2- Intro to Machine Learning 1

By: eng. Esraa Madhi

So we defined data science as: It's the process of asking interesting questions, and then answering those questions using data.

- **When these questions involve making predictions about the future, this is where machine learning comes into play.**

For example, consider a scenario where we have data on a 2015 Kia car sold for 22,000 SR and a 2017 Kia sold for 25,000 SR. If someone wants to determine the potential selling price of their 2016 Kia, machine learning can help make this prediction.

User-uploaded image: 510701_1050x520.jpg

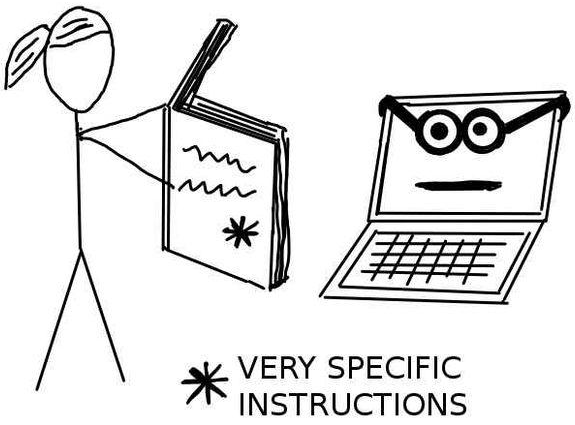
What is Machine Learning?

So we can define machine learning as: Finds patterns in data and uses them to make predictions.
Machine learning model is: a mathematical representation of the patterns hidden in data.

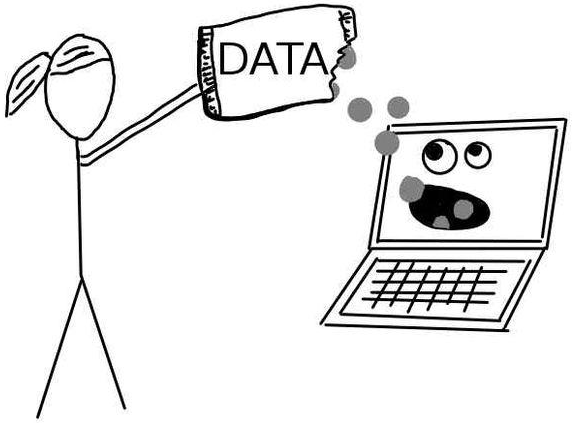
what is the difference between Machine learning and Traditional software

- **Traditional software (Coding)** is basically a set of rules, written by a human, intended to achieve a particular output.
- **Machine learning** software finds rules (patterns) on its own and tries to produce a certain output (It's software that writes software).

Without Machine Learning

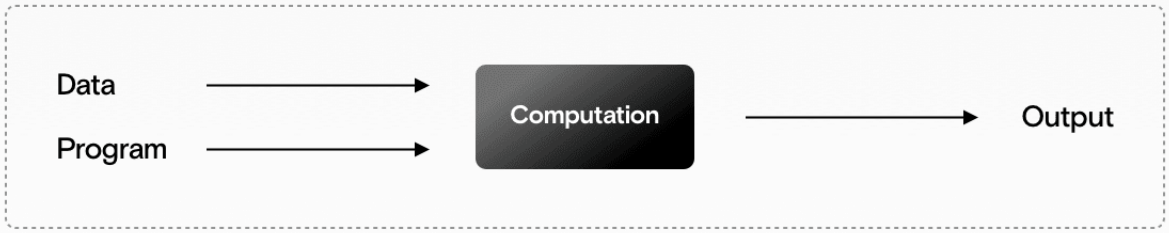


With Machine Learning



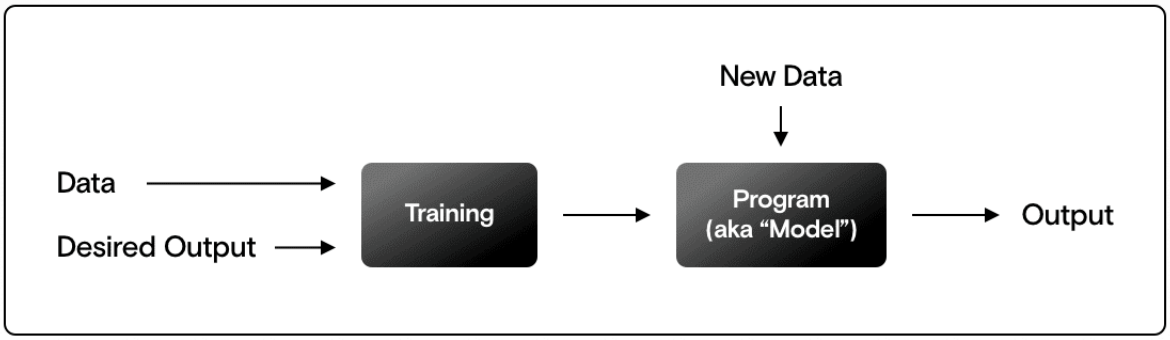
Traditional Programming

Developers write rules (program) that produce an output.



