## (EMAI-2) ASSIGNMENT-1

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

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

3. Three cards are drawn in succession, without replacement, from an ordinary deck of playing cards. Find the probability that the event ($\mathbf{A_1 \cap A_2 \cap A_3}$) occurs, where $\mathbf{A_1}$ is the event that the first card is a red ace, $\mathbf{A_2}$ is the event that the second card is a 10 or a jack, and $\mathbf{A_3}$ is the event that the third card is greater than 3 but less than 7. **(Hint: Use Product rule)**

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. **(Hint: Use Product rule)**

5. Consider a random variable X whose probability distribution is

$$f(x) = \frac{x}{2}, 0 \leq x < 2$$

$$0\,, otherwise.$$

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).

(**Formula:** $g(y) = f\left(x(y)\right)\left|\frac{dx(y)}{dy}\right|, where\ x(y)\ is\ inverse\ function\ of\ y(x).$)

6. For the probability distribution f(x) in Q.5, if the transformed probability distribution g(y) is given by

$$g(y) = 2(1-y), 0 \leq y < 1$$

$$0, otherwise.$$

Then find the transformation function y(x).

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

8. Consider a sample consisting of the values 12, 11.2, 13.5, 12.3, 13.8, and 11.9 comes from a population with the density function

$$f(x;\,\theta) = \frac{\theta}{x^{\theta+1}}, x > 1$$

$$0, elsewhere,$$

where $\boldsymbol{\theta > 0}$. Find the maximum likelihood estimate of $\boldsymbol{\theta}$.

9. Let P and Q be two distributions on a discrete random variable X which takes the values {0, 1, 2}. P is a binomial distribution with n=2 and p=0.4, and Q is a uniform distribution on {0, 1, 2}. Find the K-L divergence of Q over P and P over Q.

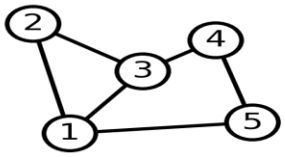
10. Consider a collection of symptoms and flu diagnoses of some patients given by the following data.

| Chills (C) | Runny nose (R) | Headache (H) | Fever (F) | Flu |
|---|---|---|---|---|
| Y | N | Mild | Y | No |
| Y | Y | No | N | Yes |
| Y | N | Strong | Y | Yes |
| N | Y | Mild | Y | Yes |
| N | N | No | N | No |
| N | Y | Strong | Y | Yes |
| N | Y | Strong | N | No |
| Y | Y | Mild | Y | Yes |

Determine whether a patient with symptoms "**Chills=Y, Runny nose =N, Headache=Mild, and Fever=Y**" has flu or not.

11. What do graphs represent in AI? Define **node features** and **edge features**.
    a. Illustrate a city's road network as a graph, where intersections are nodes and roads are edges. Label edges with features like distance, speed limits, or traffic conditions, and identify the shortest path between two points.
    b. A social media platform connects users based on friendships. Each user has attributes like age, interests, and frequency of interaction with friends. Draw a diagram representing this network, identify the **nodes** and **edges**, and list at least three **features** that influence friendships.
    c. In an online shopping platform, customers buy products, and recommendations are made based on past purchases and ratings. Represent this system as a **bipartite graph**, labelling the **nodes** (customers and products), **edges** (purchases), and at least three **features** (e.g., past purchases, ratings, browsing history).

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



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

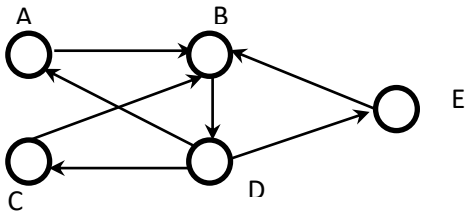
14. Consider a small web network with four pages: A, B, C, and D. The link structure is as follows: Page A links to B and C, Page B links to C, Page C links to A, and Page D links to C and A.

    a. Construct the graph and transition probability (linking) matrix for the given link structure.

    b. Compute the PageRank values after three iterations.

    c. Compute the PageRank values after three iterations when the damping factor d=0.85.

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

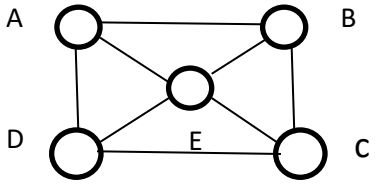



    a. Construct the transition probability (linking) matrix for the given link structure.

    b. From the matrix calculated in part a, find the rank of each page and discuss the conclusions.

    c. Perform the page rank iteration up to 2 to find the rank of nodes E and B.

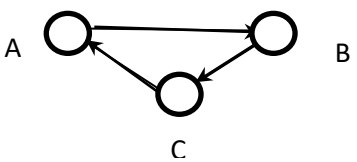
16. Consider the information vectors A[1,2,3], B[1,0,2], C[1,1,1], D[8,3,0] and E[5,1,2] are associated as the following network:




    Calculate the updated nodes after message passing, mentioning its all four step using the following as aggregated message rule:

    a. Aggregated Message(Node) = Average of node values
    b. Aggregated Message(Node) = $\frac{Product\ of\ node\ values}{sum\ of\ node\ values}$
    c. Aggregated Message(Node) = $\sqrt{sum\ of\ node\ values}$

17. Calculate the updated nodes after message passing after 2 iteration for the information vectors are A[1,0,1], B[0,2,2], C[3,1,2] associated as the following network using “Aggregated Message(Node) = Cross Product of nodes values”.



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

19. A doctor is diagnosing whether a patient has lung cancer (C) based on two symptoms: cough (S1) and fatigue (S2). The presence of lung cancer is influenced by whether the patient smokes (R). The Bayesian network structure is as follows: Smoking (R) affects Lung Cancer (C), Lung Cancer (C) affects Cough (S1) and Fatigue (S2).

    a. Draw the Directed Acyclic graph (DAG).

    b. Given the following conditional probabilities:

Smoking(R)

| P(R=T) | P(R=F) |
|---|---|
| 0.3 | 0.7 |

Lung Cancer(C)

| Smoking(R) | P(C=T) | P(C=F) |
|---|---|---|
| T | 0.4 | 0.6 |
| F | 0.05 | 0.95 |

Cough (S1)

| Lung Cancer(C) | P(S1=T) | P(S1=F) |
|---|---|---|
| T | 0.8 | 0.2 |
| F | 0.1 | 0.9 |

Fatigue(S2)

| Lung Cancer(C) | P(S2=T) | P(S2=F) |
|---|---|---|
| T | 0.7 | 0.3 |
| F | 0.2 | 0.8 |

    c. If a patient presents with cough (S1 = T) and fatigue (S2 = T), what is the probability that they have lung cancer (C)?

20. An email can be **spam** (S) based on whether it contains **suspicious words** (W) and/or an **unknown sender** (U). W(**Suspicious Words**) influences S (**Spam**) and U (**Unknown Sender**) also influences S.

    a. Draw the Directed Acyclic graph (DAG).

    b. Given probabilities:

    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)=0.1.

    If an email is marked as **spam** (S=1), what is the probability that it contained **suspicious words** (P(W=1|S=1))?