Governor's Champions' Academy 2025

AI and Machine Learning Bootcamp

July 21–25, 2025

Southern Methodist University (SMU)

Research Lead:

Dr. Inam Ullah Khan
Postdoctoral Fellow
Southern Methodist University

A hands-on residential camp for high school research champions focused on advancing skills in Artificial Intelligence and Machine Learning.

Empowering Tomorrow's AI Innovators

Introduction to Machine Learning

What is Machine Learning?

Machine Learning (ML) is a powerful technology that enables computers to learn from data and make decisions without being explicitly programmed. It powers many tools we use every day—such as voice assistants (like Siri or Alexa), recommendation engines (like those on Netflix or Spotify), and even autonomous vehicles.

In simple terms, **machine learning is the process of teaching a computer to recognize patterns and make predictions based on data**. Instead of manually writing rules to solve a problem, we give the computer a large dataset and let it learn the underlying relationships on its own.

Why Use Machine Learning?

Machine learning is useful when:

- The task is too complex to be described with rules (e.g., recognizing handwriting).
- The system needs to improve over time (e.g., spam filters).
- Predictions are based on large datasets (e.g., weather forecasting).

Example: Predicting Rainfall

Let's say we want to build an app that predicts whether it will rain tomorrow.

- **Traditional approach:** Write complex physics-based equations to simulate the atmosphere—this requires deep knowledge of meteorology and massive computational resources.
- **Machine Learning approach:** Feed historical weather data (like temperature, humidity, pressure, etc.) into an ML model. The model learns patterns in the data that typically lead to rainfall and uses those patterns to predict future rain.

This way, the ML model becomes a shortcut for recognizing patterns that are difficult to write out manually.

How Does ML Work?

Most ML systems follow three key steps:

1. **Collect Data:** Gather data relevant to the task. For rain prediction, this could include temperature, humidity, wind speed, and historical rainfall.
2. **Train a Model:** Use the data to teach the model. During training, the model adjusts itself to minimize prediction error.
3. **Make Predictions:** Once trained, the model can be given new data and will output a prediction.

Check Your Understanding

What is a *model* in machine learning?

a) A model is a piece of computer hardware.

b) A model is a mathematical relationship derived from data that an ML system uses to make predictions.

c) A model is a smaller representation of the thing you're studying.

Types of Machine Learning Systems

Machine learning (ML) systems can be categorized based on how they learn from data. The four main types are:

1. **Supervised Learning**
2. **Unsupervised Learning**
3. **Reinforcement Learning**
4. **Generative AI**

1. Supervised Learning

Supervised learning is like studying for a test using past exams that already contain the correct answers. The model learns from labeled data—input data paired with correct outputs—so it can make predictions on new, unseen data.

Common Tasks in Supervised Learning:

- **Regression:** Predicts a continuous value.
- **Classification:** Predicts a category.

Example: Regression

Regression models predict numeric values. Suppose we want to predict house prices based on features like size, location, and number of bedrooms.

Scenario	**Input Features**	**Prediction (Output)**
Home price estimate	Square footage, location, number of rooms, mortgage rate	Price of the home (in $)
Ride time prediction	Traffic data, distance, weather conditions	Time to destination (in minutes)

Example: Classification

Classification models determine which category an item belongs to.

- **Binary classification:** Only two possible outcomes (e.g., spam or not spam).
- **Multiclass classification:** More than two possible outcomes (e.g., rain, snow, or hail).

Example:

- Email filtering: Predict if a message is "spam" or "not spam".
- Image recognition: Identify whether a picture contains a "cat," "dog," or "rabbit".

Check Your Understanding

Question: If you wanted to use a machine learning model to **predict energy usage for commercial buildings**, what type of model would you use?

a) Classification

b) Regression

Unsupervised Learning

Unsupervised learning models work with data that does not include correct or labeled answers. The goal of these models is to identify meaningful patterns or structures within the data. In other words, the model is not told how to categorize the data; instead, it must discover the rules and relationships on its own.

One commonly used technique in unsupervised learning is called **clustering**. In clustering, the model groups similar data points together—revealing natural patterns that may not be immediately obvious.

