Topic – Probabilistic reasoning and uncertainity

Objective –

Outcome -

Probabilistic Reasoning

Probabilistic reasoning is a method of representation of knowledge where the concept of probability is applied to indicate the uncertainty in knowledge.

Probabilistic reasoning is **a form of knowledge representation in which the concept of probability is used to indicate the degree of uncertainty in knowledge**. In AI, probabilistic models are used to examine data using statistical codes. It was one of the first machine learning methods.

Probabilistic reasoning is using logic and probability to handle uncertain situations. An example of probabilistic reasoning is **using past situations and statistics to predict an outcome**.

Probabilistic reasoning is used in AI:

- When we're unsure of the predicates
- When the possibilities of predicates become too large to list down
- When it's known that an error occurs during an experiment

In probabilistic reasoning, there are two ways to solve problems with uncertain knowledge:

- **Bayes' rule**
- **Bayesian Statistics**

Probability: Probability can be defined as a chance that an uncertain event will occur. It is the numerical measure of the likelihood that an event will occur. The value of probability always remains between 0 and 1 that represent ideal uncertainties.

1. $0 \leq P(A) \leq 1$, where P(A) is the probability of an event A.

1. $P(A) = 0$, indicates total uncertainty in an event A.

1. $P(A) = 1$, indicates total certainty in an event A.

We can find the probability of an uncertain event by using the below formula.

$$\textbf{Probability of occurrence} = \frac{\text{Number of desired outcomes}}{\text{Total number of outcomes}}$$

- $P(\neg A)$ = probability of a not happening event.

- P(¬A) + P(A) = 1.

Event: Each possible outcome of a variable is called an event.

Sample space: The collection of all possible events is called sample space.

Random variables: Random variables are used to represent the events and objects in the real world.

Prior probability: The prior probability of an event is probability computed before observing new information.

Posterior Probability: The probability that is calculated after all evidence or information has taken into account. It is a combination of prior probability and new information.

Conditional probability:

Conditional probability is a probability of occurring an event when another event has already happened.

Let's suppose, we want to calculate the event A when event B has already occurred, "the probability of A under the conditions of B", it can be written as:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

Where P($A \wedge B$)= Joint probability of a and B

P(B)= Marginal probability of B.

If the probability of A is given and we need to find the probability of B, then it will be given as:

$$P(B|A) = \frac{P(A \wedge B)}{P(A)}$$

It can be explained by using the below Venn diagram, where B is occurred event, so sample space will be reduced to set B, and now we can only calculate event A when event B is already occurred by dividing the probability of **P(A∧*B*) by P(B)**.

Summary -

Topic – Bayes theorem and Bayesian network

Objective –

Outcome –

Bayes' theorem:

Bayes' theorem is also known as **Bayes' rule, Bayes' law**, or **Bayesian reasoning**, which determines the probability of an event with uncertain knowledge.

In probability theory, it relates the conditional probability and marginal probabilities of two random events.

Bayes' theorem was named after the British mathematician **Thomas Bayes**. The **Bayesian inference** is an application of Bayes' theorem, which is fundamental to Bayesian statistics.

It is a way to calculate the value of P(B|A) with the knowledge of P(A|B).

Example: If cancer corresponds to one's age then by using Bayes' theorem, we can determine the probability of cancer more accurately with the help of age.

Bayes' theorem can be derived using product rule and conditional probability of event A with known event B:

As from product rule we can write:

1. $P(A \wedge B) = P(A|B)\, P(B)$ or

Similarly, the probability of event B with known event A:

1. $P(A \wedge B) = P(B|A)\, P(A)$

Equating right hand side of both the equations, we will get:

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)} \quad \text{....(a)}$$

The above equation (a) is called as **Bayes' rule** or **Bayes' theorem**. This equation is basic of most modern AI systems for **probabilistic inference**.

P(A|B) is known as **posterior**, which we need to calculate, and it will be read as Probability of hypothesis A when we have occurred an evidence B.

P(B|A) is called the likelihood, in which we consider that hypothesis is true, then we calculate the probability of evidence.

