

HCIA-AI Huawei

Eng.Eman Elgallad

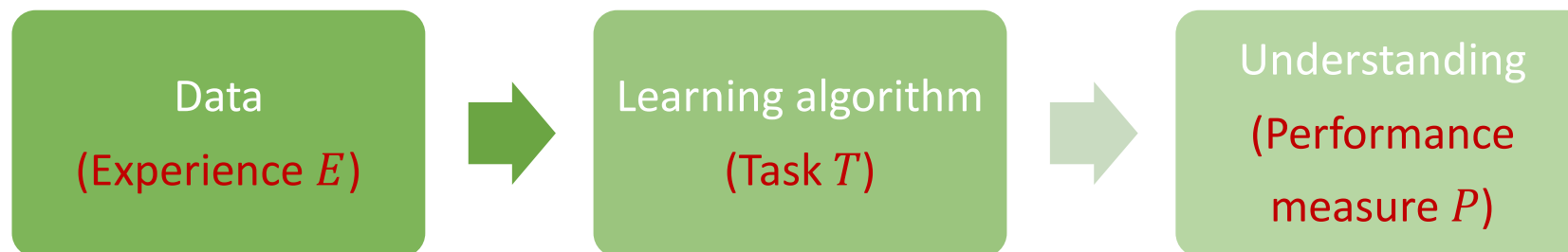
Lecture 03 – Machine Learning Overview

Contents

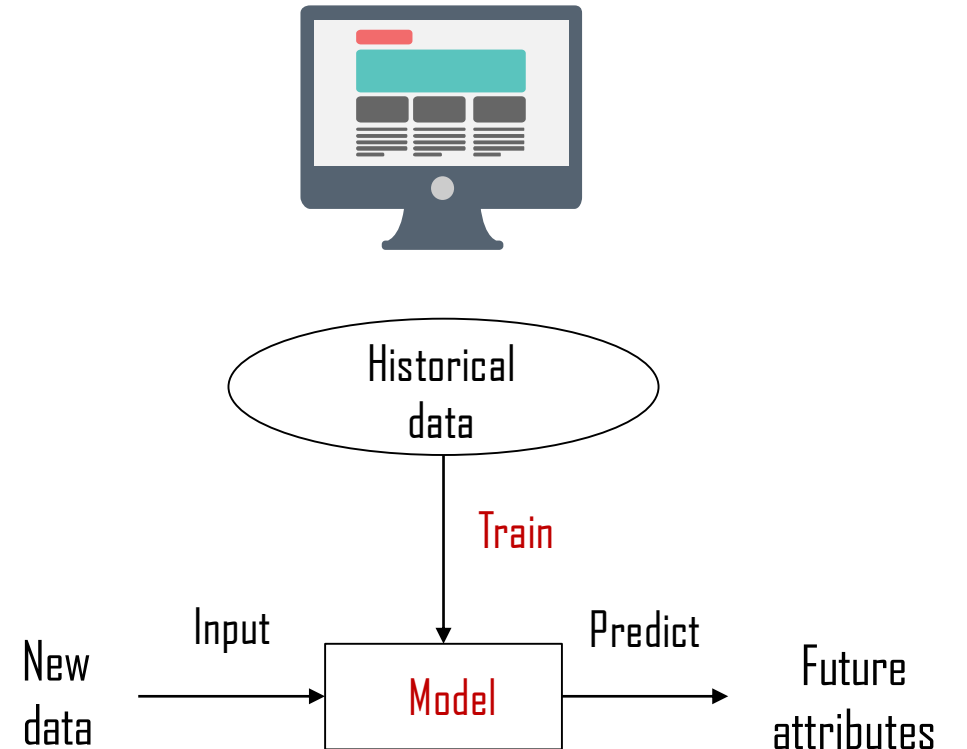
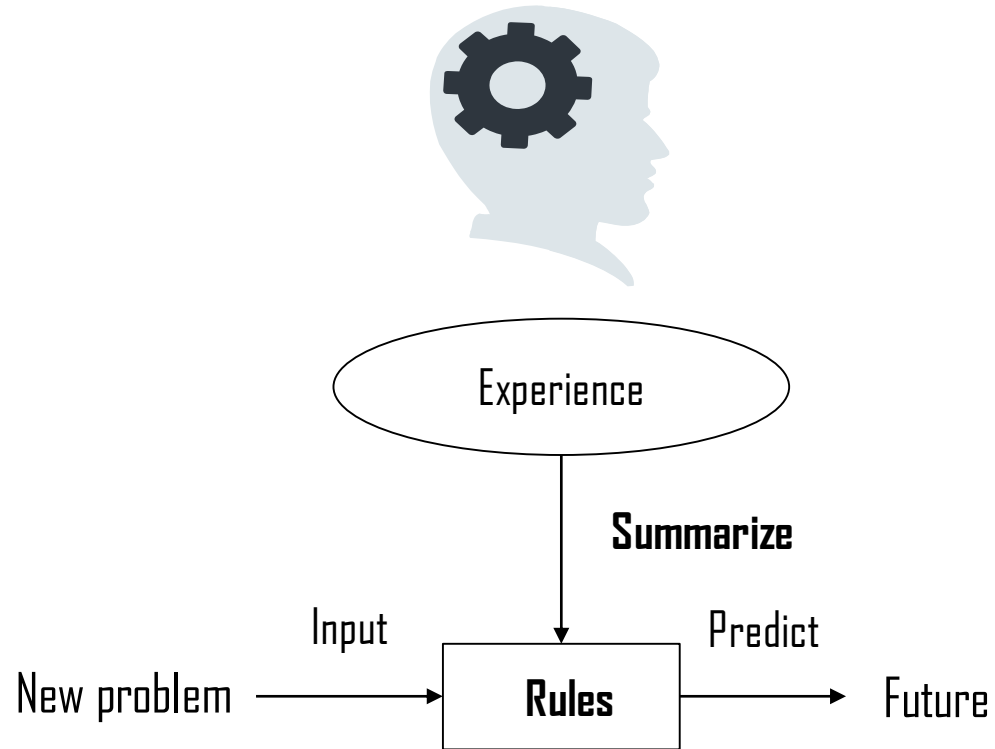
1. **Machine Learning Algorithms**
2. Types of Machine Learning
3. Machine Learning Process
4. Important Machine Learning Concepts
5. Common Machine Learning Algorithms

Machine Learning Algorithms

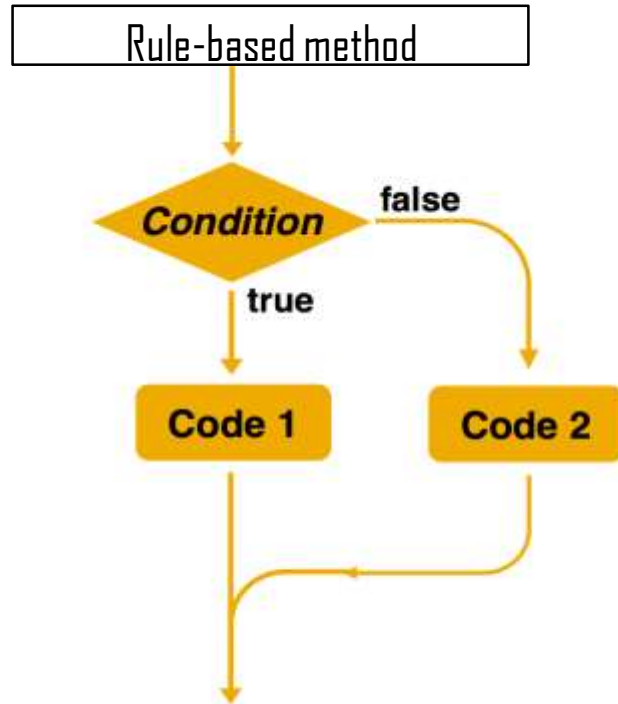
- Machine learning is often combined with deep learning methods to study and observe AI algorithms. A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .



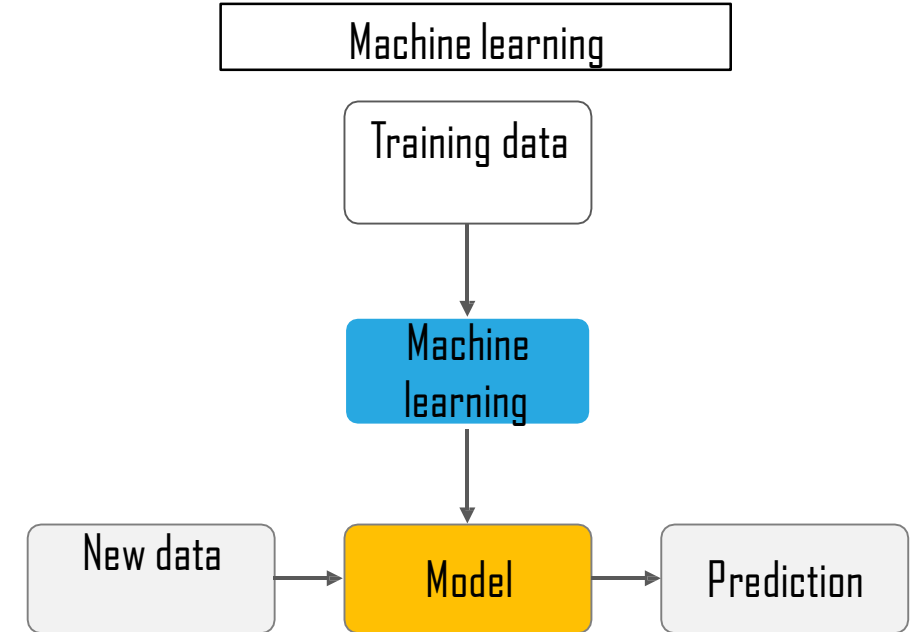
Machine Learning Algorithms Cont.



Differences Between Machine Learning Algorithms and Traditional Rule-based Methods



- Explicit programming is used to solve problems.
- Rules can be manually determined.

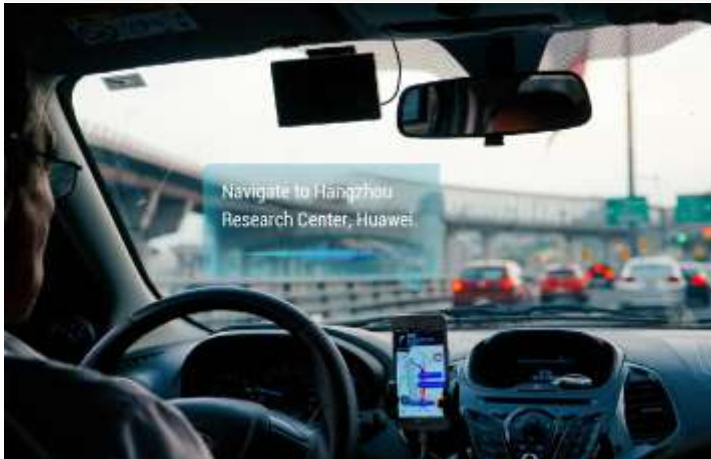


- Models are trained on samples.
- Decision-making rules are complex or difficult to describe.
- Machines automatically learn rules.

When to Use Machine Learning

- Machine learning provides solutions to complex problems, or those involving a large amount of data whose distribution function cannot be determined.
- Consider the following scenarios:

Rules are complex or difficult to describe, for example, speech recognition.



