

**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

**ans. d) Both a and b**

linear regression is a statistical method that is used to model the relationship between a dependent variable  $y$  and one or more independent variables  $x$

the dependent variable  $y$  is also called response or outcome variable, and the independent variables  $x$  are also called the predictor or explanatory variables.

the linear regression model assumes that the conditional mean of  $y$  given  $x$  is an affine function of  $x$  which means it can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_r x_r + \epsilon$$

$\beta_0, \beta_1, \dots, \beta_r$  are indeed the regression coefficients.  $\beta_0$  represents the intercept, and  $\beta_1, \dots, \beta_r$  are the coefficients corresponding to the independent variables  $x_1, \dots, x_r$ .

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

in linear regression,  $R^2$  measures the proportion of the variance in the dependent variable  $y$  that is explained by the independent variables  $x_1, \dots, x_r$ . It ranges from 0 to 1, where the value of 1, where a value of 1 indicates a perfect fit.

sum of squared residuals represent the sum of squared difference between the observed value of the dependent variable and predicted values from the regression line

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

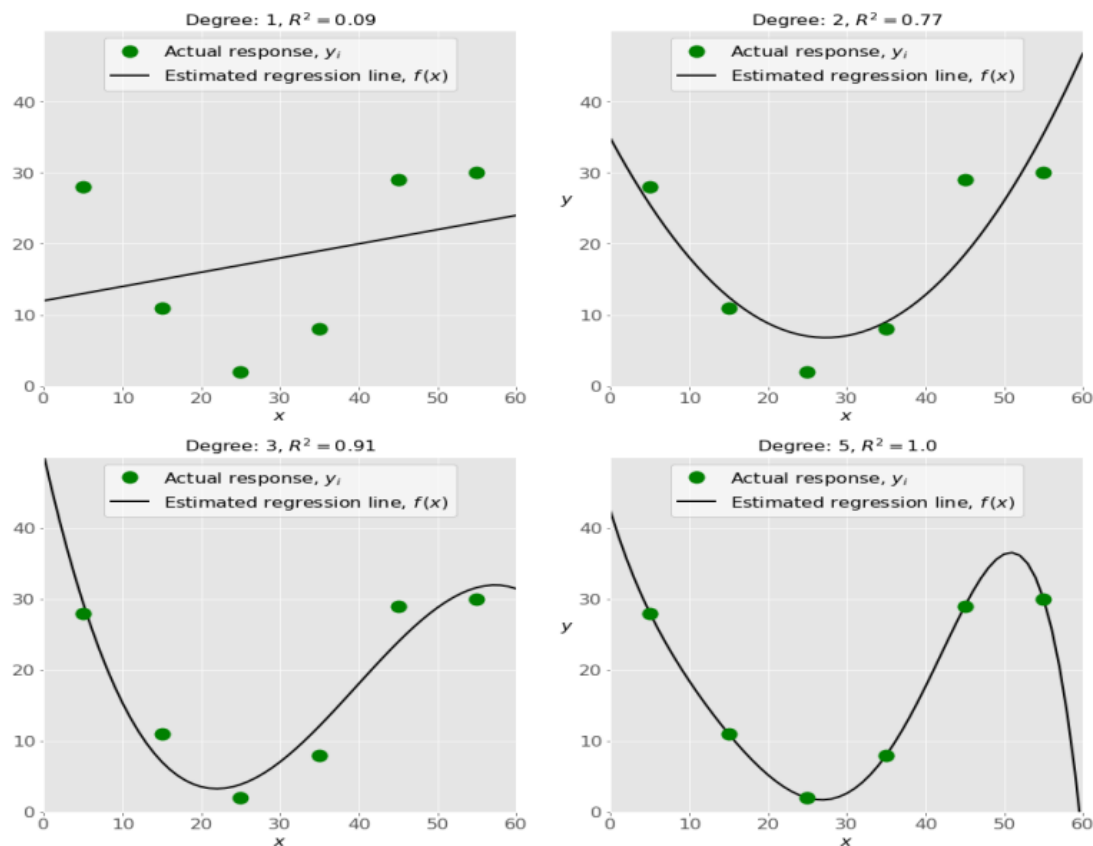
- a) Y
- b) B0
- c) B1
- d) F

ans. b) B0

in a simple linear regression, the estimated regression line is represented by the equation  $y = \beta_0 + \beta_1 x$

here,  $\beta_0$  is the intercept of the regression line, which indicate the point where the line crosses the y-axis. it's the value of dependent variable y when the indicate variable x is equal to 0.

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. d) The top-left plot**

the top-left plot shows a linear regression line that has low  $R^2$ . It is important that a straight line can't fit the data well.

**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. d) d, b, e, a, c**

**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. Fit\_intercept , normalize, copy\_X and \_jobs**

**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**

polynomial regression is a type of regression in which the relation between the independent variable X and dependent variable Y is modeled as an nth degree polynomial.

**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. d) You need to include optional parameters**

**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) Stats model
- 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**