Instance-Based Vs Model-Based Learning

Machine learning models learn from data in two primary ways, similar to how humans learn. The key difference is how the model generalises the data to make predictions.

1- Instance Based Learning

→ This method is analogous to memorising data
→ The model stores the training examples and makes predictions on new data by directly comparing it to these stored instances.

How Instance Based Learning Work:

→ This approach is heavily dependent on the training data for every prediction it makes
→ model stores the entire training dataset
→ The actual "learning" or pattern discovery is postponed until a new data point (a query) is received
→ When a new data point arrives, the model measures its similarity or distance to the stored training points
→ It then identifies the closest data points. The prediction for the new point is based on the outcomes of these neighbours.

Example: For a student, to predict if they will pass or fail based on study and play hours, the model would find the most similar students in the existing data. If the majority of those similar students passed, the model predicts "pass" for the new student.

Algorithms:

→ k-Nearest Neighbors (k-NN) is a classic example.
→ Support Vector Machines (SVM).
→ Algorithms using Kernel Functions.

When to Use:

→ Ideal for small to medium datasets where storing all the data is feasible.
→ Useful in situations that require dynamic updates, such as recommendation systems.

2 - Model Based Learning

→ This method is analogous to generalising or understanding the underlying concepts and patterns in the data.
→ The model builds a generalised representation (like a mathematical equation or a set of rules) from the training data and uses this model for future predictions.

How Instance Based Learning Work:

→ Understand the patterns in the data to create a predictive model
→ During training, the algorithm analyses the entire dataset to discover patterns, relationships, and estimate model parameters.
→ It creates a generalised model, often in the form of a mathematical function or a decision boundary.
→ For any new data point, the model simply checks which side of the boundary it falls on to make a prediction.
→ **A key feature** is that once the model is trained and the decision boundary is established, the original training data is no longer needed for making predictions. The model itself, containing all the learned patterns, is what's used.

Algorithms: Most machine learning algorithms fall into this category.

→ Linear Regression
→ Logistic Regression
→ Decision Trees
→ Neural Networks

When to Use:

→ Ideal for large datasets that have consistent patterns.
→ Well-suited for applications needing quick predictions, like spam detection or stock price forecasting.

Key Differences:

Feature	Model-Based Learning ("Generalising")	Instance-Based Learning ("Memorising")
Primary Goal	Learn patterns to build a generalised model.	Memorises training data to use for direct comparison.
Training Data	Used to train the model, but can be **discarded after training**.	**Must be kept** and is used directly for every prediction.

Model Storage	**Memory-efficient**. The model is stored in a compact, serialised format (e.g., pickle, H5 file).	**Storage-intensive**. Requires storing the entire training dataset, which can be large.
Prediction Speed	**Generally faster**, as it involves applying a pre-learned mathematical function.	**Can be slower**, as it requires computing distances to many stored data points for each new prediction.
Adaptability	Requires retraining the entire model to incorporate new data.	Adapts naturally to new data without needing retraining.