

COMPUTING FOR MEDICINE
End-semester examination

Date: 4/12/2025

Max marks: 100

Time: 60 mins

Name:

Roll Num:

Multiple choice questions:

4*15 = 60

1) When should you use PCA as a preprocessing step in the machine learning pipeline?

- a) When you want to increase the number of features
- b) When you suspect multicollinearity among features**
- c) When you want to remove outliers
- d) PCA should never be used as the preprocessing step

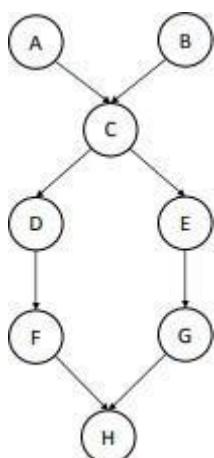
2) You have a Bayesian network with the structure:

Rain → WetGrass ← Sprinkler.

This means Rain and Sprinkler both independently influence WetGrass. If you observe that WetGrass = true, what happens to the relationship between Rain and Sprinkler?

- a) Rain and Sprinkler become independent.
- b) Rain and Sprinkler become dependent because they are “explaining away” each other.**
- c) Rain and Sprinkler remain independent because WetGrass does not affect them.
- d) Rain directly causes Sprinkler, so observing WetGrass changes only Rain.

3) Consider the following Bayesian network, and answer yes or no for the following questions,



- a) Are D and E necessarily independent given evidence about C? Yes
- b) Are A and C necessarily independent given evidence about D? No. They are directly dependent. The path A-C is not blocked**

- c) Are A and H necessarily independent given evidence about C? **Yes. All paths from A to H are blocked.**

4) Below are the two sets of statistical significance. Set-1 lists the parametric test, while set-2 provides their non-parametric counterparts. Match the two and give your answer from the following options,

Set I (Parametric Tests)		Set II (Non-parametric Tests)	
(a)	Independent t-test	(i)	Chi-square test
(b)	Dependent t-test	(ii)	Kruskal Wallis H-test
(c)	ANOVA F-test	(iii)	Mann Whitney U-test
(d)	Anand	(iv)	Wilcoxon T-test

- a) a=ii, b=i, c=iv
- b) a=iii, b=i, c=ii
- c) a=iii, b=iv, c=ii**
- d) a=ii, b=i, c=iii

5) A hospital measures systolic blood pressure in 100 patients. The standard deviation is 15 mmHg. One patient with an extremely high BP of 300 mmHg is added. Which statement best describes the impact?

- A. Both mean and standard deviation increase slightly
- B. Mean increases moderately, SD increases significantly**
- C. Mean increases drastically, SD remains almost unchanged
- D. Mean decreases, SD increases

6) Why is it necessary to standardize or normalize the predictor variables before applying Lasso or Ridge regression?

- A. To ensure the L1 and L2 penalties can be easily calculated
- B. To convert the target variable to a normal distribution
- C. To prevent the penalty from disproportionately affecting coefficients of variables with larger scales**
- D. To guarantee that the data is normally distributed

8) A Delhi hospital wants to process multilingual clinical notes (Hindi + English + Punjabi) using transformers to generate automatic medical summaries. How do embeddings help in unifying symptoms written in multiple languages?

- A. By generating separate, language-specific embedding spaces and then averaging them
- B. By focusing only on non-linguistic data and discarding descriptive text
- C. By mapping semantically similar clinical concepts into proximity within a single shared vector space**
- D. By translating all tokens into a common low-resource language before generating embeddings

9) The Ayushman Bharat Digital Mission (ABDM) wants to build a nationwide predictive healthcare system using Transformers on patient data (lab results, vitals, immunizations, prescriptions). Clinical data arrives at irregular time intervals. Which modification is crucial?

- A. Replacing standard positional encoding with a continuous, time-aware attention mechanism**
- B. Using linear interpolation of missing data to make time steps uniform
- C. Encoding the time elapsed between events as an additional time-gap embedding**
- D. Replacing multi-head attention with a simple RNN layer

10) In a dataset with highly correlated lab markers predicting diabetes, which statement is TRUE?

- A. LASSO may shrink some coefficients of correlated markers to zero, selecting one over others, while Ridge distributes weights across correlated features**
- B. Ridge may shrink some coefficients of correlated markers to zero, selecting one over others, while LASSO distributes weights
- C. Both LASSO and Ridge always zero out the same correlated features
- D. Neither LASSO nor Ridge affects correlated features

11) Bayesian Probability

Given:

$$\begin{aligned} P(\text{spots} \mid \text{smallpox}) &= 0.9 \\ P(\text{smallpox}) &= 0.001 \\ P(\text{spots}) &= 0.081 \end{aligned}$$

Find $P(\text{smallpox} \mid \text{spots})$

- A. 0.0011
- B. 0.111
- C. 0.011**
- D. 0.9

12) If $Q1 = 15$ days and $Q3 = 25$ days, what is the upper whisker limit using the $1.5 \times \text{IQR}$ rule?

