**Exercises**

**Multiple Choice Questions:**

1. **Which one of the following Python code assigns column names to a data frame in Python?**

a. columns = name([‘Col\_1’, ‘Col\_2’])

b. df.index = [‘Col\_1’, ‘Col\_2’ ]

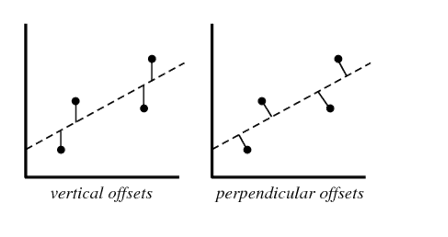
c. a = df.names([‘Col\_1’, ‘Col\_2’])

d. df.columns = ['Col\_1', 'Col\_2']

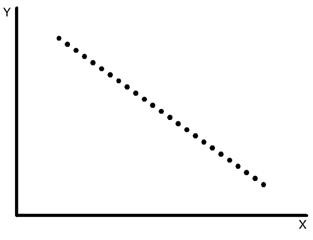
1. **What is the main difference between a Python del() and Pandas drop() function?**
2. del() operates on series only while drop() operates on data frames
3. del() operates on rows only while drop() operates only on columns
4. del() operates on only one item at a time while drop() operates on multiple items
5. del() operates on multiple items simultaneously while drop() on only one at a time
6. **Which library of Python does the merge() function belong to?**
7. Pandas
8. Matplotlib
9. SciPy
10. NumPy
11. **Which of the following methods do we use to find the best fit/Regression line for data in Linear Regression?**
12. Random Forest Regression
13. Least Square Error
14. Logarithmic Loss
15. Mean Squared Error
16. **Suppose that you found the correlation coefficient for the independent variable X with Y in a particular data set to be -0.95. Which of the following is true for X?**
17. Relation between the X and Y is weak
18. Relation between the X and Y is strong
19. Relation between the X and Y is neutral
20. Correlation can’t judge the relationship
21. **Which of the following statistical detail is not included in the describe() function of Pandas?**
22. Variance
23. Mean
24. Percentiles
25. Minimum & Maximum value
26. **What does the following code do?**

****

1. Prints the datatypes of each row of the data set
2. Returns the datatype of the whole series data
3. Prints the datatypes of each column of the data set
4. Asks the user for inputting values for the data frame
5. **If one wants to print the first and last five rows of a data frame, which of the following function would he/she prefer?**
6. data.head(5)
7. data.head(-1)
8. data.display(5, -5)
9. print(data, 5)
10. **Which of the following offsets, do we use in linear regression’s least square line fit? Suppose horizontal axis is independent variable and vertical axis is dependent variable.**

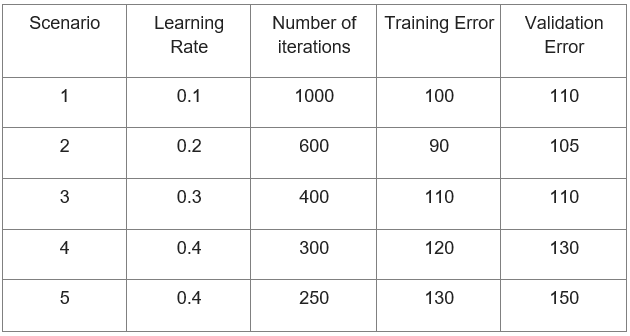


1. Vertical offset
2. Perpendicular offset
3. Both, depending on the situation
4. None of above
5. **Suppose you have fitted a complex regression model on a dataset. Now, you are using Ridge regression with penalty x. Choose the option which describes bias in best manner.**
6. In case of very large x; bias is low
7. In case of very large x; bias is high
8. We can’t say about bias
9. None of these
10. **Which of the following statement is true about outliers in Linear regression?**
11. Linear regression is sensitive to outliers
12. Linear regression is not sensitive to outliers
13. Can’t be described/not enough information
14. None of these
15. **Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and you found that there is a relationship between them. Which of the following conclusion do you make about this situation?**
16. Since the there is a relationship means our model is not good
17. Since the there is a relationship means our model is good
18. Can’t say anything/not enough information
19. None of the above
20. **Consider the following data where one input(X) and one output(Y) is given.**



**What would be the root mean square training error for this data if you run a Linear Regression model of the form (Y = A0+A1X)?**

1. Less than zero
2. Greater than zero
3. Equal to zero
4. None of the above
5. **Suppose you have been given the following scenario for training and validation error for Linear Regression.**



**Which of the following scenario would give you the right hyper parameter?**

1. 1
2. 2
3. 3
4. 4