

Machine Learning Worksheet

Q1.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

A1.Least Square Error

Q2.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

A2.Linear regression is sensitive to outliers

Q3.A line falls from left to right if a slope is _____?

- A) Positive B) Negative C) Zero D) Undefined

A3.Negative

Q4.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

A4.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

A5.Low bias and high variance

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

- A) Descriptive model B) Predictive modal C) Reinforcement learning D) All of the above

A6. Predictive Model

Q7.Lasso and Ridge regression techniques belong to _____? A) Cross validation
B) Removing outliers C) SMOTE D) Regularization

A7.Regularization

Q8.To overcome with imbalance dataset which technique can be used? A) Cross validation
B) Regularization C) Kernel D) SMOTE

A8. Cross Validation

Q9.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

A9.TPR and FPR

Q10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

A) True B) False

A10. True

Q11. 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

A11. Except forward selection we can use all other 3 methods for feature extraction that is A ,B and C .

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

Q12. 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.

A12. A,B,C are correct

Q13.Explain the term regularization?

A13. Regularization is the technique used to reduce the error by fitting a function on a given training set and avoid overfitting.

Q14.Which particular algorithms are used for regularization?

A14. There are two main algorithms used for regularization:-

1.L1 (LASSO) Regularization- It used for absolute values for Normalization

2.L2(RIDGE)Regularization- Ridge regression is one of the types of linear regression in which we introduce a small amount of bias, known as Ridge regression penalty so that we can get better long-term predictions.

Q15. Explain the term error present in linear regression equation?

A15. The term error present in linear regression equation tells the difference between theoretical value of the model and the observed results.