P(A) is called the **prior probability**, probability of hypothesis before considering the evidence

P(B) is called **marginal probability**, pure probability of an evidence.

In the equation (a), in general, we can write P (B) = P(A)*P(B|Ai), hence the Bayes' rule can be written as:

$$P(A_i | B) = \frac{P(A_i)*P(B|A_i)}{\sum_{i=1}^{k} P(A_i)*P(B|A_i)}$$

Where A_1, A_2, A_3,........, A_n is a set of mutually exclusive and exhaustive events.

Example-1:

Question: what is the probability that a patient has diseases meningitis with a stiff neck?

Given Data:

A doctor is aware that disease meningitis causes a patient to have a stiff neck, and it occurs 80% of the time. He is also aware of some more facts, which are given as follows:

- The Known probability that a patient has meningitis disease is 1/30,000.
- The Known probability that a patient has a stiff neck is 2%.

Let a be the proposition that patient has stiff neck and b be the proposition that patient has meningitis. , so we can calculate the following as:

P(a|b) = 0.8

P(b) = 1/30000

P(a)= .02

$$\mathbf{P(b|a)} = \frac{P(a|b)P(b)}{P(a)} = \frac{0.8*(\frac{1}{30000})}{0.02} = 0.001333333.$$

Hence, we can assume that 1 patient out of 750 patients has meningitis disease with a stiff neck.

Application of Bayes' theorem in Artificial intelligence:

Following are some applications of Bayes' theorem:

- It is used to calculate the next step of the robot when the already executed step is given.
- Bayes' theorem is helpful in weather forecasting.
- It can solve the Monty Hall problem.

BAYESIAN BELIEF NETWORK IN ARTIFICIAL INTELLIGENCE

Bayesian belief network is key computer technology for dealing with probabilistic events and to solve a problem which has uncertainty. We can define a Bayesian network as:

"A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph."

It is also called a **Bayes network, belief network, decision network**, or **Bayesian model**.

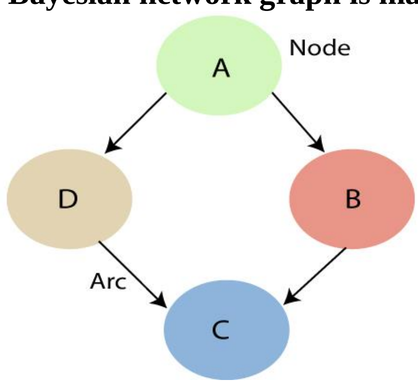
Bayesian networks are probabilistic, because these networks are built from a **probability distribution**, and also use probability theory for prediction and anomaly detection.

Bayesian Network can be used for building models from data and experts opinions, and it consists of two parts:

- **Directed Acyclic Graph**
- **Table of conditional probabilities.**

The generalized form of Bayesian network that represents and solve decision problems under uncertain knowledge is known as an **Influence diagram**.

A Bayesian network graph is made up of nodes and Arcs (directed links), where:



- Each **node** corresponds to the random variables, and a variable can be **continuous** or **discrete**.
- **Arc or directed arrows** represent the causal relationship or conditional probabilities between random variables. These directed links or arrows connect the pair of nodes in the graph.
  These links represent that one node directly influence the other node, and if there is no directed link that means that nodes are independent with each other
  - **In the above diagram, A, B, C, and D are random variables represented by the nodes of the network graph.**
  - **If we are considering node B, which is connected with node A by a directed arrow, then node A is called the parent of Node B.**
  - **Node C is independent of node A.**

Problem:

Solution:

- The Bayesian network for the above problem is given below. The network structure is showing that burglary and earthquake is the parent node of the alarm and directly affecting the probability of alarm's going off, but David and Sophia's calls depend on alarm probability.
- The network is representing that our assumptions do not directly perceive the burglary and also do not notice the minor earthquake, and they also not confer before calling.
- The conditional distributions for each node are given as conditional probabilities table or CPT.
- Each row in the CPT must be sum to 1 because all the entries in the table represent an exhaustive set of cases for the variable.
- In CPT, a boolean variable with k boolean parents contains 2^K probabilities. Hence, if there are two parents, then CPT will contain 4 probability values

