Lesson 2: Core ML Concepts

Lesson Objectives

In this lesson, you will be introduced to core ML concepts. Upon successful completion of this lesson, you should be able to understand the following:

- ☐ Dataset
- ☐ Supervised Learning
- ☐ Unsupervised Learning
- ☐ Reinforcement Learning
- ☐ Deep Learning

Dataset

A ***dataset*** in machine learning is an essential part. It is the collection of data that a model will use for training. For example, when you want to learn Spanish, you need to buy a Spanish learning book and a dictionary to give your brain the learning opportunity. Similarly, the AI models need data to learn a task.

A ***labeled dataset*** is a dataset where each data point is associated with a corresp g c label or category. For example, in an image recognition task, the label might indicate t ject o scene depicted in the image.

On the other hand, an ***unlabeled dataset*** is a dataset where the output labels are rovide this case, the machine learning algorithm must find patterns and structure in th data ts own without the aid of explicit output labels. There are several reasons why a datase ht be u eled such as labeling data can be time-consuming and expensive.

The below image provides an example of labeled and unlabele atasets. xamp shows two different forms of labeling the same dataset; one by labeling t object and a her by labeling its weight.

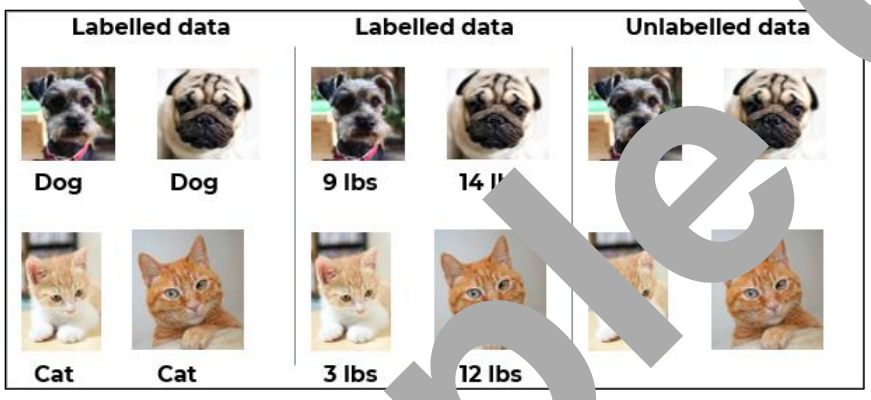

Figure 2-1: Example of labeled and un d datas

The most common form datas e:

- Text data
- Image data
- Audio data
- Vide a
- N neri

When bu ing achine learning model, it is important to split the available data into three different s ining dataset, validation dataset, and testing dataset. This is done to ensure that odel earning in the right way.

- ***T ning dataset*** is the first collection of the data that is used to train the machine learning odel

  Validation dataset is used to evaluate the performance of the model during the training process

- **Train model**: You'll use the training set to train the model
- **Evaluate model**: After the model has been trained, you'll evaluate its performance using the test set

Supervised learning is used in many different applications, ranging from image recognition to speech recognition to natural language processing. For example, you might use supervised learning to build a spam filter that can automatically detect and filter out unwanted emails based on their content. Or you might use supervised learning to build a recommendation system that can suggest movies or products based on a user's past behavior.

Learn the Skill

After collecting data, it is split into:

a. Training and testing sets
b. Training and validation sets
c. Testing and validation sets
d. Training, testing and validation sets

Unsupervised Learning

Unsupervised learning is a subfield of machine learning that enables models to identify patterns and relationships in data without explicit instruction or guidance from humans.

For example, as shown in the below image, imagine you want to train a model to recognize different types of shapes. You give the model images of shapes (hexagon, triangle, and square), and the model would learn how to group the similar shapes together.

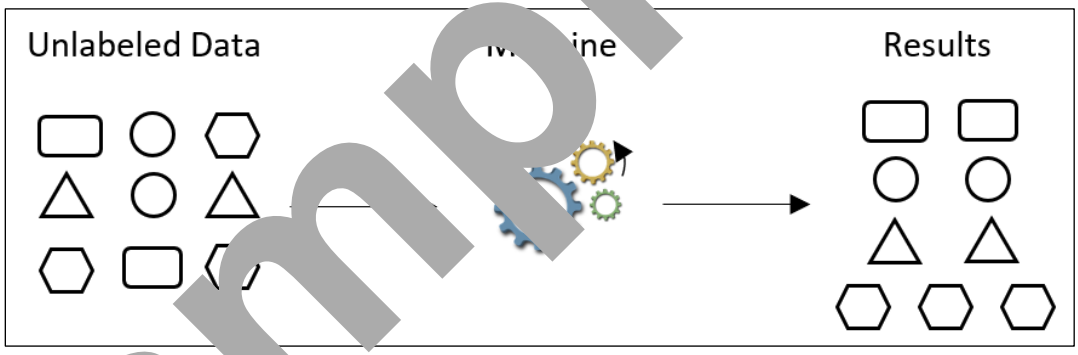

Figure 2-7 Example of unsupervised learning process

One of the most famous techniques in unsupervised learning is ***clustering***. The clustering is based on grouping data points together based on similarities in their attributes. For example, clustering can be used to group customers together based on their purchasing history or to group images together based on their visual features. Clustering algorithms work by calculating the distance between each data point and all other data points in the dataset. Points that are closer together are grouped together into clusters.

