Assignment - 1

1) What are Probabilistic Generative Models? Write some applications of Probabilistic Generative Models?

Ans→ Given a set of training data, probabilistic Generative Models learn the joint probability distribution that represents the training data and use this underlying distribution to generate new data similar to the training data.

Applications →

i) Natural Language Processing (NLP) → Text generation: Models like GPT and BERT are inspired by probabilistic generative models for text generation. Speech Recognition: HMMs are used in speech to text systems.

ii) Healthcare and Bioinformatics → Disease Diagnosis: Bayesian network model the probability of diseases given symptoms. Genomics: Used in modelling DNA sequences and predicting mutations.

2) Explain the difference between generative and deterministic models. Provide examples of each.

Ans→

Generative Model	Deterministic Model
i) Given, a set of training data, it learns the joint probability distribution that represents the training data and use this to generate data similar to the training data.	i) Given a set of training data, it constructs a deterministic function and uses this function to generate a new output for the new input.
ii) It may produce different output for the same input.	ii) It produces same output for the same input.
iii) Eg → Naive Baye's Classification Model	iii) Eg → Regression

3) Three Cards are drawn in succession, without replacement, from an ordinary deck of playing cards. . . . .

Ans→ Product Rule → $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2)$

$P(A_1) = \frac{2}{52} = \frac{1}{26}$

Now, after drawing a red ace, 51 cards remain.

- Four 10s (10 of hearts, diamonds, clubs & spades)

• Four Jacks (Jack of Heart, Diamonds, Clubs and Spades)

$\therefore$ There are 8 favourable Cards, $P(A_2 | A_1) = \frac{8}{51}$

Numbers > 3 but < 7 are 4, 5, 6.

• Four 4s, 5s & 6s.

$\therefore$ Total favourable cards = 12

So, $P(A_3 | A_1 \cap A_2) = \frac{12}{50} = \frac{6}{25}$

$\therefore P(A_1 \cap A_2 \cap A_3) = \frac{1}{26} \times \frac{8}{51} \times \frac{6}{25} = \frac{8}{5525} = 0.00145$ (Ans)

4) Find the probability of randomly selecting 4 good quarts of milk in succession from a cooler containing 20 quarts, of which 5 have spoiled.

Ans → $P(G_1 \cap G_2 \cap G_3 G_4) = P(G_1) \cdot P(G_2 | G_1) \cdot P(G_3 | G_1 \cap G_2) \cdot P(G_4 | G_1 \cap G_2 \cap G_3)$

$P(G_1) = \frac{15}{20} = \frac{3}{4}$, $P(G_3 | G_1 \cap G_2) = \frac{13}{18}$, $P(G_2 | G_1) = \frac{14}{19}$, $P(G_4 | G_1 \cap G_2 \cap G_3) = \frac{12}{17}$

$\therefore P(G_1 \cap G_2 \cap G_3 \cap G_4) = \frac{15}{20} \times \frac{14}{19} \times \frac{13}{18} \times \frac{12}{17}$

$= \frac{6552}{23256} = 0.2817$ (Ans)

5) Consider a random variable of X whose probability distribution is $f(x) = \frac{x}{2}, 0 \le x < 2$
0, otherwise

Suppose the transformation function is $y(x) = 1 - \frac{\sqrt{4-x^2}}{2}$. Find the transformed probability distribution $g(y)$ of random variable $y(x)$.

Ans → $f(x) = \frac{x}{2}, 0 \le x < 2$; $y(x) = 1 - \frac{\sqrt{4-x^2}}{2}$
0, otherwise

$f(y^{-1}(x)) = f(x(y)) = 2\sqrt{y(2-y)}$

$\therefore g(y) = 2\sqrt{y(2-y)} \times \left| \frac{d(2\sqrt{y(2-y)})}{dy} \right|$

$= 2\sqrt{y(2-y)} \times \frac{2(1-y)}{\sqrt{y(2-y)}}$

$= 4(1-y)$ (Ans)

7) Find the Empirical distribution $F_n(x)$ of the data [0,1,2,2,4,6,6,7]. Also, draw the graph of $F_n(x)$.

