Applying AI, ML and DS in Engineering Systems, and Cyber Security

Prof. K. R. Chowdhary,
Former prof. & HOD CS Dept., M B M
Email: kr.chowdhary@iitj.ac.in
http://krchowdhary.com

Friday 8th March, 2024

A scientific field is best defined by the **central question** it studies. The field of Machine Learning seeks to answer the question,

"1. How to build computer Programs that automatically improve their performance with experience?, and

2. What are the fundamental laws that govern all learning processes?"

Artificial Intelligence vs Machine Learning

- Artificial intelligence is a technology which enables a machine to simulate human behavior.
- Machine learning is a subset of AI which allows a machine to automatically learn from past data without programming explicitly.
- The goal of AI is to make a smart computer system like humans to solve complex problems

Artificial Intelligence...

The Artificial intelligence system does not require to be pre-programmed, instead, they use such algorithms which can work with their own intelligence. It involves machine learning algorithms such as Reinforcement learning and deep learning. Based on capabilities, AI can be classified into three types:

- Weak AI
- General AI
- Strong AI

Currently, we are working with weak AI and general AI. The future of AI is Strong AI for which it is said that it will be intelligent than humans

Machine Learning ...

- Machine learning enables machines to learn from past data or experiences without being explicitly programmed.
- Machine learning enables a computer system to make predictions or take some decisions without being explicitly programmed.
- Machine learning uses a massive amount of structured and semi-structured data so that a machine learning model can generate accurate result or give predictions based on that data.
- It can be divided into three types:
  - Supervised learning
  - Reinforcement learning
  - Unsupervised learning

Some examples

- How to design autonomous mobile robots that learn to navigate from their own experience,
- How to mine the data of historical medical records to learn which future patients will respond best to which treatments,
- How to build search engines that automatically customize to their user's interests.

Definition

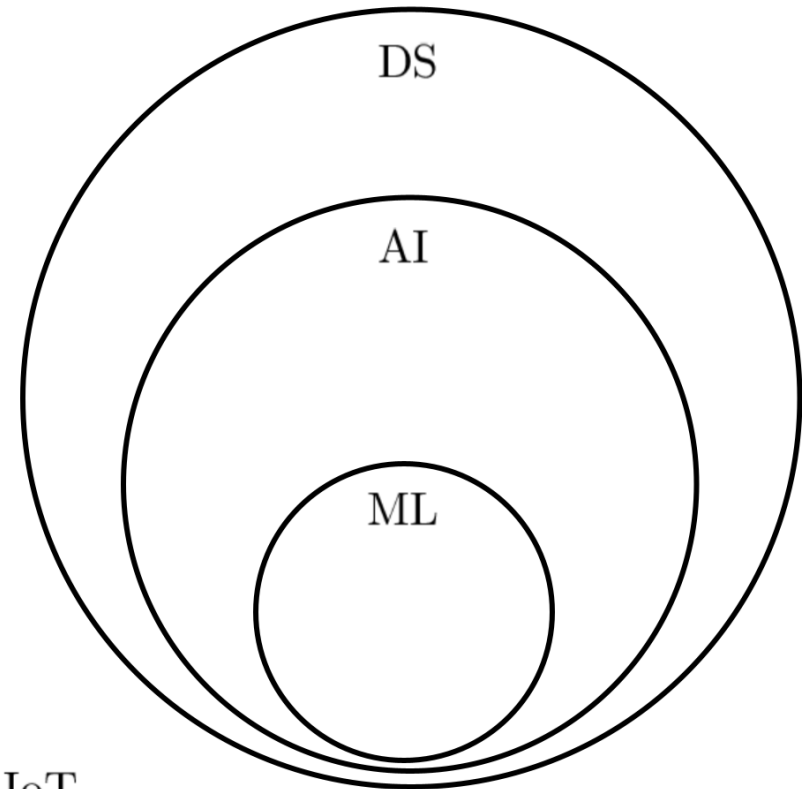
A machine learns with respect to a particular task T, performance metric P, and type of experience E, if the system improves its performance P at task T, following experience E.

AI, ML, DS and Cyber Security

DS:
Collection, preparation
and analysis of data

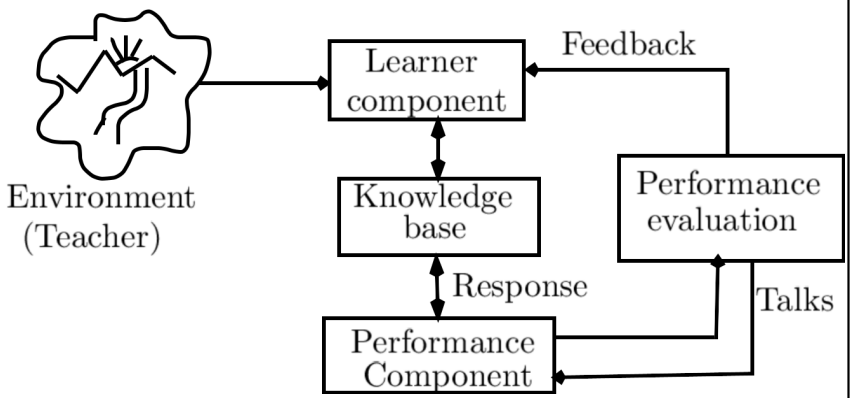
AI:
Technology for machines
to understand / interpret,
learn, and make
intelligent decisions

ML:
Algorithms of reinforcement,
Supervised and unsupervised
learning



Cyber security:
Cryptography, Cloud Security, IoT
Networks Security, AI in Cyber Security
and Information Security, Risk Management.