List of all events occurring in this network:

- **Burglary (B)**
- **Earthquake(E)**
- **Alarm(A)**
- **David Calls(D)**
- **Sophia calls(S)**

We can write the events of problem statement in the form of probability: **P[D, S, A, B, E]**, can rewrite the above probability statement using joint probability distribution:

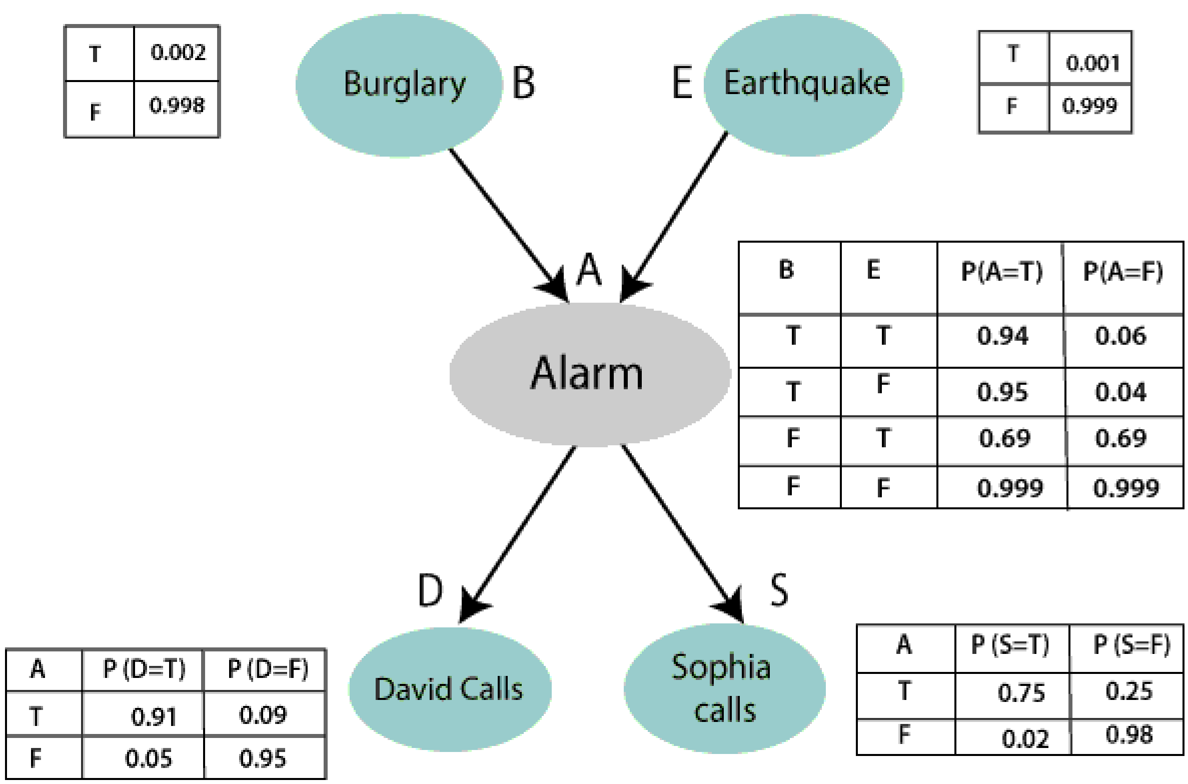
P[D, S, A, B, E]= P[D | S, A, B, E]. P[S, A, B, E]

=P[D | S, A, B, E]. P[S | A, B, E]. P[A, B, E]

= P [D| A]. P [S| A, B, E]. P[A, B, E]

= P[D | A]. P[S | A]. P[A| B, E]. P[B, E]

= P[D | A]. P[S | A]. P[A| B, E]. P[B |E]. P[E]



Let's take the observed probability for the Burglary and earthquake component:

P(B= True) = 0.002, which is the probability of burglary.

