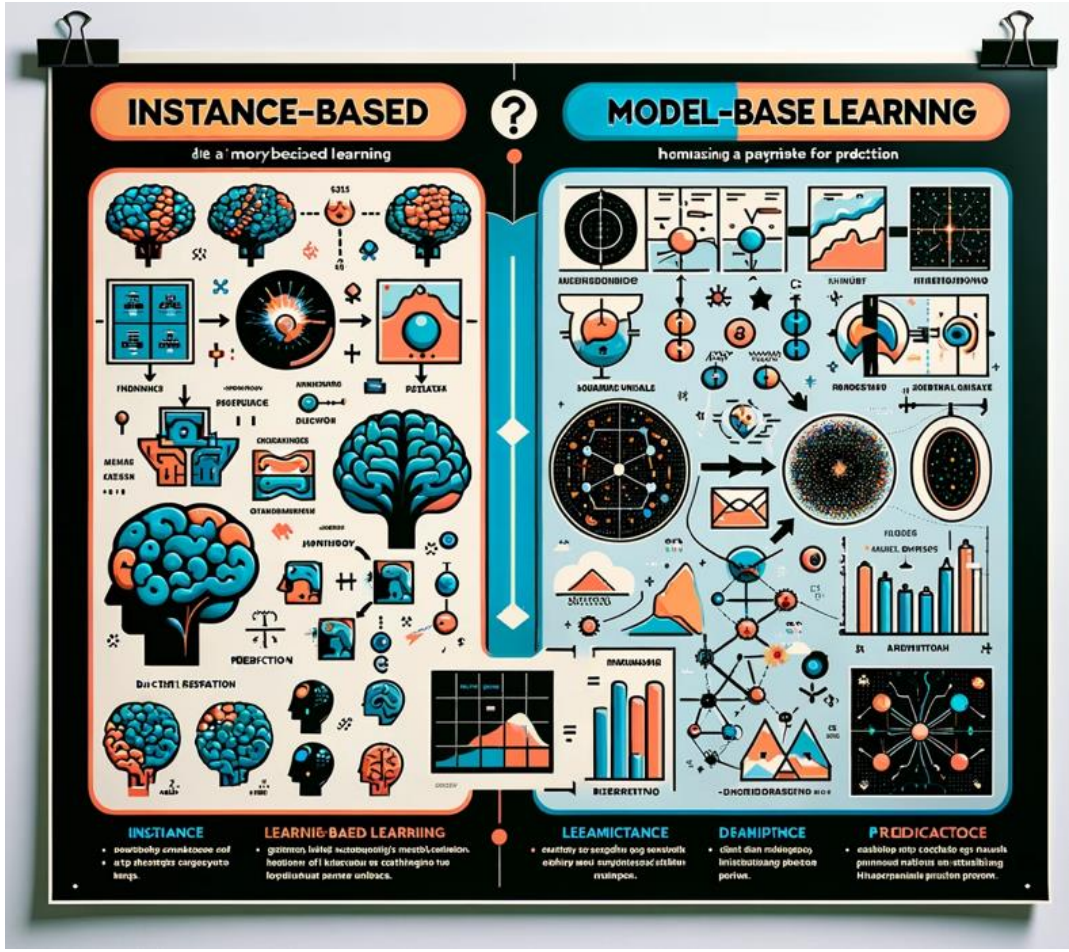
Machine Learning

| Instance-Based Vs. Model-Based

Prof. Dr. Mostafa Elhosseini

Professor of Smart Systems Engineering

https://youtube.com/drmelhosseini



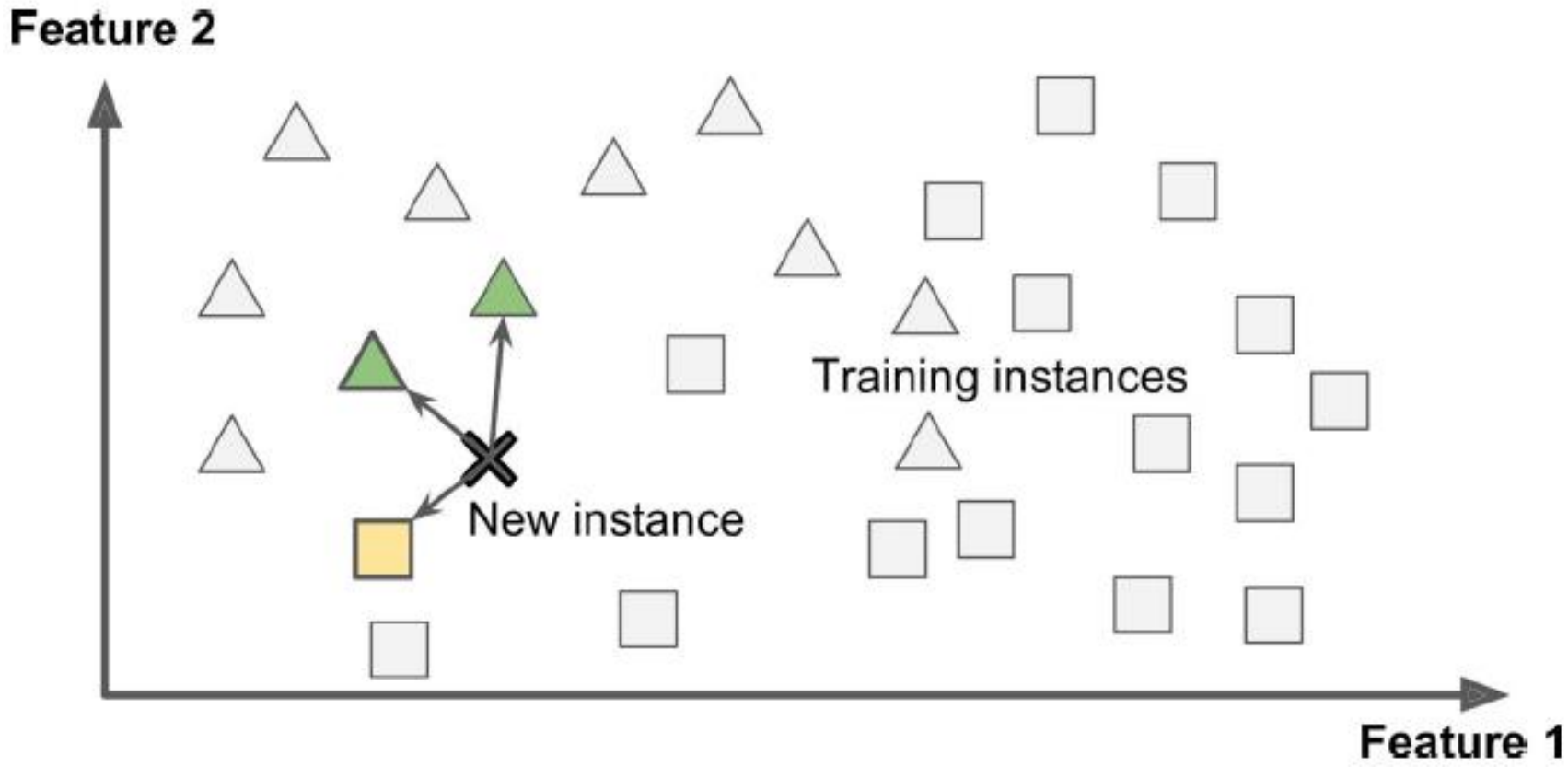
Introduction

- One more way to categorize Machine Learning systems is by **how they generalize**
  - Needs to be able to generalize to examples it has never seen before
- Having a good performance measure on the training data is good, but insufficient; the **true goal is to perform well on new instances**
- There are two main approaches to generalization:
  - Instance-based learning and
  - Model-based learning

Instance-based learning

- A type of machine learning approach where the model learns the training data by heart and makes predictions based on the similarity between new data points and instances of the training data
  - Flag emails that are very similar to known spam emails
  - Not the worst solution, but certainly not the best.
  - The system would flag an email as spam if it has many words in common with a known spam email
- This requires a **measure of Similarity** between two emails
  - Common measures include Euclidean distance for numerical data and Hamming distance for categorical data.

Instance-based learning



Instance-based learning

- **k-Nearest Neighbors (k-NN)**: This is one of the most well-known instance-based learning algorithms. For a new data point, the algorithm identifies the 'k' closest instances in the training dataset and bases the prediction on the majority label of these 'k' instances.
- **Case-Based Reasoning**: This involves solving new problems based on the solutions of similar past problems.
- **Use Cases**: Instance-based learning is often used in applications where the decision boundary is irregular and hard to model with parametric methods. It's also suitable in situations where retraining a model is not feasible, but new data keeps coming in, like in recommendation systems or anomaly detection.