A Learning Model



Basic Learning categories: 1. Supervised, 2. Unsupervised, and 3. Rein-forced learning.

1: Teaching in class, dividing students as indoor/outdoor players, recognizing a person. 2: Self study, classifying students into unknown no. of groups, discovering a new idea. 3: Doing Study to get a job, studying due to appreciation in class, a sixer shot due to applause of audience in cricket.

Reinforcement Learning

Our learning is through feedback of our actions in the real-world. Human beings always learn by interactions with the environment/teacher, which are **cause** and **effect** relations.

The environment or the world around us is teacher, but its lessons are often difficult to detect or grasp by the mind or analyze, hence hard to learn.

The best example is learning by a dog or a young child, where **good actions are rewarded and bad actions are discouraged**.

The programs that improve their performance at some task by rewards and punishments from the environment, are example of **Reinforcement Learning** (RL).

For many automatic learning procedures for real-world tasks, like, job-shop scheduling and elevator scheduling,

Reinforcement Learning

There are four components of RL:

1. a policy,
2. a **reward function**,
3. a **value mapping**, and
4. a model of environment.

An agent recognizes itself in some state $p \in S$, takes some action $a \in A$, and then recognizes it self in a new state q.

$$T(S \times A) \rightarrow S. \tag{1}$$

the agent receives a reward $r \in \mathbb{R}$ for arriving to state q,

$$R(S \times A) \rightarrow \mathbb{R}. \tag{2}$$

Unsupervised vs. Supervised Learning

- The main distinction between the two approaches is the use of **labeled datasets**.
- In supervised learning, the **algorithm learns** from the training dataset by iteratively making predictions on the data and adjusting for the correct answer.
- While supervised learning models tend to be **more accurate** than unsupervised learning models, they require upfront human intervention to label the data
- Unsupervised learning models, in contrast, **work on their own to discover the inherent structure** of unlabeled data.

Unsupervised Learning

- Unsupervised learning **uses machine learning algorithms to analyze and cluster unlabeled data sets**.
  Used for three main tasks:
  - **Clustering** is a data mining technique for grouping unlabeled data based on their similarities or differences.
  - **Association** Learning method uses different rules to find relationships between variables in a given dataset.
  - **Dimensionality** reduction is a learning technique used when the number of features (or dimensions) in a given dataset is too high

Main difference: Sup. vs. Unsup: Labeled data

- In supervised learning, the algorithm **"learns" from the training dataset** by iteratively making predictions on the data and adjusting for the correct answer.
- Unsupervised learning models, in contrast, work **on their own to discover the inherent structure of unlabeled data**.

Key differences: Supervised vs. unsupervised learning

- **Goals**: In supervised learning, the goal is to predict outcomes for new data.
- **Applications**: Supervised learning models are ideal for spam detection, sentiment analysis,..
- **Complexity**: Supervised learning is a simple method for machine learning, typically calculated through the use of programs like R or Python
- **Drawbacks**: Supervised learning models can be time-consuming to train, and the labels for input and output variables require expertise

Which is best?

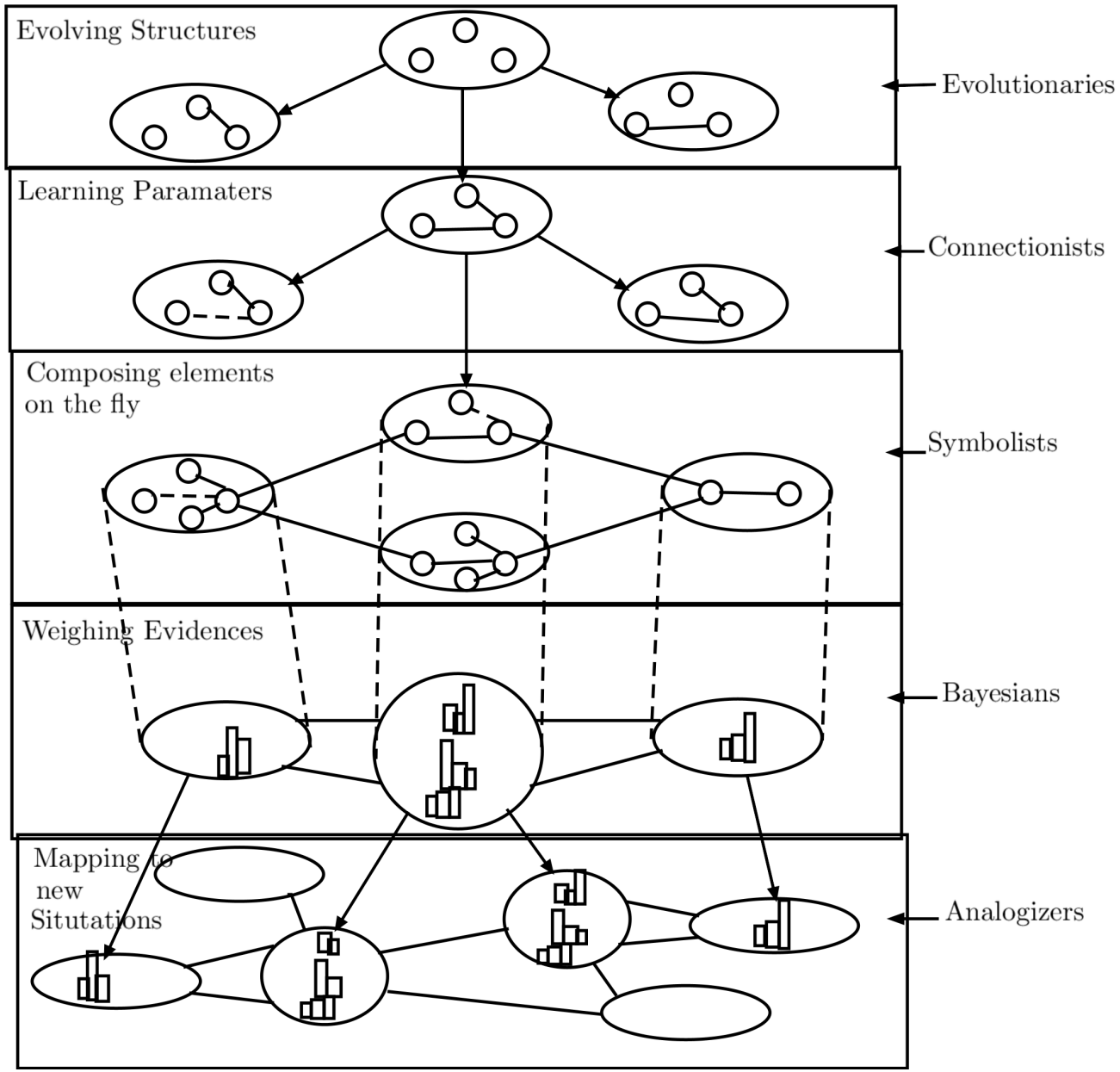
Choosing the right approach for your situation depends on how your data scientists assess the structure and volume of your data, as well as the use case. To make your decision, be sure to do the following:

