

MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

- A) Least Square Error
- B) Maximum Likelihood
- C) Logarithmic Loss
- D) Both A and B

To find the best fit line for data in linear regression we use least square method . this method determine the best fitting line for the noted data by reducing the sum of the square of the vertical deviation from each point to be line

option A) Least Square Error

2. Which of the following statement is true about outliers in linear regression?

- A) Linear regression is sensitive to outliers
- B) linear regression is not sensitive to outliers
- C) Can't say
- D) none of these

In linear regression, **outliers** are observations that fall far from the “cloud” of points. These points are especially important because they can have a strong influence on the least squares line

Linear regression is **sensitive to outliers** because it tries to minimize the sum of the squared distances between the predicted values and the actual values. If there are outliers in the data, the line of best fit will be pulled towards them, resulting in a poor fit for the rest of the data ².

option A) Linear regression is sensitive to outliers.

2. A line falls from left to right if a slope is _____?

- A) Positive
- B) Negative
- C) Zero
- D) Undefined

A line falls from left to right if the slope is **negative**. A positive slope indicates that the line rises from left to right, while a slope of zero indicates a horizontal line, and an undefined slope indicates a vertical line .

option B) Negative.

3. Which of the following will have symmetric relation between dependent variable and independent variable?

- A) Regression
 - B) Correlation
 - C) Both of them
 - D) None of these
-

MACHINE LEARNING

Neither regression nor correlation have a symmetric relationship between dependent and independent variables. In regression, the dependent variable is predicted by the independent variable, while in correlation, the relationship between two variables is measured

option D) None of these.

4. Which of the following is the reason for over fitting condition?
- | | |
|--------------------------------|------------------------------|
| A) High bias and high variance | B) Low bias and low variance |
| C) Low bias and high variance | D) none of these |

C) Low bias and high variance. Overfitting occurs when a model is too complex and fits the training data too well, leading to poor generalization on new data. This happens when the model has **low bias** and **high variance**. Low bias means that the model is flexible enough to fit the training data well, while high variance means that the model is too sensitive to the noise in the training data and captures the random fluctuations instead of the underlying pattern.

5. If output involves label then that model is called as:
- | | |
|---------------------------|---------------------|
| A) Descriptive model | B) Predictive modal |
| C) Reinforcement learning | D) All of the above |

B) Predictive model. If a machine learning model's output involves a target variable, it is called a predictive model.

6. Lasso and Ridge regression techniques belong to _____?
- | | |
|---------------------|----------------------|
| A) Cross validation | B) Removing outliers |
| C) SMOTE | D) Regularization |

D) Regularization. Regularization is a technique used to prevent overfitting in machine learning models. It involves adding a penalty term to the loss function that the model tries to minimize. Lasso and Ridge regression are two popular regularization techniques used in linear regression models. Lasso regression adds a penalty term proportional to the absolute value of the coefficients, while Ridge regression adds a penalty term proportional to the square of the coefficients.

MACHINE LEARNING

7. To overcome with imbalance dataset which technique can be used?
- | | |
|---------------------|-------------------|
| A) Cross validation | B) Regularization |
| C) Kernel | D) SMOTE |

D) SMOTE. SMOTE stands for Synthetic Minority Over-sampling Technique. It is an oversampling technique used to balance the class distribution of a dataset by creating synthetic minority class samples. SMOTE is a type of data augmentation technique that generates new synthetic samples by interpolating between existing minority-class samples.

8. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?
- | | |
|--------------------------------|------------------------------|
| A) TPR and FPR | B) Sensitivity and precision |
| C) Sensitivity and Specificity | D) Recall and precision |

The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses **True Positive Rate (TPR)** and **False Positive Rate (FPR)** to make the graph. The TPR is also known as sensitivity, while FPR is equal to 1 - specificity. The ROC curve is a plot of TPR vs FPR at different classification thresholds, and the AUC-ROC curve is the area under this curve.

Option A TPR and FPR

9. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
- | | |
|---------|----------|
| A) True | B) False |
|---------|----------|

The statement “In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less” is **False**.

In fact, the opposite is true. The **higher** the area under the curve (AUC), the **better** the model's performance at distinguishing between the positive and negative classes

10. Pick the feature extraction from below:
- | |
|---|
| A) Construction bag of words from a email |
| B) Apply PCA to project high dimensional data |
| C) Removing stop words |
| D) Forward selection |

The feature extraction method from the given options is **A) Construction bag of words from an email**¹²³.