Ans → Arranging the data in ascending order = [0,1,2,2,4,6,6,7]. $n=8$

$x<0 : F_n(x) = \frac{\text{Count}}{8} = 0$

$x=0 : F_n(x) = 1/8 = 0.125$

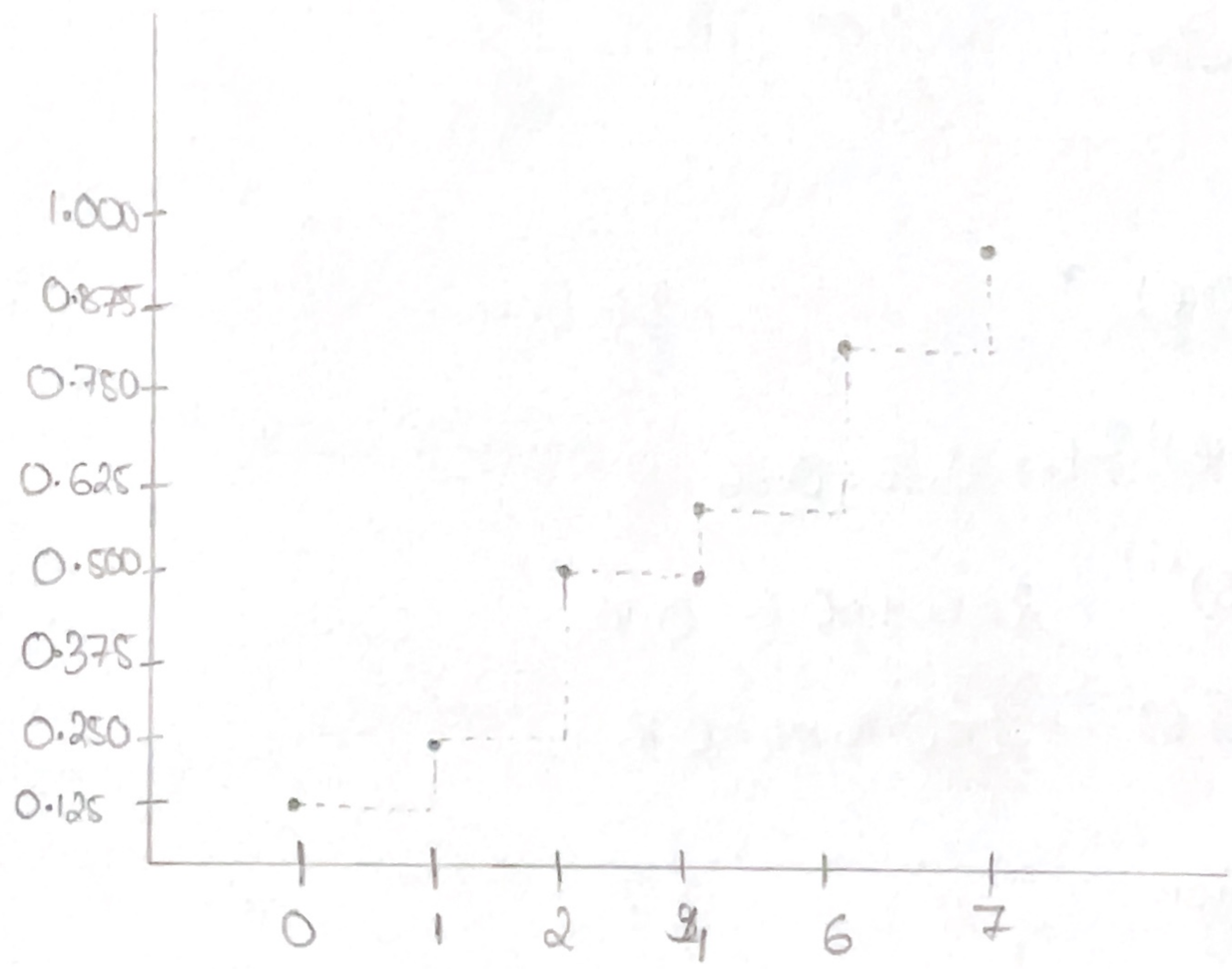
$x=1 : F_n(x) = 2/8 = 0.25$

$x=2 : F_n(x) = 4/8 = 0.5$

$x=4 : F_n(x) = 5/8 = 0.625$

$x=6 : F_n(x) = 7/8 = 0.875$

$x=7 : F_n(x) = 8/8 = 1$



8) Consider a sample consisting of the values 12, 11.2, 13.5, 12.3, 13.8 and 11.8 comes from a population with the density function $f(x;\theta) = \frac{\theta}{x^{\theta+1}}, x>1$ where $\theta>0$. Find the maximum likelihood estimate of θ. 0, elsewhere

