

# Worksheet 1 answers

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

Ans- least square method

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

Ans - Linear regression is sensitive to outliers

- 3) A line falls from left to right if a slope is \_\_\_\_\_?

Ans – negative

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

Ans – correlation

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

Ans - Low bias and high variance

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

Ans - Predictive model

- 7) Lasso and Ridge regression techniques belong to \_\_\_\_\_?

Ans – Regularization

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

Ans – SMOTE

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

Ans - TPR and FPR

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

Ans – False

11) Pick the feature extraction from below:

Ans - 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?

- Ans - We don't have to choose the learning rate
- Ans - It becomes slow when number of features is very large.

13) Explain the term regularization?

Ans – One of the major aspect of training your machine learning model is avoiding overfitting. The model will have low accuracy if it is overfitting. Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting. Regularization is a form of regression that regularizes or shrinks the coefficient estimates towards zero.

The commonly used regularization techniques are:

- **L1 regularization** – A regression model that uses L1 regularization technique is called **LASSO** (Least absolute shrinkage and selection operator) regression.
- **L2 regularization** – A Regression model that uses L2 regularization technique is called **RIDGE REGRESSION**

14)

Which particular algorithms are used for regularization?

- Ans – **LASSO (Least absolute shrinkage and selection operator):**

It is used over regression methods for a more accurate prediction. This model uses shrinkage (where data point values are shrunk towards a central point as the mean). This regularization type can result in sparse model with few coefficients might become zero and get eliminated from the model.

#### **Mathematical Equation for LASSO**

Residual Sum Of Squares +  $\lambda$  \* (Sum of the absolute value of the magnitude of coefficients)

$$\sum_{i=1}^n (y_i - \sum_j x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Where

$\lambda$  = amount of shrinkage

- **RIDGE REGRESSION:**

**Ridge regression is a technique** Which is used for analyzing multiple regression where the data suffers from Multicollinearity. The problem that occurred occurred due to multicollinearity is that the basic linear regression model becomes unbiased and the variance become so high that the predicted values are far from the true value.

the advantage of using ridge regression is to avoid overfittings.

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

Ans -  $\hat{Y} = mx + b$

$Y = mx + b + \text{error}$

Error = actual – predicted

The difference between the actual value and the predicted value. When the actual y differs from the predicted y in the model during test then the error term does not equal to 0.

