

21) Answer: A) and B) are True

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

In linear regression, R^2 (R-squared) is a statistical measure that indicates the proportion of the variance in the dependent variable that can be explained by the independent variables. It ranges from 0 to 1, with 1 indicating a perfect fit.

SSR (Sum of Squared Residuals) is a measure of the total unexplained variation in the dependent variable. It represents the sum of the squared differences between the actual values of the dependent variable and the predicted values.

When $R^2 = 1$, it means that the regression model perfectly explains all the variance in the dependent variable. In other words, the model fits the data perfectly, and there are no residuals or errors left unexplained. This corresponds to $SSR = 0$ since there is no variation remaining to be explained.

Therefore, the statement (d) The value $R^2 = 1$, which corresponds to $SSR = 0$ indicates a perfect fit in linear regression.

23) Answer: (b) B_0

In simple linear regression, the relationship between a dependent variable (y) and a single independent variable (x) is modeled using a straight line equation of the form:

$$y = B_0 + B_1x$$

Here, B_0 represents the y-intercept, which is the value of y when the independent variable (x) is zero. It indicates the point where the regression line intersects the y-axis. The y-intercept represents the starting value or baseline level of the dependent variable.

24) Answer: (a) The bottom-left plot

The bottom-left plot, which represents an underfitted model.

25) Answer: (b) e, d, b, a, c

The correct order of the five basic steps when implementing linear regression is **e, d, b, a, c**

26) Answer:

The optional parameters to LinearRegression in scikit-learn are as follows:

- a) Fit: It is not a parameter but a method used to fit the linear regression model to the data.
- b) fit_intercept: Specifies whether to calculate the intercept for the model. It is set to True by default.
- c) normalize: If set to True, the input variables will be normalized before regression by subtracting the mean and dividing by the l2-norm. It is set to False by default.
- d) copy_X: Determines whether a copy of the input X should be made. It is set to True by default.
- e) n_jobs: Specifies the number of parallel jobs to use during the model fitting. If set to -1, it will use all available processors. By default, it is set to 1.
- f) reshape: This is not an optional parameter for LinearRegression in scikit-learn. It seems to be unrelated to the LinearRegression class.

So, the correct options are: b) fit_intercept, c) normalize, d) copy_X, and e) n_jobs.

27) Answer: (c) Polynomial regression

In polynomial regression, the relationship between the input variables (features) and the target variable is modeled as an nth degree polynomial. This allows for capturing nonlinear relationships between the variables by including higher-order terms like x^2 , x^3 , etc

28) Answer: You need more detailed results

When I want graphical representations of my data. Statsmodels provides a range of statistical models and methods specifically designed for statistical analysis, hypothesis testing, and econometric modeling. It offers detailed statistical output with comprehensive information about model parameters, p-values, confidence intervals, and more. It also supports graphical representations such as regression plots, residual plots, and diagnostic plots, which are helpful for visualizing and interpreting the results. Additionally, statsmodels allow for including optional parameters and working with nonlinear terms, making it a suitable choice for advanced statistical analysis and research purposes.

29) Answer: Numpy

NumPy, full form (**Numerical Python**), is a fundamental package for scientific computing with Python. It is a powerful library that provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays efficiently. NumPy forms the foundation for many other scientific computing libraries in Python.

30) Answer: Seaborn

Seaborn is a powerful Python data visualization library built on top of matplotlib. It offers a high-level interface for creating visually appealing and informative statistical graphics, making it easier to explore and comprehend your data. Seaborn simplifies the process of creating complex visualizations by providing intuitive functions and default settings that result in visually appealing plots. It seamlessly integrates with pandas data structures, allowing for efficient data manipulation and analysis. Seaborn capabilities include generating various types of plots such as scatter plots, line plots, bar plots, histograms, and heatmaps, among others. With Seaborn, users can effectively communicate insights and patterns within their data.