Visualizing Clusters

Clustering in Action: A Visual Walkthrough

To understand how unsupervised learning (clustering) works, consider the four-panel figure below that shows the progression from raw data to meaningful clusters.

Explanation of Each Panel:

- **Top-left (Figure 1):** The model begins with unlabeled data—dots distributed in a 2D space with no predefined groupings.
- **Top-right (Figure 2):** A clustering algorithm groups the data into three regions based on similarity: Cluster 1 (blue), Cluster 2 (pink), Cluster 3 (yellow).
- **Bottom-left (Figure 3):** These clusters may represent real-world categories such as *RAIN, SNOW, SLEET*, and *NO RAIN*.

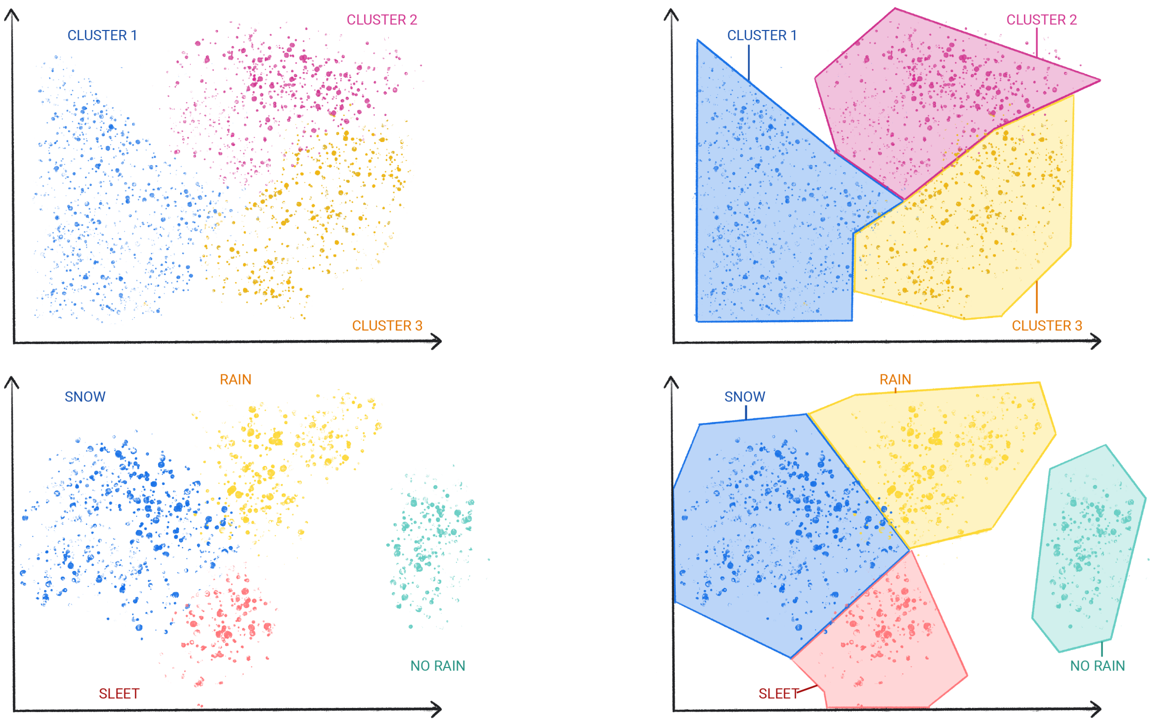


Figure 1: Top-left: Unlabeled data. Top-right: Three detected clusters. Bottom-left: Weather categories emerge. Bottom-right: Final labeled data points.

- **Bottom-right (Figure 4):** The result is a clear separation of data into meaningful clusters—a classic outcome of unsupervised learning.

This step-by-step clustering process illustrates how unsupervised models can discover hidden structure in data, even when no labels or guidance are provided.

Key Takeaway

Unsupervised learning helps us explore hidden structures in data and is especially useful when we lack labeled examples. It is commonly used for:

- Customer segmentation in marketing.
- Grouping news articles by topic.
- Identifying genetic clusters in biology.

Check Your Understanding

Question: What distinguishes a *supervised* approach from an *unsupervised* approach?

a) A supervised approach typically uses clustering.

b) An unsupervised approach knows how to label clusters of data.

c) A supervised approach is given data that contains the correct answer.

