

-:Feature Engineering:-

1. What is a parameter?

Ans - A parameter is a variable inside a machine learning model that is learned from the training data. Parameters define how the model makes predictions. Examples include weights and bias in linear regression. Parameters are adjusted during training to minimize the loss function.

2. What is correlation? What does negative correlation mean?

Ans- Correlation:-

Correlation is a statistical measure that describes the strength and direction of the relationship between two variables. It shows how one variable changes with respect to another. The correlation value ranges from -1 to +1.

Negative Correlation:-

Negative correlation means that when one variable increases, the other variable decreases. For example, as the speed of a vehicle increases, the time taken to reach a destination decreases. The correlation value lies between 0 and -1.

3. Define Machine Learning. What are the main components in Machine Learning?

Ans- Machine Learning is a subset of Artificial Intelligence that enables systems to learn patterns from data and make predictions without being explicitly programmed.

Main components are:

- Data
- Features
- Model
- Loss function
- Optimizer

4. How does loss value help in determining whether the model is good or not?

Ans- The loss value measures the difference between the predicted output and the actual output. A lower loss value indicates better model performance. By monitoring loss during training and testing, we can judge whether the model is learning correctly or overfitting.

5. What are continuous and categorical variables?

Ans- Continuous variables are numerical values that can take any value within a range, such as height or temperature.

Categorical variables represent categories or labels, such as gender, color, or city.

6. How do we handle categorical variables in Machine Learning? What are the common techniques?

Ans- Categorical variables must be converted into numerical form. Common techniques include:

- Label Encoding
- One-Hot Encoding
- Ordinal Encoding

7. What do you mean by training and testing a dataset?

Ans- Training data is used to train the machine learning model and learn patterns.

Testing data is used to evaluate the model's performance on unseen data to check generalization.

8. What is sklearn.preprocessing?

Ans- `sklearn.preprocessing` is a module in Scikit-learn used for data preprocessing. It provides tools for scaling, normalization, encoding categorical variables, and handling missing values.

9. What is a Test set?

Ans- A test set is a portion of the dataset that is not used during training. It is used only to evaluate the final performance of the trained model on unseen data.

10. How do we split data for model fitting (training and testing) in Python? How do you approach a Machine Learning problem?

Ans- We use `train_test_split()` from `sklearn.model_selection`.

Example -

```
from sklearn.model_selection import train_test_split  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

Approach a Machine Learning problem:-

Ans- The steps are:

- Understand the problem.
- Collect and clean data.
- Perform EDA.
- Feature engineering.
- Model selection and training.
- Evaluation and optimization.

11. Why do we have to perform EDA before fitting a model to the data?

Ans- EDA helps understand data distribution, detect outliers, missing values, and relationships between variables. It ensures data quality and helps in selecting suitable features and models.

12. What is correlation?

Ans- Correlation measures the relationship between two variables and indicates whether they move together or in opposite directions. It does not imply causation.

13. What does negative correlation mean?

Ans- Negative correlation indicates an inverse relationship between variables. As one variable increases, the other decreases.

14. How can you find correlation between variables in Python?

Ans- We use the corr() method in Pandas.

Ex- df.corr()

15. What is causation? Explain difference between correlation and causation with an example.

Ans- Causation means one variable directly affects another.

Correlation only shows a relationship, not cause.

Example: Ice cream sales and drowning cases are correlated, but heat causes both, not ice cream.

16. What is an Optimizer? What are different types of optimizers? Explain each with an example.

Ans- An optimizer updates model parameters to minimize loss.

Types include:

- Gradient Descent
- Stochastic Gradient Descent (SGD)
- Adam Optimizer

Example: Adam is widely used in neural networks due to fast convergence.

17. What is sklearn.linear_model ?

Ans- sklearn.linear_model is a Scikit-learn module that provides linear models such as Linear Regression, Logistic Regression, Ridge, and Lasso.

18. What does model.fit() do? What arguments must be given?

Ans- `model.fit()` trains the model using training data.

Arguments:

- Input features (`X_train`)
- Target values (`y_train`)

19. What does model.predict() do? What arguments must be given?

Ans- `model.predict()` generates predictions for new or test data.

Argument:

- Input features (`X_test`)

20. What are continuous and categorical variables?

Ans- Continuous variables are numeric and measurable.

Categorical variables represent groups or categories

21. What is feature scaling? How does it help in Machine Learning?

Ans- Feature scaling standardizes feature values to a common range. It helps models converge faster and improves performance, especially for distance-based algorithms.

22. How do we perform scaling in Python?

Ans- Using Scikit-learn scalers such as:

- StandardScaler
- MinMaxScaler

Ex- `from sklearn.preprocessing import StandardScaler`

```
scaler = StandardScaler()  
X_scaled = scaler.fit_transform(X)
```

23. What is sklearn.preprocessing?

Ans- It is a module that provides tools for scaling, encoding, normalization, and transformation of data before model training.

24. How do we split data for model fitting (training and testing) in Python?

Ans- By using `train_test_split()` from Scikit-learn to divide data into training and testing sets.

25. Explain data encoding?

Ans- Data encoding converts categorical data into numerical format so that machine learning models can process it. Common encoding methods include Label Encoding and One-Hot Encoding.