

### Machine Learning -worksheet 1

Q.1. 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

Solution : a) Least Square Error

Q.2. 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

Solution : a) Linear regression is sensitive to outliers

Q.3. A line falls from left to right if a slope is \_\_\_\_\_?

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

Solution : b) Negative

Q.4. 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

Solution : b) Correlation

Q.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

Solution : c) Low bias and High Variance

Q.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

Solution : b) Predictive Model

Q.7. Lasso and Ridge regression techniques belong to \_\_\_\_\_?

- A) Cross validation
- B) Removing outliers
- C) SMOTE
- D) Regularization

Solution : d) Regularization

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

- A) Cross validation
- B) Regularization
- C) Kernel
- D) SMOTE

Solution : a) Cross Validation

Q.9. 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

Solution : a) TPR and FPR.

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

- A) True
- B) False

Solution : b) False

Q.11. 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

Solution : b) Apply PCA to project high dimensional data

Q.12. 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.

Solution : a) and b)

Q.13. Explain the term regularization?

Solution : Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it. Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique.

It mainly regularizes or reduces the coefficient of features toward zero. In simple words, "In regularization technique, we reduce the magnitude of the features by keeping the same number of features."

Q.14. Which particular algorithms are used for regularization?

Solution : There are two main types of regularization techniques: Ridge Regularization and Lasso Regularization.

Ridge Regularization :

Also known as Ridge Regression, it modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

This means that the mathematical function representing our machine learning model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added. Ridge Regression performs regularization by shrinking the coefficients present.

Lasso Regression :

It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients.

Lasso regression also performs coefficient minimization, but instead of squaring the magnitudes of the coefficients, it takes the true values of coefficients. This means that the coefficient sum can also be 0, because of the presence of negative coefficients.

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

Solution : It is often said that the error term in a regression equation represents the effect of the variables that were omitted from the equation.

The general linear regression model can be stated by the equation:

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i$$

where  $\beta_0$  is the intercept,  $\beta_i$ 's are the slope between  $Y$  and the appropriate  $X_i$ , and  $\epsilon$  (pronounced epsilon), is the error term that captures errors in measurement of  $Y$  and the effect on  $Y$  of any variables missing from the equation that would contribute to explaining variations in  $Y$ .