P(B= False)= 0.998, which is the probability of no burglary.

P(E= True)= 0.001, which is the probability of a minor earthquake

P(E= False)= 0.999, Which is the probability that an earthquake not occurred.

We can provide the conditional probabilities as per the below tables:

Conditional probability table for Alarm A:

The Conditional probability of Alarm A depends on Burglar and earthquake:

B	E	P(A= True)	P(A= False)
True	True	0.94	0.06
True	False	0.95	0.04
False	True	0.31	0.69
False	False	0.001	0.999

Conditional probability table for David Calls:

The Conditional probability of David that he will call depends on the probability of Alarm.

A	P(D= True)	P(D= False)
True	0.91	0.09
False	0.05	0.95

Conditional probability table for Sophia Calls:

The Conditional probability of Sophia that she calls is depending on its Parent Node "Alarm."

A	P(S= True)	P(S= False)

True	0.75	0.25
False	0.02	0.98

From the formula of joint distribution, we can write the problem statement in the form of probability distribution:

$P(S, D, A, \neg B, \neg E) = P(S|A) * P(D|A) * P(A|\neg B \wedge \neg E) * P(\neg B) * P(\neg E)$.

$= 0.75 * 0.91 * 0.001 * 0.998 * 0.999$

$= 0.00068045$.

Hence, a Bayesian network can answer any query about the domain by using Joint distribution.

The semantics of Bayesian Network:

There are two ways to understand the semantics of the Bayesian network, which is given below:

1. To understand the network as the representation of the Joint probability distribution.

It is helpful to understand how to construct the network.

2. To understand the network as an encoding of a collection of conditional independence statements.

Summary

Topic – Expert system

Objective –

Outcome –

What is Expert System?

Expert System is an interactive and reliable computer-based decision-making system which uses both facts and heuristics to solve complex decision-making problems. It is considered at the highest level of human intelligence and expertise. The purpose of an expert system is to solve the most complex issues in a specific domain.

The Expert System in AI can resolve many issues which generally would require a human expert. It is based on knowledge acquired from an expert. Artificial Intelligence and Expert Systems are capable of expressing and reasoning about some domain of knowledge. Expert systems were the predecessor of the current day artificial intelligence, deep learning and machine learning systems.

Examples of Expert Systems

Following are the Expert System Examples:

- **MYCIN:** It was based on backward chaining and could identify various bacteria that could cause acute infections. It could also recommend drugs based on the patient's weight. It is one of the best Expert System Example.
- **DENDRAL:** Expert system used for chemical analysis to predict molecular structure.
- **PXDES:** An Example of Expert System used to predict the degree and type of lung cancer
- **CaDet:** One of the best Expert System Example that can identify cancer at early stages

Characteristics of Expert System

Following are the important Characteristics of Expert System in AI:

- **The Highest Level of Expertise:** The Expert system in AI offers the highest level of expertise. It provides efficiency, accuracy and imaginative problem-solving.

- **Right on Time Reaction:** An Expert System in Artificial Intelligence interacts in a very reasonable period of time with the user. The total time must be less than the time taken by an expert to get the most accurate solution for the same problem.

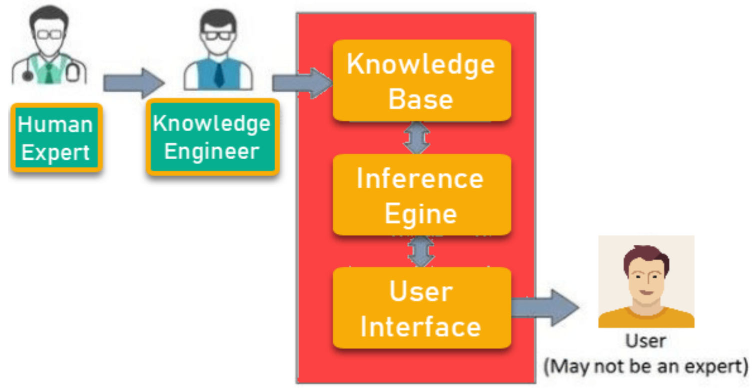
- **Good Reliability:** The Expert system in AI needs to be reliable, and it must not make any a mistake.

