MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
   A) Least Square Error B) Maximum Likelihood
   C) Logarithmic Loss D) Both A and B
   Answer: A) Least Square Error

2. Which of the following statement is true about outliers in linear regression?
   A) Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers
   C) Can't say D) none of these
   Answer: A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is_____?
   A) Positive B) Negative
   C) Zero D) Undefined
   Answer: B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?
   A) Regression B) Correlation
   C) Both of them D) None of these
   Answer: B) Correlation

5. Which of the following is the reason for over fitting condition?
   A) High bias and high variance B) Low bias and low variance
   C) Low bias and high variance D) none of these
   Answer: C) Low bias and high variance

6. If output involves label then that model is called as:
   A) Descriptive model B) Predictive model
   C) Reinforcement learning D) All of the above
   Answer: B) Predictive model

7. Lasso and Ridge regression techniques belong to_____?
   A) Cross validation B) Removing outliers
   C) SMOTE D) Regularization
   Answer: D) Regularization

8. To overcome with imbalance dataset which technique can be used?
   A) Cross validation B) Regularization
   C) Kernel D) SMOTE
   Answer: D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses_____to make graph?
   A) TPR and FPR B) Sensitivity and precision
   C) Sensitivity and Specificity D) Recall and precision
   Answer: A) TPR and FPR-- It is a probability curve that plots the **TPR** against **FPR** at various threshold values and essentially **separates the 'signal' from the 'noise.'**

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
    A) True B) False
    Answer: B) False-- The higher the AUC, the better the model's performance at distinguishing between the positive and negative classes.

MACHINE LEARNING

11. Pick the feature extraction from below:
    A) Construction bag of words from a email
    B) Apply PCA to project high dimensional data
    C) Removing stop words
    D) Forward selection

Answer: B) Apply PCA to project high dimensional data

In Q12, more than one options are correct, choose all the correct options:

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
    A) We don't have to choose the learning rate.
    B) It becomes slow when number of features is very large.
    C) We need to iterate.
    D) It does not make use of dependent variable.

Answer: A), B)-- there is no need to iterate but in the option c) it states we need to iterate which is not correct so only --A) We don't have to choose the learning rate (**and**) B) It becomes slow when number of features is very large are correct options.

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Answer: The term "regularization" describes methods for calibrating machine learning models to reduce the adjusted loss function and avoid overfitting or underfitting. Regularization allows us to properly adapt our machine learning model to a specific test set, hence lowering the errors in the test set. The machine learning model may occasionally perform well with training data but poorly with test data. When dealing with unseen data by introducing noise in the output, it means the model is unable to anticipate the result and is therefore referred to as being overfitted. The use of a regularization approach can solve this issue.

By lowering the magnitude of the variables, this strategy can be applied to keep all variables or features in the model. Consequently, it keeps the model's generalization and accuracy. The coefficient of features is mostly regularized or reduced towards zero. To put it plainly, "In regularization technique, we reduce the magnitude of the features by keeping the same number of features."

14. **Which particular algorithms are used for regularization?**

Answer: There are 2types of regularization Lasso regularization and ridge regularization.

Ridge Regularization: It is also referred to as Ridge Regression and modifies over- or under-fitted models by applying a penalty equal to the sum of the squares of the coefficient magnitude. As a result, coefficients are calculated and the mathematical function that represents our machine learning model is minimized. The coefficients' magnitudes are squared and summed. Ridge Regression applies regularization by reducing the number of coefficients.

Cost function = Loss + λ x∑||w||^2

Loss= Sum of the squared residuals

λ= Penalty for the errors

W= slope of the curve/line

Lasso Regularization: By imposing a penalty equal to the total of the absolute values of the coefficients, it alters the models that are either overfitted or underfitted. Lasso regression likewise attempts coefficient minimization, but it uses the actual coefficient values rather than squaring the magnitudes of the coefficients. As a result of the occurrence of negative coefficients, the coefficient sum can also be 0.

MACHINE LEARNING

Cost function = Loss + λ x ∑‖w‖

Loss= Sum of the squared residuals

λ= Penalty for the errors

W= slope of the curve/line

15.Explain the term error present in linear regression equation?

In regression equation, an error term is a measure of how the observed data differs from the actual data. It may also represent a variable which reflects how a statistical model differs from reality. The error term is also written as ε. In scatter plots, we look for the line $y = \beta_0 + \beta_1 x$ that best fits the data. "Best fit" in this context refers to the error term and the separation between each point and the line.

The graph won't actually cover all of our data points because the relationship between variables is probably not entirely linear and because there are other factors. The distance between graph and each term is the error term. We can write the equation as, $R^B=\beta_0 + \beta_1 E^x + \varepsilon$ where β_0 and β_1 are constants and ε is an (non-constant) error term. Everything that differentiates our model from reality is included in the error term. As a result, it will take into account nonlinearities, unpredictable outcomes, measurement flaws, and omitted factors.

Even though the words "error" and "residual" are sometimes used interchangeably, there is a significant formal distinction. A residual shows how observed data differs from sample data, whereas an error term shows how observed data differs from the actual data. As a result, a residual is frequently considerably simpler to quantify. A residual can be seen, whereas errors are typically invisible.
You may think of the residual as an estimation of the real error term.

MACHINE LEARNING