File2

21

When implementing linear regression of some dependent variable y on the set of independent

variables $\mathbf{x} = (x1, ..., xr)$, where r is the number of predictors, which of the following statements will

be true?

a) is **not true:** as we have (x1, ⋯., xr), we should also have (β1, ⋯..., βr)

b) is True: Linear regression is about determining the **best predicted weights** by using the **method of ordinary least squares**.

c) E is the error term, the error in predicting the value of label y, knowing features (x1, ⋯., xr)

22)

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

d) The value $R^2 = 1$, which corresponds to SSR = 0 (SSR = sum squared regression = sum of residuals squared. R^2 = 1-SSR)

23)

In simple linear regression, the value of **what** shows the point where the estimated regression line

crosses the y axis?

b) B0 is true (Y = B0 + B1*X: for X = 0, Y = B0)

24)

Check out these four linear regression plots:

Which one represents an **underfitted** model?

d) The top-left plot: the model doesn't capture the relationship shown in the dataset

25)

There are five basic steps when you're implementing linear regression:

However, those steps are currently listed in the wrong order. What's the correct order?

d) d, b, e, a, c is the correct order

26)

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

b) fit_intercept

c) normalize

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 $x2$?

c) Polynomial regression

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

c) You need more detailed results.

d) You need to include optional parameters.

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.

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 pandas data structures.

b) Seaborn