Ans → $f(x;\theta) = \frac{\theta}{x^{\theta+1}}, x<1$, $n=6$ [12, 11.2, 13.5, 12.3, 13.8 and 11.8]

$\ln L(x_1, x_2, \ldots, x_6; \theta) = \ln(f(x_1,\theta) \cdot f(x_2,\theta) \ldots f(x_6,\theta))$

$= \ln f(x_1,\theta) + \ln f(x_2,\theta) + \ldots + \ln f(x_6,\theta)$

$$= \ln\left(\frac{\theta}{x_1^{\theta+1}}\right) + \ln\left(\frac{\theta}{x_2^{\theta+1}}\right) + \dots + \ln\left(\frac{\theta}{x_6^{\theta+1}}\right)$$

$$= 6\ln(\theta) - (\theta+1)\sum_{i=1}^{6} \log x_i$$

Now, $\frac{d}{d\theta} 6\ln(\theta) - (\theta+1)\sum_{i=1}^{6} \log x_i = \frac{n}{\theta} - \sum_{i=1}^{6} \log x_i$

$$\frac{n}{\theta} - \sum \log x = 0$$

$$\Rightarrow \theta = \frac{n}{\sum_{i=1}^{6} \log x_i}$$

$$= \frac{6}{15.1143} = 0.397$$

9) Let P & Q be two distributions on a discrete random variable ....

Ans → $P(x)$: Binomial Distribution

$n = 2,\ p = 0.4$

$$P(x) = \binom{2}{x}(0.4)^x(0.6)^{2-x}$$

$$P(0) = \binom{2}{0}(0.4)^0(0.6)^{2-0} = 1 \times 1 \times 0.36 = 0.36$$

$$P(1) = \binom{2}{1}(0.4)^1(0.6)^{2-1} = 2 \times 0.4 \times 0.6 = 0.48$$

$$P(2) = \binom{2}{2}(0.4)^2(0.6)^{2-2} = 1 \times 0.16 \times 1 = 0.16$$

$Q(x)$: Uniform Distribution

$$Q(0) = Q(1) = Q(2) = 1/3$$

$$\therefore D_{KL}(P \| Q) = P(0)\log\frac{P(0)}{Q(0)} + P(1)\log\frac{P(1)}{Q(1)} + P(2)\log\frac{P(2)}{Q(2)}$$

$$= 0.36\log\frac{0.36}{1/3} + 0.48\log\frac{0.48}{1/3} + 0.16\log\frac{0.16}{1/3}$$

$$= 0.0853 \text{ (Ans)}$$

$$\therefore D_{KL}(Q \| P) = Q(0)\log\frac{Q(0)}{P(0)} + Q(1)\log\frac{Q(1)}{P(1)} + Q(2)\log\frac{Q(2)}{P(2)}$$

$= \frac{1}{3}\log\frac{1/3}{0.36} + \frac{1}{3}\log\frac{1/3}{0.48} + \frac{1}{3}\log\frac{1/3}{0.16}$

$= 0.975$ (Ans)

10) Consider ~~the~~ a collection of symptoms and flu diagnoses of some patients given by the following data.....

Ans → Given, Chills (C) = Y, Runny Nose (R) = N, Headache (H) = Mild, Flu (F) = Y

$P(F=Y) = \frac{5}{8}$, $P(F=N) = 3/8$

For F = Y :

$P(C=Y \mid F=Y) = 3/8$ $\quad$ $P(H=\text{Mild} \mid F=Y) = \frac{2}{5}$

$P(R=N \mid F=Y) = \frac{2}{5}$ $\quad$ $P(F=Y \mid F=Y) = \frac{4}{5}$

$\therefore \frac{\frac{5}{8}\times\frac{3}{8}\times\frac{2}{5}\times\frac{2}{5}\times\frac{4}{5}}{P(B)} = \frac{0.048}{P(B)}$