- **Flexible:** It is vital that it remains flexible as it the is possessed by an Expert system.

- **Effective Mechanism:** Expert System in Artificial Intelligence must have an efficient mechanism to administer the compilation of the existing knowledge in it.

- **Capable of handling challenging decision & problems:** An expert system is capable of handling challenging decision problems and delivering solutions.

Components of Expert System



The Expert System in AI consists of the following given components:

User Interface

The user interface is the most crucial part of the Expert System Software. This component takes the user's query in a readable form and passes it to the inference engine. After that, it displays the results to the user. In other words, it's an interface that helps the user communicate with the expert system.

Inference Engine

The inference engine is the brain of the expert system. Inference engine contains rules to solve a specific problem. It refers the knowledge from the Knowledge Base. It selects facts and rules to apply when trying to answer the user's query. It provides reasoning about the information in the knowledge base. It also helps in deducting the problem to find the solution. This component is also helpful for formulating conclusions.

Knowledge Base

The knowledge base is a repository of facts. It stores all the knowledge about the problem domain. It is like a large container of knowledge which is obtained from different experts of a specific field.

Thus we can say that the success of the Expert System Software mainly depends on the highly accurate and precise knowledge.

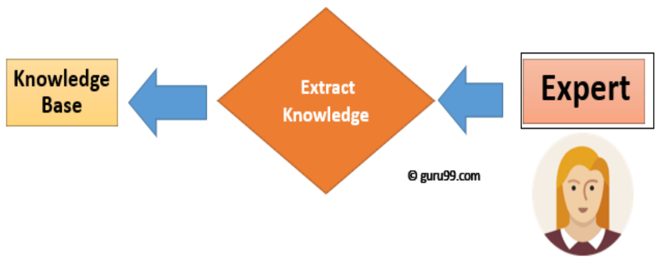
Other Key terms used in Expert Systems

Facts and Rules

A fact is a small portion of important information. Facts on their own are of very limited use. The rules are essential to select and apply facts to a user problem.

Knowledge Acquisition

The term knowledge acquisition means how to get required domain knowledge by the expert system. The entire process starts by extracting knowledge from a human expert, converting the acquired knowledge into rules and injecting the developed rules into the knowledge base.



Knowledge Extraction Process

Participant in Expert Systems Development

Participant	Role
Domain Expert	He is a person or group whose expertise and knowledge is taken to develop an expert system.
Knowledge Engineer	Knowledge engineer is a technical person who integrates knowledge into computer systems.
End User	It is a person or group of people who are using the expert system to get to get advice which will not be provided by the expert.

The process of Building An Expert Systems

- Determining the characteristics of the problem
- Knowledge engineer and domain expert work in coherence to define the problem
- The knowledge engineer translates the knowledge into a computer-understandable language. He designs an inference engine, a reasoning structure, which can use knowledge when needed.
- Knowledge Expert also determines how to integrate the use of uncertain knowledge in the reasoning process and what type of explanation would be useful.

Conventional System vs. Expert System

Conventional System	Expert System
Knowledge and processing are combined in one unit.	Knowledge database and the processing mechanism are two separate components.
The programme does not make errors (Unless error in programming).	The Expert System may make a mistake.
The system is operational only when fully developed.	The expert system is optimized on an ongoing basis and can be launched with a small number of rules.
Step by step execution according to fixed algorithms is required.	Execution is done logically & heuristically.
It needs full information.	It can be functional with sufficient or insufficient information.

