Assignment = 2
4.5.2023 – 7.5.23

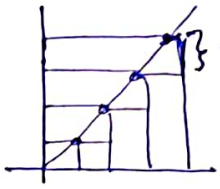
21 When implementing linear regression of some dependent variable y on the set of independent variables $\mathbf{x} = (x_1, \ldots, x_r)$, where r is the number of predictors, which of the following statements will be true?

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



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

R^2 = Coefficient of determination

$R^2 = 1$ = 100% No error

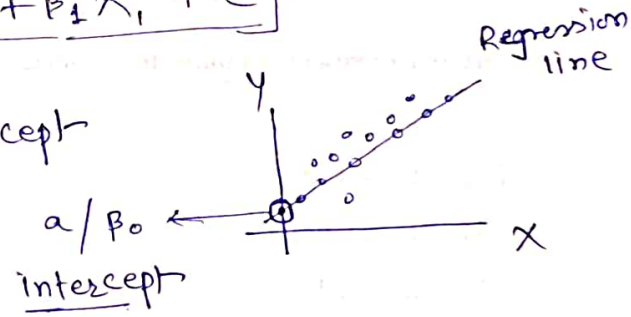
23)

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

a) Y
b) B0 — intercept — or
c) B1
d) F

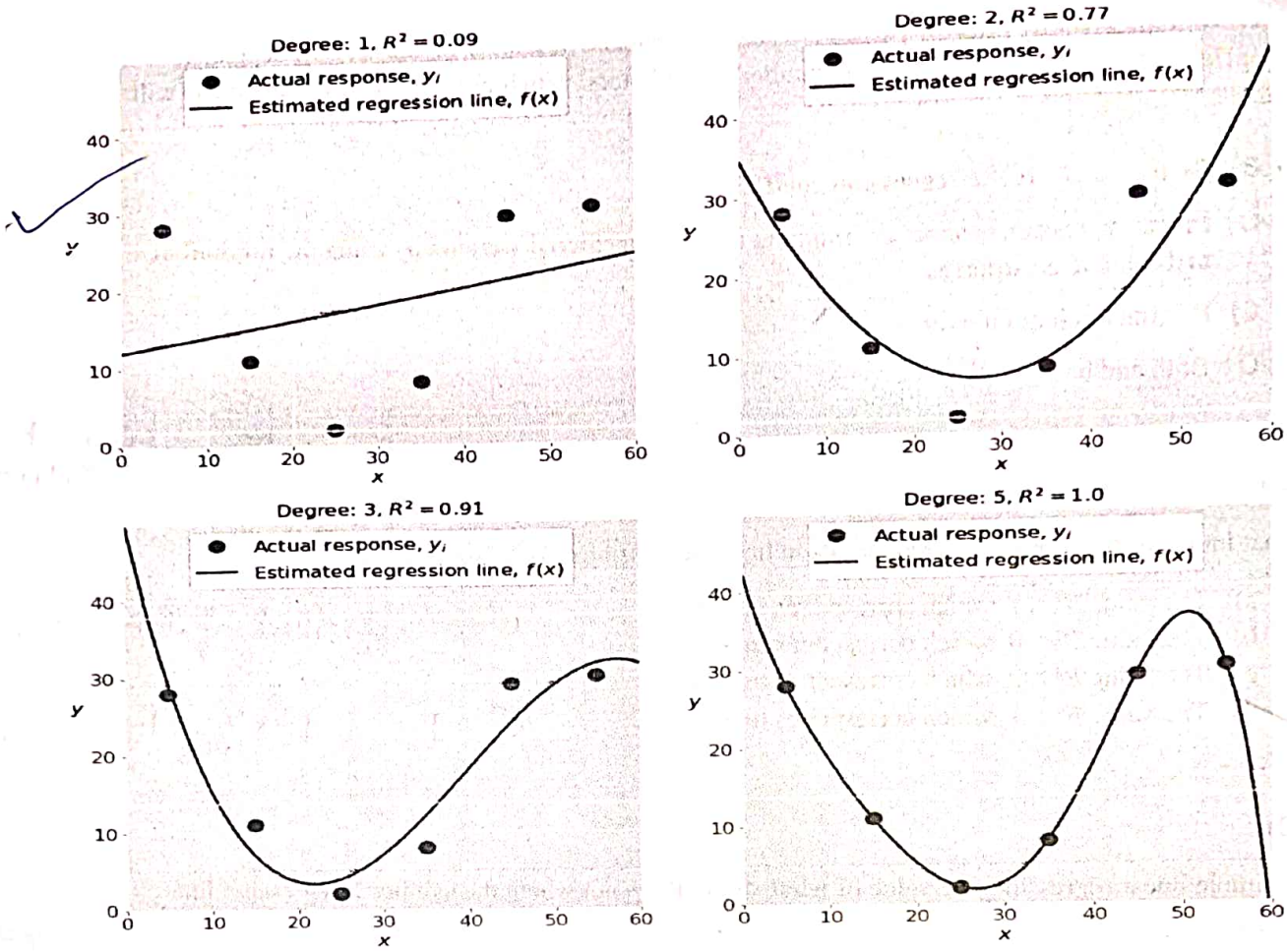
$y = a + bx + e$ — both are same
$y = \beta_0 + \beta_1 x_1 + e$

↓ intercept



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

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) c, c, a, b, d
b) e, d, b, a, c
c) d, e, c, b, a
d) d, b, e, a, c

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

a) Fit
b) fit_intercept : bool, default=True — No intercept will be used, data expected to be centered
c) normalize
d) copy_X : bool, default=True — X will be copied
e) n_jobs : int, default=None — For large data
f) reshape
(g) Positive : bool, default=False

sklearn.linear_model

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

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.

29) 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.

a) Pandas
b) Numpy
c) Statsmodel
d) scipy

30) 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. It integrates closely with pandas data structures.

(a) Bokeh (b) seaborn (c) Matplotlib (d) Dash.