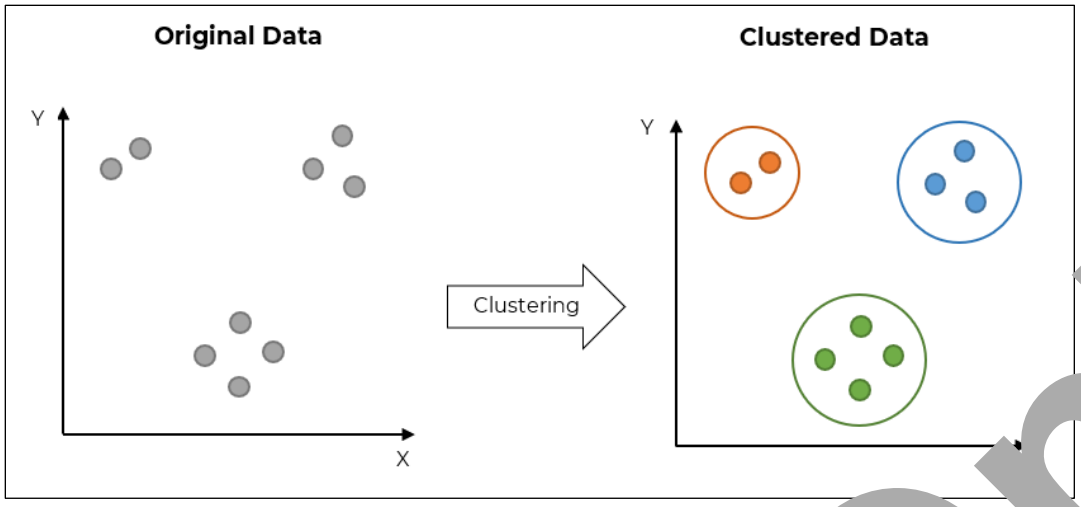

Figure 2-4: Clustering process

Unsupervised learning has many applications in real life. For example, it is used in target marketing campaigns to find the patterns and to target the right audience. Also, it is used in the genetics field by clustering DNA patterns to analyze evolutionary biology.

Learn the Skill

_____ is one of the most famous techniques in unsupervised learning.

Reinforcement Learning

Reinforcement learning is another subfield of machine learning. Reinforcement learning is based on trial and error using feedback from the model's actions and experiences.

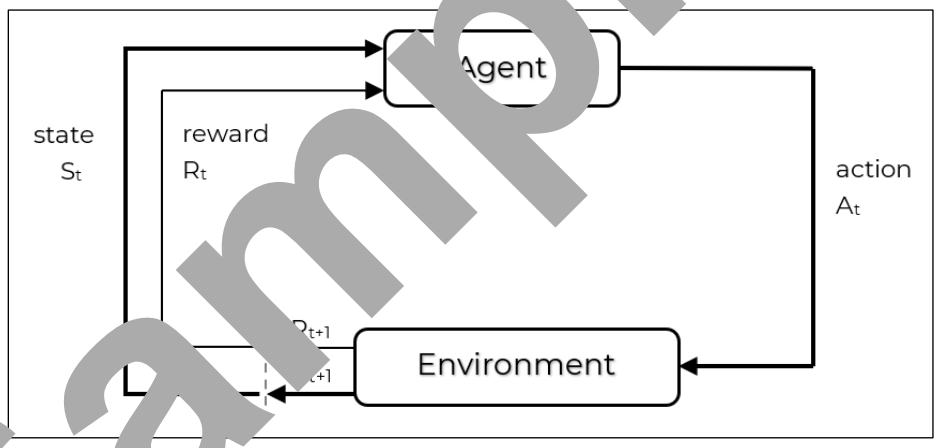

Figure 2-5: Reinforcement learning process

Reinforcement learning has five basic elements:

- **Environment**: Place where the model is trying to learn
- **State**: Situation of the model
- **Rewards**: Feedback from the environment
- **Policy:** Rule of how the environment gives the rewards
- **Value:** Future reward

Practice Exercise

Match the following ML concepts to the description:

a. Supervised
b. Unsupervised
c. Reinforcement
d. Training
e. Validation
f. Testing

1. Type of machine learning where a model is trained using a labeled da
2. Dataset used to train the machine learning model.
3. Type of machine learning that is based on trial and er using fe ck fr he model actions and experiences.
4. Dataset used to evaluate the performance of the mod uring the tra ng process.
5. Type of machine learning that enables models to identify terns a elationships in data without explicit instruction or guidance from humans.
6. Dataset used to evaluate the final performance of the machine learning model.

Practice Questions

1. An unlabeled dataset is a dataset where the output labels are provided.
   a. True
   b. False
2. You want to classify between types of shapes, the labels are:
   a. Hexagon, triangle, and square
   b. Images of the shapes
   c. Features of each shape
   d. Predictions made by the model
3. If you want to create an email spam filter which type of learning [illegible] you choose?
   a. Supervised learning
   b. Unsupervised learning
   c. Reinforcement learning
   d. Deep learning
4. Which of the following is a type of unsupervised learning?
   a. Clustering
   b. Regression
   c. Classification
5. In which of the following fields is reinforcement learning mostly used?
   a. E-commerce
   b. Gaming
   c. Medical
   d. 3D construction
6. Which of the following is not a part of a deep learning network?
   a. Input layer
   b. Hidden layer
   c. Output layer
   d. [illegible] brain
7. Neural networks are computer systems that are designed to mimic the way the human brain works.
   a. True
   b. False