Machine Learning Basics - 2

1. What are Overfitting and Underfitting?

- **Overfitting**
  The model learns the training data too well, including noise and random details.
  It performs great on training data but poorly on new/unseen data.
  It memorizes instead of learning patterns.
- **Underfitting**
  The model is too simple and fails to learn the important patterns in the data.
  It performs poorly on both training and testing data.
  It doesn't learn enough.

2. What is the difference between Supervised and Unsupervised Learning?

Supervised Learning	**Unsupervised Learning**
Uses labeled data (input + correct output)	Uses unlabeled data (no correct answers)
Learns to predict outputs	Learns to find hidden patterns or groups
Example: Email spam detection	Example: Customer segmentation

3. What is a training dataset and a testing dataset? Why is data splitting important?

- **Training Dataset**
  The data used to train the model so it can learn patterns.
- **Testing Dataset**
  The data used to evaluate the model's performance on unseen data.
- **Why Splitting is Important**
  If we test on the same data we trained on, the model may look very accurate but actually

fail in real-world use.
Splitting helps us measure how well the model generalizes.

4. What is feature scaling and why is it needed in some algorithms?

Feature scaling means bringing different features to a similar range of values (like 0–1 or mean 0, std 1).

It is important because some algorithms (like KNN, SVM, Gradient Descent-based models) are sensitive to the scale of data.
Without scaling, features with large values can dominate and mislead the model.

5. How does a Linear Regression model work?

Linear Regression predicts a value using a straight-line relationship between input features and the target.

It tries to find the best-fitting line:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

Where:

- **y** = predicted value
- **x_1, x_2…** = features
- **b_0** = intercept
- **b_1, b_2…** = coefficients (how much each feature affects y)

The model chooses these coefficients by minimizing the error between predicted and actual values (usually using Least Squares).