

Classical Data Analysis. Linear Regression Test

1) Which of the following problems cannot be modeled as a regression problem?

- a. Predict the value (price) of a car based on its characteristic;
- b. Predict the volume of gas that will be consumed in the next two months;
- c. Predict if a car will be sold within the next 3 months.
- d. None of the above problems can be modeled as a regression problem.

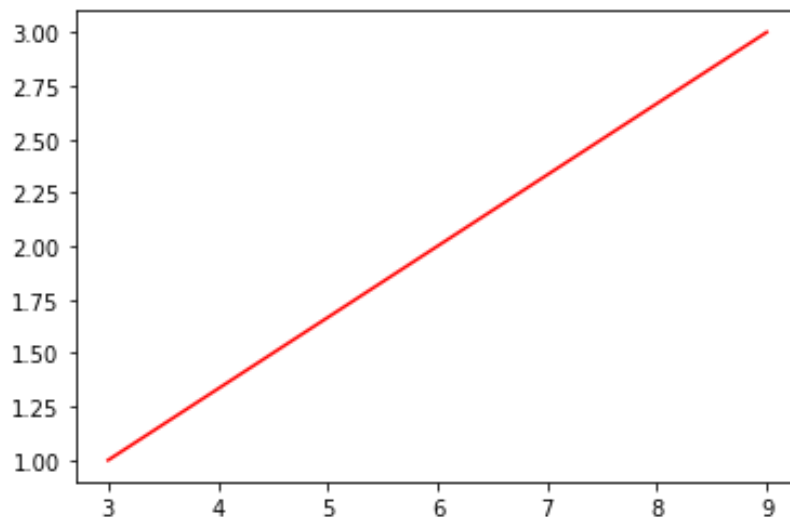
2) Consider the following dataset:

X	y
3	1
6	2
12	4
30	10

Considering the linear regression model $f(x) = a + b \cdot x$, what are the values of a and b you would expect to better fit the data?

- a. $a=0, b=0$
- b. $a=3, b=0$
- c. $a=1, b=3$
- d. $a=0, b=1/3$

3) Which of the following linear model correspond to that represented in the plot?



- a. $y = 1 + 3 \cdot x$
- b. $y = (1/3) \cdot x$

- c. $y = 3/x$
- d. None of the above

4) Considering the linear regression model $y = a + b \cdot x$, which is the dependent variable and which is the independent variable?

- a. y and x are both dependent
- b. y is the dependent variable and x is the independent variable
- c. x is the dependent variable and y is the independent variable
- d. None of the above

5) In the linear regression model $f(x) = 5 + 6 \cdot x$ which is the intercept and which is the slope of the model?

- a. 5 is the intercept, 6 is the slope
- b. 6 is the intercept, 5 is the slope
- c. The intercept is 0 and the slope is $5/6$

6) In the linear regression model of the exercise 3, is there a relationship between the dependent and the independent variable?

- a. Yes, there is a negative relationship
- b. No, there is no relationship
- c. Yes, there is a positive relationship

7) If you represent your linear model through a linear function having a negative slope, the relationship between the dependent variable and the independent variable will be:

- a. There is no relationship between the variables
- b. I don't have enough information to reply
- c. There will be a negative relationship between dependent and independent variables.
- d. There will be a positive relationship between dependent and independent variables.

8) Suppose we set $a = -1$ and $b = 5$. If your linear model is $f(x) = a + b \cdot x$, which is the value of $f(5)$?

- a. 25
- b. 24
- c. 26
- d. There are not enough information to compute that value

9) Which of the following metrics can be used to evaluate regression models (multiple choices):

- a. R Squared
- b. RMSE / MSE / MAE
- c. nDCG
- d. Accuracy
- e. F Statistics

10) What are the units of Mean Absolute Error (MAE)?

- a. The same units as your target variable (the Y).
- b. It is just an error, it does not have units.

- c. The units of the target squared.
- d. None of the above.

11) How many parameters do you need to estimate in a univariate (one independent variable) linear regression model?

- a. 1
- b. 2
- c. 0
- d. 3

12) In a simple linear regression model (one independent variable) $y = 2 + 0.5x$, if we change the input variable by 1 unit, which is the change expected in the output variable? (Multiple choices)

- a. 0.5
- b. 1 unit
- c. the value of the slope
- d. the value of the intercept

13) Suppose to run the following python code:

```
x = 2 - 3 * np.random.normal(0, 1, 20)
```

```
y = x - 2 * (x ** 2) + 0.5 * (x ** 3) + np.random.normal(-3, 3, 20)
```

14) Which model would you expect would fit best the data points?

- a. $f(x) = -3 + 3x - 5x^2 + x^3 + 3x^4$
- b. $f(x) = -3 + 4x$
- c. $f(x) = -3 + x - 2x^2 + 0.5x^3$
- d. $f(x) = -3 + x - 2x^2 + 5x^3$

15) What does the intercept represent in a regression linear model?

- a. the expected mean value of y when $x=0$.
- b. A measure of the relationship between x and y
- c. the expected mean value of x when $y=0$.
- d. None of the above