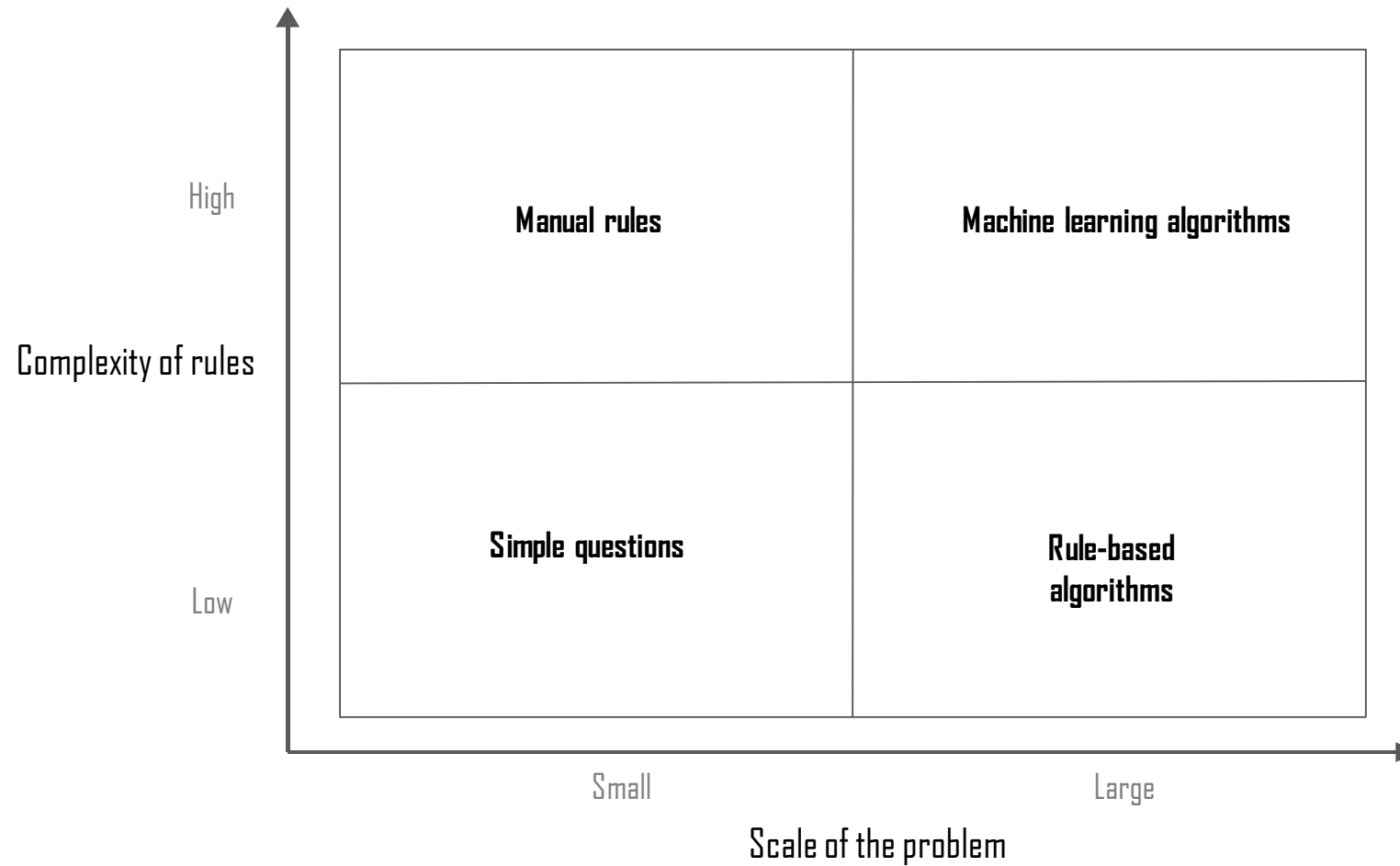
Task rules change over time, for example, part-of-speech tagging, in which new words or word meanings can be generated at any time.



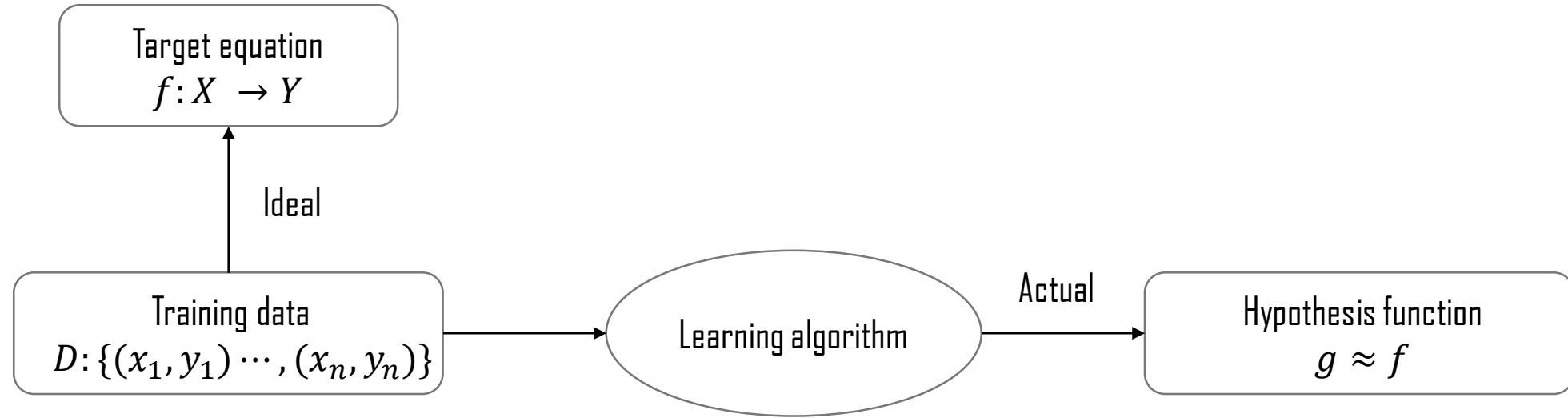
Data distribution changes over time and programs need to adapt to new data constantly, for example, sales trend forecast.



When to Use Machine Learning Cont.



Rationale of Machine Learning Algorithms



- The objective function f is unknown, and the learning algorithm cannot obtain a perfect function f .
- Hypothesis function g approximates function f , but may be different from function f .

Main Problems Solved by Machine Learning

- Machine learning can solve many types of tasks. Three most common types are:
 - **Classification:** To specify a specific one of the k categories for the input, the learning algorithm usually outputs a function $f: R^n \rightarrow (1, 2, \dots, k)$. For example, image classification algorithms in computer vision solve classification tasks.
 - **Regression:** The program predicts the output for a given input. The learning algorithms usually output a function $f: R^n \rightarrow R$. Such tasks include predicting the claim amount of a policy holder to set an insurance premium or predicting the security price.
 - **Clustering:** Based on internal similarities, the program groups a large amount of unlabeled data into multiple classes. Same-class data is more similar than data across classes. Clustering tasks include search by image and user profiling.
- **Classification and regression are two major types of prediction tasks. The output of classification is discrete class values, and the output of regression is continuous values.**

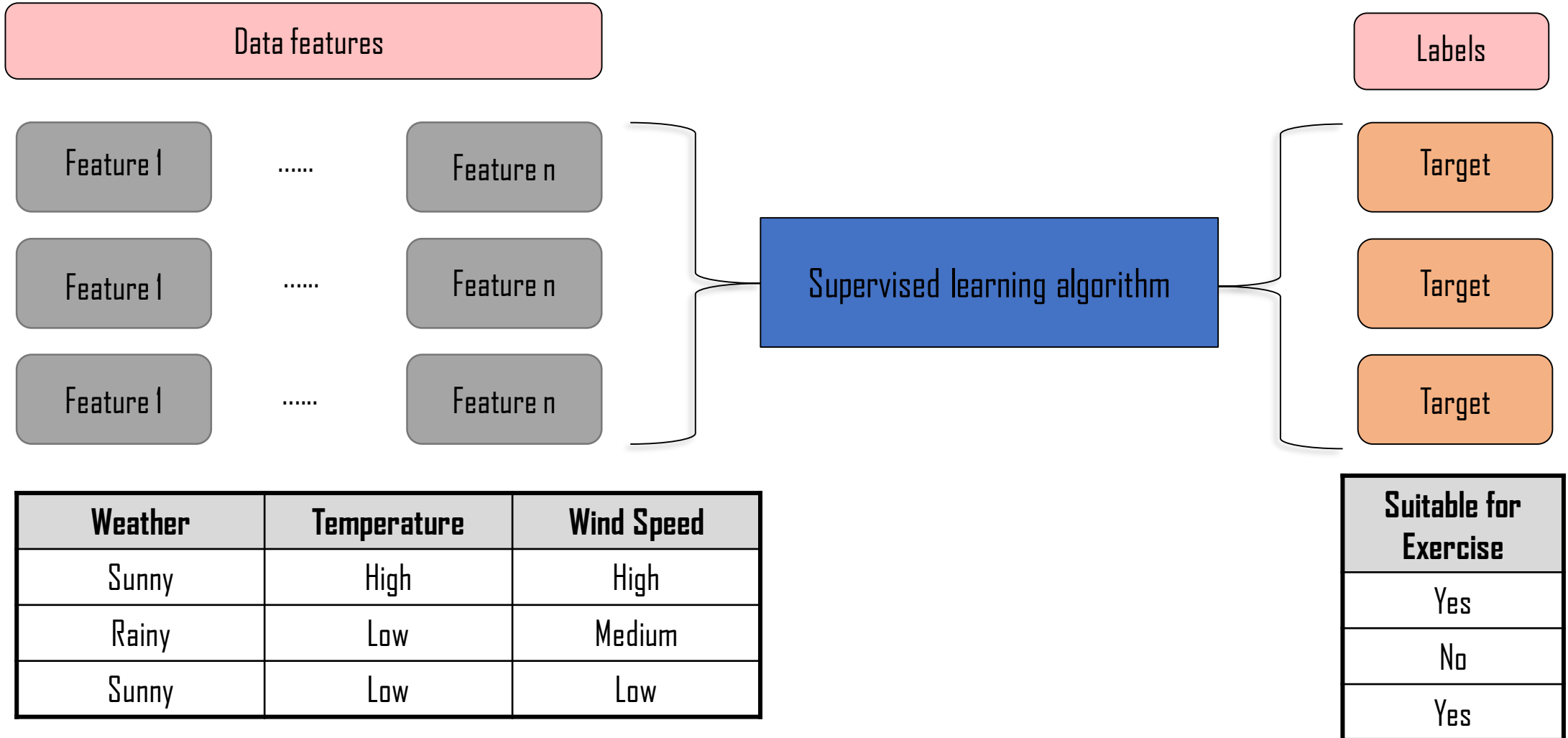
Contents

1. Machine Learning Algorithms
- 2. Types of Machine Learning**
3. Machine Learning Process
4. Important Machine Learning Concepts
5. Common Machine Learning Algorithms

Types of Machine Learning

- **Supervised learning:** The program takes a known set of samples and trains an optimal model to generate predictions. Then, the trained model maps all inputs to outputs and performs simple judgment on the outputs. In this way, unknown data is classified.
- **Unsupervised learning:** The program builds a model based on unlabeled input data. For example, a clustering model groups objects based on similarities. Unsupervised learning algorithms model the highly similar samples, calculate the similarity between new and existing samples, and classify new samples by similarity.
- **Semi-supervised learning:** The program trains a model through a combination of a small amount of labeled data and a large amount of unlabeled data.
- **Reinforcement learning:** The learning systems learn behavior from the environment to maximize the value of reward (reinforcement) signal function. Reinforcement learning differs from supervised learning of connectionism in that, instead of telling the system the correct action, the environment provides scalar reinforcement signals to evaluate its actions.
- Machine learning evolution is producing new machine learning types, for example, self-supervised learning, contrastive learning, generative learning.

Supervised Learning



Supervised Learning - Regression

- **Regression** reflects the features of sample attributes in a dataset. A function is used to express the sample mapping relationship and further discover the dependency between attributes. Examples include:
 - How much money can I make from stocks next week?
 - What will the temperature be on Tuesday?

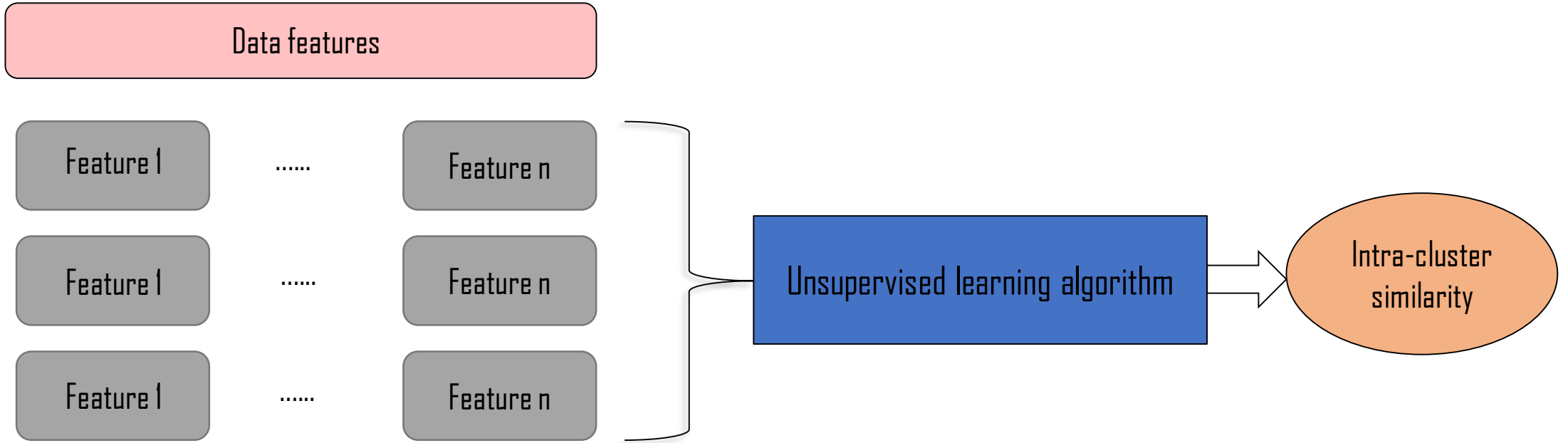
Monday	Tuesday
38°	?