- **Evaluate your input data**: Is it labeled or unlabeled data? Do you have experts that can support additional labeling?
- **Define your goals:** Do you have a recurring, well-defined problem to solve? Or will the algorithm need to predict new problems?
- **Review your options for algorithms:** Are there algorithms with the same dimensionality you need (number of features, attributes or characteristics)? Can they support your data volume and structure?
- **Semi-supervised learning: The best of both worlds**

Major Learning Algorithms

- Need to provide enough of right kind of data
- Five Tribes of Machine Learning Algorithms:
  1. Symbolists (knowledge in Symbols)
  2. Connectionists (... in connections between neurons)
  3. Evolutionaries (... in Evolution)
  4. Bayesians (... through Bayes Rule)
  5. Analogizers (... in Analogy)
- The true master learning algorithm must solve all five problems

Single Algorithm with these capabilities



1. Symbolists: Principle of Inverse Deduction

Classical example of **Inverse Deduction:**

1. Socrates is man.

2. All men are mortal. (Conclusion)

Therefore .....? (Inference)

Induction:

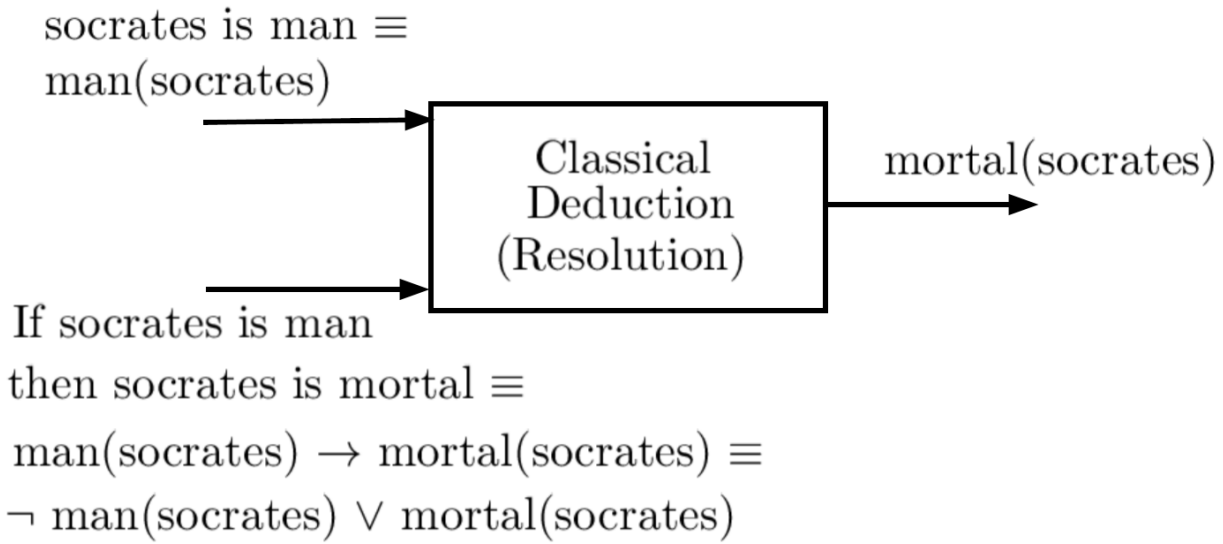
Socrates is man.

