Practice Questions

Q1 – Imagine a graph where the horizontal axis represents an independent variable, and the vertical axis represents a dependent variable. In the methodology of linear regression, specifically when fitting a line using the least squares method, which of the following offsets do we use in linear regression least squares line fit?

*A. Vertical offset
B. Perpendicular offset
C. Both, depending on the situation
D. None of the above

Q2 – What does the Mean Absolute Error (MAE) in a linear regression model signify?

*A. The average of the absolute values of the vertical offsets between the predicted and actual values.
B. The average squared difference between the predicted and actual values.
C. The square root of the average squared differences between the predicted and actual values.
D. The total sum of all errors in the prediction model.

Q3 – Within Python's machine learning libraries, select from the following the purpose of the 'random_state' parameter in the train_test_split function found in the scikit-learn (sklearn) Python library.

A. It defines the proportion of the dataset to include in the test split.
B. It determines the variables that will be used as predictors in the model.
*C. It sets a seed for the random number generator to allow reproducibility, ensuring the same result regardless of the number of executions.
D. It specifies the number of folds to use in cross-validation.

Q4 – Data was collected on two variables: x, and y, and fitted with a least squares regression line. The resulting equation is y=?2 + 2x ; what is the residual error for point (3, 7)?

A. -2
B. 1
*C. 3
D. 7

Q5 – In multiple linear regression, what does the term 'multicollinearity' refer to?

A. The condition where multiple regression coefficients are equal to zero.
B. The occurrence where the independent variables are not correlated with the dependent variable.
*C. The phenomenon where two or more independent variables in a model are highly correlated with each other.
D. The situation where the regression model perfectly fits the data.

Q6 – In simple linear regression, which of the following best describes the role of the coefficient of determination (R^2)?
*A. It measures the proportion of variance in the dependent variable that can be predicted from the independent variable.
B. It indicates the strength of the correlation between the independent and dependent variables.
C. It is the slope of the regression line and indicates the strength of the relationship.
D. It represents the average distance of the data points from the fitted regression line.

Q7 – Consider a scenario with a high-variance model. Which of the following issues is it most likely to encounter?
A. Inability to capture any underlying patterns in the data.
B. Tendency to oversimplify the model's relationship with the data.
*C. Overfitting to the training data, capturing noise along with the underlying pattern.
D. Consistently underperforming across both training and new data.

Q8 – When choosing between L1 and L2 Loss Functions, which of the following statements is most accurate regarding their sensitivity to outliers?
A. Both L1 and L2 are equally sensitive to outliers.
B. L1 is more sensitive to outliers because it considers the squared differences.
*C. L2 is more sensitive to outliers due to the squared term in its calculation.
D. Neither L1 nor L2 is affected by outliers in any significant way.

Q9 – In regression analysis, what is the primary goal of employing feature selection methods like best subset selection, forward stepwise selection, and backward stepwise selection?
A. To enhance the speed of the regression model.
B. To construct a model with the highest number of variables.
C. To identify which variables are most closely dependent variables.
*D. To select the most relevant features that ensure a balance between a model's accuracy and its complexity.

Q10 – In a robotic learning system where the task is observing the environment, making decisions, taking actions, and getting rewards, what is the primary objective of the learning agent?
A. To minimize the reward signal over time through a series of actions.
*B. To learn a policy that maximizes the positive reward by observing the environment and performing actions.
C. To label the data manually before processing.
D. To simplify the data into a single, manageable category.