Instructions

1. **Exam Duration: 1.5 hour**
2. **Maximum Marks: 50**
3. **Fill the front page with the correct details**
4. **Only one option in MCQs is correct.**
5. **MCQs must be submitted on the Answer Sheet; responses marked on the Question Paper will NOT be considered.**

MULTIPLE CHOICE QUESTIONS (1 Mark Each)   (10 Marks)

1. _____FHIR resources describe how a system does or should work.
a) Administration
b) Infrastructure
c) Conformance
d) Financial

2. What does a Z score of 3 signify in the context of the Z score?
a. The data point is three times the mean
b. The data point is three times the standard deviation
c. The data point is 3 standard deviations below the mean
d. The data point is 3 standard deviations above the mean

3. Which statement is an example of a Type I error in hypothesis testing?
a. H0 is true, and H0 is accepted.
b. H0 is true, and H0 is rejected
c. H1 is true, and H1 is accepted
d. H1 is true, and H1 is rejected

4. _____is the second order moment of a distribution.
a. Mean
b. Mode
c. Standard deviation
d. Kurtosis

5. _____is used to compare two models in Bayesian model selection
a. Posterior odds
b. Marginal Likelihood
c. Likelihood ratio
d. Bayes factor

6. Arun wanted to know the correlation between rain levels and workplace attendance. He found a correlation coefficient of -0.93. What does this tell Arun?

a. Rain levels and workplace attendance are strongly inversely correlated

b. Rain levels and workplace attendance are weakly inversely correlated

c. Rain levels and workplace attendance are strongly positively correlated

d. Rain levels and workplace attendance are not related.

7. What is the atomic unit for word representation in GloVe pre-trained word embedding?

a. Context

b. Word

c. Sentence

d. Character

8. Which of the following is not protected information in e-health records?

a. Age

b. Gender

c. Pin Code

d. State

9. In the sentence "The enigmatic Mona Lisa, housed in the Louvre Museum in Paris, captivated millions with her enigmatic smile," which of the following is NOT a named entity?

a. Louvre Museum

b. Paris

c. Enigmatic

d. Smile

10. Dimensionality reduction techniques such as PCA aid in

a. reducing the number of features without losing much information

b. An increasing number of features without losing much information

c. reducing the number of features while losing much information

d. An increasing number of features while losing much information

SHORT ANSWER QUESTIONS (10 marks)

1. **Quantiles are rank statistics. Explain. What is the special name given to 50th quantile? [2+1]**

 Answer: Quantiles are rank statistics that divide the dataset into intervals of equal probability or frequency or division of observations into four defined intervals based on the data values and how they are compared to a set of observations. the special name given to the 50th quantile is median, which separates the data into two halves.

2. **Enumerate two challenges of representing text using One Hot encoding representation. [2]**

Answer: One-hot encoding is used to represent categorical data as binary vectors. however, it has some disadvantages. Here are two challenges representing text using One Hot encoding representation:
   a. **High dimensionality**-one-hot encoding results in a high-dimensional vector representation of text data, which can cause overfitting.
   b. **Lack of Semantic Information**- It treats each word independently and ignores the semantic relationship between words.
   c. **Out of Vocabulary words, spelling mistakes etc cannot be accounted for**
   d. **Inefficienct for storing**
   e. **Sparsity** of data matrix

3. **Expand the abbreviation: FHIR. Enumerate at least two key differences between FHIR and its previous avatar of HL7 used for syntactic interoperability. [1 + 2]**

   Answer: FHIR stands for Fast Healthcare Interoperability Resources.

The key difference between FHIR and HL7 used for syntactic interoperability:
**Granularity**:
1.   **FHIR-** it is granular as it is resource-centric rather than complete documents.
2.   **HL7**-It is documentation-centric centric allows the exchange of entire clinical documents.
3.   HL7 Was meant for machine-to-machine communication. FHIR can incorporate any healthcare information as long as a resource exists. e.g. environmental factors, SDOHs etc
4.   The format of data are different. FHIR uses JSON/XML objects, HL7 is a pipe delimited text document

API Design:
   1. **FHIR-** incorporates restful API, a developer-friendly approach to the healthcare system**.**
   2. **HL7-** it works on traditional protocols, which are less flexible than the FHIR approach.