Human expert vs. Expert System

Human Expert	Artificial Expertise
Perishable	Permanent
Difficult to Transfer	Transferable
Difficult to Document	Easy to Document
Unpredictable	Consistent
Expensive	Cost effective System

Advantages of Expert System

Below are the main advantages/benefits of Expert Systems in Artificial Intelligence (AI):

- It improves the decision quality
- Cuts the expense of consulting experts for problem-solving
- It provides fast and efficient solutions to problems in a narrow area of specialization.
- It can gather scarce expertise and used it efficiently.
- Offers consistent answer for the repetitive problem

- Maintains a significant level of information
- Helps you to get fast and accurate answers
- A proper explanation of decision making
- Ability to solve complex and challenging issues
- Artificial Intelligence Expert Systems can steadily work without getting emotional, tensed or fatigued.

Limitations of Expert System

Below are the disadvantages/limitations of Expert System in AI:

- Unable to make a creative response in an extraordinary situation
- Errors in the knowledge base can lead to wrong decision
- The maintenance cost of an expert system is too expensive
- Each problem is different therefore the solution from a human expert can also be different and more creative

Applications of Expert Systems

Some popular Application of Expert System:

- Information management
- Hospitals and medical facilities
- Help desks management
- Employee performance evaluation
- Loan analysis
- Virus detection
- Useful for repair and maintenance projects
- Warehouse optimization
- Planning and scheduling
- The configuration of manufactured objects
- Financial decision making Knowledge publishing
- Process monitoring and control
- Supervise the operation of the plant and controller
- Stock market trading
- Airline scheduling & cargo schedules

Summary

Topic – MYCIN Expert system

Objective –

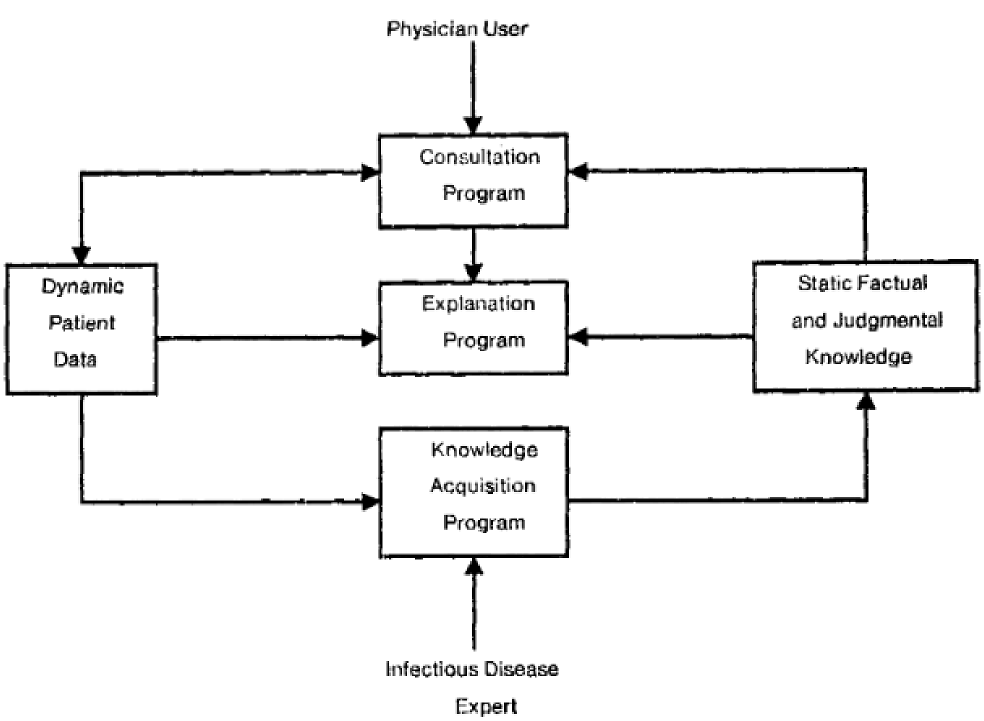
Outcome –

Introduction:

MYCIN was an early backward chaining expert system that used artificial intelligence to identify bacteria causing severe infections and to recommend antibiotics, with the dosage adjusted for patient's body weight.

- MYCIN was developed over five or six years in the early 1970s at Stanford University.
- It was written in Lisp
- MYCIN operated using fairly simple inference engine, and a knowledge base of approximately 600 rules.