For F = N :

$P(C=Y \mid F=N) = \frac{2}{3}$ $\quad$ $P(H=\text{Mild} \mid F=Y) = 1/3$

$P(R=N \mid F=N) = \frac{2}{3}$ $\quad$ $P(F=Y \mid F=N) = \frac{2}{3}$

$\therefore \frac{\frac{3}{8}\times\frac{2}{3}\times\frac{2}{3}\times 1/3\times 2/3}{P(B)} = \frac{0.037}{P(B)}$

Now, since $\frac{0.037}{P(B)} < \frac{0.048}{P(B)}$ → The patient does not have flu.

11) What ~~does~~ graphs ~~feature~~ represents in A.I? Define node features & edge features.....

Ans → Graphs are powerful data structures used in AI to represent relationships between entities. They consist of nodes representing and edges representing connections between them.

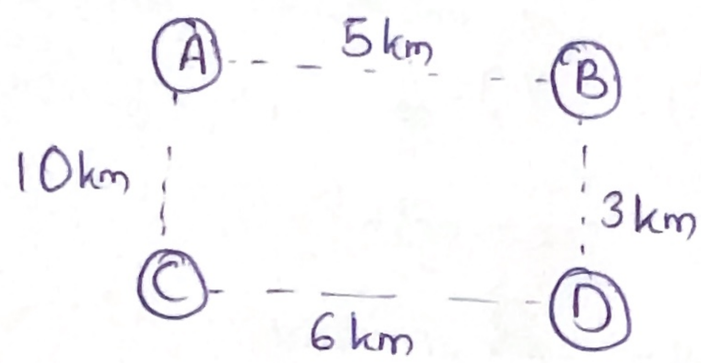
- Node features → Attributes of a node in a graph.
  Eg → In a social network, a node (user) may have features like age, interests & location

- Edge Features → Attributes of an edge representing relationships between nodes.

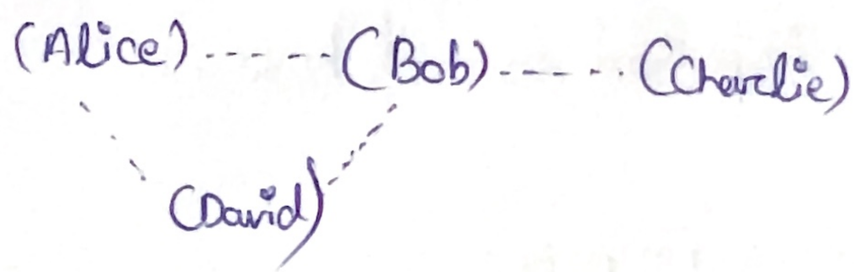
Eg → In a road network, an edge may have features like distance, speed limit or traffic conditions.

a) City's Road Network as a Graph →
- Nodes : Intersections or locations in the city
- Edges : Roads connecting intersections
- Edge Features : Distance (km), Speed Limit (km/h), Traffic Congestion Level

To find the shortest path between two points, algorithms like Dijkstra's Algorithm can be used.



b) Social Media Network as a Graph
- Nodes : Users on the platform
- Edges : Friendships between users
- Node Features : Age, Interests, Location
- Edge Features : Frequency of Interactions (messages, likes), Strength of friendship (years known), Mutual friends count

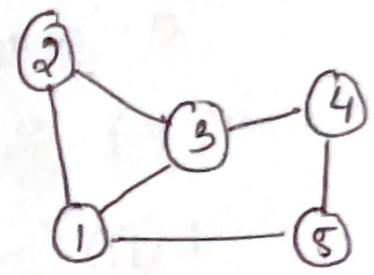


c) A bipartite graph has two distinct types of nodes (customers and products) and edges represent interactions (purchases).
- Nodes : Customers and products
- Edges : Represent purchases made by customers
- Edge Features : Purchase Frequency, Rating given by customer, Browsing history

(Customer A) - - - - - - (Product X)

(Customer B) - - - - - - (Product Y)

12) What are the different types of graph representations? Express any two graph representations of the following graph.