Supervised Learning - Classification

- **Classification** uses a classification model to map samples in a dataset to a given category.
 - What category of garbage does the plastic bottle belong to?
 - Is the email a spam?



Unsupervised Learning



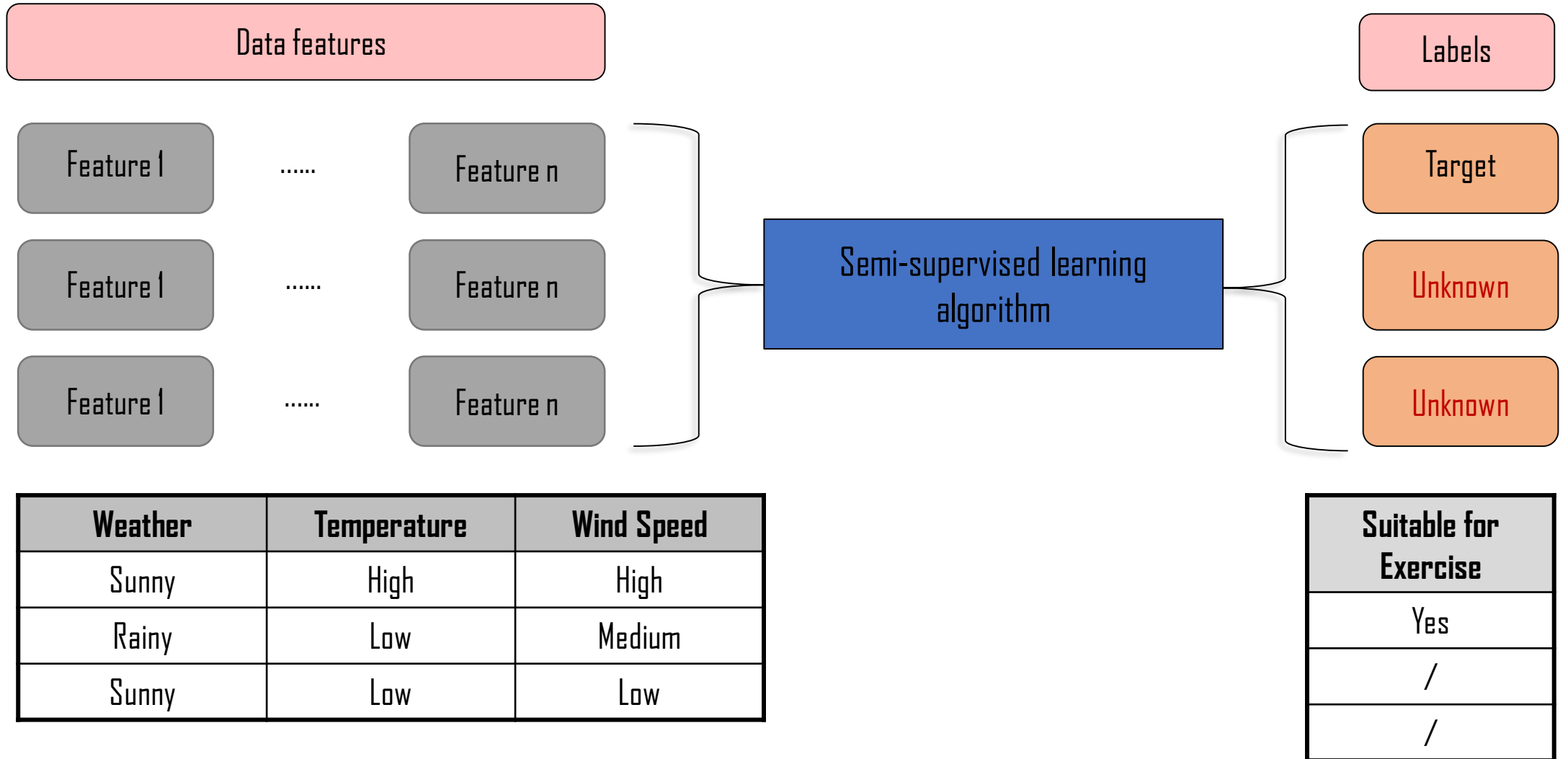
Monthly Sales Volume	Product	Sale Duration	Category
1000-2000	Badminton racket	6:00-12:00	Cluster 1
500-1000	Basketball	18:00-24:00	Cluster 2
1000-2000	Game console	00:00-6:00	

Unsupervised Learning - Clustering

- **Clustering** uses a clustering model to classify samples in a dataset into several categories based on similarity.
 - Defining fish of the same species.
 - Recommending movies for users.

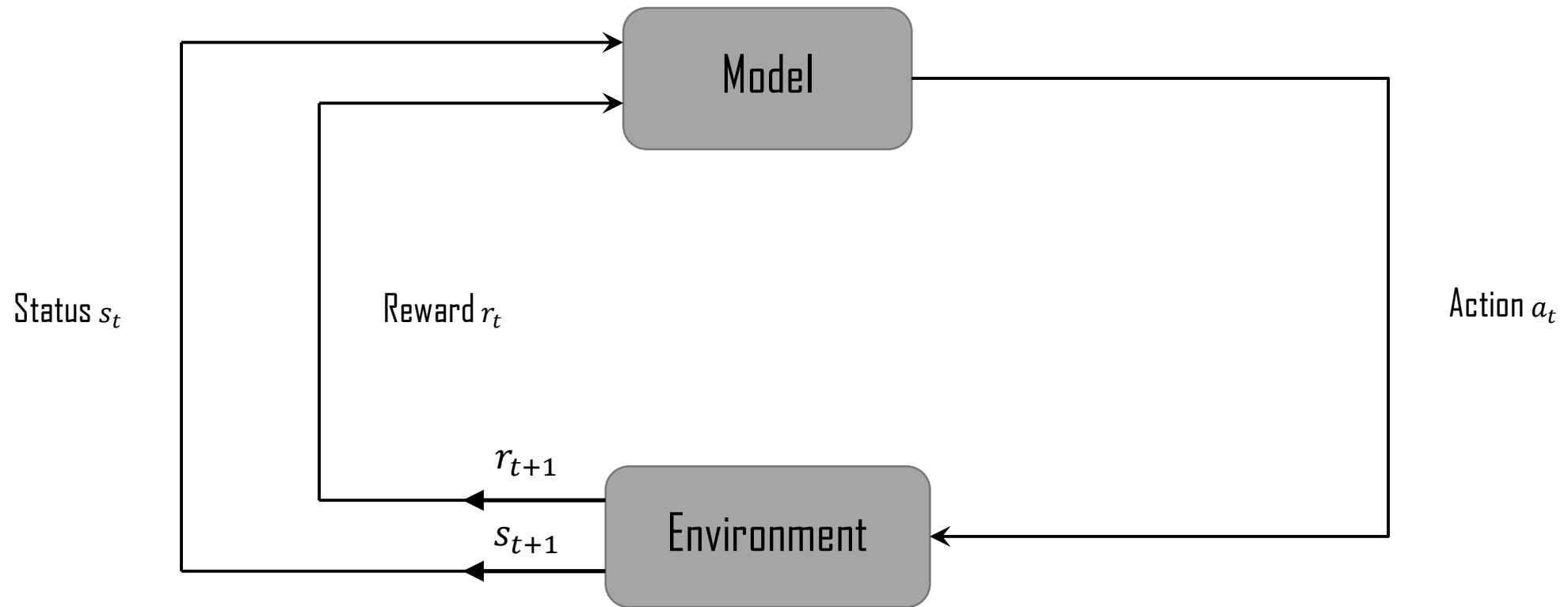


Semi-supervised Learning



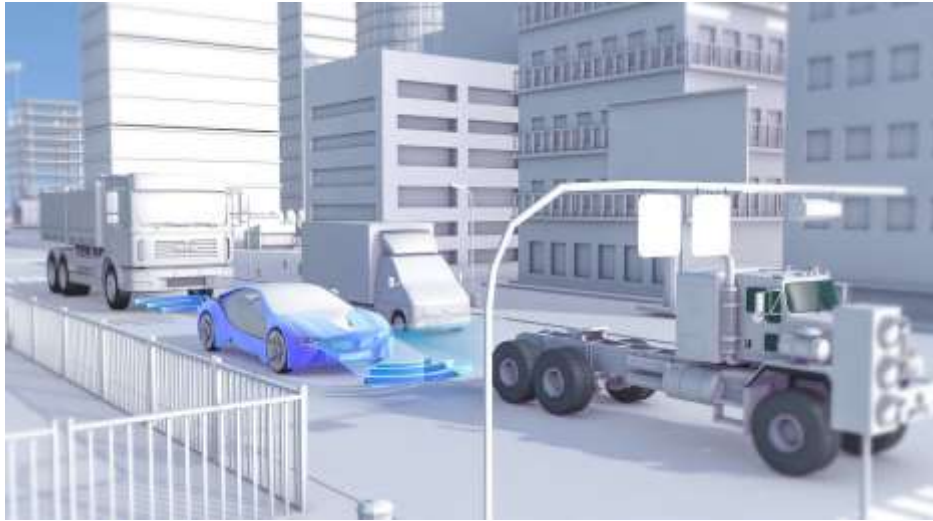
Reinforcement Learning

- A reinforcement learning model learns from the environment, takes actions, and adjusts the actions based on a system of rewards.



Reinforcement Learning - Best Action

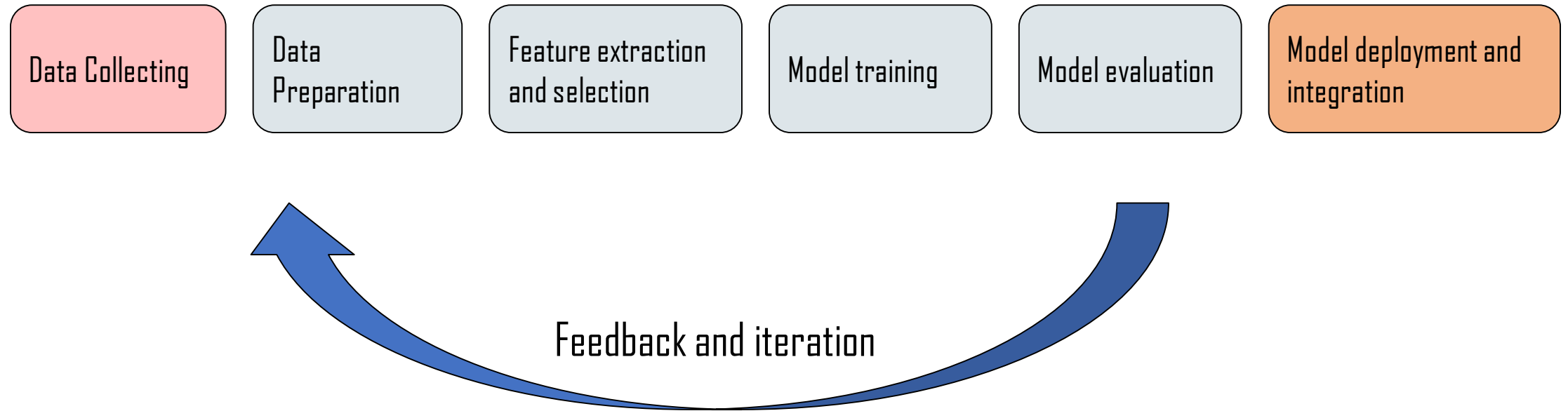
- **Reinforcement learning** always tries to find the best action.
 - Autonomous vehicles: The traffic lights are flashing yellow. Should the vehicle brake or accelerate?
 - Robot vacuum: The battery level is 10%, and a small area is not cleaned. Should the robot continue cleaning or recharge?