.....? (Inference)

Therefore, Socrates is mortal. (Conclusion)

Classical Deduction

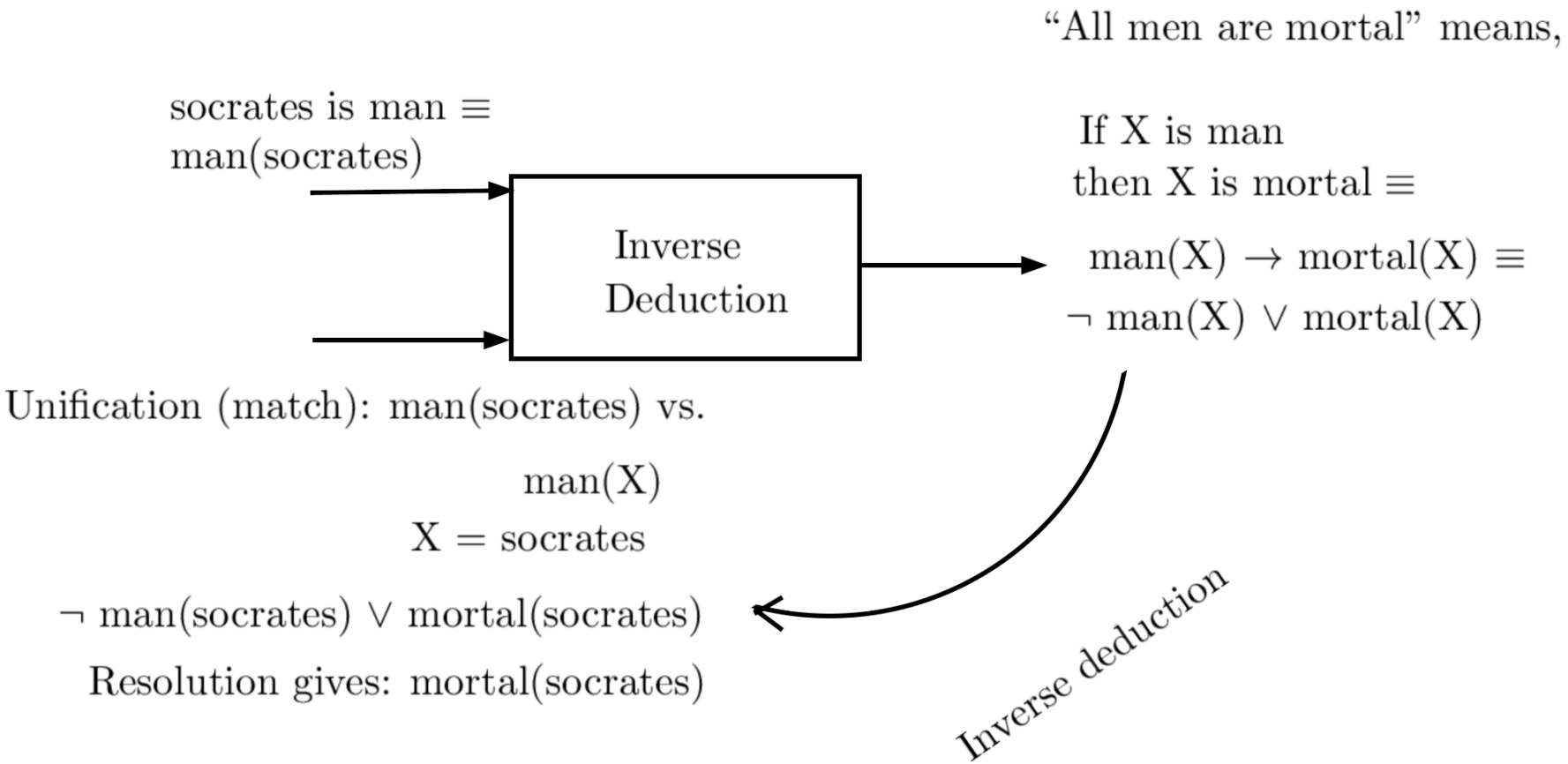
- Socrates is man
- if Socrates is man then Socrates is mortal



- There is no "All"

Inverse Deduction

- Socrates is man
- if X is man then X is mortal



Learning to cure Cancer (Symbolist)

- Inverse deduction is a great way to discover new knowledge in Biology,
- Curing the cancer means stopping the bad cells from reproducing, without harming the good cells. Given,
  1. *If the temperature is high, then gene A is expressed.*
  2. *If gene A is expressed and gene B is not, then gene C is expressed.*
  3. *If C is expressed, then gene D is not.*

  if we knew the 1st and 3rd rule, but not the 2nd, and we have micro-array data, where at high temperature, B and D are not expressed, we could induce the 2nd rule by inverse deduction.
- Disadvantage: Computation intensive.

Learning through Decision-Trees...

- An unseen example is classified starting from the root, testing attributes of internal nodes, to a leaf.
- The decision-tree states that examples belonging to different classes have different values in at least one of their attributes.
- Each internal node of a decision-tree is labeled by an attribute and links from a node are labeled by the possible values of the attribute, as shown in Figure 1.

