**What is Machine Learning?**

ML is the process of training a piece of software, called a model, to make useful predictions or generate content from data.

**Types of ML Systems**

ML systems fall into one or more of the following categories based on how they learn to make predictions or generate content:

* Supervised learning
* Unsupervised learning
* Reinforcement learning
* Generative AI

**Supervised learning**

Supervised learning models can make predictions after seeing lots of data with the correct answers and then discovering the connections between the elements in the data that produce the correct answers. This is like a student learning new material by studying old exams that contain both questions and answers. Once the student has trained on enough old exams, the student is well prepared to take a new exam. These ML systems are "supervised" in the sense that a human gives the ML system data with the known correct results.

Two of the most common use cases for supervised learning are regression and classification.

**Regression**

A regression model predicts a numeric value. For example, a weather model that predicts the amount of rain, in inches or millimeters , is a regression model.

See the table below for more examples of regression models:

|  |  |  |
| --- | --- | --- |
| **Scenario** | **Possible input data** | **Numeric prediction** |
| Future house price | Square footage, zip code, number of bedrooms and bathrooms, lot size, mortgage interest rate, property tax rate, construction costs, and number of homes for sale in the area. | The price of the home. |
| Future ride time | Historical traffic conditions (gathered from smartphones, traffic sensors, ride-hailing and other navigation applications), distance from destination, and weather conditions. | The time in minutes and seconds to arrive at a destination. |

**Classification**

Classification models predict the likelihood that something belongs to a category. Unlike regression models, whose output is a number, classification models output a value that states whether or not something belongs to a particular category. For example, classification models are used to predict if an email is spam or if a photo contains a cat.

Classification models are divided into two groups: binary classification and multiclass classification. Binary classification models output a value from a class that contains only two values, for example, a model that outputs either rain or no rain. Multiclass classification models output a value from a class that contains more than two values, for example, a model that can output either rain, hail, snow, or sleet.

**Unsupervised learning**

Unsupervised learning models make predictions by being given data that does not contain any correct answers. An unsupervised learning model's goal is to identify meaningful patterns among the data. In other words, the model has no hints on how to categorize each piece of data, but instead it must infer its own rules.

A commonly used unsupervised learning model employs a technique called clustering. The model finds data points that demarcate natural groupings.

|  |  |
| --- | --- |
| An image showing colored dots in clusters. | An image showing colored dots in clusters that are enclosed in a shape and border each other. |
| **Figure 1**. An ML model clustering similar data points. | **Figure 2**. Groups of clusters with natural demarcations. |

Clustering differs from classification because the categories aren't defined by you. For example, an unsupervised model might cluster a weather dataset based on temperature, revealing segmentations that define the seasons. You might then attempt to name those clusters based on your understanding of the dataset.

|  |  |
| --- | --- |
| An image showing colored dots in clusters that are labeled as snow, rain, hail, and no rain. | An image showing colored dots in clusters that are labeled as snow, rain, hail, and no rain that are enclosed in a shape and border each other. |
| **Figure 3**. An ML model clustering similar weather patterns. | **Figure 4**. Clusters of weather patterns labeled as snow, sleet, rain, and no rain. |

**Reinforcement learning**

Reinforcement learning models make predictions by getting rewards or penalties based on actions performed within an environment. A reinforcement learning system generates a policy that defines the best strategy for getting the most rewards.

Reinforcement learning is used to train robots to perform tasks, like walking around a room, and software programs like AlphaGo to play the game of Go.

**Generative AI**

Generative AI is a class of models that creates content from user input. For example, generative AI can create unique images, music compositions, and jokes; it can summarize articles, explain how to perform a task, or edit a photo.

Generative AI can take a variety of inputs and create a variety of outputs, like text, images, audio, and video. It can also take and create combinations of these. For example, a model can take an image as input and create an image and text as output, or take an image and text as input and create a video as output.

We can discuss generative models by their inputs and outputs, typically written as "type of input"-to-"type of output." For example, the following is a partial list of some inputs and outputs for generative models:

* Text-to-text
* Text-to-image
* Text-to-video
* Text-to-code
* Text-to-speech
* Image and text-to-image

The following table list examples of generative models, their input, and an example of their possible output:

|  |  |  |
| --- | --- | --- |
| **Model** | **Input** | **Example output** |
| Text-to-text | Who invented the Le Mans race? | The 24 Hours of Le Mans was founded by the Automobile Club de l'Ouest (ACO), a motoring organization founded in 1906. The first race was held on 26 and 27 May 1923 on the Sarthe Circuit, a 13.426-kilometre (8.345-mile) road circuit located near the town of Le Mans in France. The race was originally conceived as a way to promote the ACO and the Sarthe region. |
| Text-to-image | An alien octopus floats through a portal reading a newspaper. | Image of a cartoon octopus. |
| Text-to-video | A photorealistic teddy bear is swimming in the ocean at San Francisco. The teddy bear goes under water. The teddy bear keeps swimming under the water with colorful fishes. A panda bear is swimming under water. | Video of a teddy bear swimming underwater. |
| Text-to-code | Write a Python loop that loops over a list of numbers and prints the prime numbers. |  |
| Image-to-text | Image of a flamingo. | This is a flamingo. They are found in the Caribbean. |

**Supervised Learning**

Supervised learning's tasks are well-defined and can be applied to a multitude of scenarios—like identifying spam or predicting precipitation.