Contents

1. Machine Learning Algorithms
2. Types of Machine Learning
- 3. Machine Learning Process**
4. Important Machine Learning Concepts
5. Common Machine Learning Algorithms

Machine Learning Process



Data Collecting

- Data in the real world can be of numerous types & formats
 - Structured or unstructured
 - Readable or obfuscated
 - Small or big
- **Goal:**
 - Have a good plan for data gathering to save time
 - Beneficial during analysis & modeling
- **A good data collection plan should include:**
 - Data Formats
 - Data Sources
 - Data Types

Data Format:

- Many data formats are acceptable for storing information

1. **CSV or TXT:** Most widely used data exchange formats for storing tabular data containing many rows & columns

2. **Microsoft Excel format (.xls or .xlsx):** Most popular data file format in the business world.

3. **Markup languages:** Provide a system for annotating a document with a specific syntactic structure. Includes TeX, HTML, XML & XHTML forms

4. **JSON:** Widely used data interchange format in many application programming interfaces, e.g., Facebook Graph API, Google Maps, and Twitter API.

Data Sources:

- **Structured:**

- Every row & column is well defined within a connected schematic tables in a database.
- E.g., CSV & Excel files

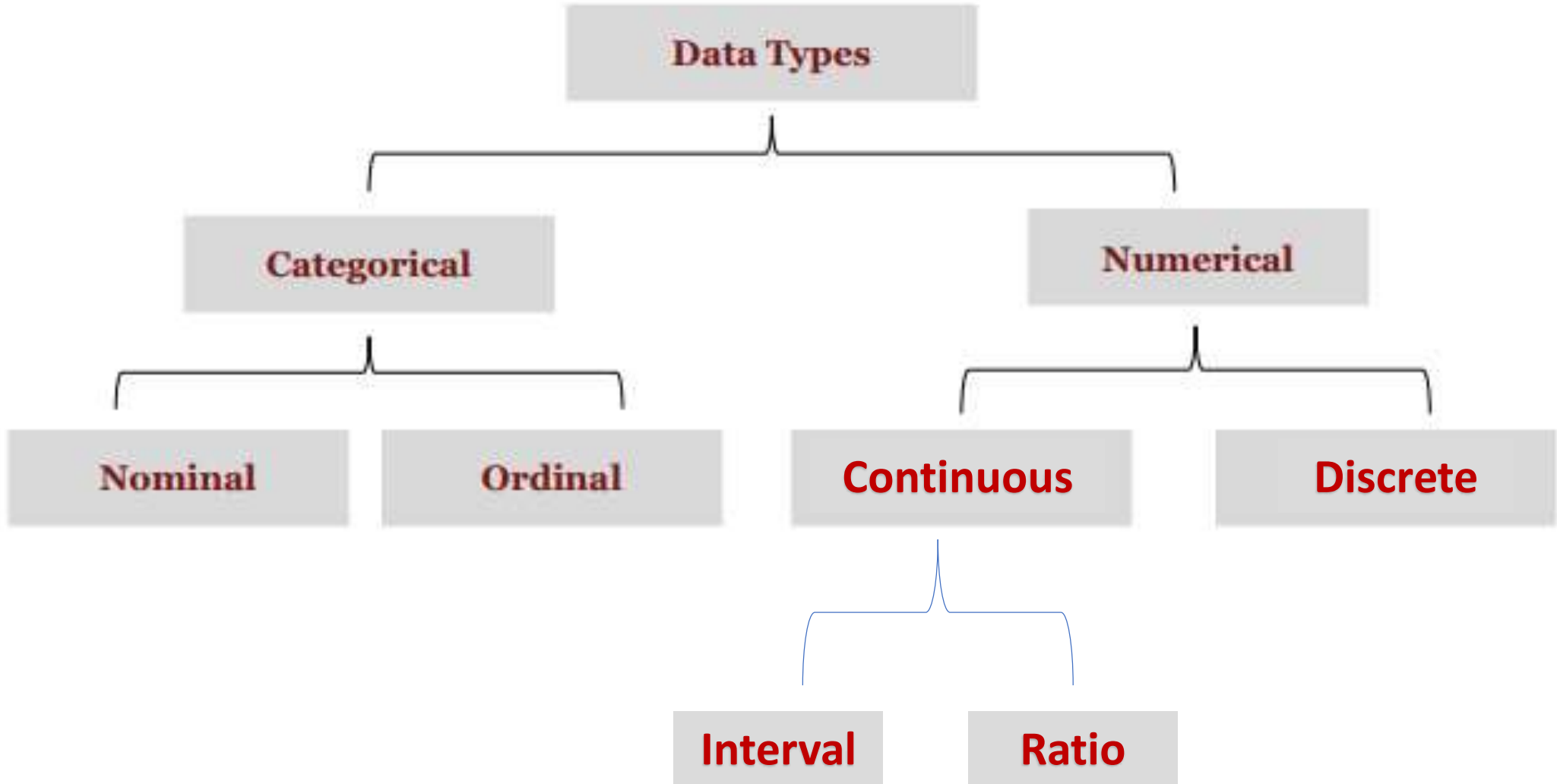
- **Semi-structured:**

- A significant move away from the traditional Relational Database Management System (RDBMS) paradigms, where data is stored with predefined schema
- E.g., JSON & XML formats

- **Unstructured:**

- Images, videos, web logs & click stream
- Textual data generated from human interactions, e.g., [Twitter Sentiment Analysis on tweets](#)

Data Types :



Data Types : Qualitative (Categorical data)

▪ Nominal:

- Uses names or symbols to distinguish objects, e.g., Hair color {auburn, black, blond, brown, grey, red, white}.
- Has no order
- Binary: Nominal attribute with only 2 states {yes or no}
 - Symmetric binary: Outcomes are equally valuable & carry same weight, e.g., gender
 - Asymmetric binary: when outcomes are not equally important, e.g., medical test (positive vs. negative)

▪ Ordinal:

- Uses values to provide meaningful order (ranking) to objects but magnitude between successive values is not known
- E.g., Size {small, medium, large}, grades, army rankings

Data Types : Quantitative (Numerical)

Continuous vs. Discrete Data

Discrete

- Countable finite set of values, e.g., houses, workers
- Represented as integers

Continuous

- Are measured, e.g., height, weight
- Typically represented as floating-point
 - **Interval:**
 - Uses equal-sized units of measurement
 - Values have order
 - Has no true zero point or fixed beginning, i.e., total absence of the variable of interest, e.g., calendar dates
 - Operations: +, -
 - **Ratio:**
 - Numeric attributes with a true zero point
 - Values are ordered, e.g., age, distance covered by a car
 - Operations: +, -, /, x

Machine Learning Basic Concept - Dataset

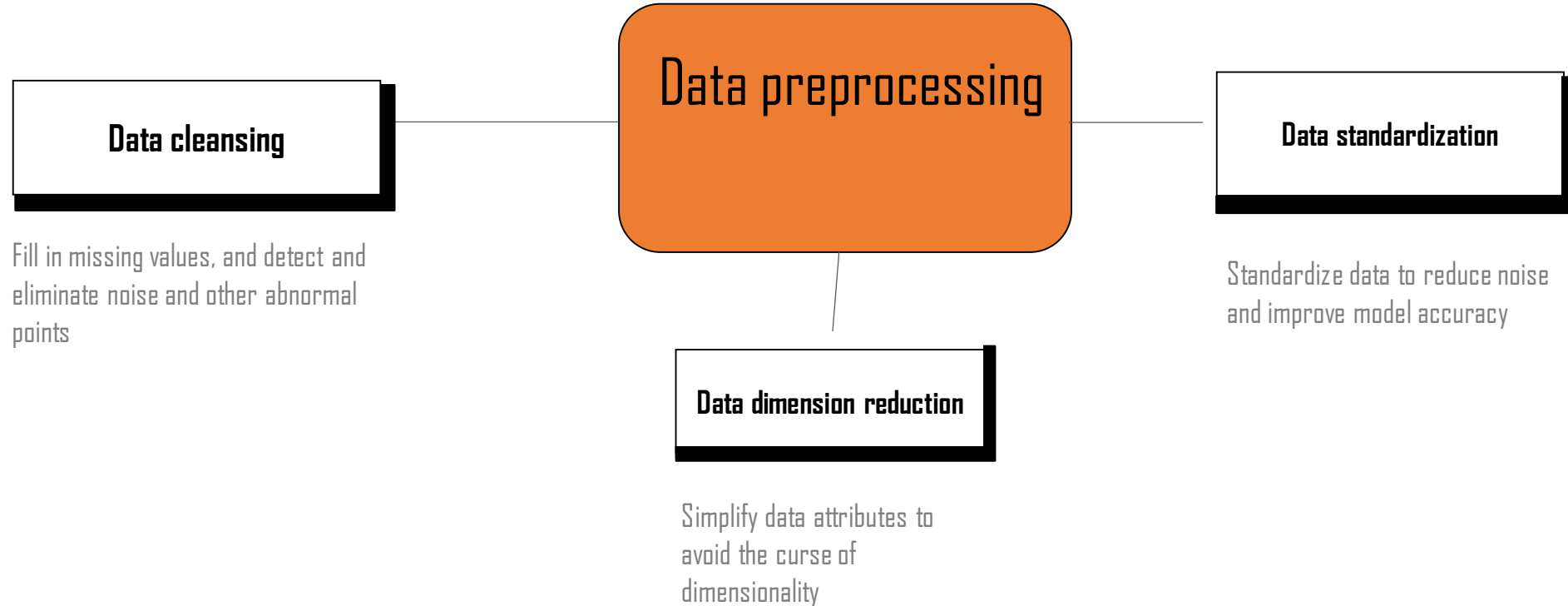
- **Dataset:** collection of data used in machine learning tasks, where each piece of data is called a sample. Items or attributes that reflect the presentation or nature of a sample in a particular aspect are called **features**.
 - **Training set:** dataset used in the training process, where each sample is called a training sample. **Learning (or training)** is the process of building a model from data.
 - **Test set:** dataset used in the testing process, where each sample is called a test sample. Testing refers to the process, during which the learned model is used for prediction.

- Typical dataset composition

		Feature 1	Feature 2	Feature 3	Label
Training set	No.	Area	Location	Orientation	House Price
	1	100	8	South	1000
	2	120	9	Southwest	1300
	3	60	6	North	700
	4	80	9	Southeast	1100
Test set	5	95	3	South	850

Importance of Data Processing (Preparation)

- Data is crucial to models and determines the scope of model capabilities. All good models require good data.



Data Cleansing

- Most machine learning models process features, which are usually numeric representations of input variables that can be used in the model.
- In most cases, only preprocessed data can be used by algorithms. Data preprocessing involves the following operations:
 - Data filtering
 - Data loss handling
 - Handling of possible error or abnormal values
 - Merging of data from multiple sources
 - Data consolidation

Data Cleansing- Dirty Data

- Raw data usually contains data quality problems:
 - Incompleteness: Incomplete data or lack of relevant attributes or values.
 - Noise: Data contains incorrect records or abnormal points.
 - Inconsistency: Data contains conflicting records.

#	Id	Name	Birthday	Gender	IsTeacher?	#Students	Country	City
1	111	John	31/12/1990	M	0	0	Ireland	Dublin
2	222	Mery	15/10/1978	F	1	15	Iceland	
3	333	Alice	19/04/2000	F	0	0	Spain	Madrid
4	444	Mark	01/11/1997	M	0	0	France	Paris
5	555	Alex	15/03/2000	A	1	23	Germany	Berlin
6	555	Peter	1983-12-01	M	1	10	Italy	Rome
7	777	Calvin	05/05/1995	M	0	0	Italy	Italy
8	888	Roxane	03/08/1948	F	0	0	Portugal	Lisbon
9	999	Anne	05/09/1992	F	0	5	Switzerland	Geneva
10	101010	Paul	14/11/1992	M	1	26	Ytali	Rome

Missing value

Invalid value

Misfielded value

Misspelling

Invalid duplicate items

Incorrect format

Dependent attributes

How to Handle Missing Data

1. Ignore :

- Usually done when class label is missing (during classification) — not effective when the % of missing values/attribute varies considerably.

