Files

1) When implementing linear regression of some dependent variable y on the set of independent variables $\mathbf{x} = (x_1, \ldots, x_r)$, where r is the number of predictors, which of the following statements will be true?

a) $\beta_0, \beta_1, \ldots, \beta_r$ are the regression coefficients.

b) Linear regression is about determining the best predicted weights by using the method of ordinary least squares.

c) E is the random interval

d) Both and b

Answer: Both A and B are correct ($\beta_0, \beta_1, \ldots, \beta_r$ are the regression coefficients and Linear regression is about determining the best predicted weights by using the method of ordinary least squares.)

2) What indicates that you have a perfect fit in linear regression?

a) The value $R^2 < 1$, which corresponds to SSR = 0

b) The value $R^2 = 0$, which corresponds to SSR = 1

c) The value $R^2 > 0$, which corresponds to SSR = 1

d) The value $R^2 = 1$, which corresponds to SSR =0

Answer: The value $R^2 = 1$, which corresponds to SSR =0 since the values of predicted and actual responses fit completely to each other.

3) In simple linear regression, the value of what shows the point where the estimated regression line crosses the y axis?

a) Y

b) B0

c) B1

d) F

Answer: The correct answer is (b) BO

4) Check out these four linear regression plots:

Which one represents an under fitted model?

a) The bottom-left plot

b) The top-right plot

c) The bottom-right plot

d) The top-left plot

Answer: The top left plot shows a linear Regression line that has a low R^2

5) There are five basic steps when you're implementing linear regression:

• a. Check the results of model fitting to know whether the model is satisfactory.

• b. Provide data to work with, and eventually do appropriate transformations.

• c. Apply the model for predictions.

• d. Import the packages and classes that you need.

• e. Create a regression model and fit it with existing data.

However, those steps are currently listed in the wrong order. What's the correct order?

a) e, c, a, b, d

b) e, d, b, a, c

c) d, e, c, b, a

d) d, b, e, a, c

Answer: The correct answer is (d) d,b,e,a,c

6) Which of the following are optional parameters to Linear Regression in scikit-learn?

a) Fit

b) fit_intercept

c) normalize

d) copy_X

e) n_jobs

f) reshape

Answer: from sklearn.linear _model import LinearRegression

Mode1=LinearRegression (fit_intercept=True, normalize=false, copy_X=True, n_jobs=None, reshape=True)

7) While working with scikit-learn, in which type of regression do you need to transform the array of inputs to include nonlinear terms such as x^2?

a)Multiple linear regression

b) Simple linear regression

c) Polynomial regression

Answer: b) Simple linear regression

8) You should choose stats models over scikit-learn when:

A) You want graphical representations of your data.

b) You're working with nonlinear terms.

c) You need more detailed results.

d) You need to include optional parameters.

Answer: A) You want graphical representations of your data.

9) _____ is a fundamental package for scientific computing with Python. It offers comprehensive mathematical functions, random number generators, linear algebra routines, Fourier transforms, and more. It provides a high-level syntax that makes it accessible and productive.

a) Pandas

b) Numpy

c) Statsmodel

d) scipy

Answer: b) Numpy

10) _____ is a Python data visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics that allow you to explore and understand your data. It integrates closely with pandas data structures.

a) Bokeh

b) Seaborn

c) Matplotlib

d) Das

Answer: b) Seaborn

11) Among the following identify the one in which dimensionality reduction reduces.

a) Performance

b) Statistics

c) Entropy

d) Collinearity

Answer: d) Collinearity

12) Which of the following machine learning algorithm is based upon the idea of bagging?

a) Decision Tree

b) Random Forest

c) Classfication

d) SVM

Answer: b) Random Forest

13) Choose a disadvantage of decision trees among the following.

a) Decision tree robust to outliers

b) Factor analysis

c) Decision Tree are prone to over fit

d) all of the above

Answer: c) Decision Tree are prone to over fit

14) What is the term known as on which the machine learning algorithms build a model based on sample data?

a) Data Training

b) Sample Data

c) Training data

d) None of the above

Answer: c) Training data

15) Which of the following machine learning techniques helps in detecting the outliers in data?

a) Clustering

b) Classification

c) Anomaly detection

d) All of the above

Answer: c) Anomaly detection

16) Identify the incorrect numerical functions in the various function representation of machine learning.

a) Support Vector

b) Regression

c) Case based

d) Classification

Answer: a) Support Vector and c) Case based

17) Analysis of ML algorithm needs

a) Statistical learning theory

b) Computational learning theory

c) None of the above

d) Both a and b

Answer: d) Both a and b

18) Identify the difficulties with the k-nearest neighbor algorithm.

a) Curse of dimensionality

b) Calculate the distance of test case for all training cases

c) Both a and b

d) None

Answer: c) Both a and b

19) The total types of the layer in radial basis function neural networks is _____

a) 1

b) 2

c) 3

d) 4

Answer: c) 3 layers

20) Which of the following is not a supervised learning

a) PCA

b) Naïve bays

c) Linear regression

d) K Means

Answer: a) PCA