Reinforcement Learning

Reinforcement learning (RL) models learn by interacting with an environment. They make decisions and receive feedback in the form of **rewards** (for good actions) or **penalties** (for bad actions). Over time, the model improves by learning which actions lead to better outcomes.

The result of this learning process is a **policy**—a strategy that tells the model which action to take in different situations to maximize rewards.

Examples of Reinforcement Learning:

- Training a robot to walk by rewarding it when it moves correctly.
- Teaching a computer program, like AlphaGo, to play the game of Go by giving points for winning moves.

Generative AI

Generative AI refers to models that can **create new content** based on user input. These models are capable of producing things like:

- Images
- Music
- Jokes or stories
- Summaries of articles
- Edited or enhanced media (photos, videos, etc.)

Generative models can handle various combinations of inputs and outputs. These are often described as *"input-type to output-type"* tasks.

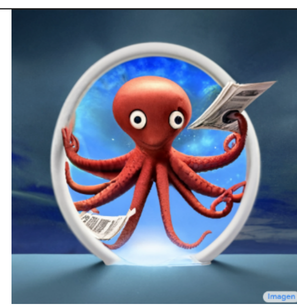
Examples of Input-Output Types

- Text-to-text (e.g., summarizing or answering questions)
- Text-to-image (e.g., generating a drawing from a description)
- Text-to-video
- Text-to-code
- Text-to-speech
- Image-and-text to image

Model Examples

Model	Input	Example Output
Text-to-text	What is the Governor's Champions Academy?	The Governor's Champions Academy is a summer residential program in Texas that brings together top high school science fair winners. Hosted at universities like SMU, it allows students to explore real-world research in areas such as artificial intelligence, drones, and biological signal processing under the guidance of university faculty.

Text-to-code	Write a Python loop to print prime numbers.	`for number in numbers:`
		`# Check if the number is prime.`
		`is_prime = True`
		`for i in range(2, number):`
		`if number % i == 0:`
		`is_prime = False`
		`break`
		`if is_prime:`
		`print(number)`

Text-to-image	An alien octopus floats through a portal while reading a newspaper.	
Image-to-text	ITT.png	This is a flamingo. They are found in the Caribbean.

How Does Generative AI Work?

Generative models learn patterns in existing data so they can generate new content that looks or sounds similar. Think of it like:

- A comedian imitating someone's voice and gestures.
- An artist painting in the style of Van Gogh after studying many of his artworks.
- A band mimicking the music of a famous group after listening to their songs.

Generative AI usually starts by learning with **unsupervised learning**, then may be refined with supervised or reinforcement learning to perform specific tasks—like summarizing a text or editing an image.

Final Note

Generative AI is advancing rapidly. New uses are discovered every day—from improving ecommerce product images to creating realistic animations or enhancing low-quality media. It's an exciting area to explore for creative and technical innovation.

Supervised Learning

Supervised learning is one of the most widely used types of machine learning. It involves learning a function that maps input data (features) to the correct output (label) using a labeled dataset. Common real-world applications include:

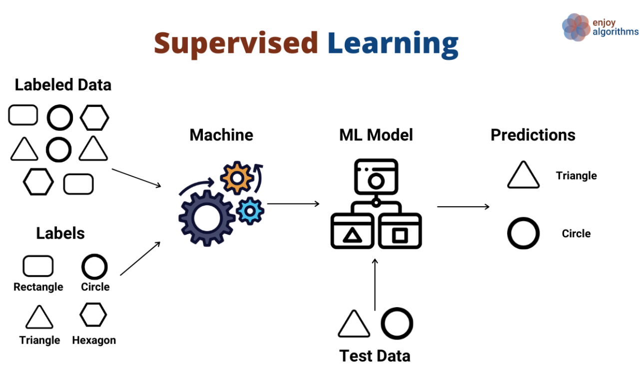


Figure 2: Examples of labeled data: features + label.

- Classifying emails as spam or not spam.
- Predicting rainfall from weather conditions.
- Estimating house prices based on location and size.