Machine Learning

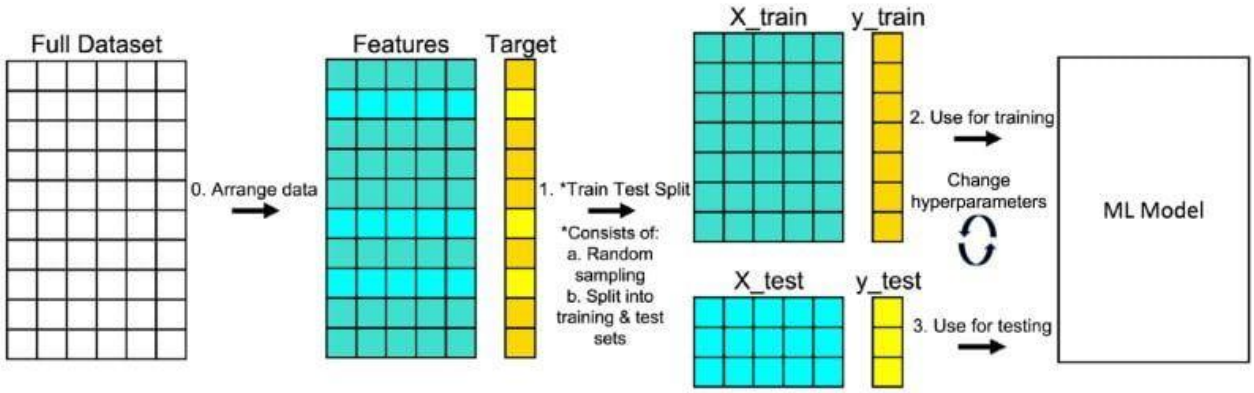
Developers write a training algorithm, that finds rules, which produce the desired output.