2. Fill in the missing value manually:

- Tedious + Infeasible

3. Fill in missing value automatically:

- Replace with a global constant such as a label like “unknown”

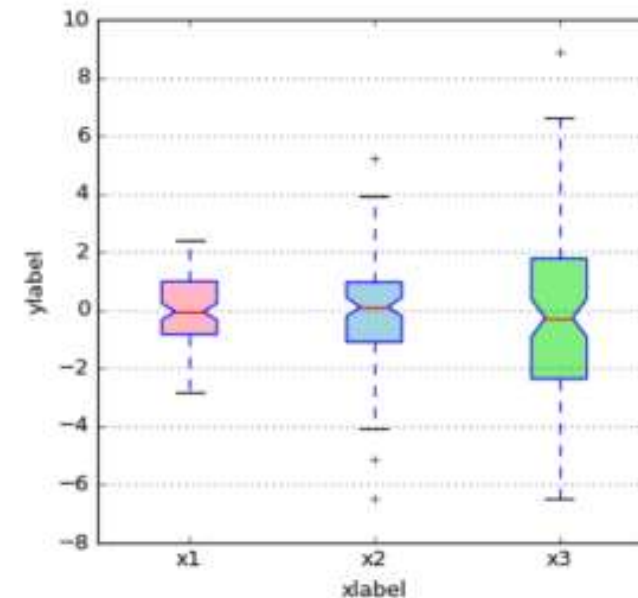
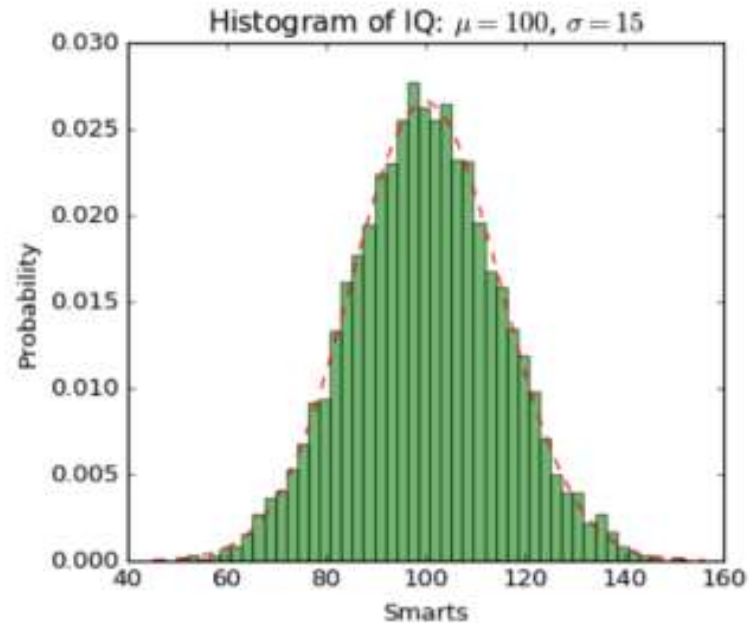
4. Fill missing value with measure of central tendency:

- For normal data distribution (Symmetric), use the mean
- For skewed data distribution (Asymmetric), use the median

Data Visualization

Data visualization Very important in Data Analysis Explore data
Report insights

https://www.w3schools.com/python/matplotlib_intro.asp
<https://matplotlib.org/stable/tutorials/index.html>



Matplotlib Library:

Install using pip:

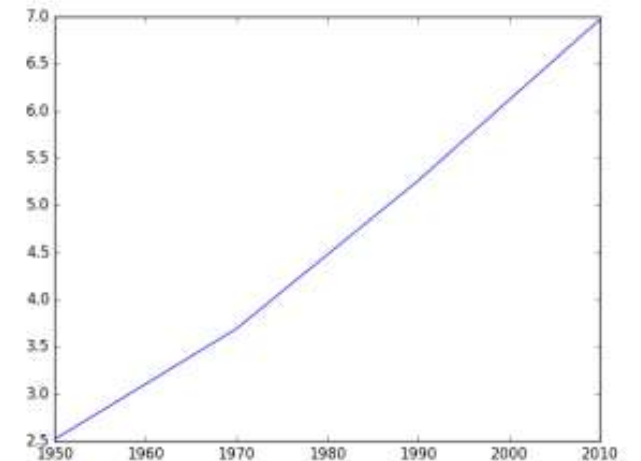
```
!pip install matplotlib
```

Matplotlib graphs your data on [Figures](#) each of which can contain one or more [Axes](#), an area where points can be specified in terms of x-y coordinates (or theta-r in a polar plot, x-y-z in a 3D plot, etc.).

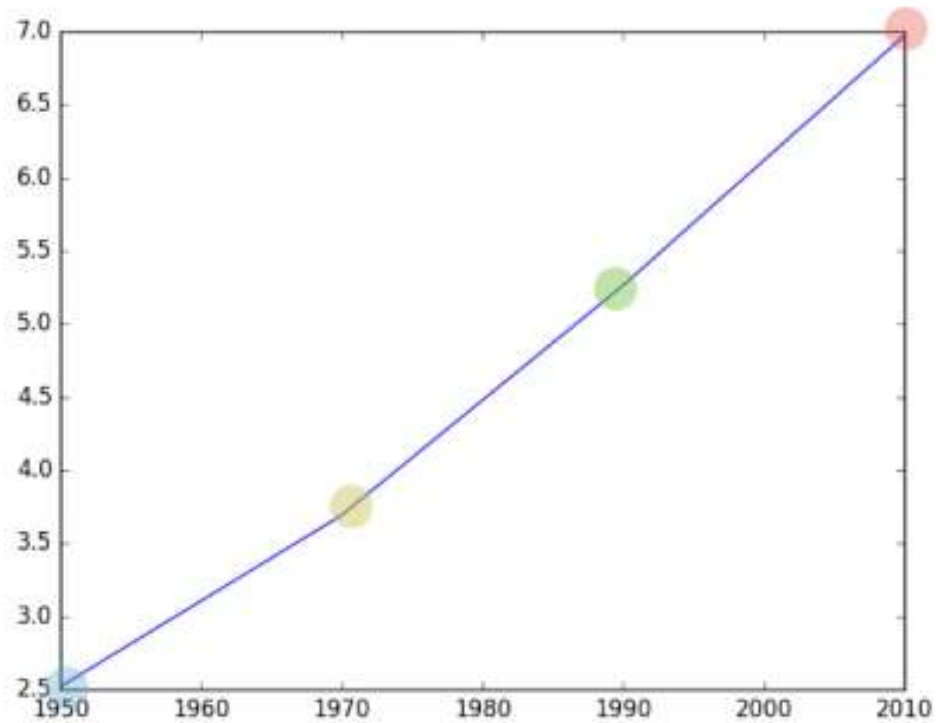
The simplest way of creating a Figure with an Axes is using [pyplot.subplots](#). We can then use [Axes.plot](#) to draw some data on the Axes.

plt.show() is used to get this Figure to display

```
import matplotlib.pyplot as plt
year = [1950, 1970, 1990, 2010]
pop = [2.519, 3.692, 5.263, 6.972]
plt.plot(year, pop)
plt.show()
```



Matplotlib Library:

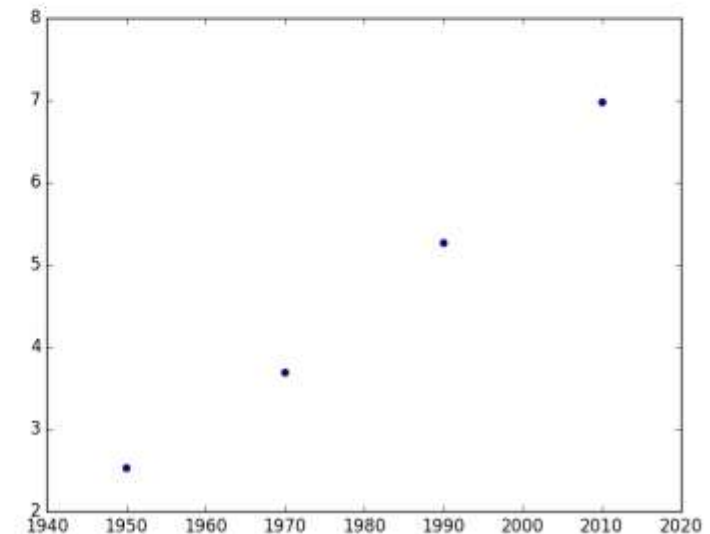


```
year = [1950 , 1970 , 1990 , 2010]  
pop  = [2.519, 3.692, 5.263, 6.972]
```

Matplotlib Library-Scatter plot:

- A scatter plot is a diagram where each value in the data set is represented by a dot.
- The Matplotlib module has a method for drawing scatter plots `.scatter(x,y)`, it needs two arrays of the same length, one for the values of the x-axis, and one for the values of the y-axis.

```
import matplotlib.pyplot as plt  
year = [1950, 1970, 1990, 2010]  
pop = [2.519, 3.692, 5.263, 6.972]  
plt.scatter(year, pop)  
plt.show()
```

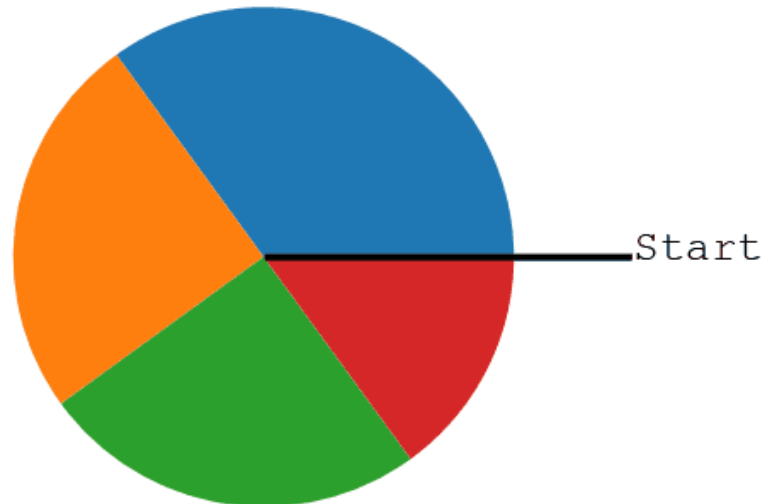


Matplotlib Library-Pie Charts:

With Pyplot, you can use the `pie()` function to draw pie charts

```
import matplotlib.pyplot as plt  
y = [35, 25, 25, 15]  
plt.pie(y)  
plt.show()
```

By default, the plotting of the first wedge starts from the x-axis and moves counterclockwise:

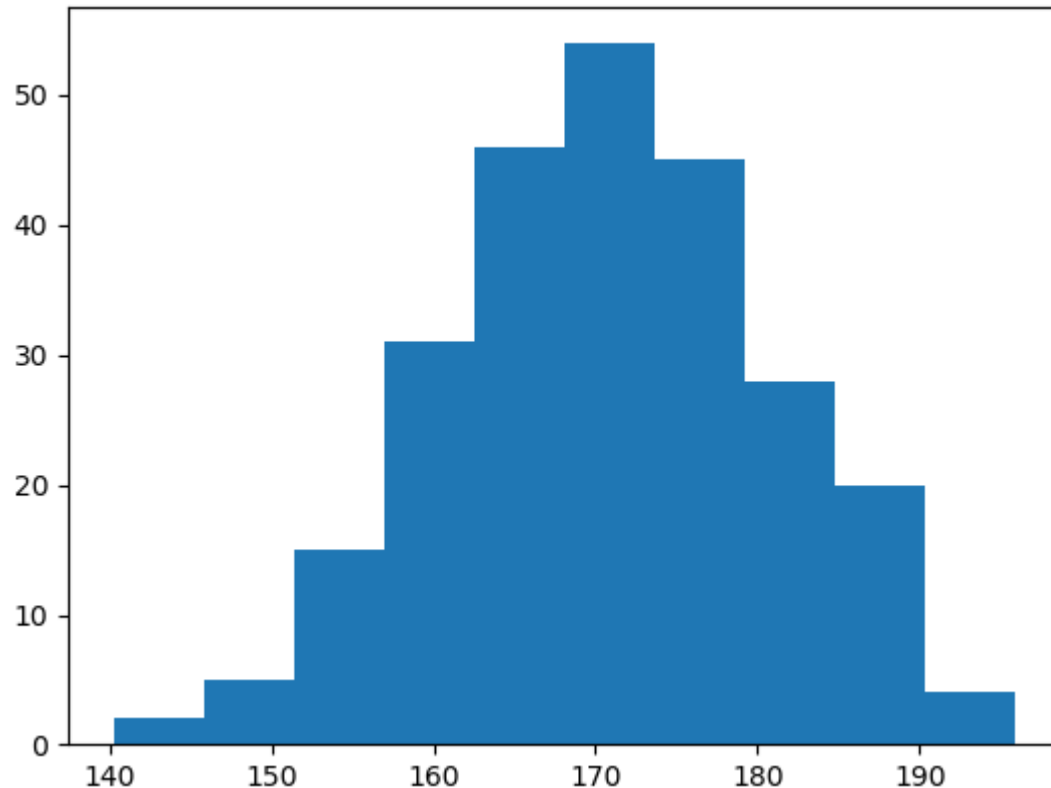


Matplotlib Library- Histogram :

A histogram is a graph showing *frequency* distributions.

