ARTIFICIAL INTELLIGENCE NEEDS REAL INTELLIGENCE

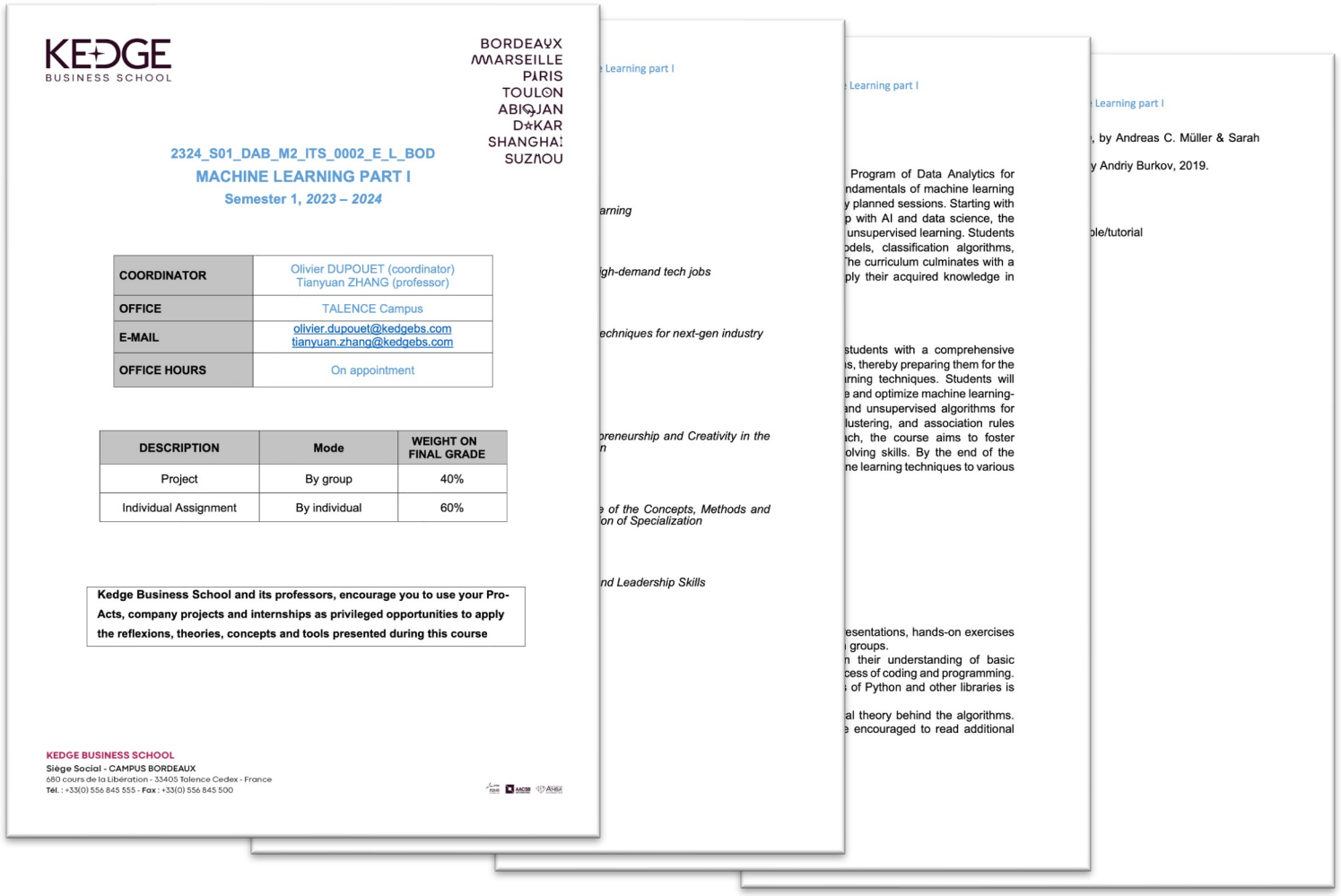
MSc DAB
Machine Learning
Part I

Professor: Tianyuan ZHANG
tianyuan.zhang@kedgebs.com

KEDGE BUSINESS SCHOOL

Syllabus

- Available on Learn



KEDGE BUSINESS SCHOOL

BORDEAUX
MARSEILLE
PARIS
TOULON
ABIDJAN
DAKAR
SHANGHAI
SUZHOU

2324_S01_DAB_M2_ITS_0002_E_L_BOD

MACHINE LEARNING PART I

Semester 1, *2023 – 2024*

COORDINATOR	Olivier DUPOUET (coordinator) Tianyuan ZHANG (professor)
OFFICE	TALENCE Campus
E-MAIL	olivier.dupouet@kedgebs.com tianyuan.zhang@kedgebs.com
OFFICE HOURS	On appointment