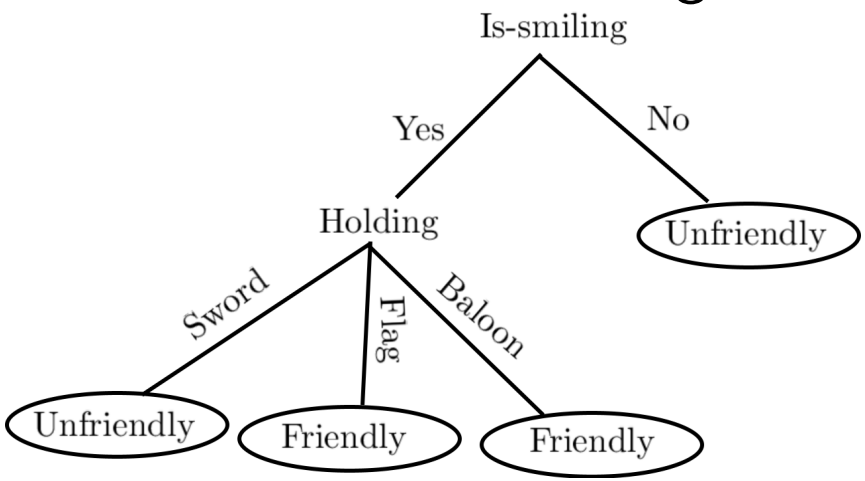


Figure 1: A decision-tree for root domain.

Learning through Decision-Trees...

- The **training database is accessed** extensively while the tree is constructed. If the training database does not fit the memory, an efficient data-access method is needed to achieve the scalability.
- The **statistics can be constructed** in memory at each node in a single scan over the corresponding database partition, that is, it **satisfies the splitting criteria**, leading to the node. Sufficient statistics data is necessary for construction of learning decision-tree.

2. Bayesians

- The answer is Bayes theorem.Sun rise causes the stars to fade and the sky to lighten,
- According to Bayes theorem, the more likely effect is given the cause, the more likely the cause is given the effect: if $P(lightening\_sky|sunrise)$ is higher than $P(fading\_stars|runrise)$,

$$P(cause|effect) = \frac{P(cause) \times P(effect|cause)}{P(effect)}$$

- A learner that uses Bayes theorem and assumes that effect are independent given the cause is called Näive Bayes classifier.

Bayesians: For speech recognistion

We can treat the acoustic input (observation) O as a sequence of individual symbols (e.g. slices of inputs $(o_1, o_2, o_3, \dots)$, say one symbol for every 10 milli-secs., and represent each slice by *frequency* or *energy* of that slice. Let us assume,

$$O = o_1, o_2, \dots o_T. \tag{3}$$

Let the sentence which might have caused this observation O, is a string of words:

$$W = w_1, w_2, \dots, w_n \tag{4}$$

The probability of occurrence of this sentence, given that acoustic observation O is evidenced, is

$$\hat{W} = argmax_{w \in \mathscr{L}} \; P(W|O) \tag{5}$$

$\mathscr{L}$ is English Language.

Bayesians: For speech recognistion

The equation (5) can be expressed as:

$$\hat{W} = argmax_{w \in \mathscr{L}} \frac{P(O|W) * P(W)}{P(O)}. \tag{6}$$

The prior probability $P(W)$ is estimated by n-gram language model. We can ignore the probability $P(O)$, because it is a common denominator for all the sentences. Hence, it reduces to simply:

$$\hat{W} = argmax_{w \in \mathscr{L}} P(O|W) * P(W) \tag{7}$$

where, W is most probable sentence, $P(W)$ is prior probability (it is called the *language model*), and $P(O|W)$ is observation likelihood.

3. Evolutionaries (Evolution): The nature's Learning Algorithm

- GA developed by Holland, are search algorithm, which search for the fittest structures, though slow in computation, but robust in nature.
- GA works based on three basic operations on population: *selection* of fittest member, *cross-over*, and *mutation*.
- The evolution searches for good structures, and neural learning fills them in