4. **What is a confounder variable? Draw the 3-node motif for the confounder effect [2]**

   Answer: It is a third variable related to both the independent and dependent variables in a study that creates a false impression of the relationship between two variables. The 3-node motif for the confounder effect is a graphical representation of three nodes for example: Smoking is the independent variable, coffee consumption acts as the confounding variable, and heart disease is the output variable.
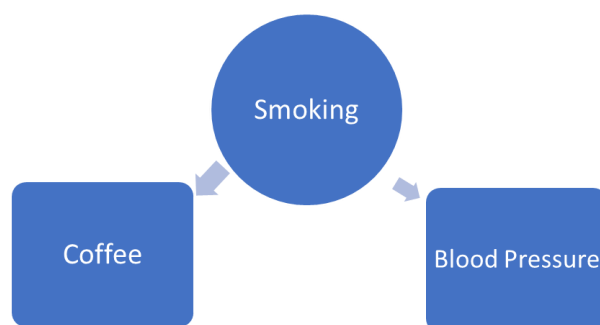
LONG ANSWER QUESTIONS (30 Marks)

1. **RT-PCR test is conducted for detection of Covid-19. The probability that tests correctly identify someone with the illness as positive is 0.99, and the probability that the test correctly identifies someone without the illness as negative is 0.95. The incidence of illnesses in the general population is 0.0001. Anita takes the test and the result is positive. What is the probability that Anita is suffering from Covid-19? [5]**

   Answer: To solve this problem, we can use Bayes' Theorem. The theorem is expressed as:

   $P(A|B) = P(B|A) \times P(A)/P(B)$ where:

   $P(A|B)$ is the probability of event A given that event B has

   occurred. $P(B|A)$ is the probability of event B given that event

   A has occurred. $P(A)$ is the probability of event A.

   $P(B)$ is the probability of event B

   $P(A)$: The incidence of Covid-19 in the general population $= 0.0001$.

   $P(B|A)$: The probability that the test correctly identifies someone with Covid-19 as positive $= 0.99$.

   $P(A')$: Anita's probability of not having Covid-19 $= 1 - P(A)$.

   $P(B'|A')$: The probability that the test correctly identifies someone without Covid-19 as negative =
   0.95. Now, we can calculate $P(A|B) = P(B|A) \times P(A)/P(B|A) \times P(A) + P(B|A') \times P(A')$

   $P(A|B) = 0.99 \times 0.0001/0.99 \times 0.0001 + (1-0.95) \times (1-0.0001)$
   $= 0.952$ is the probability that Anita has Covid-19, given the positive test result.

2. **Explain and illustrate Self Attention and Masked Attention in Transformer Models. Explain how context is modeled better by the Transformer architecture compared to the Word2Vec architecture.[5+5]**

Answer: Self-attention is a mechanism that allows each element in a sequence to focus on different elements in the same sequence, with the weights of the focus determined by the content of the elements. In the Transformer architecture, self-attention captures relationships between words in a sequence.
For example: "The cat sat on the mat." Each word can focus on every other word in self-attention, and the attention weights are dynamically computed based on the content. For instance, the word "cat" might pay more attention to "mat" in the context of this sentence.

Self attention for 'CAT':

CAT

SAT   MAT   THE
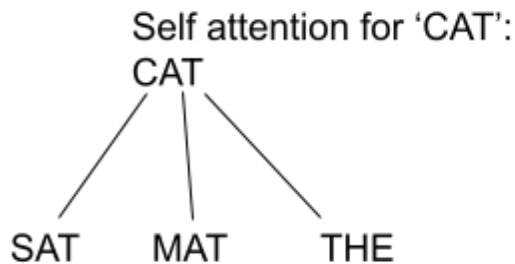
Masked attention is used during the training of Transformer models to prevent positions after the current position from being attended to. This is crucial in language modeling tasks, ensuring the model cannot access future tokens during training.

