Linear Regression Questions

TA. Ahmed Almohammed

10/01/2024

Question 1: When implementing linear regression of some dependent variable y on the set of independent variables $x = (x_1, x_2, ..., x_r)$, where r is the number of predictors, which of the following statements will be true?

- a. ϵ is the random interval.
- ***b. Linear regression is about determining the best predicted weights by using the method of ordinary least squares.***
- ***c. The estimators of the regression coefficients define the estimated regression function*** $f(x) = b_0, b_1x_1 + b_2x_2 + ... + b_rx_r$***.***
- ***d.*** $\beta_1, \beta_2, ..., \beta_r$ ***are the regression coefficients.***

Question 2: In simple linear regression, the value of what shows the point where the estimated regression line crosses the y axis?

- a. f
- ***b.*** b_0
- c. y
- d. b_1

Question 3: In polynomial regression, your regression function can include nonlinear terms such as $b_2x_1^2$, $b_3x_1^3$ or even $b_4x_1x_2$, $b_5x_1^2x_2$.

- ***a. True***
- b. False

Question 4: There are five basic steps when you're implementing linear regression:

1. Check the results of model fitting to know whether the model is satisfactory.

2. Provide data to work with, and eventually do appropriate transformations.

3. Apply the model for predictions.

4. Import the packages and classes that you need.

5. 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. 4, 5, 3, 2, 1
- b. 5, 3, 1, 2, 4
- c. 5, 4, 2, 1, 3
- ***d. 4, 2, 5, 1, 3***

Question 5: Why do we need regularisation?

- a. To penalize the model
- b. To avoid overfitting
- c. To generalise better on unseen data
- ***d. All of the above***

Question 6: Which of the following corresponds to the equation of LASSO regression?

- a. $\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$
- b. $\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2-\lambda\sum_{j=1}^{p}|\hat{\beta}|$
- ***c.*** $\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2+\lambda\sum_{j=1}^{p}|\hat{\beta}|$
- d. $\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2\times\lambda\sum_{j=1}^{p}|\hat{\beta}|$

Question 7: Underfitting occurs when a model can't accurately capture the dependencies among data, usually as a consequence of its own simplicity. True or False?

- ***a. True.***
- b. False.

Question 8: Overfitting happens when a model learns both data dependencies and random fluctuations, meaning that the model learns the data too well. True or False?

- ***a. True***
- b. False

Question 9: In the mean squared error function or cost function J, our task is to find the value of b_0 and b_1 for which $J(b_0, b_1)$ is:

- ***a. Minimum***
- b. Maximum

Question 10: What can be said about an overfitting model with respect to bias and variance?

- a. High Bias and High Variance
- b. High Bias and Low Variance
- ***c. Low Bias and High Variance***
- d. Low Bias and Low Variance