https://youtu.be/QghjaS0WQQU

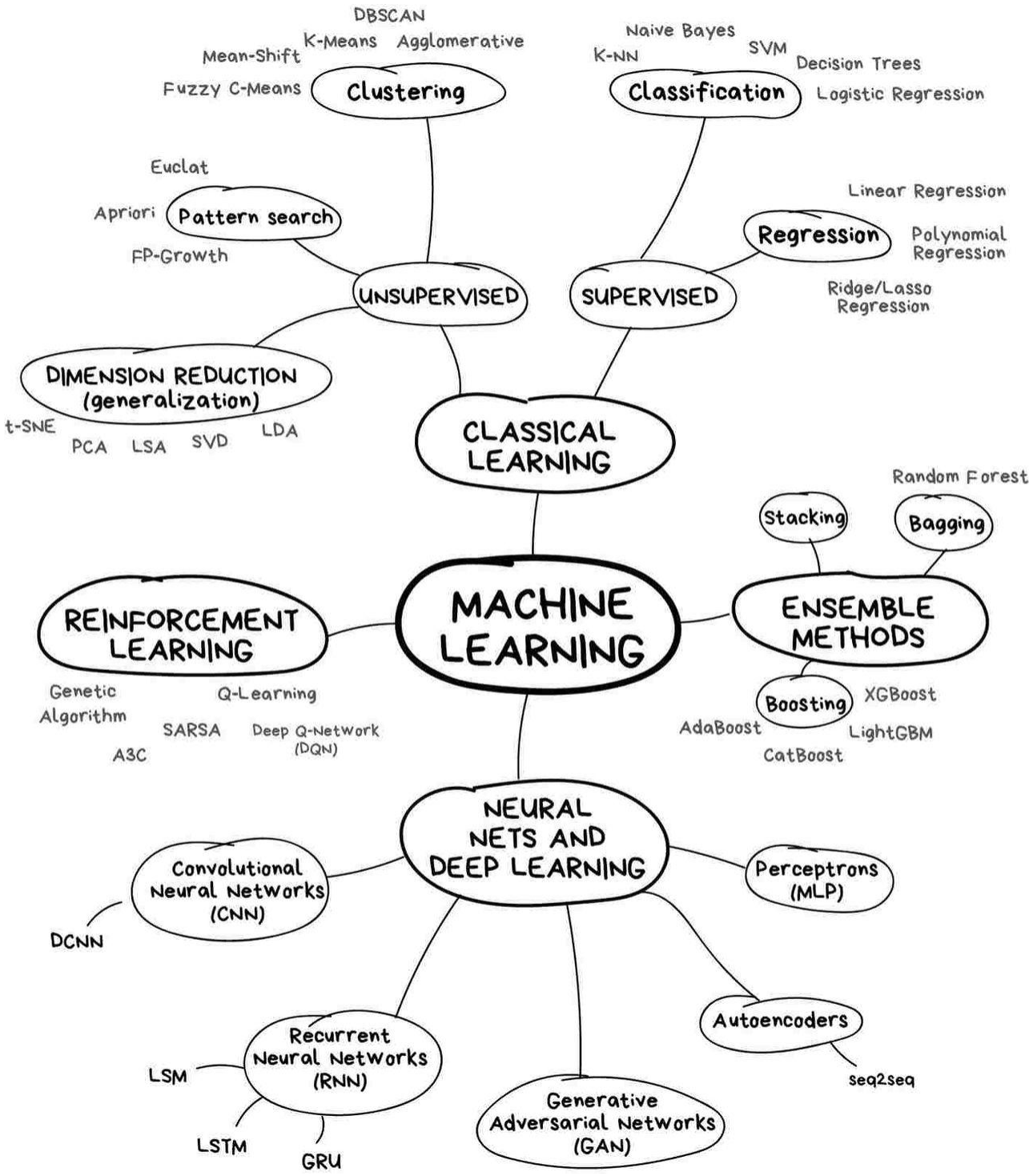
https://youtu.be/QghjaS0WQQU

How do ML algorithms learn (find rules / patterns)?



The machine learning model is trained / built / fit on some training data and then apply that model to new data, the model would be able to infer some relationship within it.

Machine learning categories:



In this course, our primary focus will be on classical learning techniques. We will also provide a broad overview of other methods for a comprehensive understanding.

Classical Learning:

1. **Supervised Learning**

In supervised learning, the algorithm is trained on a labeled dataset. This means that the input data is paired with the correct output. The model learns from this data to make predictions or decisions without being explicitly programmed to perform the task.

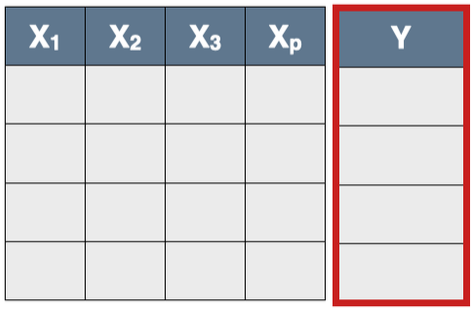
Example: Consider a dataset of housing prices. The dataset includes features (input variables) such as the number of bedrooms, number of bathrooms, square footage, and location. Each record in the dataset also has the associated house price (output variable). In supervised learning, a model could be trained to predict housing prices based on these features. The model would learn the relationship between the features of houses and their market prices.

2. **Unsupervised Learning**

Unsupervised learning involves training a model on a dataset without labeled responses. The model tries to find patterns and relationships in the data on its own. The main goal in unsupervised learning is to discover the underlying structure of the data.

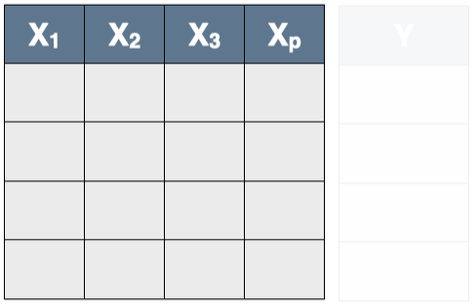
Example: Consider a dataset of customers' shopping habits. The dataset includes features such as the number and types of products purchased, the time of purchase, and the amount spent, but there are no labels telling us anything about these customers. In unsupervised learning, a model might be used to segment the customers into different groups based on their shopping patterns. This is known as clustering.