Core Concepts of Supervised Learning

Supervised learning revolves around the following five core components:

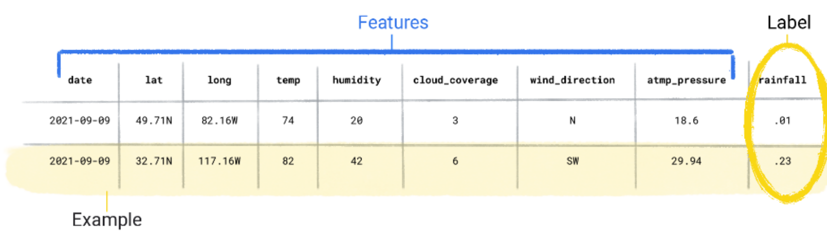
- **Data**
- **Model**

- **Training**
- **Evaluating**
- **Inference**

Data

Data is the foundation of machine learning. It can come in many forms—text, numbers, images, or audio. In supervised learning, data is organized into a **dataset** composed of many **labeled examples**.

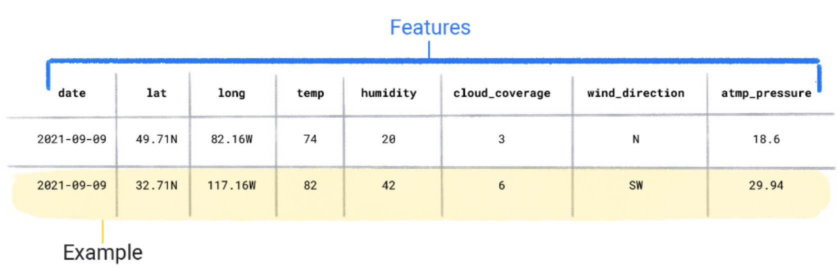
- **Features** are the inputs (e.g., temperature, humidity).
- **Label** is the expected output (e.g., rainfall amount).



date	lat	long	temp	humidity	cloud_coverage	wind_direction	atmp_pressure	rainfall
2021-09-09	49.71N	82.16W	74	20	3	N	18.6	.01
2021-09-09	32.71N	117.16W	82	42	6	SW	29.94	.23

Figure 3: Examples of labeled data: features + label.

Unlabeled examples have features but no known label—these are typically used during inference after training.



date	lat	long	temp	humidity	cloud_coverage	wind_direction	atmp_pressure
2021-09-09	49.71N	82.16W	74	20	3	N	18.6
2021-09-09	32.71N	117.16W	82	42	6	SW	29.94

Figure 4: Unlabeled examples: only features, no known label.

Dataset Characteristics

An effective dataset should be:

- **Large in size** — many examples to learn from.
- **High in diversity** — representing a wide range of conditions.

Question: Which dataset is ideal for machine learning?

- Small / Low diversity
- Small / High diversity
- Large / Low diversity

  Large / High diversity

Model

A model is a mathematical function that learns to predict labels from features. It's trained by adjusting internal values (parameters) to minimize the difference between predicted and actual labels.

Training the Model

Training is the process of feeding the model labeled examples and letting it learn the best mapping from features to labels by minimizing **loss**.

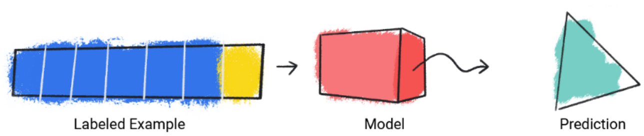


Figure 5: An ML model making a prediction from a labeled example.

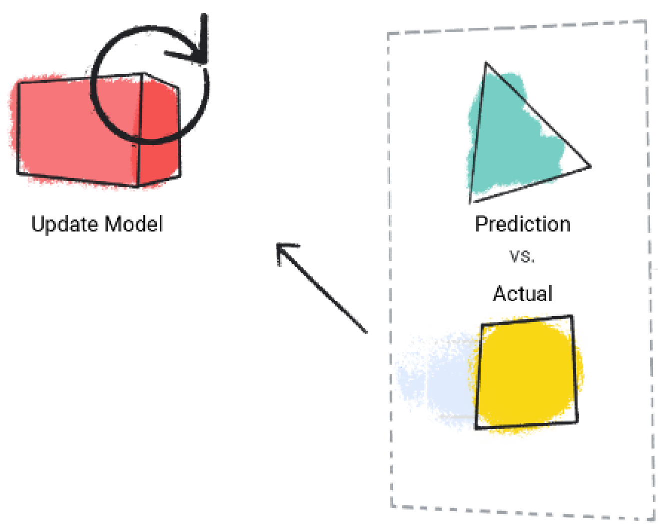


Figure 6: Comparing prediction with actual value and updating the model.

