JUNAID . GIRKAR
60004190057
TE COMPS A-4

MACHINE LEARNING
ASSIGNMENT - 1

Q1 Explain Radial Basis Function in detail.

ANS A Radial Basis function is a function whose value depends only on the distance from the origin. The function must contain only real values. Alternative forms of radial basis functions are defined as the distance from another point c, called center.

WORKING: A Radial Basis Function works by defining itself by the distance from its origin or center. This is done by incorporating the absolute value of the function.

Values of RBF are defined as: $\phi(x) = \phi(\|x\|)$.

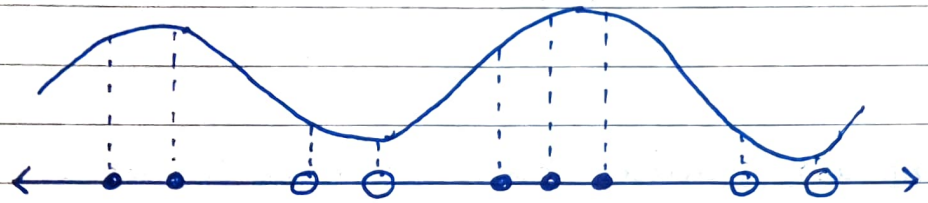
The Gaussian variation of RBF is often applied in RBF networks.

$$f(x) = \exp\left(-\frac{\|x - \bar{x}\|^2}{2\sigma^2}\right)$$

The double pipes represent the idea of distance. This is the radius aspect of radial basis function. RBF are symmetrical around the origin.

Consider a set of D data, we want to divide into 2 classes with one line.

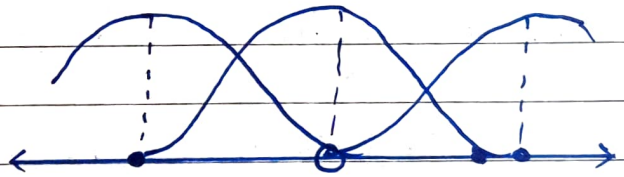
lets introduce a wavy function f(x) and map each value of x to its corresponding output.



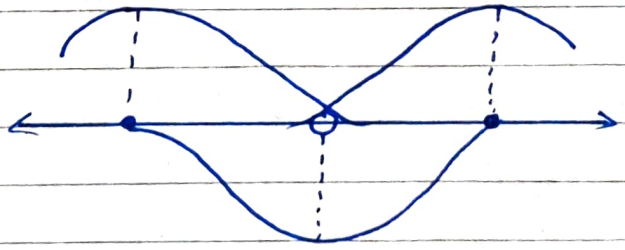
The solution seems sneaky, but it can be generalized with RBF. Gaussian function is an RBF since points are represented as number of standard deviations from the mean.

RBF KERNEL : lets take an example with three points.

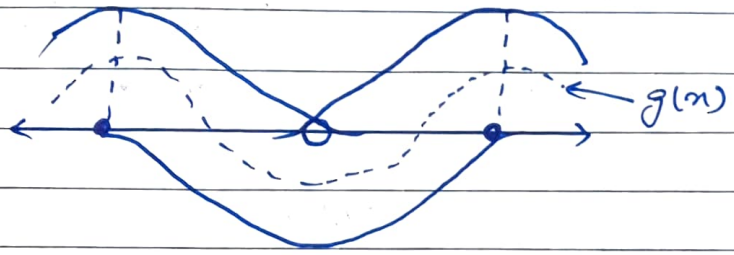
lets draw a Gaussian RBF centered at each of three points.



We can sum all RBF for data points of one class



If we add all values of RBF at each point x will get a global function.

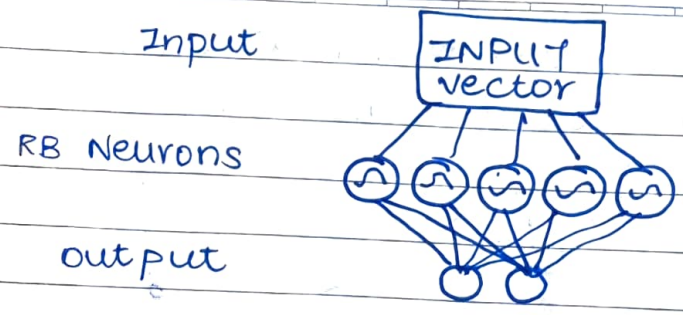


The global RBF function is dense in one central area and less in all other places. Thus it has a lot of sway in deciding value of s(x) when x are near its location, with dimensioning power as distance increases. This problem of separation is similar to SVM using RBF Kernel.

