Work sheet 2

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?

Ans; d) Both and b

22) What indicates that you have a perfect fit in linear regression?

Ans; c) The value $R^2 > 0$, which corresponds to SSR = 1

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

Ans; a) Y

24) Check out these four linear regression plots:
a) The bottom-left plot
b) The top-right plot
c) The bottom-right plot
d) The top-left plot
Which one represents an underfitted model?

Ans; 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?

Ans; d) d, b, e, a, c

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

Ans; b) fit _ intercept
d) copy _ X
e) n_ 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?

Ans; c) Polynomial regression

28) You should choose statsmodels over scikit-learn when:

Ans; c) 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.

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 panda's data structures.

Ans; b) Seaborn