- A. 30.0 days
- B. 36. days
- C. 42.5 days
- D. 40.0 days**

13) Center 1 success: young = 95%, older = 85%

Center 2 success: young = 92%, older = 80%

Overall, Center 2 appears better because it treats more young patients. Which is correct?

- A. Center 2 is objectively superior
- B. Aggregated results are misleading**

- C. Random chance explains differences
- D. Confounding cannot be identified

14) In a hospital study, Drug A appears to reduce recovery time overall.

After stratifying by age, Drug B reduces recovery time in each age group. This is an example of:

- A. Confounding
- B. Selection Bias
- C. Berkson's Bias
- D. Simpson's Paradox**

15) A hospital wants to train a Transformer-based model to extract rare drug interactions from 10 years of clinical notes. The network has 24 layers. During training, the model's performance plateaus and gradients in early layers vanish. Which modification specifically addresses this problem?

- A. Adding residual connections so gradients can flow directly across layers**
- B. Replacing token embeddings with one-hot vectors
- C. Using a smaller batch size to prevent overfitting
- D. Increasing the number of attention heads to focus on rare interactions

16) Two ICU wards record patients' heart rates. Ward A: mean≈median. Ward B: mean>median. Which statement is likely correct?

- A. Ward A has positively skewed data, Ward B has symmetric data**
- B. Ward A has symmetric data, Ward B has positively skewed data
- C. Both wards have negatively skewed data
- D. Ward A has more outliers than Ward B

Short Answer Questions

10*4= 40

1. Differentiate between correlation and causation in clinical datasets.
2. Patient vitals are recorded irregularly over time. Explain why standard machine learning models may fail on such data and one approach to model irregular time-series effectively.
3. Why is explainability crucial in AI models for healthcare? Describe one method to make predictions interpretable for clinicians.
4. Explain how Directed Acyclic Graphs (DAGs) can help identify confounders in observational healthcare studies and give a simple example.
5. What information does a boxplot contain? Show with example
6. When would you prefer a non-parametric test over a parametric test in medical data analysis? Give one example.
7. A diagnostic test has sensitivity 95% and specificity 90%. Disease prevalence is 1%. Explain how to calculate the probability that a patient actually has the disease if the test is positive.
8. Why does dividing by n in a sample underestimate the population variance? Explain in terms of "bias" and how using n-1 corrects for it.
9. What problem does the Transformer architecture solve that RNNs struggle with? What does it mean that BERT is a "bidirectional" Transformer? How does this differ from

unidirectional models like GPT? What is the function of **[CLS]** and **[SEP]** tokens in BERT? What are the main limitations of BERT.

10. A patient's EHR from Hospital A records "Type 2 Diabetes" using SNOMED CT 44054006. Hospital B uses ICD-10 codes. Explain how semantic interoperability ensures accurate understanding of this diagnosis in Hospital B. Give one potential consequence if semantic interoperability fails.

1. Differentiate between correlation and causation in clinical datasets.

In clinical data, correlation and causation describe fundamentally different relationships:

1. Nature of Relationship:

- Correlation means two variables move together statistically, but one does not necessarily cause the other.
- Causation means one variable directly produces a change in the other.

2. Evidence Required:

- Correlation is detected using statistical measures such as Pearson/Spearman coefficients.
- Causation requires controlled experiments, randomized trials, longitudinal studies, or formal causal inference.

3. Confounders:

- Correlation may arise because of hidden confounders influencing both variables.
- Causation requires identifying and adjusting for confounders to avoid biased conclusions.

4. Clinical Importance:

- Correlation helps generate hypotheses but cannot support clinical decisions.
- Causation supports safe and effective treatment decisions based on true cause–effect evidence.

2. Why standard ML models fail on irregular time-series and one approach to handle them.

Standard ML models assume evenly spaced timestamps. In clinical data, vitals such as heart rate or blood pressure are recorded at irregular intervals, creating gaps that distort model learning and temporal patterns.

Why models fail:

- They misunderstand long gaps as regular intervals.
- They cannot capture rapid deterioration or recovery trends.
- Missing or uneven points break temporal dependencies.

Effective approach:

GRU-D (Gated Recurrent Unit with Decay) incorporates time gaps directly into model computations. It learns how older measurements decay in influence and handles missing data robustly, making it ideal for irregular clinical time-series.

3. Why explainability is crucial in healthcare AI and one interpretable method.

Explainability is critical because AI-driven clinical decisions directly impact patient safety, diagnosis, and treatment. Clinicians must understand why a model makes a prediction to trust or act upon it.