Supervised Learning



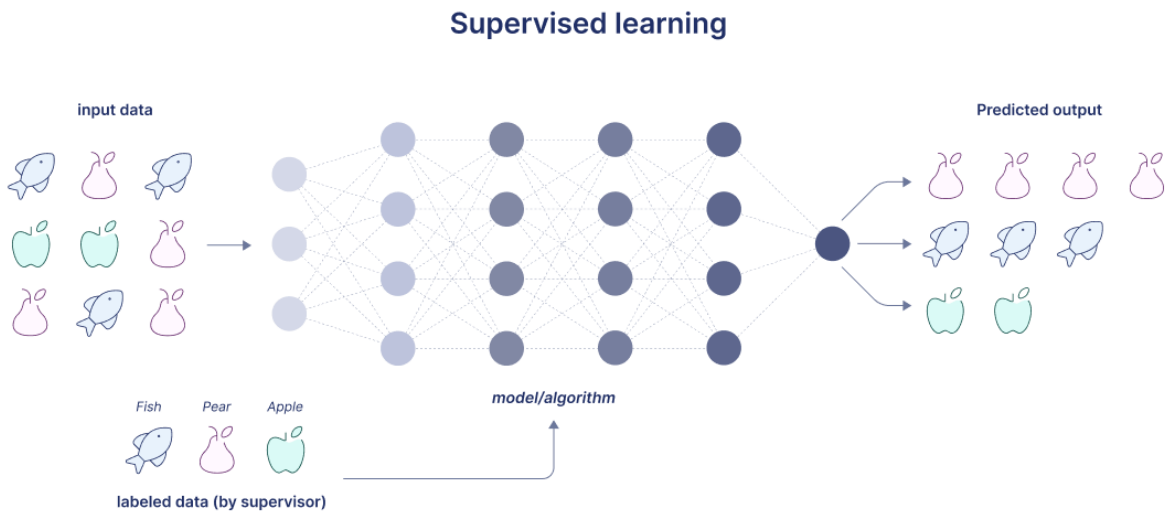
Target

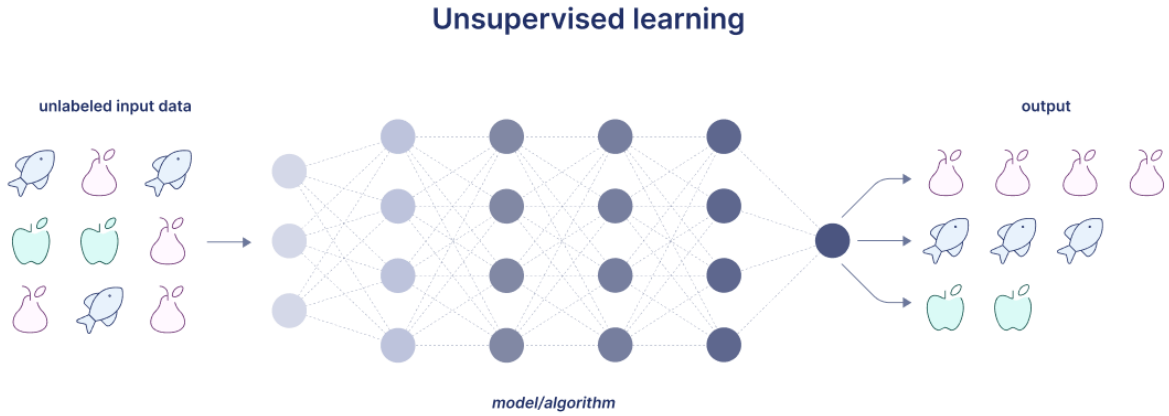
Un-Supervised Learning



No Target

Supervised vs. Unsupervised Learning





Scribbr

Key Differences	Supervised Learning	Unsupervised Learning
Labels	Uses labeled data	Uses unlabeled data
Purpose	Is used for prediction (regression) or classification	Is used for clustering, association, or dimensionality reduction.

	tasks	
Feedback	The model is corrected by the labels during training, guiding the learning process	The model receives no feedback and must identify patterns without guidance.
Complexity of Interpretation	The results of supervised learning can be easier to interpret since the desired output is known	Unsupervised learning often requires more analysis to understand the structures and patterns found in the data
Use Cases	Is commonly used for spam detection, image recognition, and medical diagnosis	Is used for market basket analysis, anomaly detection, and customer segmentation.

Supervised machine learning:

Supervised machine learning is used for two types of problems or tasks:

1. **Classification:** is used when the output variable is a category, such as "spam" or "not spam" in email filtering, or "malignant" or "benign" for a tumor diagnosis.

Output: The output is discrete. It involves assigning a class label to input data from a finite set of possible categories.

2. **Regression:** is used when the output variable is a real or continuous value, such as "salary" or "house price."

Output: The output is continuous. It involves predicting a quantity, which means that the model must predict a range of possible numbers.