It is a graph showing the number of observations within each given interval.

Example: Say you ask for the height of 250 people; you might end up with a histogram like this



You can read from the histogram that there are approximately:

2 people from 140 to 145cm

5 people from 145 to 150cm

15 people from 151 to 156cm

31 people from 157 to 162cm

46 people from 163 to 168cm

53 people from 168 to 173cm

45 people from 173 to 178cm

28 people from 179 to 184cm

21 people from 185 to 190cm

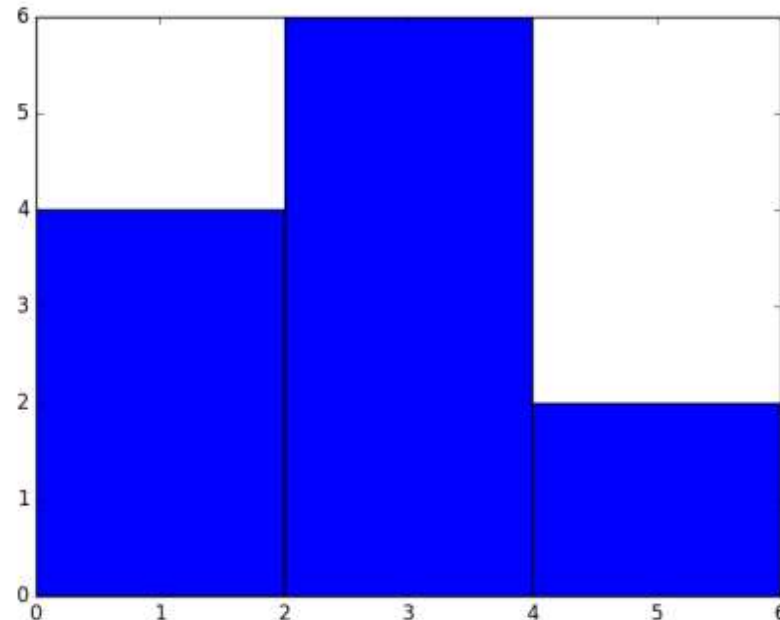
4 people from 190 to 195cm

Matplotlib Library-Histogram:

In Matplotlib, we use the `hist()` function to create histograms.

The `hist()` function will use an array of numbers to create a histogram, the array is sent into the function as an argument.

```
values = [0,0.6,1.4,1.6,2.2,2.5,2.6,3.2,3.5,3.9,4.2,6]  
plt.hist(values, bins=3)  
plt.show()
```



Matplotlib Library-Title:

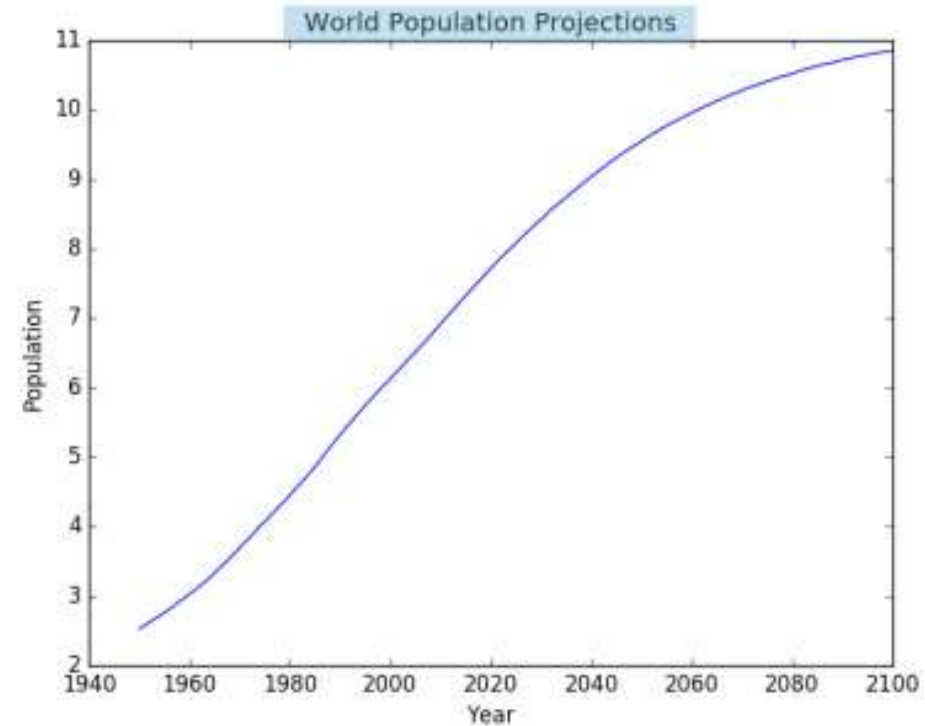
With Pyplot, you can use the `xlabel()` and `ylabel()` functions to set a label for the x- and y-axis and `title()` to set a graph title .

```
import matplotlib.pyplot as plt
year = [1950, 1951, 1952, ..., 2100]
pop = [2.538, 2.57, 2.62, ..., 10.85]

plt.plot(year, pop)

plt.xlabel('Year')
plt.ylabel('Population')
plt.title('World Population Projections')

plt.show()
```

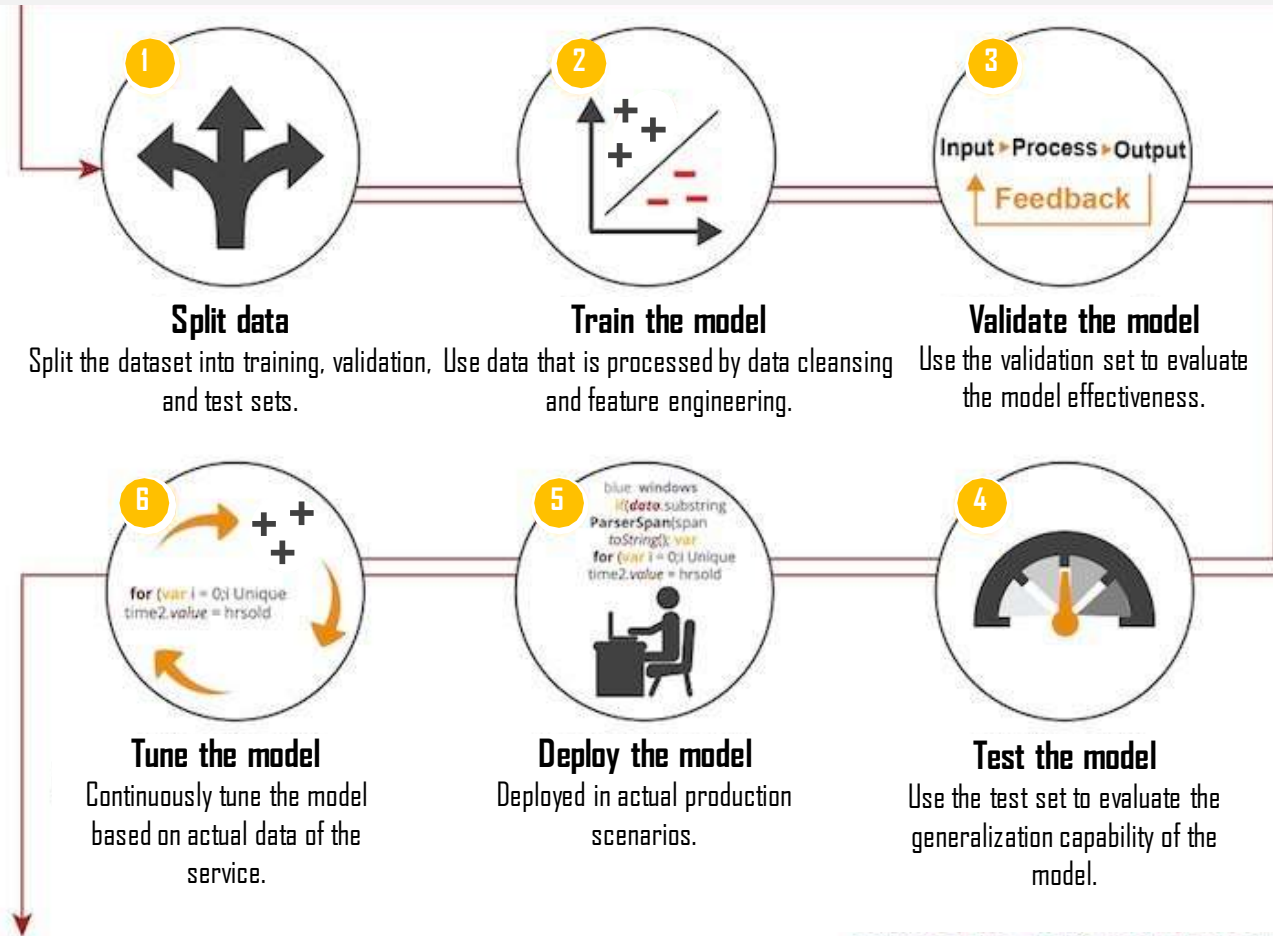


Data Conversion

- Preprocessed data needs to be converted into **a representation suitable for machine learning models**. The following are typically used to convert data:
 - Encoding categorical data into numerals for classification
 - Converting numeric data into categorical data to reduce the values of variables (for example, segmenting age data)
 - Other data:
 - Embedding words into text to convert them into word vectors (Typically, models such as word2vec and BERT are used.)
 - Image data processing, such as color space conversion, grayscale image conversion, geometric conversion, Haar-like features, and image enhancement
 - Feature engineering:
 - Normalizing and standardizing features to ensure that different input variables of a model fall into the same value range
 - Feature augmentation: combining or converting the existing variables to generate new features, such as averages.

