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

Answer: d

Explanation:-
1. $\beta_0, \beta_1, \ldots, \beta_r$ are the regression coefficients. These coefficients represent the weights assigned to each independent variable in the linear regression equation. They determine the impact of each predictor on the dependent variable.
2. Linear regression is about determining the best predicted weights by using the method of ordinary least squares. The goal is to find the weights that minimize the sum of squared differences between the predicted values and the actual values of the dependent variable. This method is commonly used to estimate the coefficients in linear regression.

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

Answer: d

Explanation:-

In linear regression, R^2 (R-squared) is a statistical measure that represents the proportion of the variance in the dependent variable that can be explained by the independent variables. It ranges from 0 to 1, where 0 indicates that the independent variables have no explanatory power, and 1 indicates a perfect fit.

SSR (Sum of Squared Residuals) measures the 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.

Therefore, when $R^2 = 1$, it indicates a perfect fit in linear regression, as all the variance in the dependent variable is explained by the independent variables, and there are no residuals or errors in the model.

23) In simple linear regression, the value of **what** shows the point where the estimated regression line crosses the y axis?
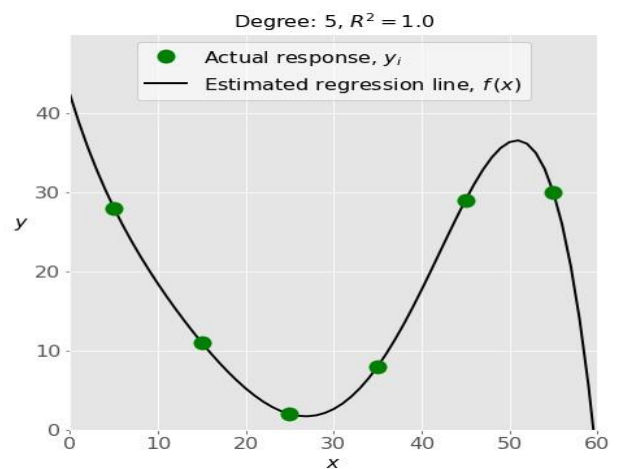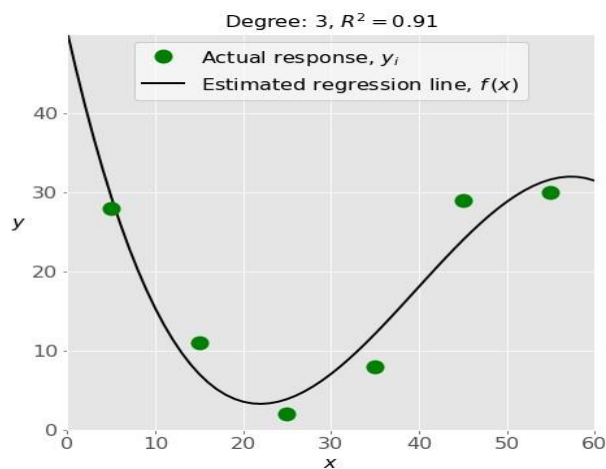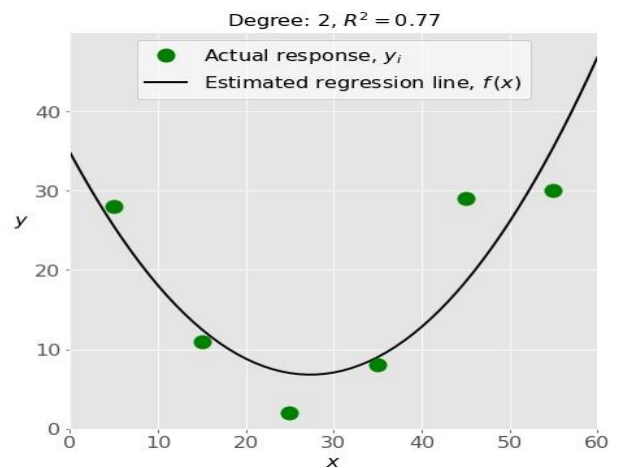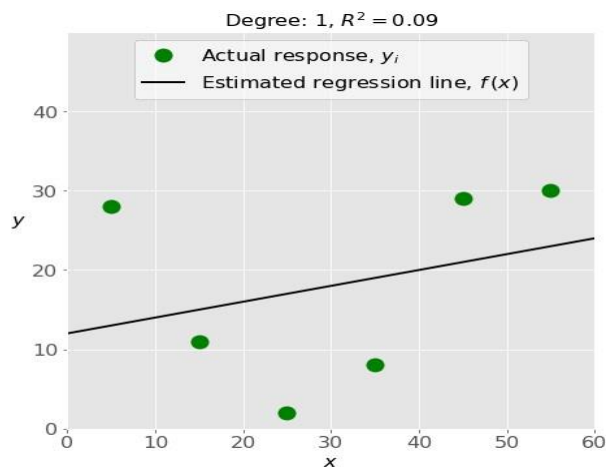
   a) Y
   b) B0
   c) B1
   d) F