Regression Data

X_1	X_2	X_3	X_p	Y
				5.2
				1.3
				23.0
				7.4

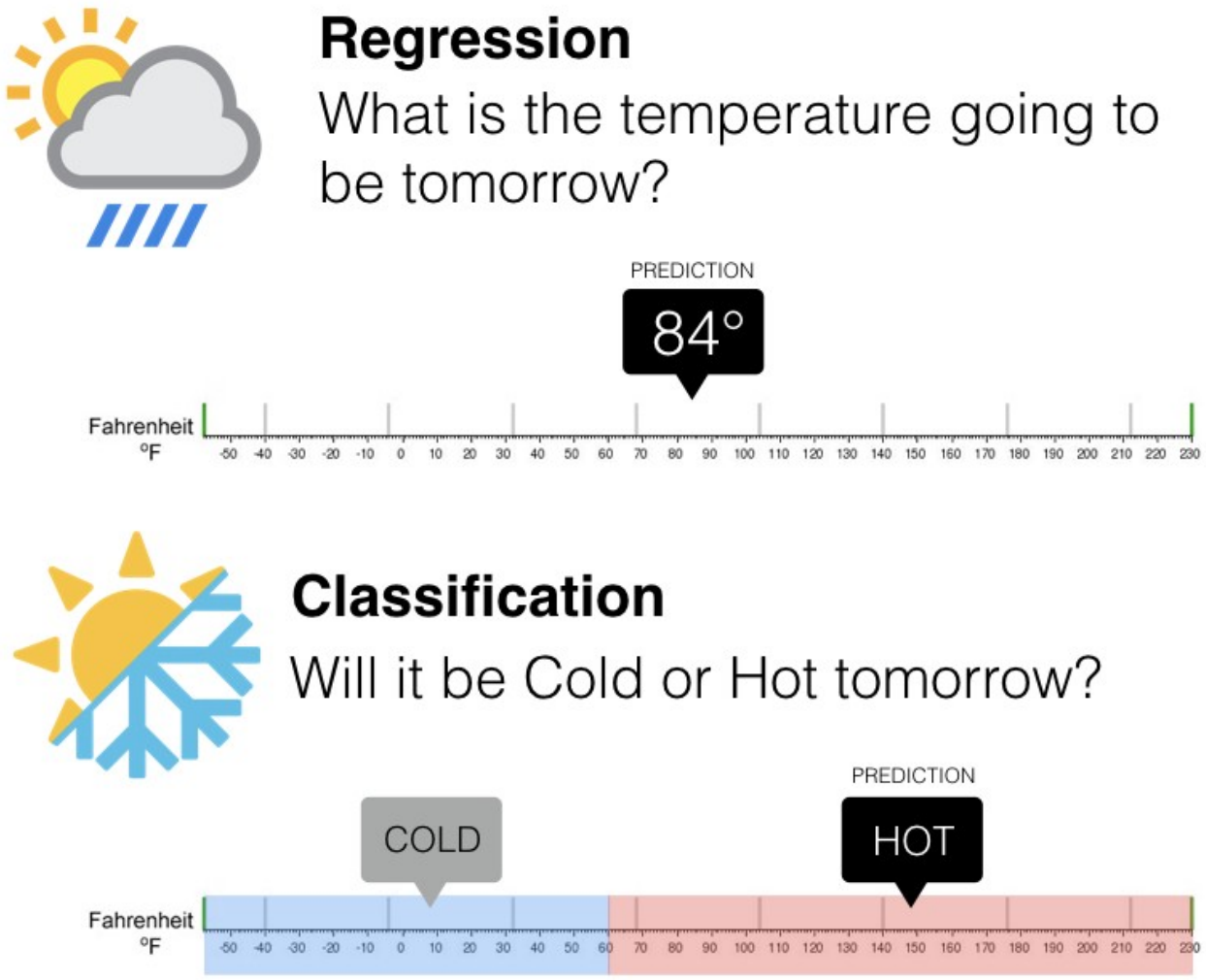
Numeric Target

Classification Data

X_1	X_2	X_3	X_p	Y
				cat
				dog
				cat
				cat

Categorical "Labels"

Key Differences	**Classification**	**Regression**
Output Type	Predicts discrete labels,	Predicts continuous quantities
Nature of Prediction	Classification assigns data into specific categories	Regression outputs a value based on input variables.



So, our Usecase 2 belong to which type?

https://youtu.be/PeMlggyqz0Y

https://youtu.be/PeMlggyqz0Y

Resources:

- https://dataaspirant.com/supervised-and-unsupervised-learning/
- https://www.sharpsightlabs.com/blog/regression-vs-classification/
- https://www.scribbr.com/ai-tools/supervised-vs-unsupervised-learning/