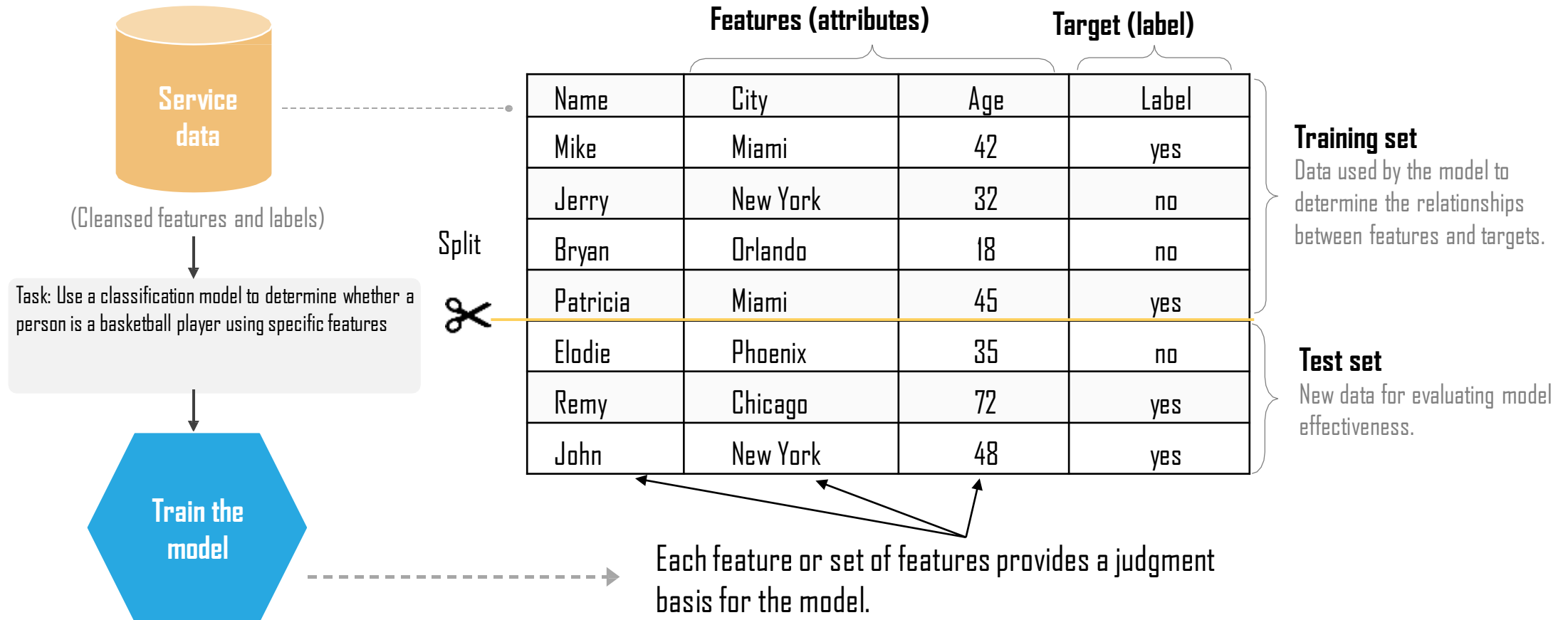
Overall Process of Building an AI Model

Model building process

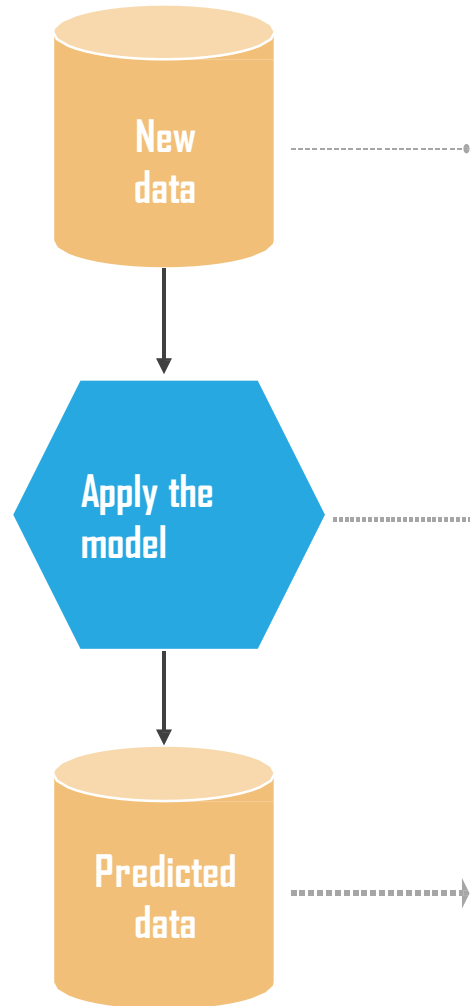


Supervised Learning Example - Learning Phase

- Use a classification model to determine whether a person is a basketball player based on specific features.



Supervised Learning Example - Prediction Phase



Name	City	Age	Label
Marine	Miami	45	?
Julien	Miami	52	?
Fred	Orlando	20	?
Michelle	Boston	34	?
Nicolas	Phoenix	90	?

Unknown data

Recent data cannot determine whether they are a basketball player.

IF city = Miami $\rightarrow$ Probability = +0.7
IF city = Orlando $\rightarrow$ Probability = +0.2
IF age > 42 $\rightarrow$ Probability = +0.05*age + 0.06
IF age <= 42 $\rightarrow$ Probability = +0.01*age + 0.02

Name	City	Age	Prediction
Marine	Miami	45	0.3
Julien	Miami	52	0.9
Fred	Orlando	20	0.6
Michelle	Boston	34	0.5
Nicolas	Phoenix	90	0.4

Predicted possibility

Use the model against new data to predict the possibility of them being basketball players.

What Is a Good Model?



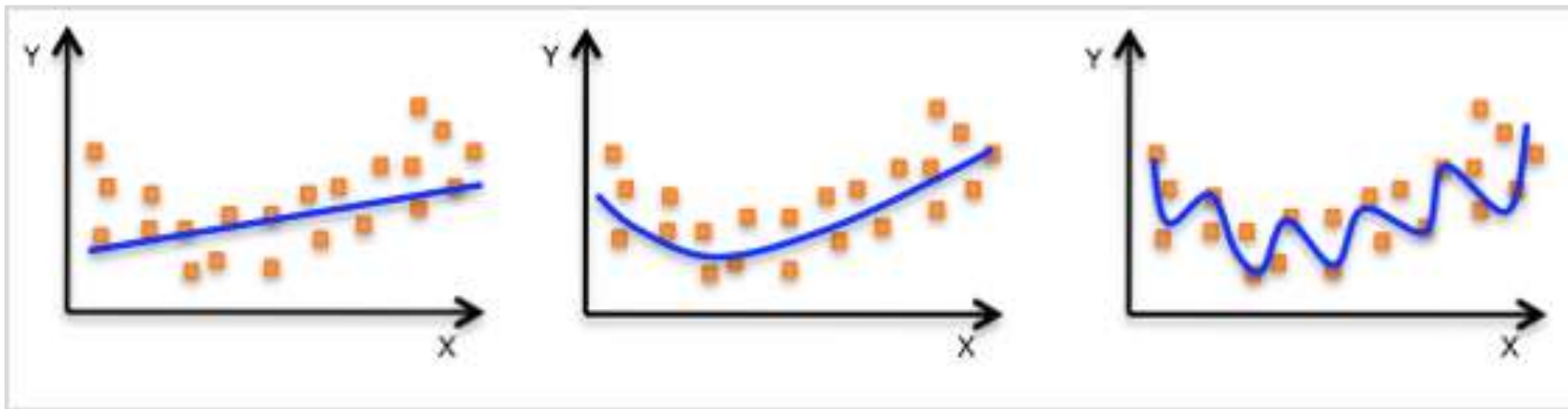
- **Generalization**
The accuracy of predictions based on actual data
- **Explain ability**
Predicted results are easy to explain
- **Prediction speed**
The time needed to make a prediction

Model Effectiveness (1)

- **Generalization capability:** Machine learning aims to ensure models perform well on new samples, not just those used for training. Generalization capability, also called **robustness**, is the extent to which a learned model can be applied to new samples.
- **Error** is the difference between the prediction of a learned model on a sample and the actual result of the sample.
 - Training error is the error of the model on the training set.
 - Generalization error is the error of the model on new samples. Obviously, we prefer a model with a smaller generalization error.
- **Underfitting:** The training error is large.
- **Overfitting:** The training error of a trained model is small while the generalization error is large.

Model Effectiveness (2)

- **Model capacity**, also known as model complexity, is the capability of the model to fit various functions.
 - With sufficient capacity to handle task complexity and training data volumes, the algorithm results are optimal.
 - Models with an insufficient capacity cannot handle complex tasks because **underfitting** may occur.
 - Models with a large capacity can handle complex tasks, but **overfitting** may occur when the capacity is greater than the amount required by a task.



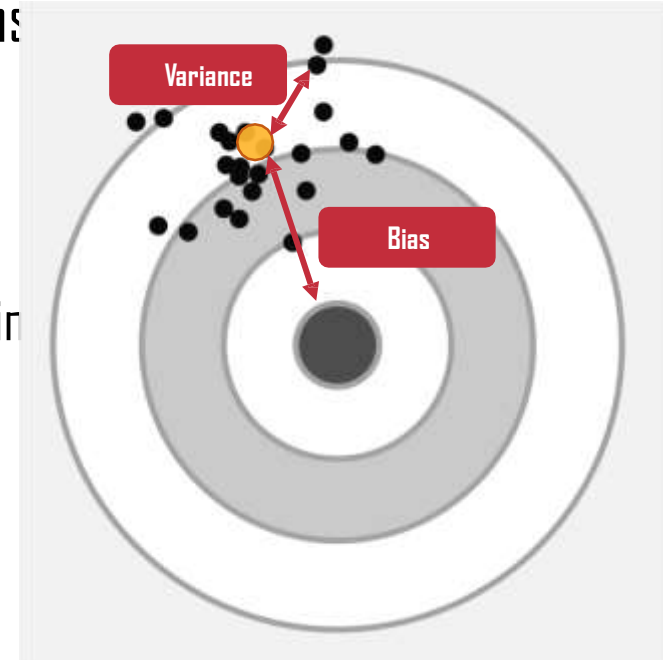
Underfitting:
features not learned

Good fitting

Overfitting:
noises learned

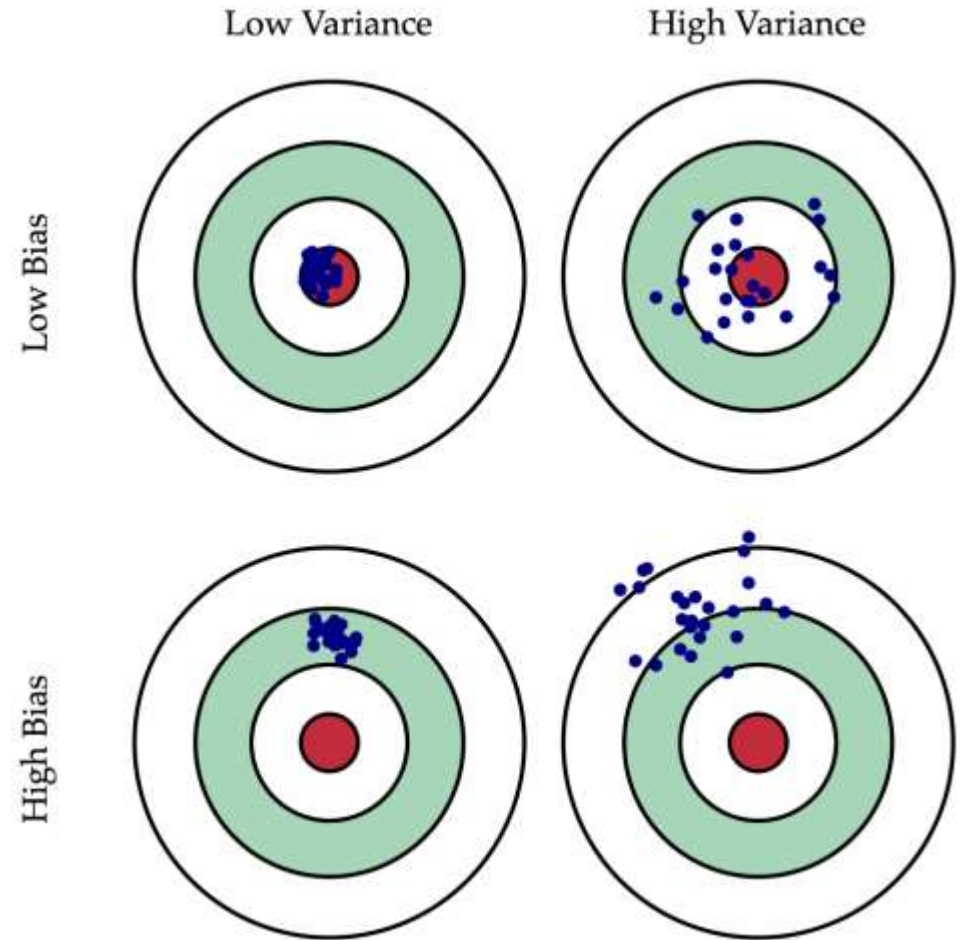
Cause of Overfitting - Errors

- Prediction error = **Bias**² + **Variance** + Ineliminable error
- In general, the two main factors of prediction error are variance and bias
- Variance:
 - How much a prediction result deviates from the mean
 - Variance is caused by the sensitivity of the model to small fluctuations in a training set.
- Bias:
 - Difference between the average of the predicted values and the actual values.



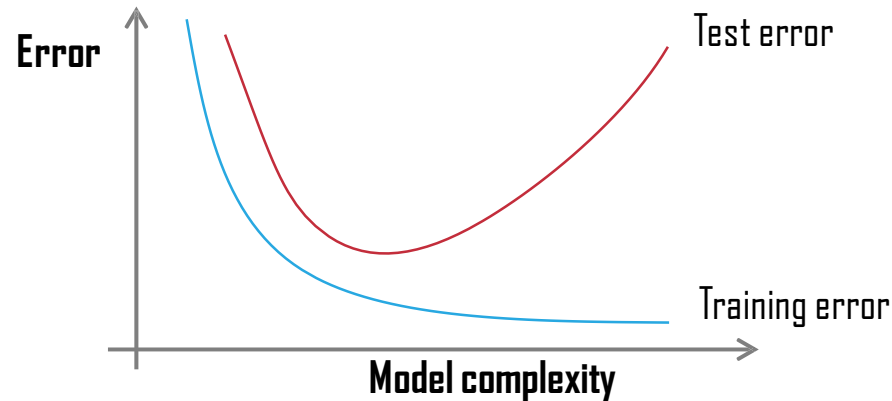
Variance and Bias

- Different combinations of variance and bias are as follows:
 - Low bias & low variance → **good model**
 - Low bias & high variance → **inadequate model**
 - High bias & low variance → **inadequate model**
 - High bias & high variance → **bad model**
- An ideal model can accurately capture the rules in the training data and be generalized to invisible (new) data. However, it is impossible for a model to complete both tasks at the same time.