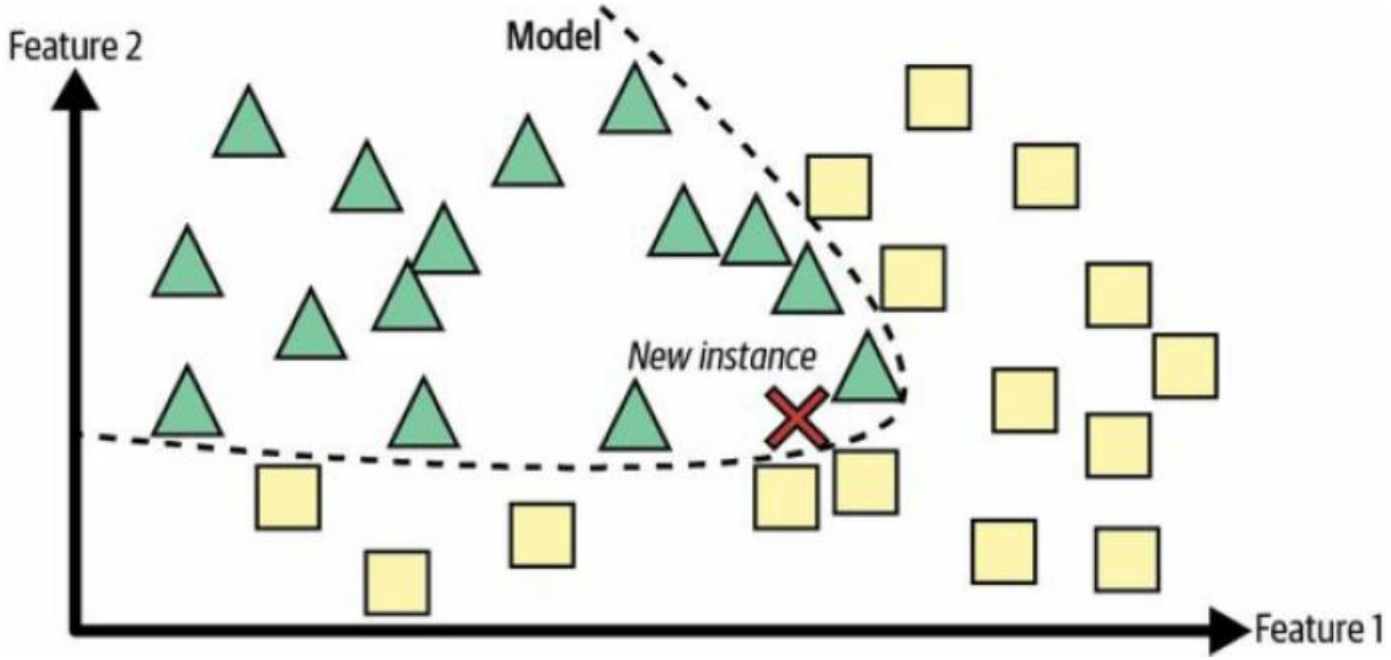
Model-based learning

- The focus is on constructing a model that can make predictions by understanding and generalizing from the underlying structure of the data.
- Model-based learning involves creating a model that encapsulates the relationships in the data
- The first step is to **choose a suitable model or algorithm** that fits the nature of the problem and the data. This model could be a simple linear model, a decision tree, a neural network, or any other algorithm that can learn from data.

Model-based learning

- The model is then **trained on a dataset**, where it tries to find patterns or relationships within the data. During this training process, the model **adjusts its parameters to minimize the error** between its predictions and the actual outcomes.
- The ultimate goal of model-based learning is to create a model that not only performs well on the training data but also generalizes well to new, unseen data
- Once trained, the model can make predictions about new data

Model-based learning



Model-based learning

- Model-based learning is widely used in various applications from simple regression tasks to complex deep learning applications in computer vision, natural language processing, and more. The choice of model and learning algorithm depends heavily on the specific problem, the nature of the data, and the desired outcome.