

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

- a) $\beta_0, \beta_1, \dots, \beta_r$ 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 a and b

Answer: d) Both a and b 1. $\beta_0, \beta_1, \dots, \beta_r$ are the regression coefficients. These represents the weight assigned to each independent variable in the linear regression equation. They determine the impact of each other predictor on dependent variable.

2. linear regression is about determining the best predicted weight by using method of ordinary least squares. Goal is to find weights that minimize sum of squared distance between predicted value and actual value of dependent variable. This method commonly used to determine coefficients in linear regression.

Both a and b is about to determine the best predicted weights by using method of ordinary least squares.

22)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 $SSR = 0$

Ans: option d) the value $R^2 = 1$, which corresponds to $SSR=0$

In linear regression, R^2 is statical measure that represents proportion of variance in dependent variable can be explained by independent variable. It ranges from 0 to 1 where 0 is independent variable and 1 indicate perfect fit.

SSR (Sum of Squared Residuals) measure sum of the squared differences between the observed and predicted values in the regression model. A perfect fit would mean that the predicted values perfectly match the observed values, resulting in $SSR = 0$.

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

- a) Y
- b) B_0
- c) B_1
- d) F

Ans: b) The value of B_0 also called the intercept, shows the point where the estimated regression line crosses the y axis.

24) Check out these four linear regression plots:

Which one represents an **underfitted** model?

- a)The bottom-left plot
- b) The top-right plot
- c) The bottom-right plot

d) The top-left plot

Ans: b) The top-right plot represent the underfitted model.

25) 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

Ans: Correct option is c)

Step1: Import the packages and classes that you need.

Step2: Create a regression model and fit it with existing data.

Step3: Apply the model for predictions.

Step4: Provide data to work with, and eventually do appropriate transformations.

Step5: Check the results of model fitting to know whether the model is satisfactory.

26) 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

Ans: b) and c) are The optional parameters to Linear Regression in scikit-learn are fit_intercept and normalize.

27) 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

Ans: c) Polynomial regression

28) You should choose statsmodels 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.

Ans: c) You need more detailed results.

- 29) _____ 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

Ans: b)Numpy

- 30) _____ 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) Dash

Ans: b) Seaborn

- 41) Among the following identify the one in which dimensionality reduction reduces.
- a) Performance
 - b) statistics
 - c) Entropy
 - d) Collinearity

Ans: d) Dimensionality reduction reduces Collinearity.

- 42) Which of the following machine learning algorithm is based upon the idea of bagging?
- a) Decision Tree
 - b) Random Forest
 - c) Classification
 - d) SVM

Ans: b) Random Forest is based on the idea of bagging.

- 43) Choose a disadvantage of decision trees among the following.
- a) Decision tree robust to outliers
 - b) Factor analysis
 - c) Decision Tree are prone to overfit
 - d) all of the above

Ans. C) Decision Tree are prone to overfit

44) 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

Ans: C) Training Data

45) 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

Ans: c) The machine learning technique which helps in detecting the outliers in data is Anomaly detection.

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

- a) Support Vector
- b) Regression
- c) Case based
- d) Classification

Ans: a) To summarize the incorrect numerical functions in various representations of machine learning are Support Vector Regression.

47) Analysis of ML algorithm needs

- a) Statistical learning theory
- b) Computational learning theory
- c) None of the above
- d) Both a and b

Ans: d) Both a and b Analysis of machine learning algorithms requires both statistical learning theory and computational learning theory.

48) 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

Ans: Both a and b k-nearest neighbor algorithm has few difficulties: a) Curse of dimensionality : As the number of dimensions or features increases, the distance between data points becomes less meaningful.

b) Calculating the distance of the test case for all training cases: In order to classify a new data point, the k-NN algorithm needs to calculate the distance between the test case and all the training cases.

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

- a) 1
- b) 2
- c) 3
- d) 4

Ans: b) The total types of layer in radial basis function of neural network is 2.

50) Which of the following is not a supervised learning

- a) PCA
- b) Naïve bayes
- c) Linear regression
- d) KMeans

Ans: d) KMeans is not supervised learning.