Answer: b

Explanation:-

In simple linear regression, the value that shows the point where the estimated regression line crosses the y-axis is represented by the coefficient B_0. This is the intercept term in the linear regression equation $y = B_0 + B_1x$, where y is the predicted value of the dependent variable y for a given value of the independent variable x.

24) Check out these four linear regression plots:

Degree: 1, $R^2 = 0.09$ (top-left)
Degree: 2, $R^2 = 0.77$ (top-right)
Degree: 3, $R^2 = 0.91$ (bottom-left)
Degree: 5, $R^2 = 1.0$ (bottom-right)

Legend for each plot:
- Actual response, y_i
- Estimated regression line, $f(x)$

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

Answer: d

Explanation:-

In the bottom-left plot, the model appears to be too simple and unable to capture the underlying patterns in the data. This is characteristic of an underfitted model, where the model's complexity is insufficient to represent the relationships present in the data. Underfitting typically results in poor performance on both the training and testing data.

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

Answer: d

Explanation:-

d) Import the packages and classes that you need.

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

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

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

c) Apply the model for predictions.

26) Which of the following are optional parameters to LinearRegression in scikit-learn?

a) Fit

b) fit_intercept
c) normalize
d) copy_X
e) n_jobs
f) reshape

Answer: b,c,d,e,f
Explanation:-
fit_intercept: This parameter determines whether to calculate the intercept for the linear regression model. By default, it is set to True.

normalize: This parameter determines whether to normalize the input features before fitting the model. By default, it is set to False.

copy_X: This parameter determines whether to make a copy of the input features. By default, it is set to True.

n_jobs: This parameter determines the number of parallel jobs to use for the computation. By default, it is set to None, which means it will use one job.

reshape: This parameter determines whether to reshape the target variable. By default, it is set to True.

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

Answer: c

Explanation:-

 In scikit-learn, when you need to transform the array of inputs to include nonlinear terms such as x^2, you would use Polynomial Regression. Polynomial regression extends the simple linear regression model by including polynomial features, allowing the model to capture nonlinear relationships between the independent variables and the dependent variable.

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.

Answer: c

Explanation:-

Scikit-learn, is more focused on machine learning tasks and predictive modeling.

Statsmodels provides comprehensive statistical information, including p-values, confidence intervals, and other detailed statistics.

So, if you prioritize detailed statistical results, you would choose statsmodels over scikit-learn.

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

Answer: b

Explanation:-

Numpy 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.

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

Answer: b

Explanation:-

Seaborn is designed to work well with pandas data structures and is particularly useful for creating visually appealing and informative plots for statistical analysis. It simplifies the process of creating complex visualizations and is often used in conjunction with Matplotlib for enhanced data visualization capabilities.