The model adjusts its internal parameters over many examples until it consistently makes good predictions.

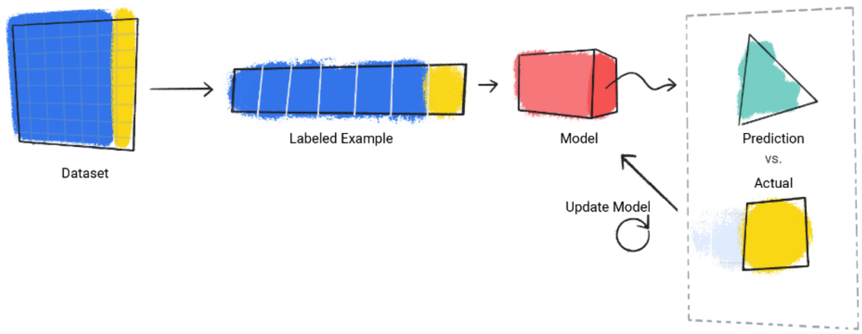


Figure 7: Repeating the process for all labeled examples in the dataset.

Evaluating the Model

Once trained, we evaluate the model using a separate labeled dataset (called the test set). The model sees only the features and predicts the label, which is then compared to the actual value.

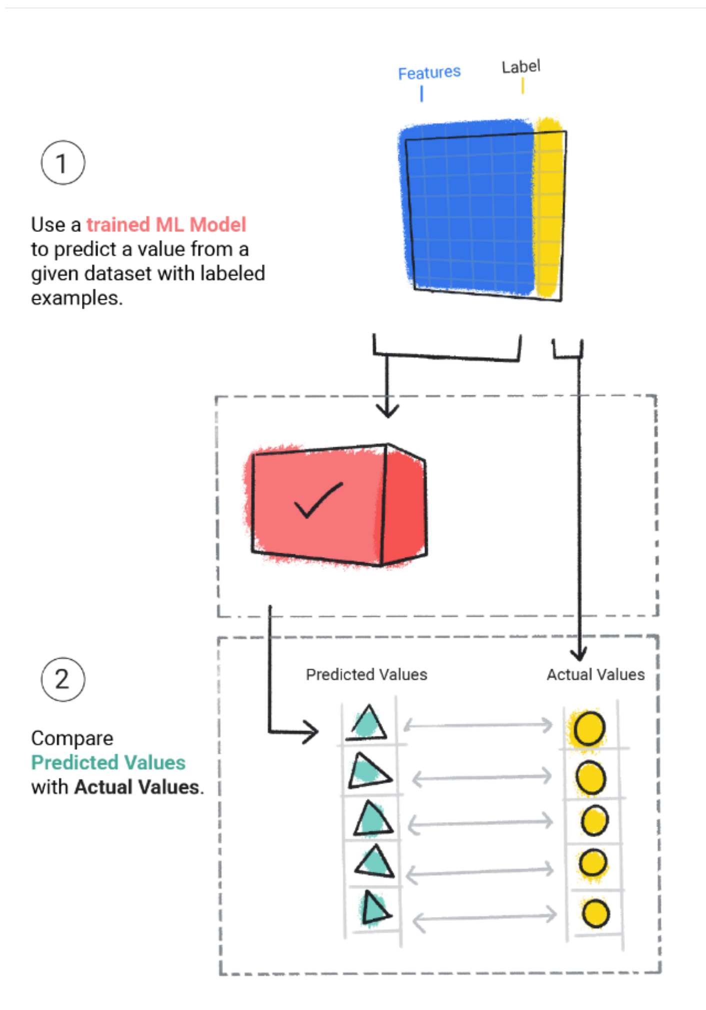


Figure 8: Evaluating the model by comparing predicted vs. actual labels.

Goal: A model that generalizes well—not just memorizing the training data but making good predictions on unseen data.