DESCRIPTION	Mode	WEIGHT ON FINAL GRADE
Project	By group	40%
Individual Assignment	By individual	60%

Kedge Business School and its professors, encourage you to use your Pro-Acts, company projects and internships as privileged opportunities to apply the reflexions, theories, concepts and tools presented during this course

KEDGE BUSINESS SCHOOL
Siège Social - CAMPUS BORDEAUX
680 cours de la Libération - 33405 Talence Cedex - France
Tél. : +33(0) 556 845 555 - Fax : +33(0) 556 845 500

KEDGE BUSINESS SCHOOL

Objectives & goals

- **Fundamental terms & concepts** of machine learning
- **Classical** machine learning tools & techniques
  - Supervised learning
    - Regression
    - Classification
  - Unsupervised learning
    - Clustering
    - Association rules mining
- The **whole process** of the machine learning project
  - From problem framing to model deployment

Pedagogy methods

- This course won’t focus on the mathematical theory behind algorithms.
- For each session:
  - Lecture to present key topics and algorithms
  - Hands-on exercises with Jupyter Notebook

Pedagogy methods

- Programming language

- Python library

Pedagogy methods

- For debug:
  - Read **official documentations**
  - Read **official tutorials**

  - Search on stackoverflow, etc
  - Ask ChatGPT

Pedagogy methods

- For debug:
  - Read **official documentations**
  - Read **official tutorials**
  - Search on stackoverflow, etc
  - Ask ChatGPT

Decide the search term & question to ask

Simply copy & paste won't work in most cases

Evaluation

- Individual evaluation → Assigements of each session → 60%
- Collective evaluation → Group project → 40%

MON	*TUE*	*WED*	THU	FRI	SAT	SUN
06/11	*07/11*	*08/11*	09/11	10/11	11/11	12/11
13/11	*14/11*	*15/11*	16/11	17/11	18/11	19/11
20/11	*21/11*	*22/11*	23/11	24/11	25/11	26/11
27/11	*28/11*	*29/11*	30/11	01/12	02/12	03/12
04/12	*05/12*	*06/12*	07/12	08/12	08/12	10/12

Evaluation

- Individual evaluation → Assigements of each session → 60%
- Collective evaluation → Group project → 40%

MON	*TUE*	*WED*	THU	*FRI*	SAT	SUN
06/11	*07/11*	*08/11*	09/11	*10/11*	11/11	12/11
13/11	*14/11*	*15/11*	16/11	*17/11*	18/11	19/11
20/11	*21/11*	*22/11*	23/11	*24/11*	25/11	26/11
27/11	*28/11*	*29/11*	30/11	*01/12*	02/12	03/12
04/12	*05/12*	*06/12*	07/12	08/12	08/12	10/12

Evaluation

- Individual evaluation → Assigements of each session → 60%
- Collective evaluation → Group project → 40%

MON	TUE	WED	THU	FRI	SAT	SUN
06/11	07/11	08/11	09/11	10/11	11/11	12/11
13/11	14/11	15/11	16/11	17/11	18/11	19/11
20/11	21/11	22/11	23/11	24/11	25/11	26/11
27/11	28/11	29/11	30/11	01/12	02/12	03/12
04/12	05/12	06/12	07/12	08/12	08/12	10/12

If you want to fly by yourself

- Online course:
  - Microsoft Learn: Machine Learning for Beginners
  - Google Machine Learning Education
- Don’t forget to complete the assignments before the deadline.

ARTIFICIAL INTELLIGENCE NEEDS REAL INTELLIGENCE

Introduction to Machine Learning

Professor: Tianyuan ZHANG
tianyuan.zhang@kedgebs.com

KEDGE BUSINESS SCHOOL

Outline

- **Machine Learning & Artificial Intelligence & Data Science**
- What is Machine Learning?
- Basic terms & Foundational concepts
- Types of Machine Learning
- Workflow of Machine Learning

Machine Learning & Artificial Intelligence & Data Science

- Artificial Intelligence (AI)
  - A science of getting machines to accomplish tasks that typically require human level intelligence.
  - Alan Turing in 1950
    - Machines that think
    - Turing test
  - Examples in the early stage
    - ELIZA chatbot
    - Shakey the robot

