COMPUTING FOR MEDICINE
RUBRIC QUIZ - 2 : 21/11/2023

MULTIPLE CHOICE QUESTIONS [10*1]

1: What best describes a z-score:
a) It is the average of all raw scores in a normal distribution
b) It is the measure of dispersion in a distribution of scores
c) It is the position of a score relative to the mean
d) It is the frequency of a score in standardized units

2 : What is the formula for calculating self attention:
a) O(n)
b) O(n2)
c) n
d) O(n log n)

3 : Calculate the 20% trimmed mean for the following values: 3,6,9,12,15,5,7,16,4,11.
a) 7.5
b) 8.3
c) 9
d) 8.5

4: Find the interquartile range for the following : Q1=46 and Q3 = 58.
a) 12
b) 10
c) 14
d) 24

5: _____ is the test performed when the data is not normalized.
a) Z-Test
b) T-test
c) Kruskal-Wallis Test
d) ANOVA

6: Which method is used to control the rate of reporting of false positives.
a) Multiple Hypothesis Testing
b) Null Hypothesis
c) Alternative Hypothesis
d) Reverse Hypothesis

7: Rhea wanted to find if there's a relationship between the foot-length and height of the population of her district. She found a correlation value of -0.69. What does this signify?

a) **Height and foot-length are strongly inversely correlated**
b) Height and foot-length are weakly inversely correlated.
c) Height and foot-length are strongly positively correlated
d) Height and foot-length are not related.

8: What is the relationship between mean, median, mode for positive skewness.
a) Mean < Median < Mode
b) Mean > Median > Mode
c) Mean = Median = Mode
d) Mean = Median > Mode

9: What does kurtosis explain in terms of distribution :
a) It explains the amount of variability in the data.
b) It measures the flatness of the distribution of data.
c) It explains the symmetry of the data.
d) It explains the average of the distribution of the data.

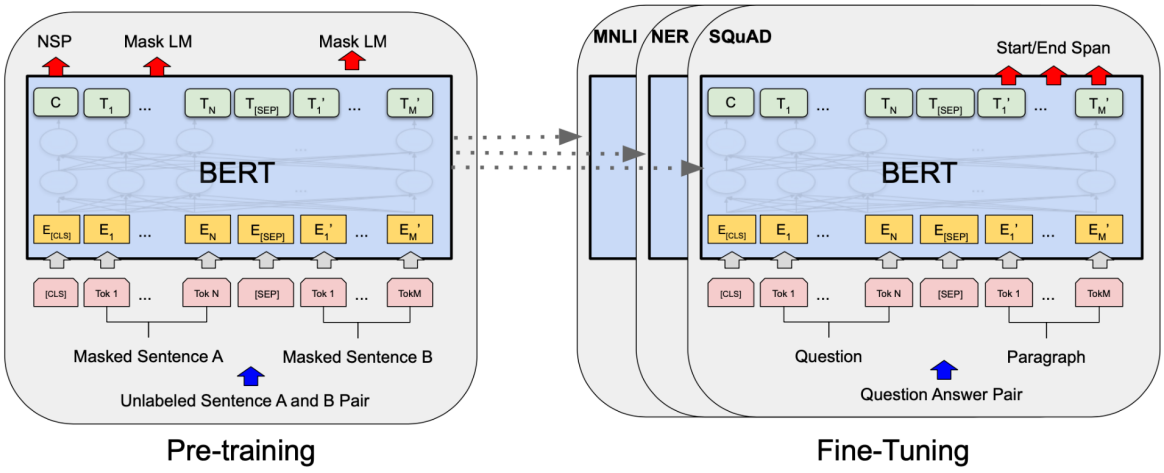
10: Which of the following is not true about the 'dropout' ?
a) Dropout is applied to the output of each sub-layer.
b) It is done before being added to the sub-layer and normalized.
c) It is applied dropout to the sums of embeddings and positional encodings in both encoder and decoder stacks.
d) It is applied dropout only to the sums of embeddings.

11: Explain the difference between query, key, and value matrices. (2+2+2)

Query (Q): Represents information about the input that is used to attend to the relevant parts of the sequence.
Key (K): Represents information about the input that is used to determine the importance of different parts of the sequence.
Value (V): Represents the actual content or information for each position in the sequence.

These matrices are essential components of the self-attention mechanism in transformers, allowing the model to capture dependencies and relationships between different parts of the input sequence.

12: Draw a schematic diagram for the basic architecture of a BERT model. Also give an example of the BERT model. (3+1)

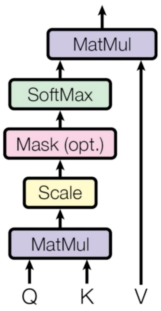


Example: Bio-BERT, Clinical-BERT, DNA-BERT,etc.

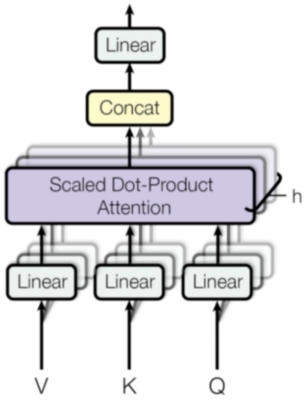
13: Short answer questions: (2+2+2+2+2)

a) Softmax Function

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$



b) Multi-head attention



c) Positional Encoding with example.

Positional encoding is a technique used in the self-attention mechanism of transformer models to inject information about the positions of words or tokens in a sequence. Positional encoding is designed to supplement these embeddings with information about the positions of the tokens in the sequence. The most common approach for positional encoding involves adding a fixed vector to the original word embeddings. The added vector is calculated based on the position of the token in the sequence.

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/dmodel})$$
$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/dmodel})$$

d) Feature Engineering

- FE is a representation problem
- Adjust the representation of the data to improve the efficacy of the ML algorithms. Combination of domain knowledge and ML method itself.
- It is difficult, expensive, and time-consuming
- Also called data munging or data wrangling

e) Parametric vs Non-parametric tests:

Parametric tests are based on the assumption that the data comes from a population with a known distribution, such as a normal distribution. This means that the parameters of the distribution, such as the mean and standard deviation, are known. Example: T-test, ANOVA, etc. Non-parametric tests do not make any assumptions about the distribution of the data. This means that they can be used with data that is not normally distributed or with data for which the distribution is unknown. Example: Kruskal-Wallis Test, etc.