**Foundational supervised learning concepts**

Supervised machine learning is based on the following core concepts:

* Data
* Model
* Training
* Evaluating
* Inference

**Data**

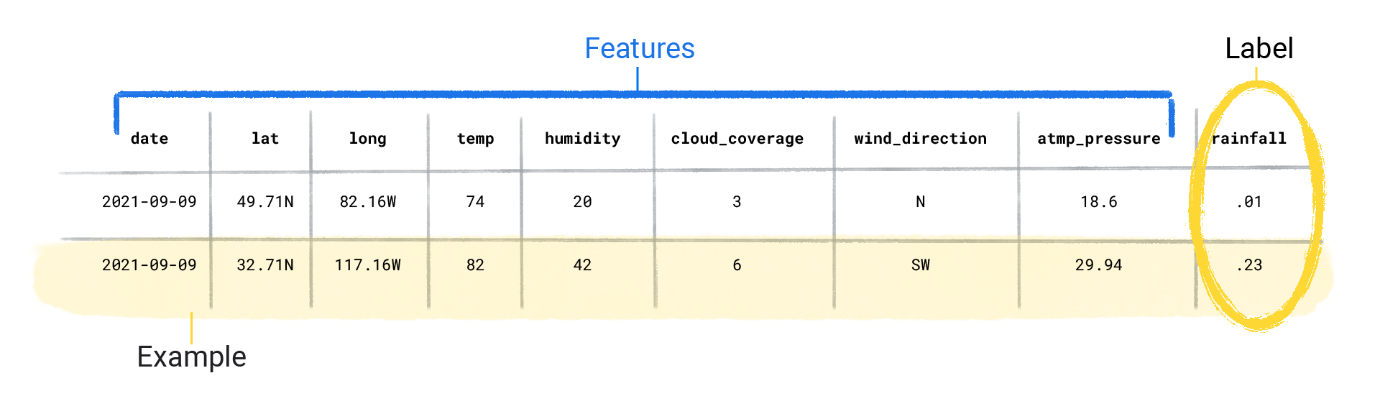
Data is the driving force of ML. Data comes in the form of words and numbers stored in tables, or as the values of pixels and waveforms captured in images and audio files. We store related data in datasets. For example, we might have a dataset of the following:

* Images of cats
* Housing prices
* Weather information

Datasets are made up of individual examples that contain features and a label. You could think of an example as analogous to a single row in a spreadsheet. Features are the values that a supervised model uses to predict the label. The label is the "answer," or the value we want the model to predict. In a weather model that predicts rainfall, the features could be *latitude*, *longitude*, *temperature*, *humidity*, *cloud coverage*, *wind direction*, and *atmospheric pressure*. The label would be *rainfall amount*.

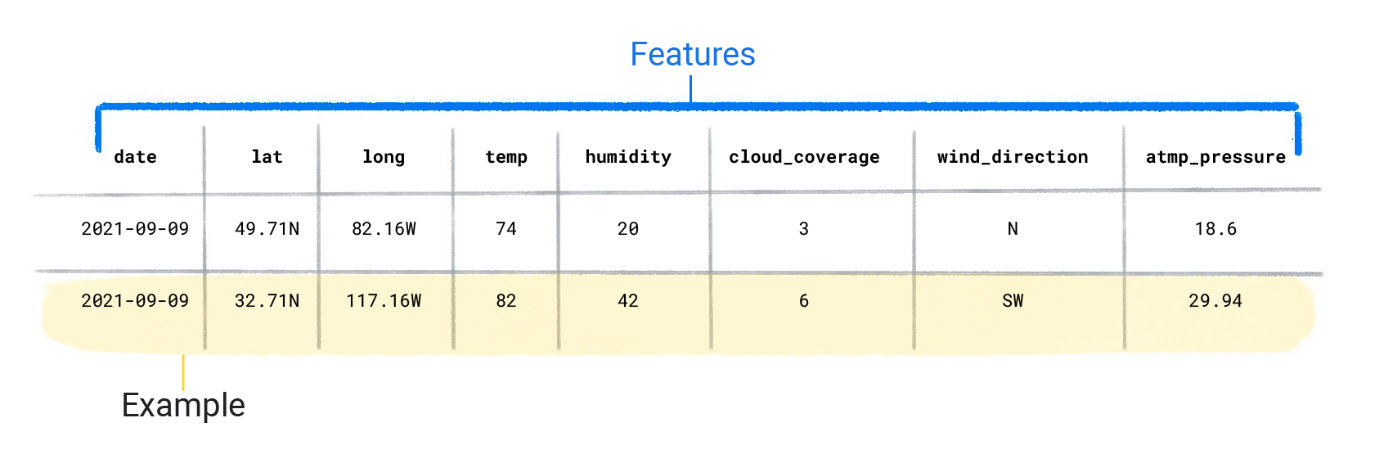
Examples that contain both features and a label are called labeled examples.

Two labeled examples



In contrast, unlabeled examples contain features, but no label. After you create a model, the model predicts the label from the features.

**Two unlabeled examples**



**Dataset characteristics**

A dataset is characterized by its size and diversity. Size indicates the number of examples. Diversity indicates the range those examples cover. Good datasets are both large and highly diverse.

Some datasets are both large and diverse. However, some datasets are large but have low diversity, and some are small but highly diverse. In other words, a large dataset doesn’t guarantee sufficient diversity, and a dataset that is highly diverse doesn't guarantee sufficient examples.

For instance, a dataset might contain 100 years worth of data, but only for the month of July. Using this dataset to predict rainfall in January would produce poor predictions. Conversely, a dataset might cover only a few years but contain every month. This dataset might produce poor predictions because it doesn't contain enough years to account for variability.