Importance:

- Ensures transparency and prevents hidden biases.
- Enables clinicians to verify correctness and identify unsafe predictions.
- Supports ethical and regulatory requirements.

Interpretable Method: SHAP

SHAP (SHapley Additive exPlanations) decomposes a prediction into individual feature contributions, showing how each clinical variable increases or decreases risk. This provides clear, case-level interpretability for clinicians.

4. How DAGs help identify confounders in healthcare studies.with example

Directed Acyclic Graphs (DAGs) visually map causal relationships, helping identify confounders—variables that influence both the exposure and the outcome. Confounders create false associations if not controlled.

How DAGs help:

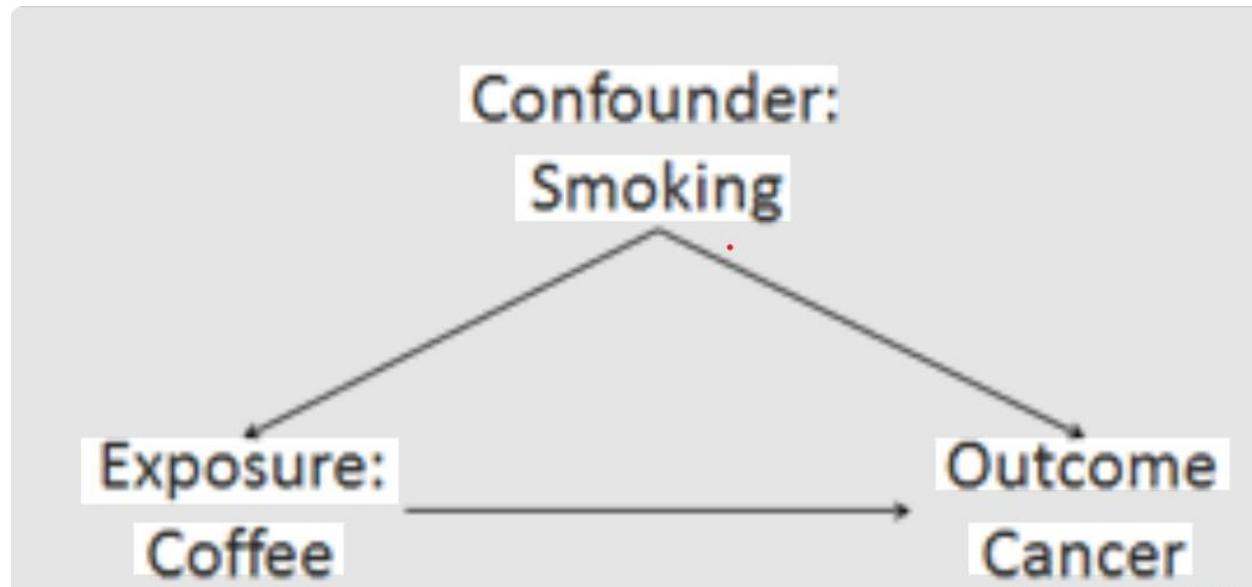
- They show causal direction using arrows.
- They identify backdoor paths that create spurious relationships.
- They guide which variables must be adjusted to obtain unbiased effects.

Example:

Studying whether exercise reduces blood sugar:

Age affects both exercise frequency and blood sugar levels. A DAG reveals age as a confounder that must be adjusted to avoid a biased estimate of exercise's true effect.

