

### **What is a parameter?**

A parameter is a variable that defines a model's configuration or behavior. In Machine Learning, parameters are internal model values learned from the data during training, such as weights in linear regression or coefficients in logistic regression.

### **What is correlation?**

Correlation is a statistical measure that describes the strength and direction of a relationship between two variables. It ranges from -1 to +1.

### **What does negative correlation mean?**

Negative correlation means that when one variable increases, the other decreases. For example, the more time spent exercising, the lower the body fat percentage.

### **Define Machine Learning. What are the main components in Machine Learning?**

Machine Learning is a subset of Artificial Intelligence that enables computers to learn from data and make predictions without being explicitly programmed. The main components are: 1) Data, 2) Model, 3) Loss Function, 4) Optimizer, 5) Evaluation Metrics.

### **How does loss value help in determining whether the model is good or not?**

Loss value indicates how well a model is performing. A lower loss means the model's predictions are closer to the actual values, implying better performance.

### **What are continuous and categorical variables?**

Continuous variables have numerical values within a range (e.g., age, salary). Categorical variables represent categories or labels (e.g., gender, color, city).

### **How do we handle categorical variables in Machine Learning? What are the common techniques?**

Categorical variables are converted to numerical form using techniques like Label Encoding, One-Hot Encoding, and Ordinal Encoding.

### **What do you mean by training and testing a dataset?**

Training a dataset means teaching the model using known data and corresponding outputs. Testing evaluates how well the trained model performs on unseen data.

### **What is sklearn.preprocessing?**

It is a module in Scikit-learn that provides tools for scaling, transforming, and encoding features before training a machine learning model.

### **What is a Test set?**

A test set is a subset of data used to evaluate the performance of a trained model. It is not used during training.

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

We use the function: `train_test_split(X, y, test_size=0.2, random_state=42)` from `sklearn.model_selection`.

### ***How do you approach a Machine Learning problem?***

- 1) Define the problem, 2) Collect and preprocess data, 3) Perform Exploratory Data Analysis (EDA),  
4) Select features, 5) Choose a model, 6) Train the model, 7) Evaluate performance, 8) Tune parameters, 9) Deploy and monitor.

### ***Why do we have to perform EDA before fitting a model to the data?***

EDA helps understand data patterns, detect outliers, handle missing values, and decide which features are important for modeling.

### ***What is correlation?***

Correlation measures the linear relationship between two variables. It helps understand how one variable moves with respect to another.

### ***What does negative correlation mean?***

It means that an increase in one variable causes a decrease in the other. Example: increase in rainfall may decrease the sale of sunglasses.

### ***How can you find correlation between variables in Python?***

You can use Pandas: `df.corr()` or NumPy: `np.corrcoef(x, y)`.

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

Causation means one variable directly affects another. Correlation only shows a relationship, not cause and effect. Example: Ice cream sales and drowning cases are correlated, but eating ice cream doesn't cause drowning; both increase in summer.

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

An optimizer adjusts model parameters to minimize the loss function. Common types: 1) Gradient Descent – updates weights based on gradient, 2) SGD – updates per batch, 3) Adam – adaptive learning rate optimization combining RMSProp and Momentum.

### ***What is `sklearn.linear_model`?***

It is a module in Scikit-learn that provides linear models like Linear Regression, Logistic Regression, Ridge, and Lasso Regression.

### ***What does `model.fit()` do? What arguments must be given?***

`model.fit()` trains the model using input features (`X`) and target labels (`y`). Arguments: `X_train` and `y_train`.

### ***What does `model.predict()` do? What arguments must be given?***

`model.predict()` uses the trained model to make predictions on new data. Argument: `X_test`.

**What are continuous and categorical variables?**

Continuous: measurable numeric values. Categorical: non-numeric labels or groups.

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

Feature scaling standardizes data to a similar range, improving algorithm performance (especially gradient-based methods).

**How do we perform scaling in Python?**

Using StandardScaler or MinMaxScaler from sklearn.preprocessing.

**What is sklearn.preprocessing?**

It provides methods for feature transformation like scaling, normalization, encoding, and imputing.

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

Using train\_test\_split(X, y, test\_size=0.2, random\_state=42).

**Explain data encoding?**

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