Ans → i) Adjacency Matrix: A m×m matrix where m is the number of nodes. If there is an edge between node i and j then A[i][j]=1, otherwise A[i][j]=0.

ii) Adjacency List: Each node maintains a list of adjacent nodes. More space-efficient for sparse graphs.

iii) Edge List: A list of edges representing connections between nodes. Useful when working with algorithms that ~~flows~~ focus on edges rather than nodes.

For the given graph,

Nodes = (1, 2, 3, 4, 5)

Edges = (1-2), (1-3), (1-5), (2-3), (3-4), (4-5)

∴ Adjacency Matrix = $\begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$

∴ Adjacency List: 1 → [2, 3, 5]
2 → [1, 3]
3 → [1, 2, 4]
4 → [3, 5]
5 → [1, 4]

13) What is the PageRank algorithm? Explain the working of the PageRank algorithm.

Ans → The PageRank algorithm is a link analysis algorithm used by Google Search to rank web pages based on their importance. The algorithm assigns a page rank score to each page, reflecting its importance in a web network.

PageRank is based on the idea that important pages are linked to by many other important pages. The algorithm simulates a random web surfer

which follows links between pages, and the PageRank score represents the probability of landing on a particular page.

$$PR(P_i) = (1-d) + d \sum_{P_i \in M(P_i)} \frac{PR(P_f)}{L(P_f)}$$

where, $PR(P_i)$ = PageRank of page P_i

d = damping factor

$M(P_i)$ = Set of pages that link to P_i

$L(P_f)$ = number of outbounds links from page P_i

14) Consider a small web network with four pages : A, B, C & D. ....

Ans → a) The directed graph is :

$A \to B$ $\quad$ $D \to C$

$A \to C$ $\quad$ $D \to A$

$B \to C$

$C \to A$

Transition Probability Matrix $= \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$

b) Without damping →

$$PR^{k+1} = M \cdot PR^k$$

∵ Since we have 4 pages, each starts with:

$$PR^{(0)} = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$$

After three iterations also, PageRank value remains the same $PR^3 = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$

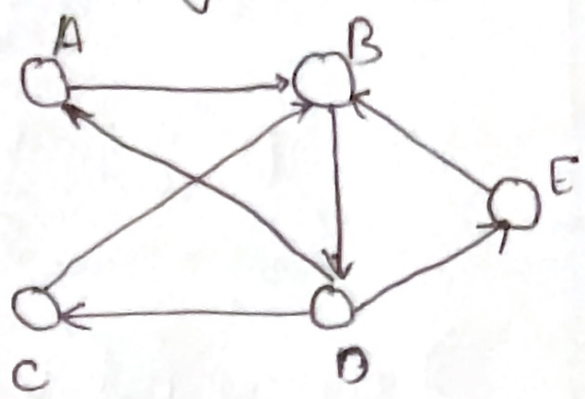
c) With damping factor, $d = 0.85$

$$PR^{k+1} = d \cdot M \cdot PR^k + (1-d) \cdot \frac{1}{N} \cdot 1$$

We can see that, after ~~that~~ three iterations, in this case also, the PageRank value remains $PR^{(3)} = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$

∴ The given graph distributes rank even across all nodes.

15) Suppose there is a graph consisting of 5 nodes, associated as the following network graph....



Ans → a) Link Structure →

A-B	C-D
A-C	D-C
B-A	D-E
B-D	E→B

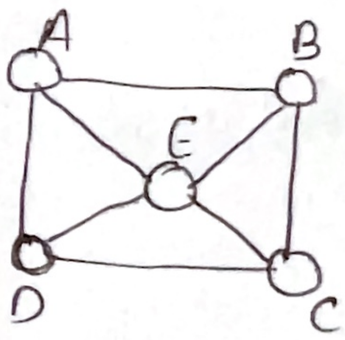
Transition Probability Matrix (M) = $\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$

b) Initial Page Rank for each page (PR^0) = $\begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}$

c) After two iterations, the Page Rank value remains the same, i.e. $PR^2 = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}$

∴ Values are still equal, the rank of E and B are the same at this stage.

