1- AI, ML, DL & Data Science (Big Picture)

Artificial Intelligence (AI)

The umbrella field.

Goal: make machines perform tasks that normally require human intelligence.

Examples: recommendation systems, spam filters, self-driving cars.

Machine Learning (ML)

Subset of AI.

Instead of hardcoding rules → we feed data.

The model learns patterns using statistics.

Deep Learning (DL)

Subset of ML.

Uses multi-layer neural networks.

Inspired by how neurons in the brain work.

Data Science

Broader field including:

ML

Statistics

Data analysis

Domain knowledge

Focuses on extracting insights from data.

Hierarchy: AI ⊃ ML ⊃ DL

2- Types of Machine Learning

Supervised Learning

Data is labeled.

We have:

Input features (X)

Output/target (Y)

➤ Regression

Predicts continuous values.

Example:

House price

Salary

Equation (single variable):

y = mx + b

m → slope (coefficient)

b → intercept

➤ Multiple Linear Regression

More than one feature:

Price = m1·Area + m2·Bedrooms + m3·Age + b

Very important in real-world datasets.

Classification

Output is categorical.

Example:

Spam / Not Spam

Pass / Fail

Unsupervised Learning

No labeled output.

Model finds patterns on its own.

➤ Clustering

Groups similar data.

Example: customer segmentation.

➤ Dimensionality Reduction

Used when we have too many features.

Reduces complexity while keeping important information.

Techniques:

PCA

LDA

3- Linear Regression (Deep Understanding)

Objective

Find the Best Fit Line that minimizes prediction error.

Cost Function

Most commonly used: Mean Squared Error (MSE)

Measures: Difference between actual values and predicted values.

Goal: Minimize cost.

Lower cost → better model.

4- Gradient Descent

Optimization algorithm used to minimize cost function.

Steps:

Start with random parameters.

Compute cost.

Update parameters in direction of negative gradient.

Repeat until minimum error is reached.

Learning Rate (α)

Very important hyperparameter.

Too large → may overshoot minimum.

Too small → slow convergence.

5- Model Performance

Overfitting

Performs very well on training data.

Poor performance on test data.

Model memorizes noise.

Characteristics:

Low bias

High variance

Underfitting

Model too simple.

Poor performance on both training & test.

Characteristics:

High bias

Low variance

Generalization

Balanced model.

Performs well on unseen data.

6- Evaluation Metrics

R-Squared (R^2)

Measures how much variance in Y is explained by X.

Value between 0 and 1.

Higher = better fit.

Adjusted R^2

Penalizes unnecessary features.

Better when using multiple variables.

7- Regularization (Prevent Overfitting)

We modify cost function by adding penalty term.

L2 Regularization (Ridge)

Adds squared value of coefficients.

Shrinks coefficients.

Doesn’t make them zero.

Used when:

All features are important.

L1 Regularization (Lasso)

Adds absolute value of coefficients.

Can make some coefficients exactly zero.

Helps in feature selection.

Used when:

We want automatic feature selection.

8- Practical Implementation (from Codebasics)

Use sklearn.linear_model.LinearRegression

Fit model using: model.fit(X, y)

Predict using: model.predict(new_data)

Also important:

Data preprocessing

Handle missing values