21) 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 a and b**

The answer is D) Both a and b. The linear regression is a type of predictive analysis that examines the relationship between a dependent variable(y) and one or more independent variables(x). It also involves estimating the regression coefficients $\beta_0, \beta_1, \ldots, \beta_r$, and it determines the best predicted weights by using the 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**

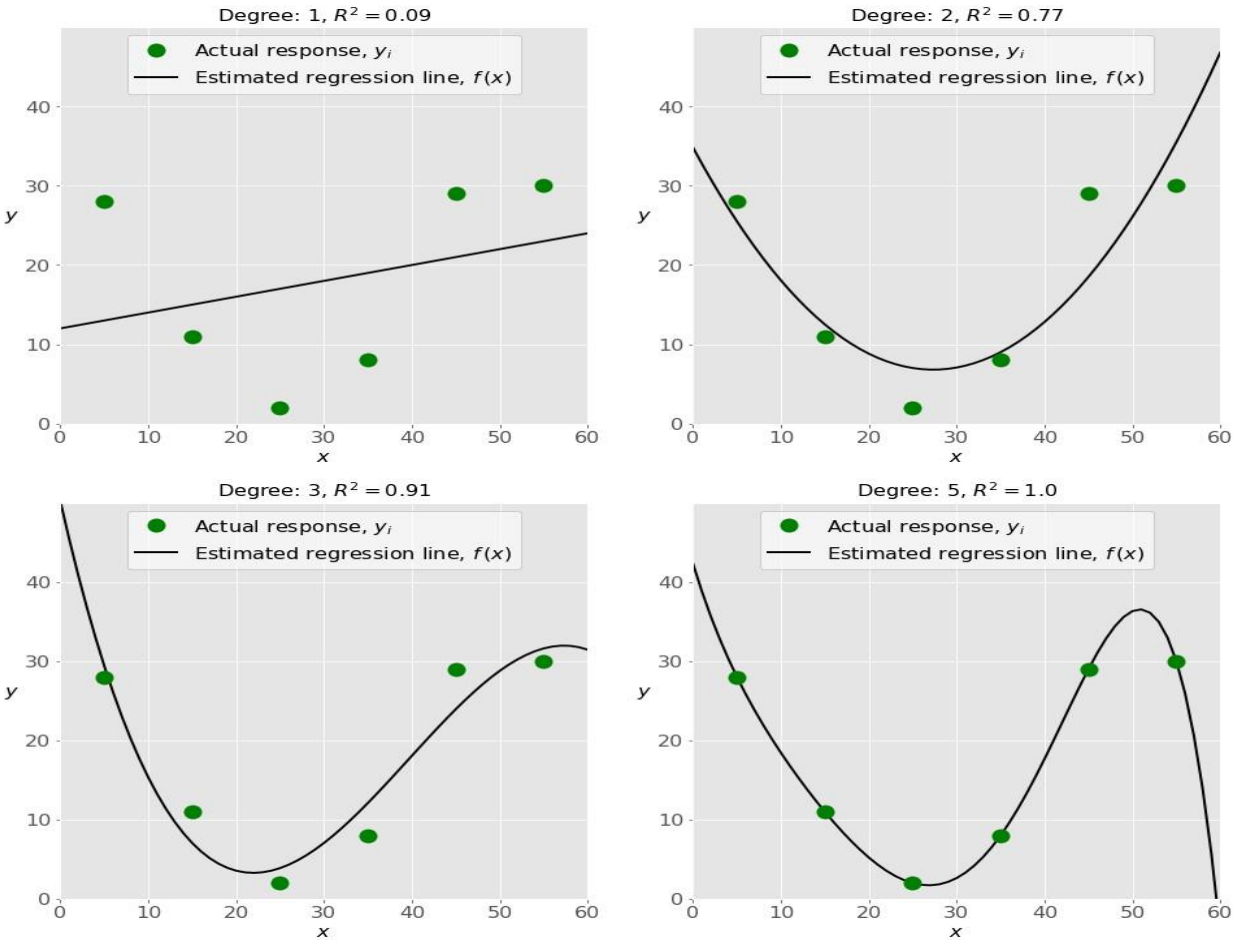
The answer is d) d) The value $R^2 = 1$, which corresponds to SSR = 0, because a perfect fit in linear regression means there is no error between the observed and predicted values of the response variable. This implies that SSR = SST and SSE = 0, where SSE is the sun of squares error, or the sum of squared differences between predicted data points (ŷi) and observed data points (yi).

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

The answer is b) b) B0 because it is another way of writing the intercept a, in the regression equation. The intercept (a) shows the point where the estimated regression line crosses the y-axis. It represents the predicted value of y when x is zero.

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

The bottom-left plot represents an underfitted model, where the linear regression line is too simple and does not 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**

The five steps of implementing a linear regression are:

1. Import the packages and classes that you need.
2. Provide data to work with, and eventually do appropriate transformations.
3. Create a regression model and fit it with existing data.
4. Check the results of model fitting to know whether the model is satisfactory.
5. Apply the model for predictions.

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

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**

The Polynomial regression is the type of regression you need to use for transforming the array of inputs to include nonlinear terms such as x^2.

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.

The answers are a) and c) because you want graphical representations of your data, and 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

The answer is b) Numpy. Numpy is a fundamental package for scientific computing with Python.

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

The answer is b) Seaborn. Seaborn 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. Seaborn is designed to make common statistical plots more intuitive and elegant, such as scatter plots, box plots, histograms, heatmaps, etc. It also supports more complex plots such as joint plots, pair plots, factor plots, etc. Seaborn also offers various themes and styles to customize the appearance of the plots