Genetic Algorithms

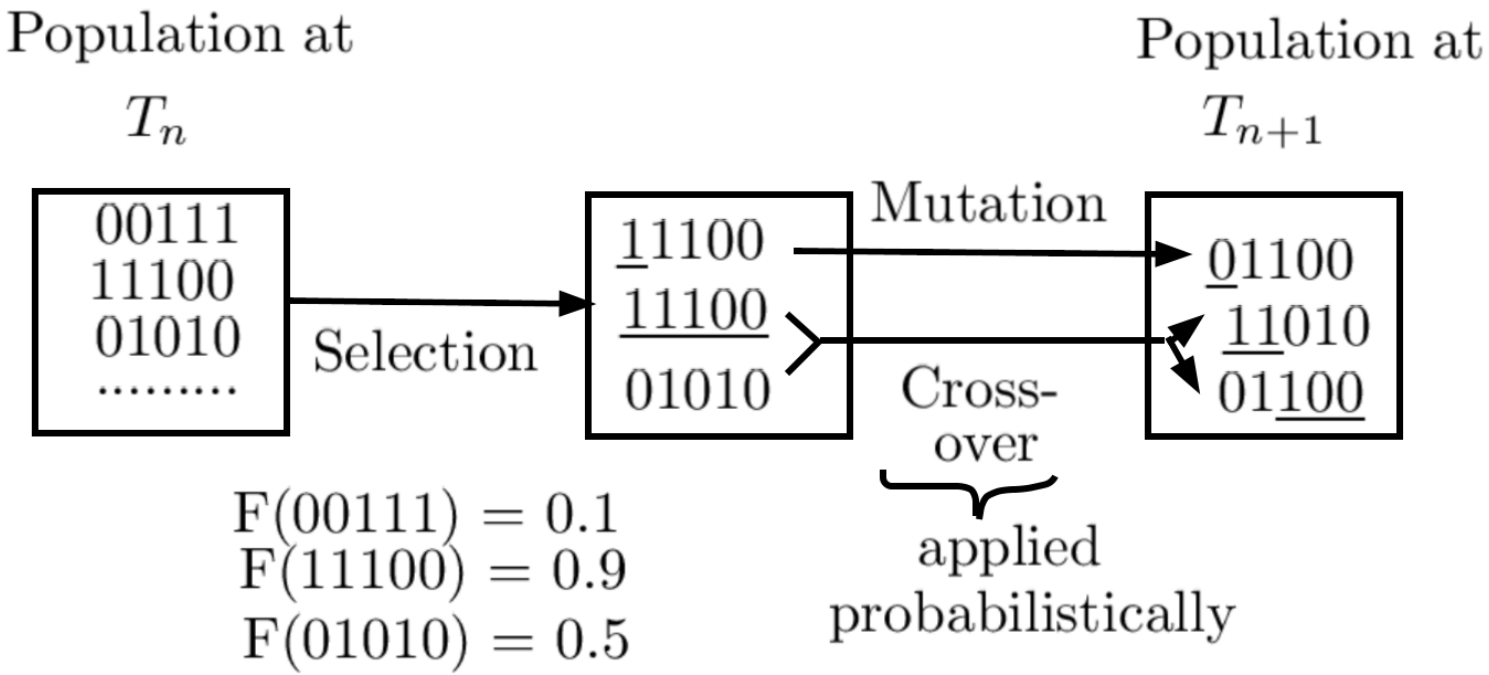


Figure 2: Sequence of operations in GA.

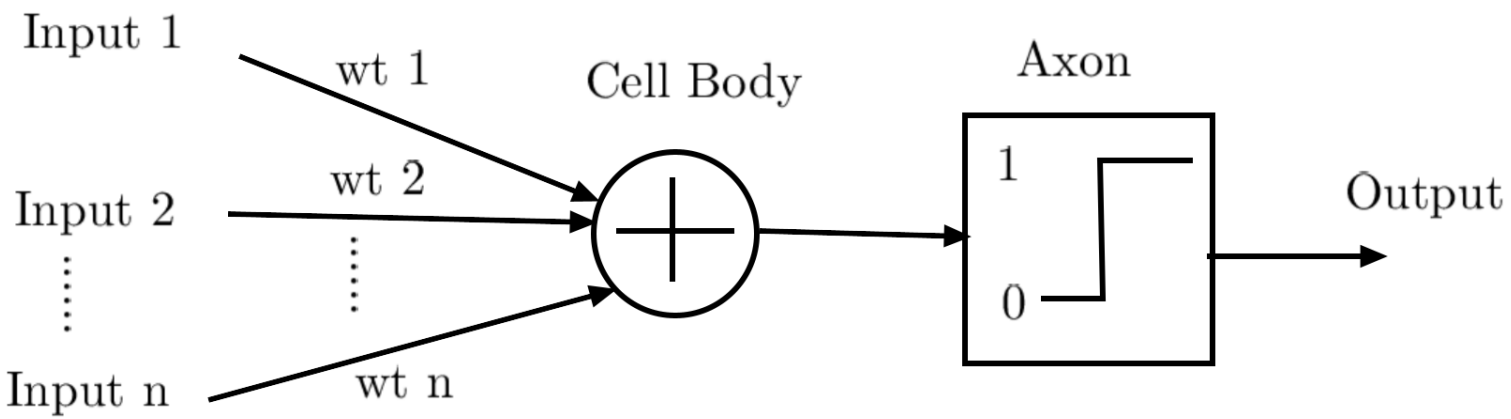
Learning through GA

Algorithm 1 Genetic-Algorithm(Input: Initial Population, fitness function, percent for mutation, selection thresh-hold)

1: **Initialize** the population with random candidate solutions
2: Apply fitness function to **Evaluate** each candidate's fitness value
3: **repeat**
4: **Select** parents based on fitness value
5: **Recombine** pairs of parents (cross-over)
6: **Mutate** resulting offspring
7: Apply fitness function to **Evaluate** new candidates' fitness value
8: **until** termination condition/goal is reached

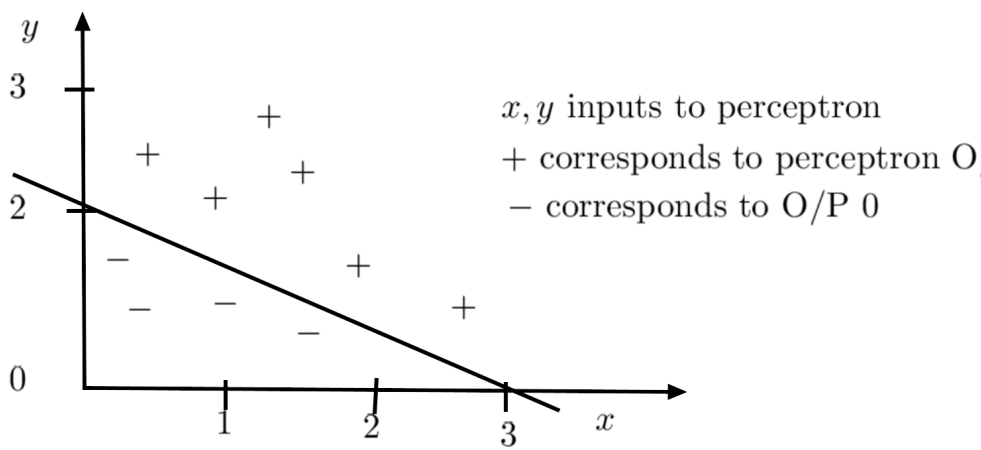
4. Neural Networks: How does our brain Learns?

