Day 06: Supervised Learning Basics - Regression & Classification

Supervised learning is a type of machine learning where the model is trained on a labeled dataset, meaning each input has a known correct output. The model learns to map inputs to outputs and can generalize to make predictions on new data.

---

* What is Regression?

Regression is used when the output variable is continuous. The goal is to predict a numerical value based on input features.

-> Linear Regression:

- Models the relationship between input features and target using a straight line.
- Equation: $y = mx + b$
- Assumes a linear relationship.

-> Evaluation Metrics:

- MSE (Mean Squared Error): average squared difference between predicted and actual values.
- MAE (Mean Absolute Error): average absolute difference.
- R^2 Score: proportion of variance explained by the model.

-> Use Cases:

- House price prediction, stock market forecasting, temperature estimation.

---

* What is Classification?

Classification is used when the output variable is categorical. The model learns to assign inputs to one of several predefined classes.

-> Logistic Regression:

- Despite the name, it's used for classification.
- Outputs a probability that a given input belongs to a certain class.
- Uses the sigmoid function.

-> Other Models:

- Decision Tree: Flowchart-like structure for decision making.
- k-NN (k-Nearest Neighbors): Classifies based on majority class among nearest data points.

-> Evaluation Metrics:

- Accuracy: Percentage of correctly classified instances.
- Precision: Correct positive predictions / total predicted positives.
- Recall: Correct positive predictions / actual positives.
- F1 Score: Harmonic mean of precision and recall.
- Confusion Matrix: Summarizes true vs predicted classifications.

-> Use Cases:

- Spam detection, image recognition, medical diagnosis.

---

* Summary

Supervised learning is essential for making predictions when labeled data is available. Regression handles continuous outputs, while classification deals with categories.

Use scikit-learn (`sklearn`) to quickly build and test these models in Python.