Structure of Mycin Program:



The MYCIN system comprises three major subprograms, as depicted in Figure above.

- The Consultation Program
- Explanation Program
- Knowledge Acquisition Program

The **Consultation Program** is the core of the system; it interacts with the physician to obtain information about the patient, generating diagnoses and therapy recommendations.

The **Explanation Program** provides explanations and justifications for the program's actions.

The **Knowledge-Acquisition** Program is used by experts to update the system's knowledge base.

EVIDENCE COMBINATION

In MYCIN it was possible that two or more rules might draw conclusions about a parameter with different weights of evidence. For example, one rule may conclude that the organism in question is E. Coli with a certainty of 0.8 whilst another concludes that it is E. Coli with a certainty of 0.5 or even -0.8. In the event the certainty is less than zero the evidence is actually against the hypothesis. In order to calculate the certainty factor MYCIN combined these weights using the formula below to yield a single certainty factor:

$$CF(x,y) = \begin{cases} X + Y - XY & \text{if } X, Y > 0 \\ X + Y + XY & \text{if } X, Y < 0 \\ \frac{X+Y}{1-\min(|X|,|Y|)} & \text{otherwise} \end{cases}$$

Where X and Y are the certainty factors. This formula can be applied more than once if more than two rules draw conclusions about the same parameter. It is commutative, so it does not matter in which order the weights were combined.

A CONSULTATION WITH MYCIN:

1. Patient's name: (first – last)

**** FRED BRAUN**

(Questions 1-3 request background patient data).

2. Sex:

**** M**

3. Age:

**** 55**

4. Are there any cultures for Fred Braun which may be related to the present illness, and from which organisms have been grown successfully in the microbiology laboratory?

**** Y**

– – – CULTURE – 1 – – –

5. From what site was the specimen for CULTURE – 1 taken?

**** BLOD**

= BLOOD

(System corrects typing error).

6. Please give the date and time when CULTURE-1 was obtained.

(mo/da/yr time)

**** JUN 20, 1977**

The first organism isolated from the blood culture of 20-JUN-77 (CULTURE-1) will be referred to as:

– – – ORGANISM – 1 – – –

(… more questions follow in order to gather sufficient information to infer the identity and significance of the infecting organism

Summary

Topic – Learning , Rote learning

Objective –

Outcome –

Learning is "**a process that leads to change, which occurs as a result of experience and increases the potential for improved performance and future learning**"

What is learning?

- According to **Herbert Simon**, learning denotes changes in a system that enable a system to do the same task more efficiently the next time.
- **Arthur Samuel stated that**, "Machine learning is the subfield of computer science, that gives computers the ability to learn without being explicitly programmed ".
- In 1997, **Mitchell** proposed that, " A computer program is said to learn from experience **'E'** with respect to some class of tasks **'T'** and performance measure **'P'**, if its performance at tasks in **'T'**, as measured by **'P'**, improves with experience **E** ".
- The main purpose of machine learning is to study and design the algorithms that can be used to produce the predicates from the given dataset.
- Besides these, the machine learning includes the agents percepts for acting as well as to improve their future performance.

The following tasks must be learned by an agent.

- To predict or decide the result state for an action.
- To know the values for each state(understand which state has high or low vale).
- To keep record of relevant percepts.

Why do we require machine learning?

- Machine learning plays an important role in improving and understanding the efficiency of human learning.
- Machine learning is used to discover a new things not known to many human beings.

Various forms of learnings are explained below:

1. Rote learning

- Rote learning is possible on the basis of memorization.
- This technique mainly focuses on memorization by avoiding the inner complexities. So, it becomes possible for the learner to recall the stored knowledge.

  For example: When a learner learns a poem or song by reciting or repeating it, without knowing the actual meaning of the poem or song.