- Hebb's rule is cornerstone of connectionism: The fact that knowledge is stored in connections between neurons. The neurons that fire together wire together.
- In **symbolist learning, there is one-to-one correspondence between symbols and the concepts** they represent. Connectionist representations are distributed, each concept is represented by many neurons.



Neural Networks: How does a Perceptron Learns?

- First formal model of **neurons** (1943, McCulloch and Walter Pitts), it switches on when number of its active input passes some threshold.
- When we give variable weights to the connection between neurons, it is called **perceptron** (Frank Rosenblatt,1950).
- Perceptron outputs 1 if sum of all the input is above the threshold, and 0 if below. By varying weights and threshold, we can change the function it computes.
- **In n dimensions**, there are n inputs and and perceptron has n weights.



Other forms of neural networks

- Perceptron is simple, yet it recognizes printed letters, speech, pictures, just by being trained with examples.
- **Hopfield Networks**: Hopfield noted a similarity between spin glasses and neural networks. An electron's spin responds to the behaviour of its neighbours much like the neurons does.
- **Boltzmann Machine**: It has a mix of sensory and hidden neurons (analogous to, for example, the retina and brain, respectively).

5. Analogizers: You are what You Resemble

- **Analogizers**: Mapping new situations (**Classification**).
- Analogy plays an important role in machine learning, ...
- **Nearest neighbor algorithm** is our first stop of analogy-based learning. Second is **support vector machines** (SVM).
- Similarity is one of the central ideas of machine learning.
- Naive Bayes is not smart enough to recognize the faces, but Nearest neighbour can do it. Human, mother, parents, use this maximum for recognizing the faces.

Analogizers: KNN

For numerical attributes, the distance metric used in nearest neighbour is **Euclidean distance**. Considering two patterns, $\mathbf{x}_1 = x_{1_1}, x_{1_2}, \ldots, x_{1_m}$ and $\mathbf{x}_2 = x_{2_1}, x_{2_2}, \ldots, x_{2_m}$, distance between two using Euclidean is given by (for 2-D, $m = 2$)

$$d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^{m} (x_{1_j} - x_{2_j})^2}. \quad (8)$$

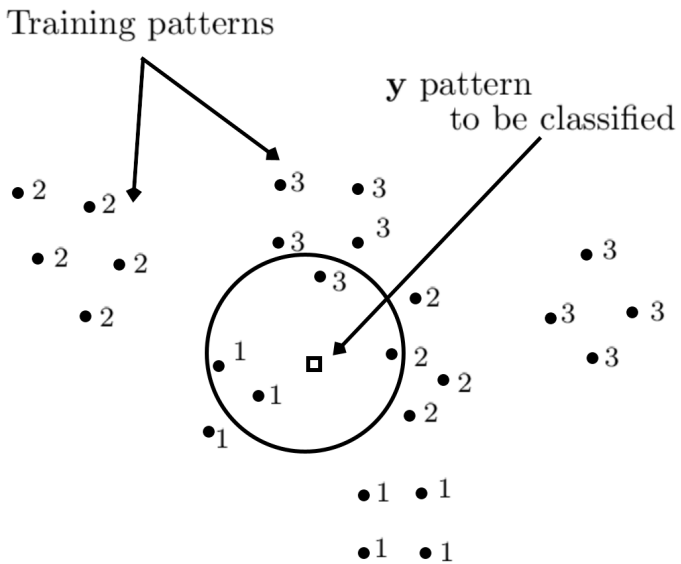


Figure 3: K-nearest neighbor problem

Analogizers: SVM

The support vector machines are falling in the category of **supervised learning** models based on **associative learning** algorithms that analyze the data and recognize patterns, hence they have applications in classification and regression analysis.

After having trained by some training examples, each belonging to one of the two categories, a SVM training algorithm builds a model that assign new examples into one of the two categories. Thus, a SVM is a *non-probabilistic linear* classifier.

They are powerful approaches to predictive modeling with success in number of applications, which includes **handwritten digit and alphabet recognition, face detection, text categorization**, etc.

Analogizers: SVM

An optimal separating surface is computed by maximizing the margin of separation.The data point with 0 attribute are labeled by circle (○s) and those with attribute 1 are labeled as squares(□s).
The variables **x, y** and $\mathbf{w}'$ are vectors.

