COMPUTING FOR MEDICINE
QUIZ - 3 : 07/12/2023

Instructions:

- **Duration: 20 mins**
- **Marks: 20**
- **Fill the credentials carefully and all answers should be written on answer sheet only.**

  - All MCQs carry 1 mark each
  - Descriptive Questions carry 12 marks

1. Anscombe's Quartet experiment exemplified that
   a. Correlation is not causation
   b. Similar correlation between two variables X and Y, implies similar data distributions
   c. Causation between x and y can be captured by visualizing point cloud distributions
   d. Similar correlation between two variables X and Y, does not imply similar data distributions

2. A strong positive correlation between two variables X and Y implies
   a. X causes Y
   b. Y causes X
   c. X causes Y and Y causes X
   d. Nothing can be inferred about a causal relationship between them

3. P(Smallpox | Spots) represents probability of
   a. finding Spots in Smallpox
   b. having Smallpox when No Spots were observed
   c. Having Smallpox when Spots were observed
   d. Having Smallpox

4. ____is the probability of a hypothesis before considering the observed data?
   a. Marginal Likelihood
   b. Prior Odds
   c. Bayes Factor
   d. Posterior Odds

5. A Bayesian network structure is a
    a. ***Directed acyclic graph***
    b. Undirected graph
    c. Directed graph with cycles
    d. Graph neural network

6. Which affine transformations are used to generate a new coordinate system in Principal Component Analysis (PCA)?
    a. Rotations
    b. Translations
    c. ***Both of the above***
    d. None of the above

7. Bayes' Rule provides a framework for:
    a. Estimating maximum likelihood
    b. ***Updating probabilities based on new evidence***
    c. Calculating standard deviation
    d. Establishing causation

8. Which of the following is NOT a representation learning approach?
    a. Principal Component Analysis
    b. Bayesian Network
    c. Word Embeddings
    d. ***None of the above***

9. **Enumerate the three types of 3-node structures possible in a Bayesian network (3 marks). Give one real-world example of each structure with diagrams (5 marks)**
    Answer:
    In a Bayesian network, there are three types of 3-node structures, often referred to as triplets. These structures are:

    1. **Causal Chain (Serial Connection):**
      - Structure: A -> B -> C
      - Real-world example:
        - Genetics -> Physical Traits -> Athletic Performance
        - Diagram:

          Genetics -> Physical Traits -> Athletic Performance

    2. **Common Cause (Converging Connection):**
      - Structure: A -> C <- B
      - Real-world example:
        - Parent1 -> Child <- Parent2

- Diagram:

  Parent1 -> Child <- Parent2

3. **Common Effect (Diverging Connection):**
   - Structure: A <- C -> B
   - Real-world example:
     - Economic Policy <- Government -> Social Welfare
     - Diagram:

       Economic Policy <- Government -> Social Welfare

Each structure represents a different type of relationship between the variables in a Bayesian network

10. Explain with a diagram, how does PCA achieve dimensionality reduction for visualizing high dimensional data. (4 marks)

Answer:

Principal Component Analysis (PCA) achieves dimensionality reduction for visualizing high-dimensional data through the following steps:

Getting the dataset

Firstly, we need to take the input dataset and divide it into two subparts X and Y, where X is the training set, and Y is the validation set.

Representing data into a structure

Now we will represent our dataset into a structure. Such as we will represent the two-dimensional matrix of independent variable X. Here each row corresponds to the data items, and the column corresponds to the Features. The number of columns is the dimensions of the dataset.

Standardizing the data

In this step, we will standardize our dataset. Such as in a particular column, the features with high variance are more important compared to the features with lower variance.
If the importance of features is independent of the variance of the feature, then we will divide each data item in a column with the standard deviation of the column. Here we will name the matrix as Z.

Calculating the Covariance of Z

To calculate the covariance of Z, we will take the matrix Z, and will transpose it. After transpose, we will multiply it by Z. The output matrix will be the Covariance matrix of Z.

Calculating the Eigen Values and Eigen Vectors

Now we need to calculate the eigenvalues and eigenvectors for the resultant covariance matrix Z. Eigenvectors or the covariance matrix are the directions of the axes with high information. And the coefficients of these eigenvectors are defined as the eigenvalues.

Sorting the Eigen Vectors

In this step, we will take all the eigenvalues and will sort them in decreasing order, which means from largest to smallest. And simultaneously sort the eigenvectors accordingly in matrix P of eigenvalues. The resultant matrix will be named as P*.

Calculating the new features Or Principal Components

Here we will calculate the new features. To do this, we will multiply the P* matrix to the Z. In the resultant matrix Z*, each observation is the linear combination of original features. Each column of the Z* matrix is independent of each other.

Remove less or unimportant features from the new dataset.

The new feature set has occurred, so we will decide here what to keep and what to remove. It means, we will only keep the relevant or important features in the new dataset, and unimportant features will be removed out.