MACHINE LEARNING

Bag of words is a technique used to represent text data as numerical features that can be used in machine learning models. It involves creating a vocabulary of unique words in the text corpus and then counting the frequency of each word in each document. The resulting feature matrix is a sparse matrix where each row represents a document and each column represents a word in the vocabulary¹².

The other options mentioned are not feature extraction methods, but rather techniques used in feature selection or dimensionality reduction. Here's a brief description of each:

- **B) Apply PCA to project high dimensional data:** Principal Component Analysis (PCA) is a technique used for dimensionality reduction. It involves projecting high-dimensional data onto a lower-dimensional subspace while retaining as much of the original variance as possible⁴.
- **C) Removing stop words:** Stop words are common words that are removed from text data before feature extraction. These words do not carry much meaning and can be safely ignored without affecting the performance of the model.
- **D) Forward selection:** Forward selection is a feature selection technique that involves starting with an empty set of features and iteratively adding the most important feature at each step until a stopping criterion is met.

In Q12, more than one options are correct, choose all the correct options:

11. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
- A) We don't have to choose the learning rate.
 - B) It becomes slow when number of features is very large.
 - C) We need to iterate.
 - D) It does not make use of dependent variable.

The Normal Equation is a closed-form solution used to find the value of θ that minimizes the cost function in Linear Regression¹². Here are the answers to the given options:

A) We don't have to choose the learning rate: This is true for the Normal Equation method. Unlike Gradient Descent, the Normal Equation method does not require the learning rate to be tuned¹².

B) It becomes slow when the number of features is very large: This is true. The Normal Equation method involves computing the inverse of a matrix, which can become computationally expensive when the number of features is very large¹².

C) We need to iterate: This is false. The Normal Equation method provides a direct solution to the optimization problem and does not require any iterations¹².

MACHINE LEARNING

D) It does not make use of the dependent variable: This is false. [The Normal Equation method uses both the independent and dependent variables to compute the optimal values of the model parameters](#)

MACHINE LEARNING

Q13 and Q15 are subjective answer type questions, Answer them briefly.

12. Explain the term regularization?

ANS- Regularization is a technique used in machine learning to prevent overfitting and improve the generalization performance of a model. Overfitting occurs when a model is too complex and fits the training data too well, but performs poorly on new data. Regularization adds a penalty term to the loss function during training, which helps to reduce the complexity of the model and avoid overfitting.

There are several types of regularization techniques, including:

- L1 regularization (Lasso): adds the absolute value of the coefficients as a penalty term to the loss function.
- L2 regularization (Ridge): adds the square of the coefficients as a penalty term to the loss function.
- Elastic Net regularization: combines both L1 and L2 regularization.

Regularization helps to find a proper balance between bias and variance, which is also known as the Bias-Variance Tradeoff. Bias refers to the errors that occur when a model is too simple and cannot capture the underlying patterns in the data, while variance refers to the errors that occur when a model is too complex and captures the noise in the data. Regularization can help to reduce the variance of a model by reducing its complexity.

13. Which particular algorithms are used for regularization?

There are several algorithms used for regularization in machine learning. Some of the commonly used ones are:

- L1 regularization (Lasso): adds the absolute value of the coefficients as a penalty term to the loss function ¹²³.
- L2 regularization (Ridge): adds the square of the coefficients as a penalty term to the loss function ¹²³.
- Elastic Net regularization: combines both L1 and L2 regularization ¹³.
- Dropout regularization: randomly drops out some neurons during training to prevent overfitting ⁴.
- Early stopping: stops the training process when the model's performance on the validation set starts to degrade

14. Explain the term error present in linear regression equation?

In linear regression, the error term represents the difference between the predicted value and the actual value of the dependent variable ¹². The goal of linear regression is to

MACHINE LEARNING

minimize the sum of the squared errors between the predicted and actual values of the dependent variable ³.

The error term is also known as the **residual** and is calculated as the difference between the observed value of the dependent variable and the predicted value of the dependent variable ². The residual is a measure of how well the model fits the data. A small residual indicates that the model is a good fit for the data, while a large residual indicates that the model is a poor fit for the data ².

The standard error of the regression (σ_S) is a measure of the average distance that the observed values fall from the regression line ⁴. It is also a measure of the precision of the predictions made by the model ⁴. The standard error of the regression can be used to calculate confidence intervals for the predicted values of the dependent variable ⁴.
