

Machine Learning

1. which of the following methods do we use to find the best fit line for data in Linear Regression?
Least Square Error
2. Which of the following statement is true about outliers in linear regression?
Linear regression is sensitive to outliers
3. A line falls from left to right if a slope is _____?
Negative
4. Which of the following will have symmetric relation between dependent variable and independent variable?
Correlation
5. Which of the following is the reason for over fitting condition?
Low bias and high variance
6. If output involves label, then that model is called as:
All of the above
7. Lasso and Ridge regression techniques belong to _____?
Regularization
8. To overcome with imbalance dataset which technique can be used?
SMOTE
9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?
TPR and FPR
10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
False
11. Pick the feature extraction from below:
Apply PCA to project high dimensional data
12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
We don't have to choose the learning rate.
It becomes slow when number of features is very large.
13. Explain the term regularization?
Regularization is a technique used in regression to reduce the complexity of the model and to shrink the coefficients of the independent features. In simple words, this technique converts a complex model into a simpler one, so as to avoid the risk of overfitting and shrinks the coefficients, for lesser computational cost.
14. Which particular algorithms are used for regularization?
 1. Ridge Regression
 2. LASSO (Least Absolute Shrinkage and Selection Operator) Regression
 3. Elastic-Net Regression
15. Explain the term error present in linear regression equation?

The linear regression model contains an error term that is represented by ϵ . The error term is used to account for the variability in y that cannot be explained by the linear relationship between x and y . If ϵ were not present, that would mean that knowing x would provide enough information to determine the value of y .