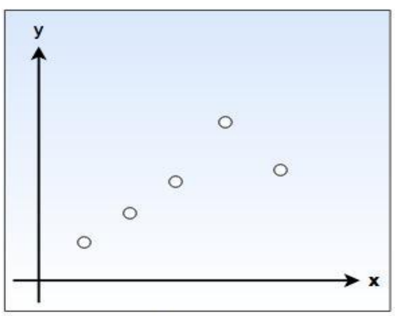
2. Induction learning (Learning by example)

- Induction learning is carried out on the basis of supervised learning.
- In this learning process, a general rule is induced by the system from a set of observed instance.
- However, class definitions can be constructed with the help of a classification method.

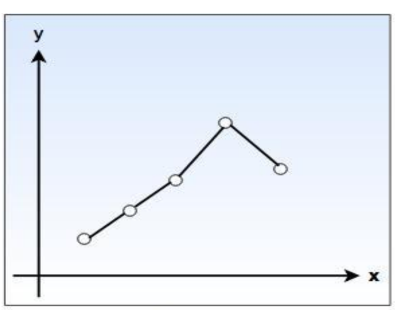
  For Example:

  Consider that **'f'** is the target function and example is a pair (x f(x)), where 'x' is input and f(x) is the output function applied to 'x'.

  Given problem: Find hypothesis h such as $h \approx f$
- So, in the following fig-a, points (x,y) are given in plane so that $y = f(x)$, and the task is to find a function h(x) that fits the point well.

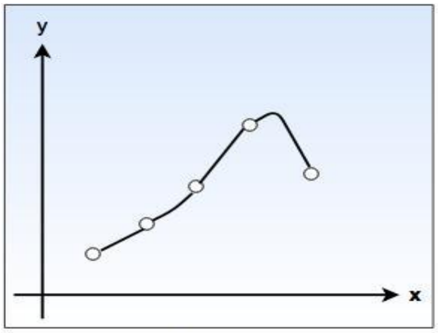


Fig(a)



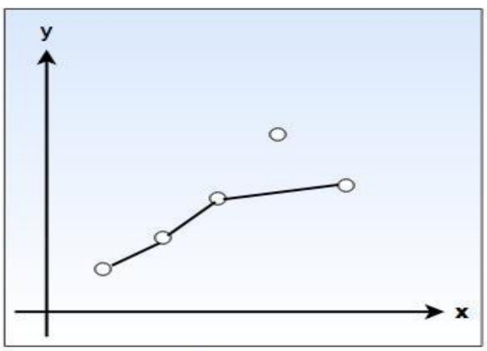
Fig(b)

- In fig-b, a piecewise-linear 'h' function is given, while the fig-c shows more complicated 'h' function.



Fig(c)

- Both the functions agree with the example points, but differ with the values of 'y' assigned to other x inputs.



Fig(d)

- As shown in fig.(d), we have a function that apparently ignores one of the example points, but fits others with a simple function. The true/ is unknown, so there are many choices for h, but without further knowledge, we have no way to prefer (b), (c), or (d).

3. Learning by taking advice

- This type is the easiest and simple way of learning.
- In this type of learning, a programmer writes a program to give some instructions to perform a task to the computer. Once it is learned (i.e. programmed), the system will be able to do new things.
- Also, there can be several sources for taking advice such as humans(experts), internet etc.
- However, this type of learning has a more necessity of inference than rote learning.

- As the stored knowledge in knowledge base gets transformed into an operational form, the reliability of the knowledge source is always taken into consideration.

Explanation based learning

- Explanation-based learning (EBL) deals with an idea of single-example learning.
- This type of learning usually requires a substantial number of training instances but there are two difficulties in this:

  I. it is difficult to have such a number of training instances

  ii. Sometimes, it may help us to learn certain things effectively, specially when we have enough knowledge.

  Hence, it is clear that instance-based learning is more data-intensive, data-driven while EBL is more knowledge-intensive, knowledge-driven.
- Initially, an EBL system accepts a training example.
- On the basis of the given goal concept, an operationality criteria and domain theory, it "generalizes" the training example to describe the goal concept and to satisfy the operationality criteria (which are usually a set of rules that describe relationships between objects and actions in a domain).
- Thus, several applications are possible for the knowledge acquisition and engineering aspects.

Summary