APPLICATION OF RADIAL BASIS FUNCTION:

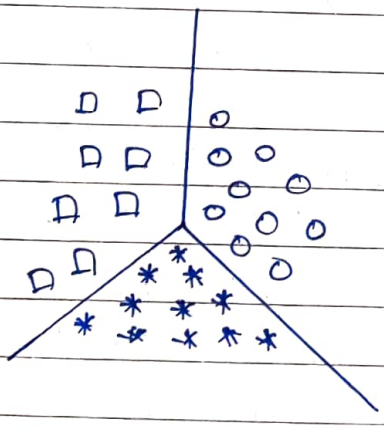
Use RBF by incorporating radial basis neurons in a simple 2 layer network.

Input vector is a n-dimensional vector on which classification or regression is to be done. Input vector is supplied to each RB neuron. Each RB neuron stores a central vector i.e., one unique vector from training set. The input vector is compared to the central vector and difference is plugged in RBF. Thus RB neuron can be thought as non-linear measure of similarity between input and central vectors. The learnings from RB neurons are weighted and provided to the output layer.

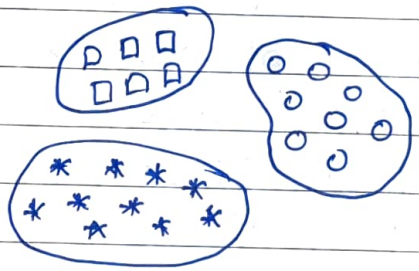


RADIAL BASIS NETWORK.

The radial basis network approaches task of classification differently than standard neural networks standard neural networks tend to separate data through linear manipulations whereas RBF tend to group the data through density based transformations.



standard NN

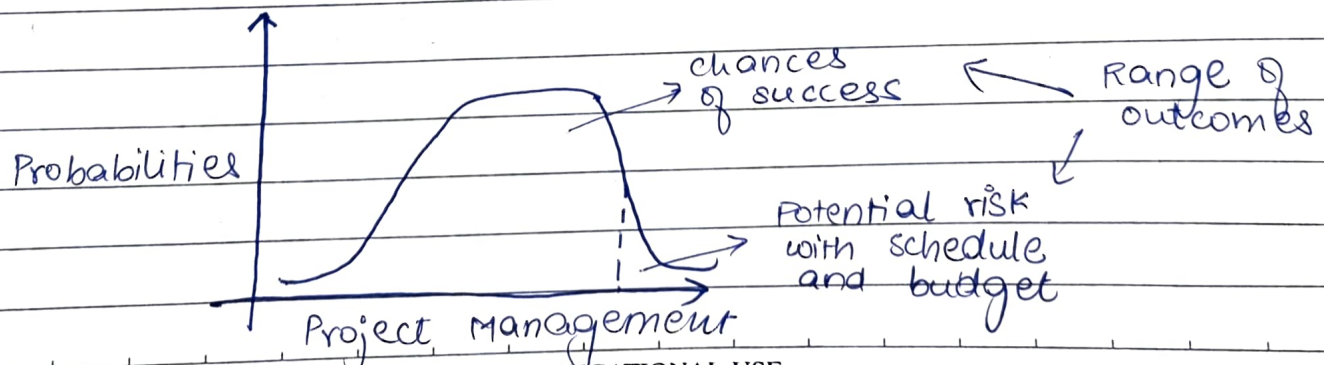


RBN

Q2 Explain Monte Carlo with suitable example ?

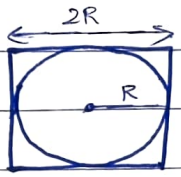
ANS Monte Carlo simulation is a mathematical way for calculating odds of multiple possible outcomes occuring in uncertain processes through repeated random sampling. This computational algorithm makes risks associated with a particular process convinient, thereby enabling better decision making.

It is a statistical technique which uses the randomness to solve probabilistic problems. Also it can perform sensitive analysis and correlation between input variables. It finds its application in prediction and forecasting models in business supply chain, project management, etc.



Consider the following examples :

If a circle of radius R is inscribed inside a square with side length 2R, then the area of circle will be πR^2 and the area of square will be $(2R)^2$. So ratio of circle area to area of square will be $\pi/4$



It also means that if we pick a random point (x,y) both x and y are between (1,-1), probability of this random point lies inside the unit circle is given as proposition between area of unit circle & square.

$$P(x^2 + y^2 < 1) = \frac{\text{Area (circle)}}{\text{Area (square)}} = \frac{\pi}{4}$$

So if you pick N points at random inside square approximately $N\pi/4$ of these points should fall inside a circle.

$$M(\text{no. of points inside the circle}) = \frac{N\pi}{4}$$

So supposing, we pick 'N' random points, out of which 'M' of those fall inside the circle, we can then calculate π by :-

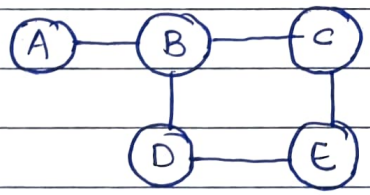
$$\pi = 4\frac{M}{N}.$$

Q3 Write a short note on :

1) Markov Random fields :-

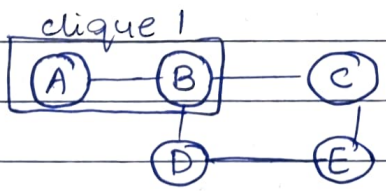
Markov Random model is a model which uses the unidirected graph. Unidirected graphical models represent edges represent the potential between two variables, syntatically, factorization distribution probabilities between variable. In each individual variable connected with edges represent a certain clique in graph ; means probability distribution of graph can factorize an individual clique potential function.