Inference

Once we're satisfied with the results from evaluating the model, we can use the model to make predictions, called inferences, on unlabeled examples. In the weather app example, we would give the model the current weather conditions—like temperature, atmospheric pressure, and relative humidity—and it would predict the amount of rainfall.

Key Terms

In this lesson, you became familiar with the following terms:

- **Classification model**: A model that predicts categories or classes.
- **Clustering**: Grouping data into natural segments.
- **Model**: The mathematical function that makes predictions.
- **Policy**: aA strategy used in reinforcement learning.
- **Prediction**: The output generated by a model.
- **Regression model**: A model that predicts continuous values.
- **Reinforcement learning**: Learning through rewards and penalties.
- **Reward**: Feedback used to guide learning in reinforcement models.
- **Supervised learning**: Learning with labeled examples.
- **Training**: The process of teaching a model using data.
- **Unsupervised learning**: Learning from data without labels.

Test Your Understanding

This section will help you reinforce core machine learning concepts through scenario-based questions.

1. Predictive Power

In supervised learning, not every feature in a dataset helps in making good predictions. The features with the strongest relationship to the label have the most **predictive power**.
Question: Which three features are likely the best predictors of a car's price?

a) Color, height, make_model

b) Make_model, year, miles

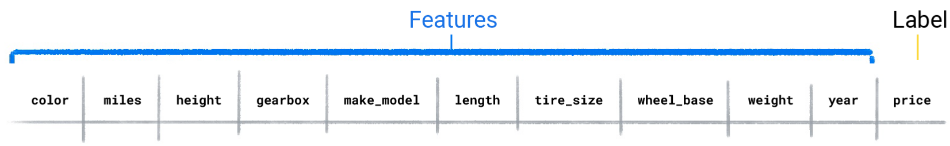


Figure 9: Sample automobile dataset. Predict car price based on features.

c) Miles, gearbox, make_model

d) Tire_size, wheel_base, year

2. Supervised vs. Unsupervised Learning

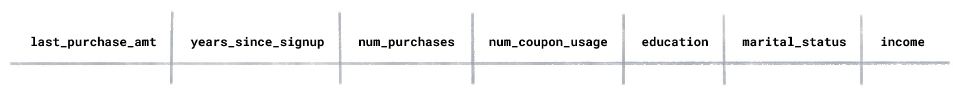


Figure 10: Sample dataset of users on an online shopping website.

Question: If you wanted to discover patterns or user types without knowing any target values in advance, which learning approach would you use?

a) Supervised learning

b) Unsupervised learning

3. Predicting Energy Usage

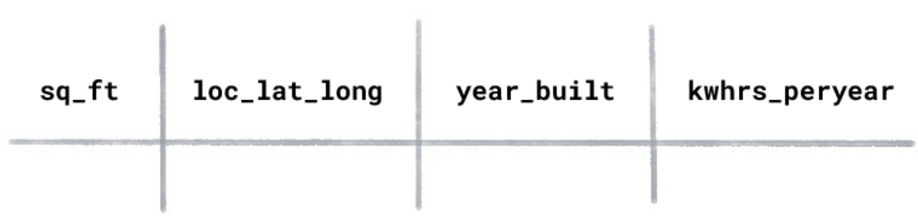


Figure 11: Sample home energy usage dataset.

Question: What type of ML would you use to predict kilowatt hours per year?

a) Supervised learning

b) Unsupervised learning

4. Regression vs. Classification

departure_airport	destination_airport	date	time_of_day	airline	fuel_price	airplane_ticket_cost

Figure 12: Flight dataset showing flight features and ticket cost.

Question A: If you want to predict the exact cost of a ticket, which technique should you use?

a) Regression

b) Classification

Question B: Could you still train a classification model to group ticket prices into “high,” “average,” and “low”?

a) No, airplane_ticket_cost is numeric, not categorical.

b) No, classification only works with two categories.

c) Yes, but you’d need to convert the cost into labeled categories.

5. Training and Evaluation

Question: If your model’s predictions are inaccurate, what are two good ways to improve it?

a) Retrain using only features with high predictive power.

b) You can’t fix it once trained.

c) Try switching to an unsupervised approach.

d) Retrain using a larger and more diverse dataset.