# WEEK : MODEL BUILDING AND DEEP LEARNING

## DAY 16 (14/07/2025)

## Supervised Learning Algorithms:

Today, I explored the intuitive logic behind three core **Supervised Learning algorithms**. Supervised learning means that the model learns from **labeled data**, where each input has a corresponding correct output. Understanding these algorithms is important because they form the foundation for many practical machine learning applications.

### 1. Simple Linear Regression

**Goal:**

- Predict a **continuous numeric value** based on input features.
- Example outputs: price of a house, temperature, rainfall, salary, or stock price.

**Mechanism:**

- Linear regression tries to find the **best fit line** through the data points.
- The “best fit line” is the line that **minimizes the total error** between predicted values and actual values.
- Once the line is computed, any new input (X) can be used to predict an output (Y).

**Example:**

- Predicting house prices:
    - Input features: area in sq. ft, number of bedrooms
    - Output: predicted price in dollars

**Advantages:**

- Very simple and interpretable.
- Works well when the relationship between input and output is roughly linear.
- Useful as a baseline model to understand trends.

## 2. K-Nearest Neighbors (KNN)

**Goal:**

- Classify a new data point based on the labels of its **closest neighbors** in the training data.
- Example outputs: classifying images, detecting spam emails, predicting whether a customer will buy a product.

**Mechanism:**

- When a new data point comes in, the algorithm calculates the **distance** (usually Euclidean distance) to all points in the training dataset.
- It then selects the **K closest neighbors**.
- For classification, the new point is assigned the **most common label** among these K neighbors. For regression, it predicts the **average value** of the neighbors' outputs.

**Example:**

- Classifying animals based on height and weight:
    - If the new animal's height and weight are closest to several dogs rather than cats, it will be classified as a dog.

**Advantages:**

- Very simple and easy to understand.
- Non-parametric: no assumption about the underlying data distribution.
- Works well for small, clean datasets.

## 3. Decision Trees

**Goal:**

- Classify data or make decisions by asking a **series of yes/no questions**.
- Example outputs: spam detection, medical diagnosis, customer segmentation.

**Mechanism:**

- The tree starts at a **Root Node**, representing the whole dataset.
- Each **Internal Node** asks a question about a feature (e.g., “Is the pet’s weight >10 ?”).
- Based on the answer, the data moves down the appropriate branch.
- At the **Leaf Node**, the final prediction (class label or value) is made.

**Example:**

- Classifying fruits:
    - Feature 1: Color (red, green)
    - Feature 2: Size (small, large)
    - Output: Apple, Cherry, or Watermelon

**Advantages:**

- Highly interpretable and easy to visualize.
- Can handle both numerical and categorical data.
- No need for feature scaling or normalization.

## Reflection:

Today, I learned that different supervised learning algorithms have different strengths:

- Linear Regression for numeric prediction
- KNN for classifying new data points based on the similarity to existing examples
- Decision Trees for stepwise decision-making

Even with small datasets, these algorithms can provide powerful insights and form the foundation for more advanced models like neural networks.

| Algorithm | Goal | Mechanism | Example |
|---|---|---|---|
| Linear Regression | Predict numeric value | Best fit line to minimize error | House price prediction |
| KNN | Predict class based on neighbors | Assign label of K closest points | Dog vs Cat classification |
| Decision Tree | Step-by-step decision | Series of yes/no questions | Fruit classification |