Lesson 4: Supervised Learning Part 2

◇ Introduction

Welcome back! In the previous lesson, we introduced supervised learning and explored its two main types: regression and classification. Now it’s time to dig deeper into the most commonly used algorithms, understand how they work, and learn how to choose the right one for your specific use case.
By the end of this lesson, you'll have a clearer view of how supervised learning algorithms operate — and why each one has strengths and trade-offs.

◇ Popular Supervised Learning Algorithms

Let’s explore some of the most well-known supervised learning algorithms:

📉 1. Linear Regression (for Regression Problems)

Use Case: Predicting continuous values (e.g., house prices, salary).
How it works: Tries to fit a straight line that best represents the relationship between features and the target.
Simple and interpretable, great for baseline models.

✅ 2. Logistic Regression (for Classification Problems)

Use Case: Binary classification (e.g., spam or not spam).
Despite the name, it's used for classification, not regression.
Outputs probabilities and makes decisions based on a threshold (e.g., > 0.5 = class 1).

🧠 3. Decision Trees

Use Case: Both classification and regression.
How it works: Splits data into branches based on feature values.
Easy to interpret and visualize but can overfit without constraints.

🌲 4. Random Forest

Use Case: Both tasks.
What it is: A collection of decision trees (ensemble method).
Advantage: More accurate and stable than a single decision tree; handles missing data well.

🔍 5. K-Nearest Neighbors (KNN)

Use Case: Simple classification and regression problems.
How it works: Classifies a data point based on the majority label of its 'K' closest neighbors.
Lazy learner: No actual model-building during training, but can be slow during prediction.

🔷 6. Support Vector Machines (SVM)

Use Case: Classification (especially for high-dimensional data).
How it works: Finds the hyperplane that best separates the data.
Powerful but computationally expensive on large datasets.

◇ How to Choose the Right Algorithm

Criteria	Considerations
Size of Dataset	Small? Try SVM or KNN. Large? Go for Random Forest or Linear Models.
Type of Output	Use regression models for numbers, classification models for categories.
Interpretability	Linear and logistic regression are most transparent.
Accuracy vs Simplicity	More complex models like Random Forest often perform better but are harder to interpret.
Speed	Linear models are fast; KNN and SVM can be slow on large data.

◇ Outro

That wraps up part 2 or our supervised learning topic, here is that what we covered:

- **Key supervised learning algorithms and how they work.**
- **Differences between them, and when to use which one.**
- **Considerations for choosing the best algorithm based on your needs.**