For example:In a language model, when predicting the next word in a sentence, the attention mechanism is masked so that a word can only attend to preceding words. Consider expecting the following word in the sentence "The cat sat on the mat."

During training, each word can only attend to the preceding words when predicting the masked token.

Masked attention during training:
The Cat Sat On The Mat (MASK)
| | | | | | |

The context is modeled better by the transformer architecture compared to Word2vec due to several key differences:

1. **Dynamic context modeling:** Transformer architecture employs a self-attention mechanism, allowing the model to consider the dynamic context of the sequence words whereas word2vec represents words by fixed size vector but do not capture the dynamic context of the word.

2. **Long-range dependencies:** word2vec considers fixed window of words around target word but Transformer architecture captures long-range dependencies of words. like self-attention do assign different weights to different positions in a sequence.

3. **Positional Informational**: Word2vec does not encode positional information inherently but the transformer model incorporates positional information through positional encodings.

4. **Bidirectional Context:** Word2vec only consider context words before the target word and limits the ability to capture bidirectional relationships whereas transformer models inherently capture bidirectional context.

3. **What is a Biomedical Ontology and its structure? How do Ontologies help in designing Decision Support and Decision Assist systems in Healthcare? Explain with a real-world example. [2+3+5]**

Answer: **A biomedical ontology** formally represents the concepts, entities, and relationships within a specific domain of biomedical knowledge.

Structure of Biomedical Ontology:
Biomedical ontologies typically consist of concepts (classes), relationships between these concepts, and associated attributes.

Role of Ontologies in Decision Support Systems:

> **Standardized Knowledge Representation:** Ontologies offer a standardized way to represent medical knowledge, ensuring that terms and relationships are consistently understood across different systems and applications.
>
> **Semantic Interoperability:** Ontologies enable semantic interoperability, allowing

different healthcare systems to exchange information and understand the meaning of data, facilitating collaboration and information sharing.

Decision Support and Decision Assist Systems: Ontologies form the foundation for decision support systems by providing a structured and standardized knowledge base. Decision support systems use this knowledge to assist healthcare professionals in making informed decisions.

For example, make a decision support model on any real-world example like cancer treatment with biomedical ontologies.

It should has

Patient Data Input:
- The system takes in patient-specific data, including genetic information, medical history, and diagnostic reports.

Ontology Integration:
- The Decision Support System integrates the biomedical ontology, mapping patient data to relevant ontology concepts.

Knowledge Inference:
- The system utilizes inference rules defined in the ontology to derive additional knowledge. For example, the system infers potential targeted therapy options if a patient has a specific genetic mutation.

Treatment Recommendations:
- The system provides personalized treatment recommendations based on the integrated data and inferred knowledge. For instance, it might recommend a combination of chemotherapy and targeted therapy for a patient with a certain cancer type and genetic profile.

Continuous Learning:
- The system updates its knowledge base based on the outcomes of treatments and ongoing research. This continuous learning loop enhances the system's ability to provide more accurate and up-to-date recommendations over time.

4. **The edges in a Bayesian network have directions. Imagine a simple network with two nodes, e.g. Blood Pressure and Blood Sugar. Explain with a Venn Diagram and probabilistic reasoning how a directed edge be empirically learned from data? [5]**

Answer:
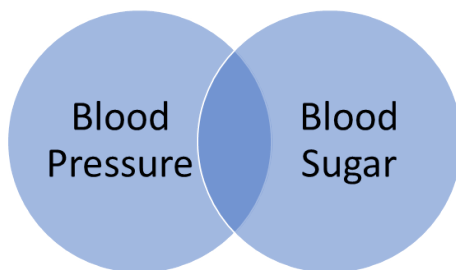
Intersection : BP ∩ BS

P(BP| BS) = P(BP ∩ BS) / P(BS)
P(BS|BP) = P(BS ∩ BP)/P(BP)
It is also observed that P(BS)< P(BP), therefore, P(BS|BP) < P(BP|BS).