Complexity and Errors of Models

- The more complex a model is, the smaller its training error is.
- As the model complexity increases, the test error decreases before increasing again, forming a convex curve.



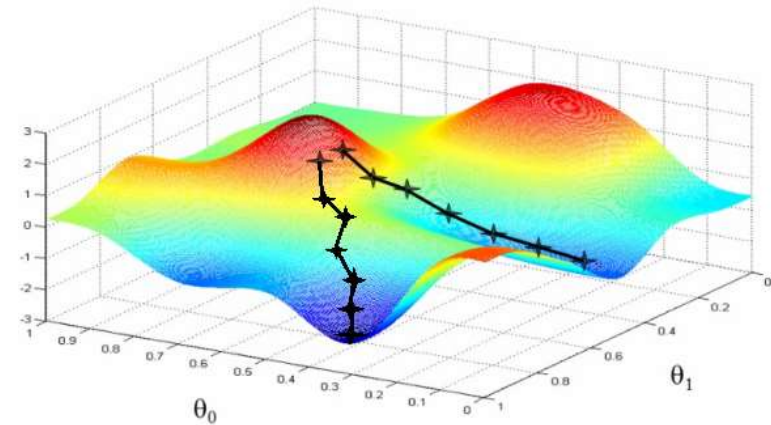
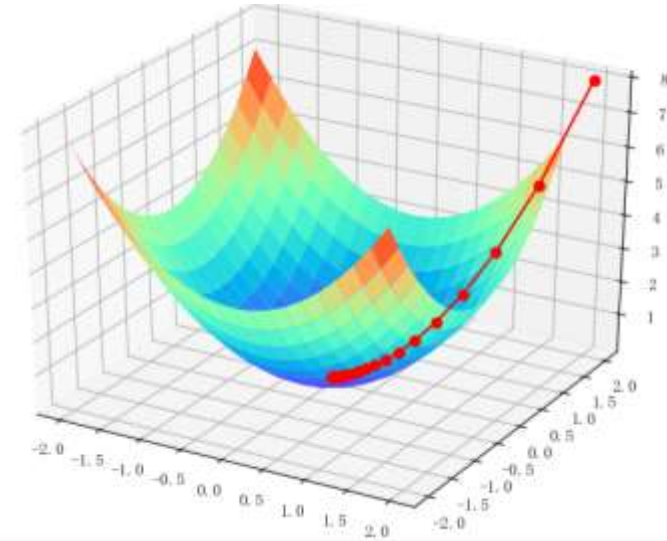
Contents

1. Machine Learning Algorithms
2. Types of Machine Learning
3. Machine Learning Process
- 4. Important Machine Learning Concepts**
5. Common Machine Learning Algorithms

Machine Learning Training Methods - Gradient Descent

- This method uses the negative gradient direction of the current position as the search direction, which is the fastest descent direction of the current position. The formula is as follows:

- $w_{k+1} = w_k - \eta \nabla f_{w_k}(x^i)$
 η is the learning rate. i indicates the i th data record. $\eta \nabla f_{w_k}(x^i)$ indicates the change of weight parameter w in each iteration.
- Convergence means that the value of the objective function changes very little or reaches the maximum number of iterations.



Machine Learning Training Methods - Gradient Descent (2)

- Batch gradient descent (BGD) uses the sum of gradients of all m samples of the dataset at the current point to update the weight parameter.

$$w_{k+1} = w_k - \eta \frac{1}{m} \sum_{i=1}^m \nabla f_{w_k}(x^i)$$

- Stochastic gradient descent (SGD) randomly uses the gradient of a random sample of the dataset at the current point to update the weight parameter in the current gradient.

$$w_{k+1} = w_k - \eta \nabla f_{w_k}(x^i)$$

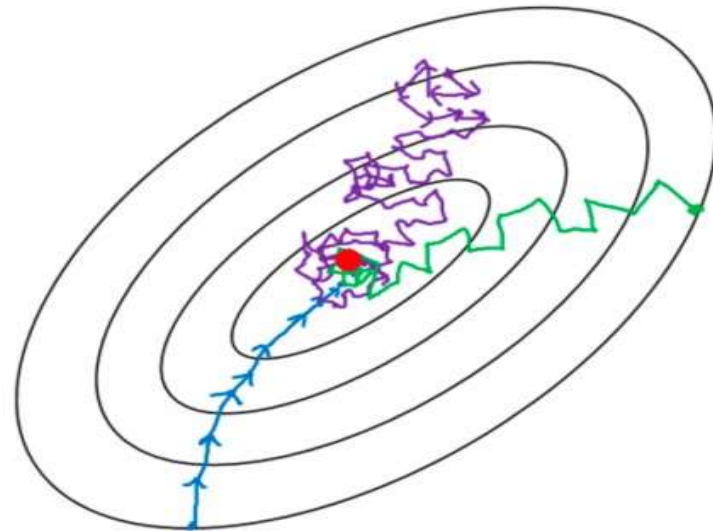
- Mini-batch gradient descent (MBGD) combines the features of BGD and SGD, and chooses the gradients of n samples in a dataset each time to update the weight parameter.

$$w_{k+1} = w_k - \eta \frac{1}{n} \sum_{i=t}^{t+n-1} \nabla f_{w_k}(x^i)$$

Machine Learning Training Methods - Gradient Descent (3)

- Comparison of gradient descent methods

- SGD randomly chooses samples for each training pass, causing instability. As a result, the loss function fluctuates or even produces reverse displacement during the process of dropping to the minimum.
- BGD is the most stable, but it consumes too many compute resources. MBGD is a balance between BGD and SGD.

**BGD**

Use **all** training samples for each training pass.

SGD

One training sample is used for each training pass.

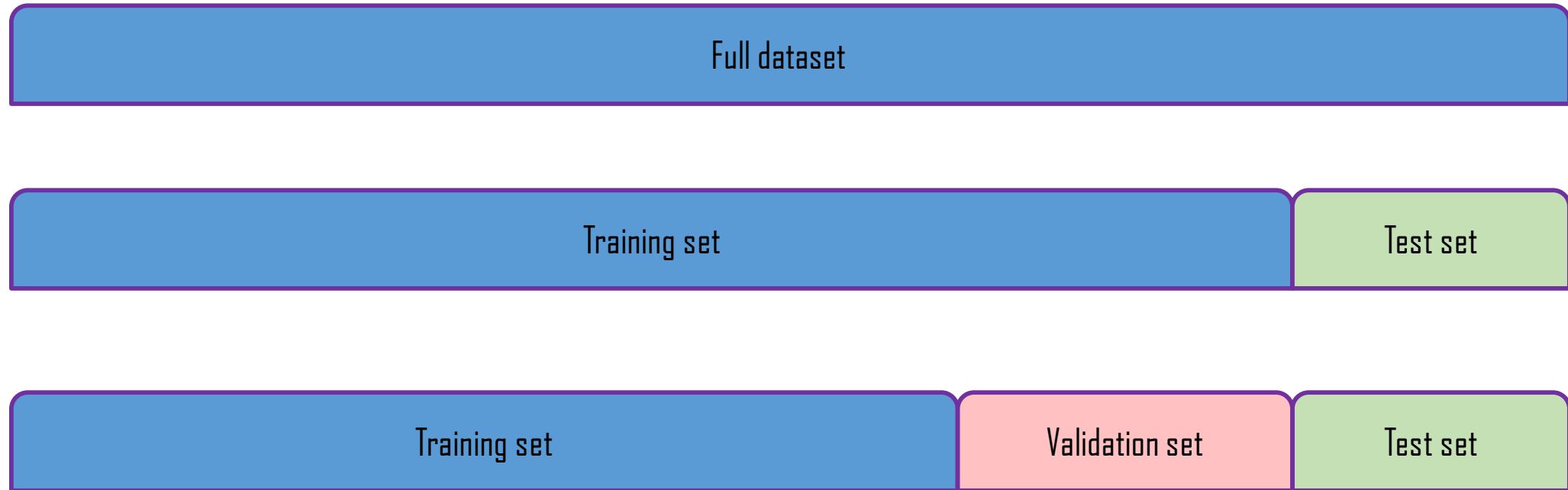
MBGD

A certain number of training samples are used for each training pass.

Cross-Validation

- **Cross-validation** is a statistical analysis method used to check the The former is used to train a classifier, whereas the latter is used to evaluate the classifier by testing the trained model.
- **k -fold cross-validation (k -fold CV):**
 - Divides the original data into k (usually equal-sized) subsets.
 - Each unique group is treated as a validation set, and the remaining $k - 1$ groups are treated as the training set. In this way, k models are obtained.
 - The average classification accuracy score of the k models on the validation set is used as the performance metric for k -fold CV classifiers.

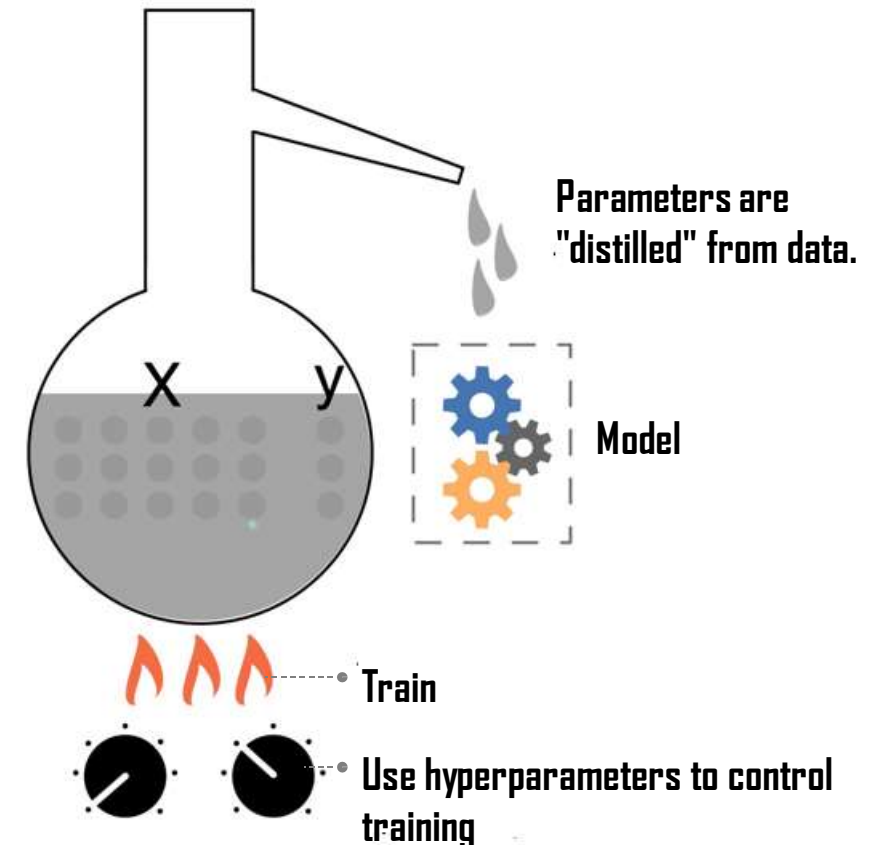
Cross-Validation (2)



- Note: k in k -fold CV is a hyperparameter.

Parameters and Hyperparameters

- A model contains not only parameters but also hyperparameters. Hyperparameters enable the model to learn the optimal configurations of the parameters.
 - Parameters are automatically learned by models.
 - Hyperparameters are manually set.



Hyperparameters

- Commonly used for model parameter estimation.
- Specified by the user.
- Set heuristically.
- Often tuned for a given predictive modeling problem.

Hyperparameters are configurations outside the model.

- λ of Lasso/Ridge regression
- Learning rate, number of iterations, batch size, activation function, and number of neurons of a neural network to be trained
- C and σ of support vector machines (SVMs)
- k in the k -nearest neighbors (k -NN) algorithm
- Number of trees in a random forest

Common hyperparameters

Hyperparameter Search Process and Methods

Hyperparameter search
general process

1. Divide a dataset into a training set, validation set, and test set.
2. Optimize the model parameters using the training set based on the model performance metrics.
3. Search for model hyperparameters using the validation set based on model performance metrics.
4. Perform step 2 and step 3 alternately until model parameters and hyperparameters are determined and assess the model using the test set.

Search algorithms
(step 3)

- **Grid search**
- **Random search**
- Heuristic intelligent search
- Bayesian search

Contents

1. Machine Learning Algorithms
2. Types of Machine Learning
3. Machine Learning Process
4. Important Machine Learning Concepts
- 5. Common Machine Learning Algorithms**

Machine Learning Algorithm Overview

