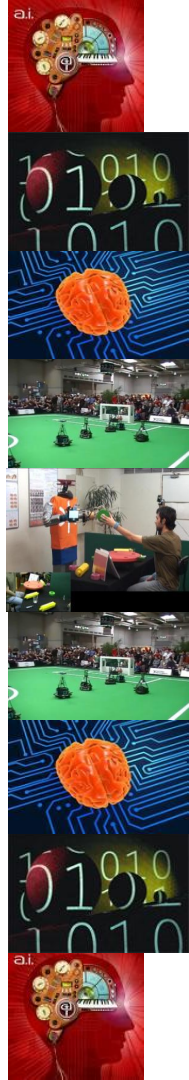
Neural Network

- Mathematical models good at pattern association and classification
- Single layer neural network
- Multilayer neural network
- Weights and learning algorithm
- Activations functions
- Convolutional neural networks and deep learning



Ensemble Learning

- An ensemble performs the classification or regression
- Each ensemble is a set of classifiers or regression models
- The output from each element of the ensemble is combined to produce a single output
  - Random forests.
  - Boosting - weak classifiers.



Next Lecture

Artificial Neural Networks

QUESTIONS