Example DAG Showing a Confounder

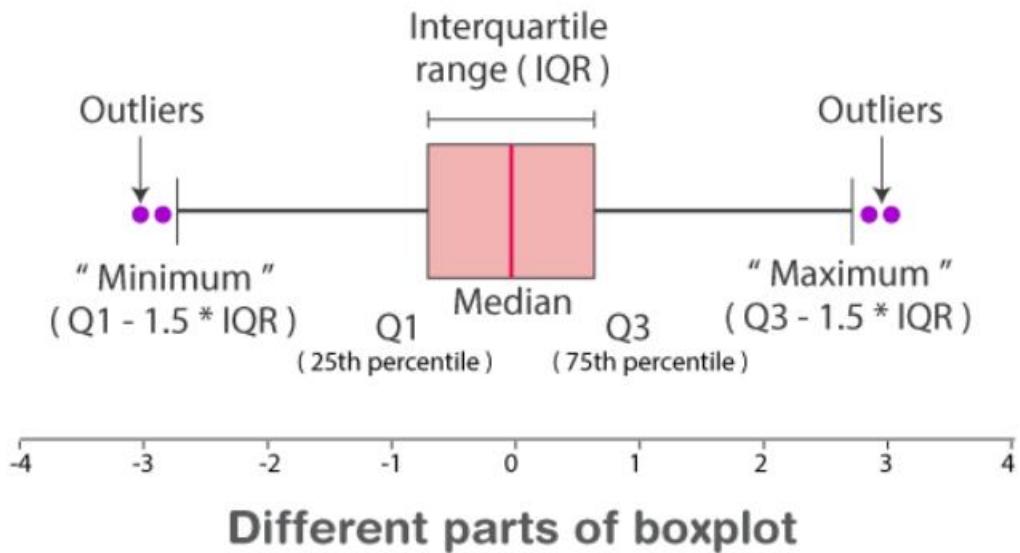


5. What information does a boxplot contain? Provide an example.

A boxplot provides a compact visual summary of numerical data using the five-number summary:

- Minimum
- First quartile (Q1)
- Median (Q2)
- Third quartile (Q3)
- Maximum

The box shows the Interquartile Range ($IQR = Q3 - Q1$), and points beyond $1.5 \times IQR$ are marked as outliers. Boxplots also help detect skewness and variability in clinical variables such as blood pressure or lab values.



Example:

In a clinical dataset of patient systolic blood pressure, a boxplot may show $Q1 = 110$ mmHg, median = for key in `llama_all_recipe_scores.keys()`:

```
print(f'{key}: {np.sum([item[0] for item in llama_all_recipe_scores[key]])}'")
```

125 mmHg, and $Q3 = 140$ mmHg. A few patients with values above 180 mmHg would appear as outliers, indicating possible cases of severe hypertension.

6. When to prefer a non-parametric test over a parametric test (with example).

Parametric tests such as the t-test or ANOVA assume that the data are normally distributed, have equal variances, and are measured on an interval or ratio scale. When these assumptions are violated, parametric results may become inaccurate or misleading.

Non-parametric tests are especially useful when the dataset is **small**, **skewed**, contains **outliers**, or is measured on an **ordinal scale** (such as Likert ratings). They do not rely on distributional assumptions and instead use ranks, making them more robust and flexible for real-world healthcare data.

Use cases:

- Data contains outliers.
- Sample size is small.
- Data are ordinal or non-normal.

Example:

Comparing pain scores (1–10) between two patient groups is better done with the **Mann–Whitney U test** instead of the t-test.

7. Calculating probability that a patient has the disease after a positive test (Bayes' theorem).

Given:

- Sensitivity = 95%
- Specificity = 90%
- Prevalence = 1%

Step 1: True positives

$$TP = 0.95 \times 0.01 = 0.0095$$

Step 2: False positives

$$FP = (1 - 0.90) \times 0.99 = 0.099$$

Step 3: Total positive tests

$$\text{Total} = 0.0095 + 0.099 = 0.1085$$

Step 4: Positive Predictive Value (PPV)

$$PPV = 0.0095 \div 0.1085 \approx 0.088$$

Final Answer: Only about **8.8%** of patients who test positive actually have the disease.

8. Why dividing by n underestimates population

variance and how $n-1$ corrects the bias.

Dividing by n underestimates variance because the sample mean is computed from the same data, making the squared deviations smaller than they would be relative to the true population mean. This downward shift results in a biased variance estimate.

Using $n - 1$ (Bessel's correction) compensates for the loss of one degree of freedom after estimating the mean, producing an unbiased estimate of population variance. This adjustment slightly increases the variance, correcting the bias.

9. BERT and Transformer Concepts.

1. What problem Transformers solve:

Transformers eliminate the long-range dependency problem that RNNs struggle with. Their self-attention mechanism allows every token to interact with all others in parallel, improving context understanding.

2. Why BERT is bidirectional:

BERT reads text from both left and right simultaneously using Masked Language Modeling, enabling deeper semantic understanding.

3. Difference from GPT:

GPT is unidirectional (left-to-right), optimized for text generation. BERT is bidirectional, optimized for understanding tasks like classification and QA.

4. Use of [CLS] and [SEP]:

- [CLS] represents the whole input sequence for classification tasks.
- [SEP] separates two sentences or marks the end of one.

5. Limitations:

- High computation cost.
- Cannot generate natural text.
- Limited sequence length (typically 512 tokens).
- Requires separate fine-tuning for each task.

10. Semantic Interoperability in Healthcare Coding Systems.

Semantic interoperability ensures that clinical meaning remains consistent across systems using different coding standards. For example, SNOMED CT code 44054006 (Type 2 Diabetes) must map accurately to ICD-10 code E11 so that the diagnosis is interpreted correctly in Hospital B.

If interoperability fails:

- The diagnosis may be misclassified.
- Incorrect treatment or medication errors may occur.
- Patient safety and clinical record accuracy may be compromised.