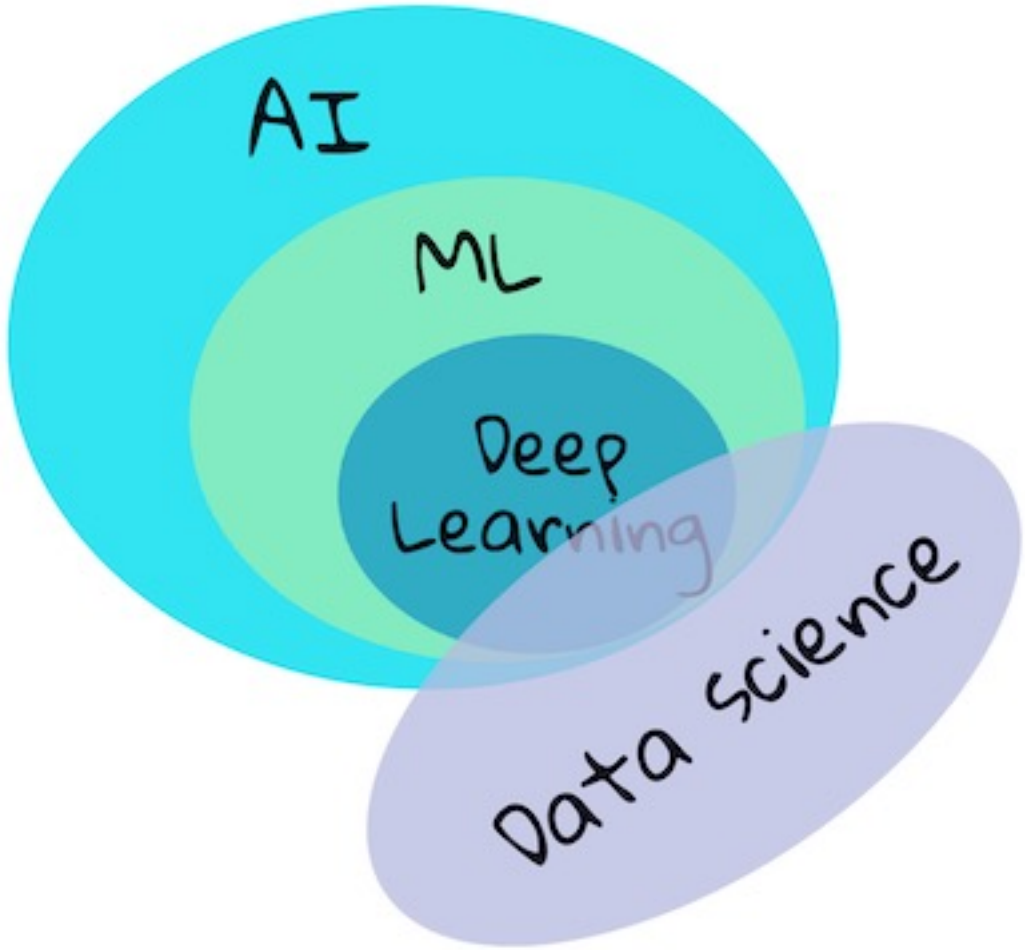
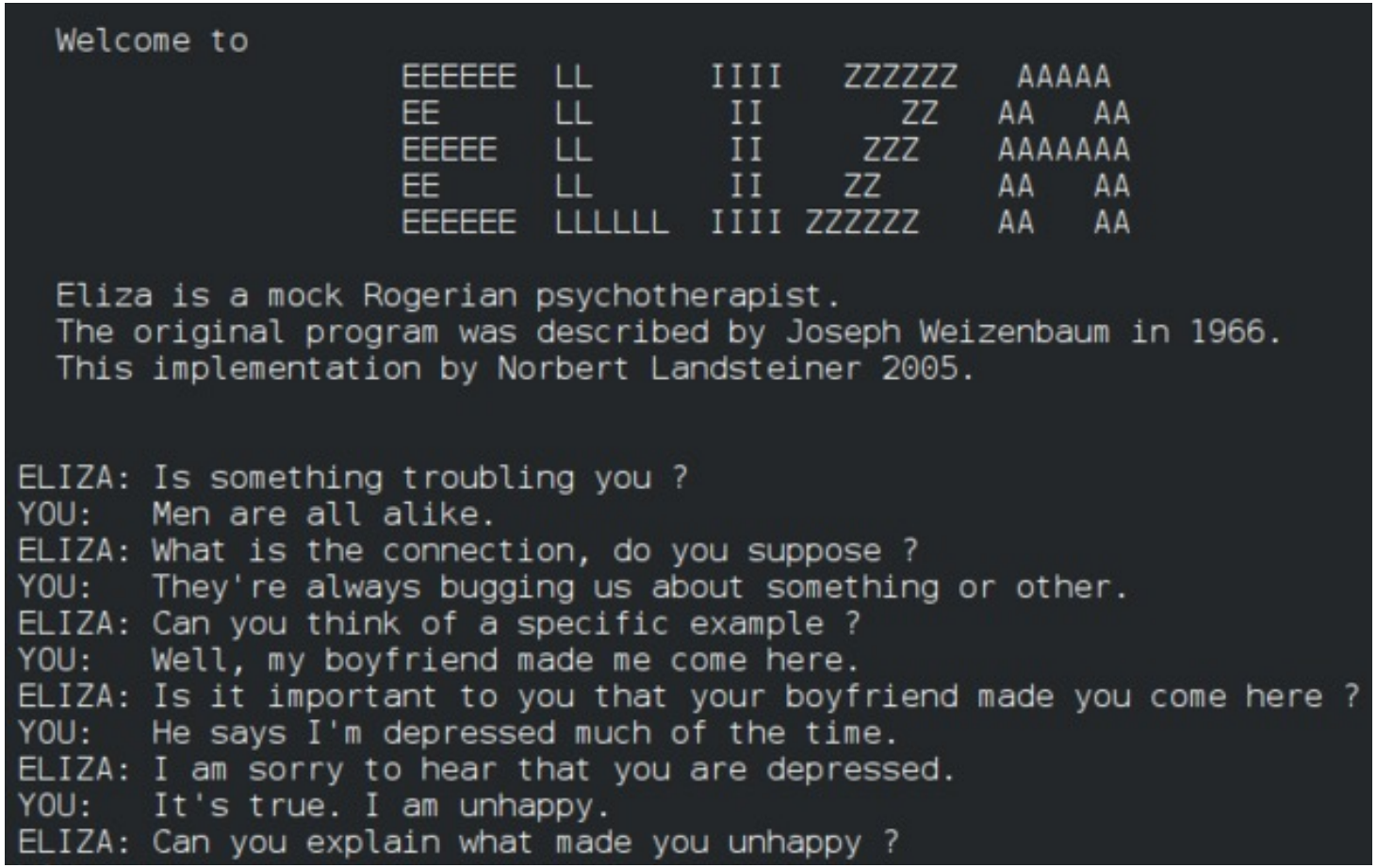