16) Consider the information vectors A [1, 2, 3], B [1, 0, 2], C [1, 1, 1], D [8, 3, 0] and E [5, 1, 2] are associated as the following network....



a) Aggregated Message = Average of Node Values

$A_{new} = \frac{1+2+3}{3} = 2.00$

$B_{new} = \frac{1+0+2}{3} = 1.00$

$C_{new} = \frac{1+1+1}{3} = 1.00$

$D_{new} = \frac{8+3+0}{3} = 3.67$

$E_{new} = \frac{5+1+2}{3} = 2.67$

b) Aggregated Message = $\frac{\text{Product of Node values}}{\text{Sum of Node values}}$

$A_{new} = \frac{1 \times 2 \times 3}{1+2+3} = 1.00$ $\qquad$ $B_{new} = \frac{1 \times 0 \times 2}{1+0+2} = 0$

$C_{new} = \frac{1 \times 1 \times 1}{1+1+1} = 0.33$

$E_{new} = \frac{5 \times 1 \times 2}{5+1+2} = 10/8 = 1.25$

$D_{new} = \frac{8 \times 3 \times 0}{8+3+0} = 0.00$

c) Aggregated Values = $\sqrt{\Sigma \text{node values}}$

$A_{new} = \sqrt{1+2+3} = \sqrt{6} = 2.45$

$B_{new} = \sqrt{1+0+2} = 1.73$

$C_{new} = \sqrt{1+1+1} = 1.73$

$D_{new} = \sqrt{8+3+0} = 3.32$

$E_{new} = \sqrt{5+1+2} = 2.83$

17) Calculate the updated nodes after message passing after 2 iterations for the information vectors are A [1,0,1], B [0,2,2], C [3,1,2] associated.....

Ans → A [1,0,1], B [0,2,2], C [3,1,2]

Aggregation Rule = Element-wise Cross Product of Neighbouring nodes

Ist iteration: $A_{new} = B \times C = [0,2,2] \times [3,1,2] = [0,2,4]$

$B_{new} = A \times C = [1,0,1] \times [3,1,2] = [3,0,2]$

$C_{new} = A \times B = [1,0,1] \times [0,2,2] = [0,0,2]$

2nd iteration: $A_{new}' = B_{new} \times C_{new}' = [3,0,2] \times [0,0,2] = [0,0,4]$

$B_{new}' = A_{new} \times C_{new} = [0,2,4] \times [0,0,2] = [0,0,8]$

$C_{new}' = A_{new} \times B_{new} = [0,2,4] \times [3,0,2] = [0,0,8]$

18) What is Bayesian Network? Discuss various components of Bayesian Network.

Ans → A Bayesian Network is a graphical model that represents probabilistic relation among a set of variables. Each node in the network represent a random variable, and edges represent probabilistic dependencial.

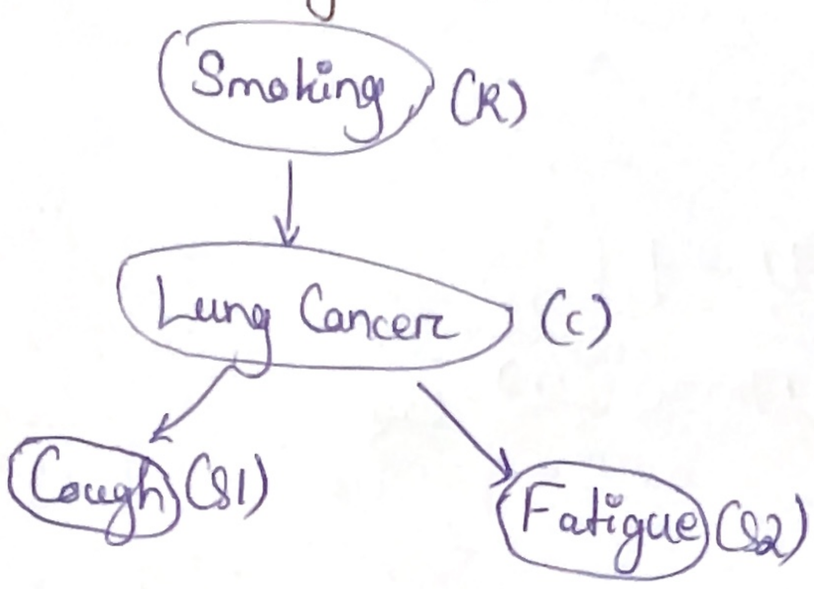
There are two different components of Bayesian Network:

i) Directed ~~Acrye~~ Acyclic Graph : This ensures that there is no closed path in the graph and the direction of edges captures the probabilistic relation between the random variables nodes.

ii) Conditional Probability Table : Each node has a conditional probability table that specifies the probability distribution of the node.