Just as we had CPD's for Bayesian networks, we have to incorporate relationship between nodes in Markov networks. However there are 2 crucial difference between the tables and CPD's.



clique in graph theory it is a subset of vertices of an unidirected graph.

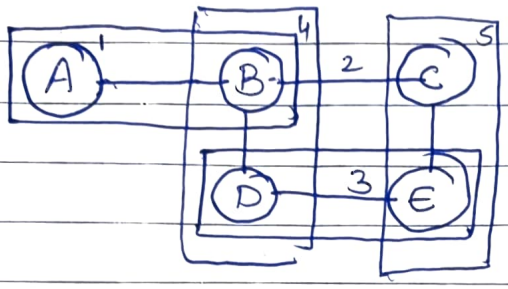
$$P(A,B,C,D,E) \propto \phi(A,B)\ \phi(B,C)\ \phi(B,D)\ \phi(C,E)\ \phi(D,E)$$

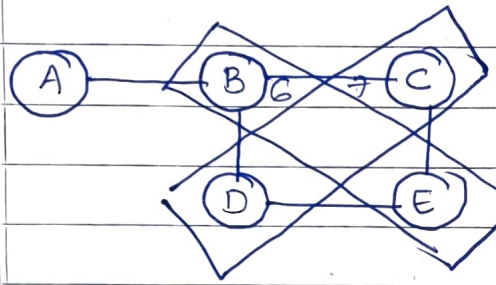


$$P(x) = \frac{1}{2} \prod_{\text{clique}} \phi_c(x_c)$$

(Potential function)

Such that it includes subgraph in every vertices in a clique is adjacent. So clique in this graph adjust adjacenty one by one.

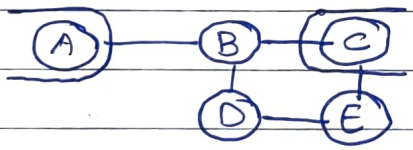




There is some difference if we join D,C & B,E clique over here, then it changes its probability.

$$P(A,B,C,D,E) \propto \phi(A,B)\ \phi(B,C,D)\ \phi(C,D,E)$$

Some undirected graphic model has Markov Random field. In MRF, certain paths between A and C



$A \rightarrow B \rightarrow C$

$A \rightarrow B \rightarrow D \rightarrow E \rightarrow C$

Unlike the Bayesian model, Markov networks do not need to be acyclic.

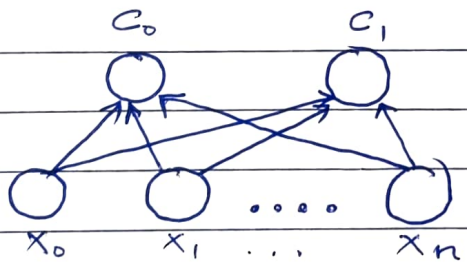
Independence properties such as Markov properties:

- Any two subsets if variables are conditionally independent given a separating subset.
- If we take 'A' as a subset and 'c' as one subset then there is a wall between them. So there is no way to go between 'A' and 'c' without getting through the subset. So we are

using (A, B) then B, C, D, E. Therefore A and C are separating subset.

2 Self organizing Maps :- (SOM)

It is a type of Artificial Neural Network which is inspired by biological models of neural systems from 1970's. It follows an unsupervised learning approach and has trained its network through a competitive learning algorithm. SOM is used for clustering and mapping techniques to map multidimensional which allows people to reduce complex problems for easy interpretation. SOM has 2 layers : input layer and output layer. The architecture of SOM with 2 clusters and n input feature of any sample is given below.



ALGORITHM :

1) Weight Initialization
2) For 1 to n number of epochs
3) Select training example.

4) Compute winning vector
5) Update the winning vector
6) Repeat steps 3,4,5 for all training example.
7) Clustering the test sample.

Lets say an input data of size (m,n) where 'm' is number of training examples and 'n' is number of features in each example. First it initializes the weights of size (n, c) where 'c' is the number of clusters. Then iterating over the input data, for each training example, it updates the winning vector.

weight updation rule is given by :-

$$w_{ij} = w_{ij}(old) + \alpha(t) * (x_i^k - w_{ij}(old))$$