CIA – 1 MACHINE LEARNING ASSINGMENT

NAME : Brejesh V D

ROLL NUMBER : 22011103007

1. Line: In machine learning, a "line" could refer to a decision boundary or a linear model. A decision boundary separates different classes in classification problems, while a linear model represents a relationship between input features and output in a linear form.

2. Box: In machine learning, a "box" might refer to a decision tree or a bounding box. A decision tree is a hierarchical model that makes decisions based on the values of input features, and a bounding box can be used in object detection tasks to enclose the area of interest around an object.

3. Histogram: In machine learning, a "histogram" might be used to visualize the distribution of a feature or target variable in a dataset. It helps in understanding the frequency of different values.

4. Distribution: In machine learning, a "distribution" refers to the probability distribution of a random variable. Understanding the distribution of data is essential for making informed decisions about which machine learning models and algorithms to use.

5. Heat Map: In machine learning, a "heat map" could be used to visualize the correlation matrix between features or to highlight areas of interest in a matrix, such as feature importance in a model.

6. Mean, S.D., and Variance: In machine learning, the "mean" is often used to describe the average performance or accuracy of a model. Standard deviation (S.D.) and variance can be used to measure the variability or dispersion of model performance across different runs or datasets.

7. Covariance: In machine learning, "covariance" can be important for understanding relationships between different features in a dataset. High covariance between features might indicate redundancy.

8. Correlation: In machine learning, "correlation" is used to quantify the strength and direction of a linear relationship between two variables. It helps in feature selection and understanding feature importance.

9. P-value: While p-values are more common in traditional statistics, in machine learning, they might be used in hypothesis testing or model evaluation to determine the significance of certain features or model improvements.

10. Bias-Variance: In machine learning, "bias" refers to the error introduced by approximating a real-world problem, and "variance" refers to the model's sensitivity to variations in the training data. The bias-variance trade-off is crucial in model selection and training.

11. SMOTE (Synthetic Minority Over-sampling Technique): In machine learning, SMOTE is a technique used to address class imbalance by generating synthetic samples for the minority class during training.

12. Standardization (Min-Max Scaler): In machine learning, "standardization" or "Min-Max scaling" is a preprocessing step to bring different features to a similar scale, ensuring that no feature dominates others during model training.

13. Normalization (Z-score): Normalization, often using Z-score, is a technique in machine learning to scale features to have zero mean and unit variance, aiding in convergence during training.

14. Pandas Syntax: In machine learning, Pandas is a popular Python library for data manipulation and analysis. Pandas syntax includes various functions and methods for loading, cleaning, and exploring datasets.

15. Encoding Syntax: In machine learning, encoding syntax refers to techniques used to convert categorical data into numerical format. Common methods include one-hot encoding, label encoding, or ordinal encoding.

16. Outlier: In machine learning, an "outlier" is a data point that significantly deviates from the rest of the data. Handling outliers is important to prevent them from unduly influencing model training.

17. Unilayer Multilayer: In machine learning, "unilayer" and "multilayer" are not standard terms. However, they could potentially refer to single-layer and multi-layer neural networks, respectively, where layers represent the number of hidden layers in the network.

18. Regression Line: In machine learning, a "regression line" represents the relationship between input features and the predicted output in regression problems. It is the line that best fits the data according to the chosen regression model.

19. Simple Linear Regression (Dependent and Independent): In machine learning, "simple linear regression" involves predicting a dependent variable based on a single independent variable. The dependent variable is the one being predicted, and the independent variable is the one used for prediction.

20. R-Square: In machine learning, R-squared (R^2) is a metric used to measure the goodness of fit of a regression model. It indicates the proportion of the variance in the dependent variable that is predictable from the independent variables.

21. Least Squares: In machine learning, "least squares" is a method used in linear regression to minimize the sum of the squared differences between the observed and predicted values.

22. Homoscedasticity: In machine learning, "homoscedasticity" refers to the assumption that the variance of the errors (residuals) is constant across all levels of the independent variables in a regression model. It is an important assumption for reliable model predictions.