① a) I Hypothesis testing is an essential procedure in statistics. A hypothesis test evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data. When we say that a finding is statistically significant, with help of hypothesis test.

- Null Hypothesis → It is a general statement or default position that there is no relationship between two measured phenomena or no association among groups.
- Alternative Hypothesis → It is used in testing that is contrary to the null hypothesis. It is usually taken to be that the observations are the result of a real effect.

II Performance Analysis & In machine learning, it is very essential part of any project.

- Classification Accuracy & It is what we usually mean when we use term accuracy. It is ratio of no of correct predictions to the total no. of inputs samples.
- Confusion Matrix → It gives us a matrix as output and describes complete performace of model. Let assume we have binary classification problem. On testing our model on 165 samples, we got the following result

n=165	Predict No	Predict Yes
Actual No	50	10
Actual Yes	5	100

1) I) Bootstrapping : It is a method of sample reuse. The idea is to use the observed sample to estimate the population distribution. Then samples can be drawn from the estimated population and the sampling distribution of any type of estimator can itself be estimated.

There are three types of bootstrapping :

i) Nonparametric : A sample of same size as the data is taken from the data with replacement. What does this mean. It means that if you measure 10 samples, you create a new sample of size 10 by replicating some of the sample that you have already seen and omitting others.

ii) Semiparametric : It can only reproduce the items that were in original sample. It assume that the population includes other items that are similar to observed sample by sampling from smoothed version of the sample histogram.

iii) Parametric : It assumes that data comes from a known distribution with unknown parameters. You estimate the parameters from the data that you have and then you use the estimated distributions to simulate the samples.

All above samples methods are simulation based ideas.

I **Cross Validation** & It is a method that uses the same data to both train the model and obtain a less biased estimate of prediction error than the direct estimate. The basic idea is to split the training data into two subsets - one is used to train the prediction rule and then the other subset is used to assess prediction error. To use the data efficiently, this is repeated with multiple splits of the data.

There is a small problem with this method for assessing prediction error. The final predictor will be based on N, but estimated prediction error is based on predictor developed on smaller sample $N > N-N/K$. So cross-validation estimate of prediction error might actually be pessimistic - might have slightly better prediction error than you think. However with 10-folds cross-validation cant be too far off because you are using at least 90% of your sample.

c) Since Bayes theorem provides a principled way to calculate the posterior probability of each hypothesis given the training data, we can use it as the basis for a straightforward learning algorithm that calculates the probability for each possible hypothesis, then outputs the most probable. This section considers such a brute-force bayesian concept learning algorithm then compares it to concept learning algorithm. As we shall see, one interesting result of this comparison is that under certain conditions several algorithms output the same hypothesis as this brute force Bayesian.

- Product Rule → $P(A \land B) = P(A|B)P(B) = P(B|A)P(A)$
- Sum Rule → $P(A \lor B) = P(A) + P(B) - P(A \land B)$
- Bayes Theorem → The Posterior Probability $P(h|D)$ of h given D

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

- Theorem Of Total Probability → If events $A_1 \ldots A_n$ are mutually exclusive with $\sum_{i=1}^{n} P(A_i) = 1$

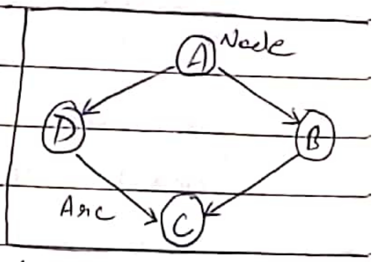
$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

d) A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph. Bayesian Networks are probabilistic because these networks are built from probability distribution and also use probability theory for prediction and anomaly detection.