19) A doctor is diagnosing whether a patient has lung cancer (C) based on two symptoms cough (S1) and fatigue (S2).....

Ans→ a)



b) Given, $P(R=T) = 0.3$, $P(R=F) = 0.7$

Conditional Probability of lung cancer given ~~something~~ smoking :

$P(C=T \mid R=T) = 0.4$ $\quad$ $P(C=F \mid R=T) = 0.6$

$P(C=T \mid R=F) = 0.05$ $\quad$ $P(C=F \mid R=F) = 0.95$

Conditional probability of cough given lung cancer:

$P(S_1=T \mid C=T) = 0.8$, $P(S_1=F \mid C=T) = 0.2$

$P(S_1=T \mid C=F) = 0.1$, $P(S_1=F \mid C=F) = 0.9$

Conditional probability of fatigue given lung cancer:

$P(S_2=T \mid C=T) = 0.7$, $P(S_2=F \mid C=T) = 0.3$

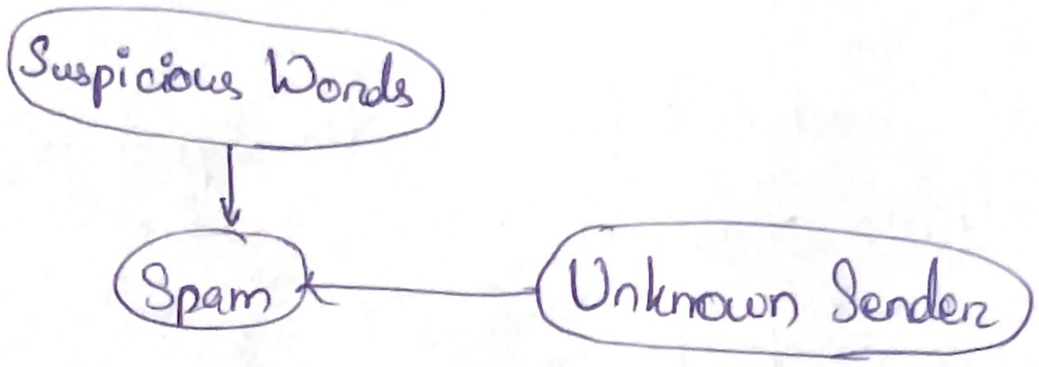
$P(S_2=T \mid C=F) = 0.2$, $P(S_2=F \mid C=F) = 0.8$

c) $$P(C=T \mid S_1=T, S_2=T) = \frac{P(S_1=T, S_2=T \mid C=T) \cdot P(C=T)}{P(S_1=T, S_2=T)} = \frac{0.56 \times 0.155}{(0.56 \times 0.155) + (0.02 \times 0.845)}$$

$$= \frac{0.0868}{0.1637} = 0.837$$

20) An email can be spam (S) based on whether it contains suspicious words (W) and/or an unknown sender (U).....

Ans → a)



b) $P(W=1) = 0.4$, $P(W=0) = 0.6$

$P(U=1) = 0.3$, $P(U=0) = 0.7$

$P(S=1 | W=1, U=1) = 0.95$, $P(S=1 | W=1, U=0) = 0.8$

$P(S=1 | W=0, U=1) = 0.7$, $P(S=1 | W=0, U=0) = 0.01$

Using Baye's Theorem,

$$P(W=1 | S=1) = \frac{P(S=1 | W=1) P(W=1)}{P(S=1)}$$

$$= \frac{(P(S=1 | W=1, U=1) P(U=1) + P(S=1 | W=1, U=0) P(U=0)) P(W=1)}{P(S=1)}$$

$$= \frac{(0.95 \times 0.3 + 0.8 \times 0.7)(0.4)}{0.506}$$

$$= \frac{0.845 \times 0.4}{0.506} = 0.668 \text{ (Ans)}$$