

MACHINE LEARNING

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

Ans: D) Both A and B

2. Which of the following statement is true about outliers in linear regression?

Ans: A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?

Ans: B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?

Ans: C) Both of them

5. Which of the following is the reason for over fitting condition?

Ans: C) Low bias and high variance

6. If output involves label then that model is called as

Ans: B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

Ans: D) Regularization

8. To overcome with imbalance dataset which technique can be used?

Ans: D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

Ans: A) TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less

Ans: B) False

11. Pick the feature extraction from below:

Ans: A) Construction bag of words from a email

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

Ans: B) It becomes slow when number of features is very large.

13. Explain the term regularization?

Ans: Regularization is a technique in machine learning that adds a penalty to the model's complexity during training. It prevents over fitting by discouraging the model from fitting the training data too closely. This helps the model generalize better to new, unseen data. There are different types of regularization, such as L1 (Lasso) and L2 (Ridge), which control the impact of coefficients in the model and help balance between accuracy and simplicity. Regularization is used to create more robust and reliable models.

14. Which particular algorithms are used for regularization?

Ans: Algorithms commonly used for regularization include:

- **Lasso Regression:** Uses L1 regularization for feature selection.
- **Logistic Regression (with L1/L2):** Prevents over fitting in classification.
- **Support Vector Machines (SVM):** Regularizes decision boundaries.
- **Decision Trees:** Simplifies trees for better generalization.
- **Gradient Boosting (XGBoost, LightGBM):** Includes regularization parameters.

These techniques balance model complexity and data fitting, enhancing generalization.

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

Ans: In linear regression, "error" refers to the difference between the actual observed values and the values predicted by the regression equation. It measures how well the model fits the data. The goal is to minimize these errors to find the best-fitting line that represents the relationship between variables.