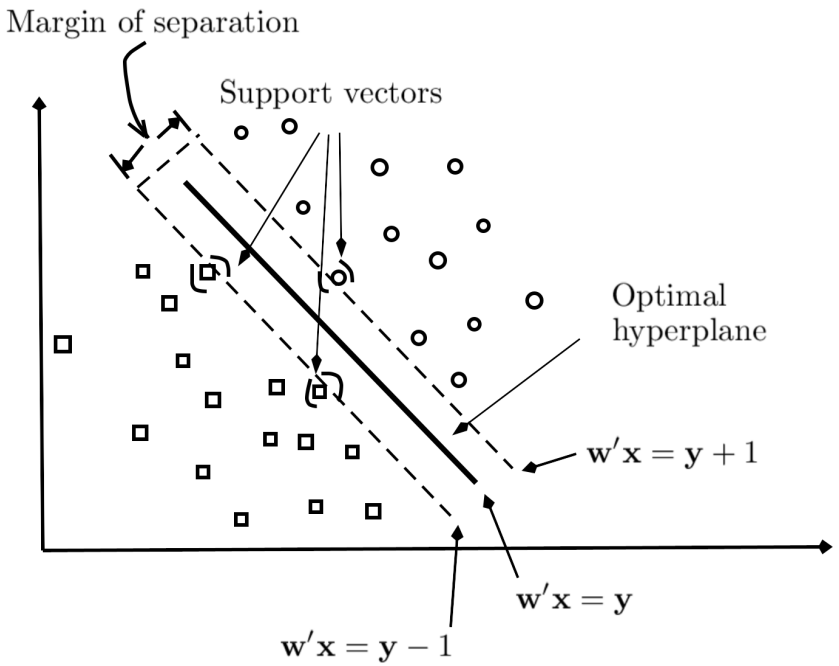


Figure 4: Classification through SVM

Classification in Big-Data

The basic idea of data **classification problem** can be simply described as follows: given a training data with known labels or classes in Table, we would like to learn a model, so that it can be used to predict data with unknown labels.

Record ID	Employment	Age	Salary	Group(label)
1	Self	30	30K	*C*
2	Industry	35	40K	*C*
3	Self	35	60K	*A*
4	Self	30	70K	*A*
5	Industry	35	40K	*C*
6	Academia	50	70K	*D*
7	Self	45	60K	*D*
8	Academia	30	70K	*B*
9	Industry	35	60K	*B*

Let us consider that we have identified some customers through **clustering of** the aggregated purchase information about the currently existing customers for certain company. Further, we have also acquired the mailing list of potential customers out of these, with their demographic information. As next step, we would like to assign each person in the mailing list to one of three groups: A , B, C, shown in Table.

The later is for the purpose of mailing them a catalog of items tailored to the individual's buying patterns.

Association Rule Mining

The Association rules are set of significant correlations, frequent patterns, associations, or causal structures from data sets found in various types of databases.

Such databases are transactional databases, relational databases, and other forms of data repositories. Mining of **association rules** is capturing those correlations, patterns, rules and representing them in the form of some *if*. . . *then* rules.

For example, given a set of transactions, each transaction comprising a set of items, an *association rule* can be an implication, $X \Rightarrow Y$, where X and Y are sets of items, indicating that presence of itemset X implies the itemset Y.

Association Rule Mining...

Consider that, an insurance company finds a strong correlation between two sets of policies X and Y of the form $X \Rightarrow Y$, which may be an indicator that customers holding policy set X were also likely to hold policy set Y, where X and Y may have one or more elements.

Now the company could more effectively target marketing the policy Y through those clients who hold policy X but not Y, to motivate them to buy policy Y. In effect, the association rule represents the knowledge about purchasing behavior of customers.

The Association rule mining has been effectively applied to many different domains that includes: **market basket analysis** in commercial environments, **astrophysics, crime prevention, fluid dynamics and to counter terrorism**,

Cyber Security

Concerns:

- Cryptography
- Cloud Security
- IoT and Ad hoc Networks Security
- AI in Cyber Security
- AI Information Security
- Distributed Systems
- Spams
- Malware
- Viruses
- Propagation and stop propagation of faults, threads
- Micro-blogging its effects
- Risk Management

Supervised Learning Algorithms

Since we already fed the machine with labeled data, so its prediction algorithm is based on **supervised learning**.
Example of supervised learning algorithms :

1. *Linear Regression*
2. *Logistic Regression*
3. *K-Nearest Neighbors*
4. *Decision Tree*
5. *Random Forest*
6. *Support Vector Machine*

Supervised Learning Applications

- Weather forecasting
- Pattern recognition
- Speech recognition
- Fraud detection
- Risk Analysis

Unsupervised Learning Algorithms

1. Dimension Reduction
2. Density Estimation
3. Market Basket Analysis
4. Generative adversarial networks (GANs)
5. Clustering

Unsupervised Learning Applications

- Important: Data are unlabeled, since data context is not available
- Massive amount of data: like twitter, Instagram, Facebook, Snapchat,
- APP; Email spam detection: far more variable

References

Chowdhary K.R. (2020) Machine Learning. In: Fundamentals of Artificial Intelligence. Springer, New Delhi.
https://doi.org/10.1007/978-81-322-3972-7_13

Chowdhary K.R. (2020) Statistical Learning Theory. In: Fundamentals of Artificial Intelligence. Springer, New Delhi.
https://doi.org/10.1007/978-81-322-3972-7_14

Chowdhary K.R. (2020) Data Mining. In: Fundamentals of Artificial Intelligence. Springer, New Delhi.
https://doi.org/10.1007/978-81-322-3972-7_17

https://www.springer.com/gp/book/9788132239703

https://www.phindia.com/Books/BookDetail/9788120350748/fundamentals-of-discrete-mathematical-structures-chowdhary