Internship

Linear regression Asignment

Ayiman Bekowei

**Ds0523**

21).

(a). B0, B1….Br are the regression coefficients.

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

22).

(d).

in linear regression, R2 (coefficient of determination) is a measure of how well he independent variables explain the variability of dependent variable. A perfect fit wound mean that the model explains all the variability, resulting in R2 = 1. When R2 is 1, it indicates that the sum of squared residuals (SSR) is O, meaning that the model perfectly predicts the dependent variable without any errors.

23).

In simple linear regression, equation of the regression line is given by:

Y = B0 + B1X

(b). B0

Where:

Y is the dependent variable,

X is the independent variable

B0 is the intercept (the point the both regression line cross the Y axis)

B1 is the slop of the regression line.

So, the value B0 represents the point where the estimated regression line crosses the Y- axis.

24).

(a). The bottom -left plot

It has the characteristic of underfitted model.

(25).

(d). d, b, e, a, c

d. First import the packages and classes that you need.

b. Provide data to work with, 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). These are optional parameters to linear regression in scikit-learn.

(b) Fit_intercept

(c) Normalize

(d) copy_X

(e) n_jobs

27). To include nonlinear terms such as X2, to transform the array, you need polynomial regression type of regression.

(b) Polynomial regression.

28).

(c)

You should choose stats models over scikit-learn when you need more detailed results.

(29)

(b) Numpy is a fundamental package for scientic computing with python. It offers comprehensive mathematical functions.

30).

(b). Seaborn is a python data visualization library based on matplotlib.

41).

(d). Collinearity, Dimensionality reduction techniques, such as principal component analysis (PCA), are used to reduce collinearity among features in a dataset.

42).

(b). Random forest is a machine learning algorithm based on the idea of bagging (Bootstrap Aggregating).

43).

(c). the disadvantage of a decision tree are prone to overfitting.

44).

(a). data training is the term for building a model based on sample data in machine learning.

45).

(c). Anomaly detection is machine learning technique that specifically focuses on detecting outliers in data.

46).

c). case based this term is not a standard representation of machine learning function.

(47).

(d) both a and b

48).

(c). both a and b

49).

(c). 3, output layer, This layer produces the final output.

50).

(d). Kmeans is a clustering algorithm and belongs to the category of unsupervised learning.