Baysian 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

A Bayesian graph is made up of nodes and Arc (directed link)

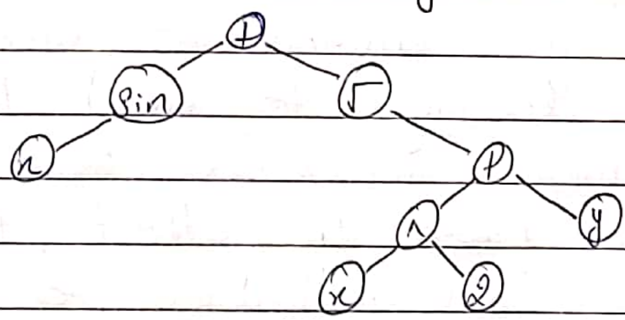


The Bayesian Network has mainly two components:

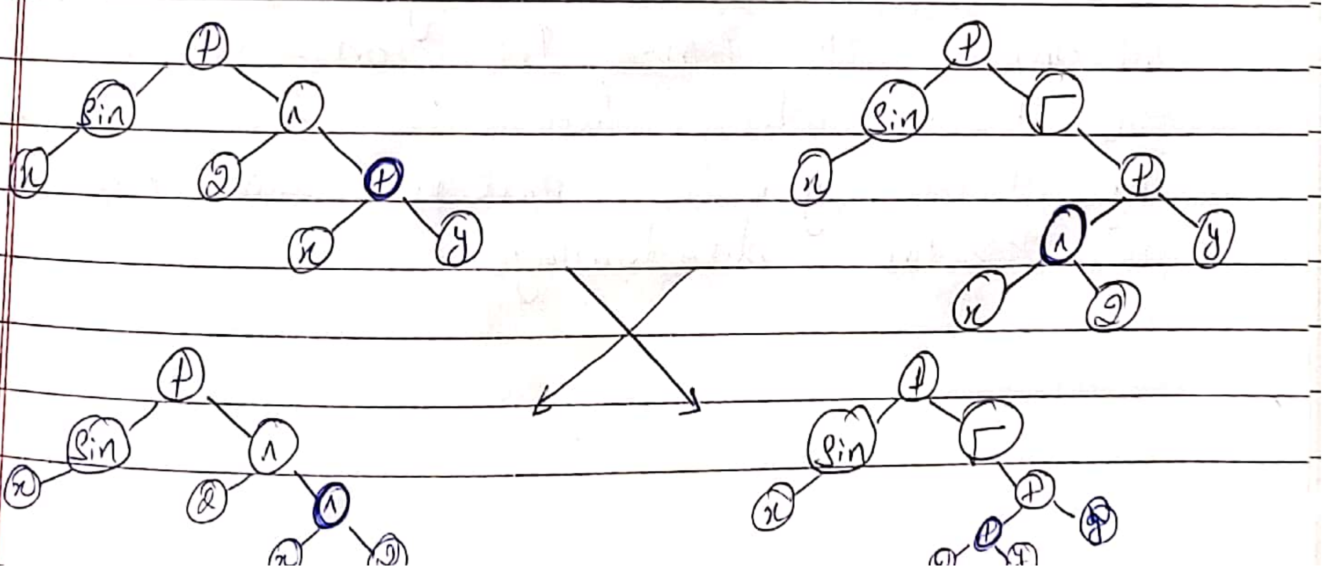
- Casual Component
- Actual numbers

) a) Genetic Programming is a branch of genetic algorithms. Genetic programming creates computer programs in the lisp or scheme computer languages as the solution.

Example → $\sin(x) + \sqrt{x^2 + y}$



Crossover



3) i) Lamarckian Evolution :- He proposed that experiences of a single organism directly affected the genetic makeup of their offspring. If an individual learned during its lifetime to avoid some toxic food, it could pass this trait on genetically to its offspring, which therefore would not need to learn the trait. This is an attractive conjecture, because it would presumably allow for more efficient evolutionary progress than a generate and test process that ignores the experience gained during an individual's lifetime.

ii) Baldwin Effect :- If a species is evolving in a changing environment, there will be evolutionary pressure to favor individuals with the capability to learn during their lifetime. For example, if a new predator appears in the environment, then individuals capable of learning to avoid the predator will be more successful than individuals who cannot learn. In effect, the ability to learn allows an individual to perform a small local search during its lifetime to maximize its fitness. In contrast, non learning individuals whose fitness is fully determined by their genetic makeup will operate at a relative disadvantage.