Image source: https://github.com/microsoft/ML-For-Beginners/blob/main/1-Introduction/1-intro-to-ML/README.md

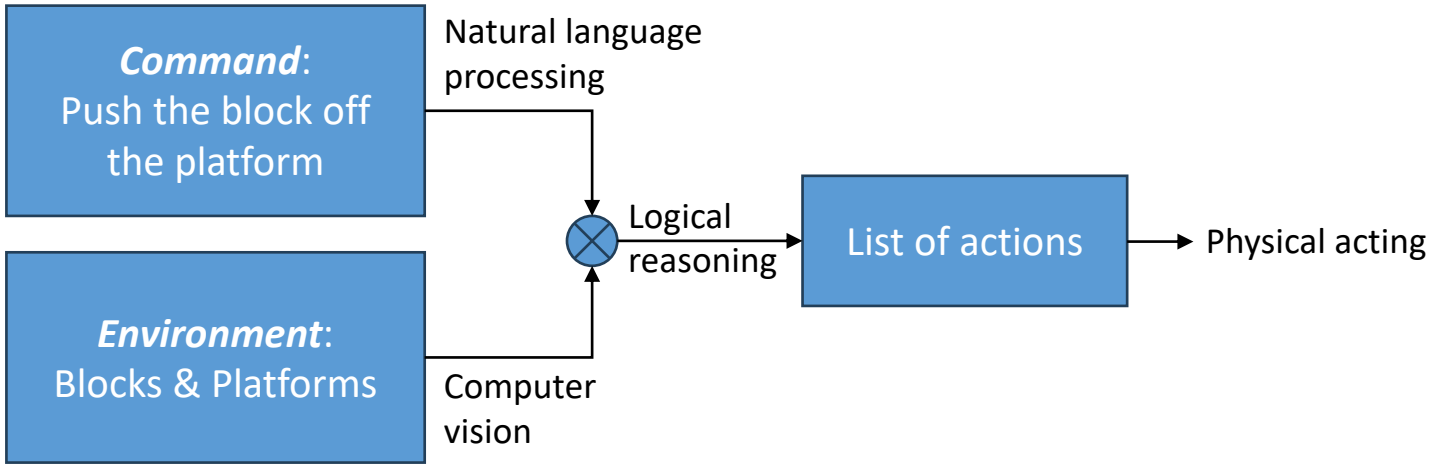
Machine Learning & Artificial Intelligence & Data Science

- Artificial Intelligence (AI)
  - ELIZA chatbot
    - From 1964 to 1967
    - Natural language processing
    - By MIT
    - Designed to act like a therapist
      - Reflect back the patient's words to the patient
      - Without really understanding what was being said by either party



Machine Learning & Artificial Intelligence & Data Science

- Artificial Intelligence (AI)
  - Shakey the robot
    - From 1966 through 1972
    - Developed by the Artificial Intelligence Center of Stanford Research Institute
    - The first general-purpose mobile robot



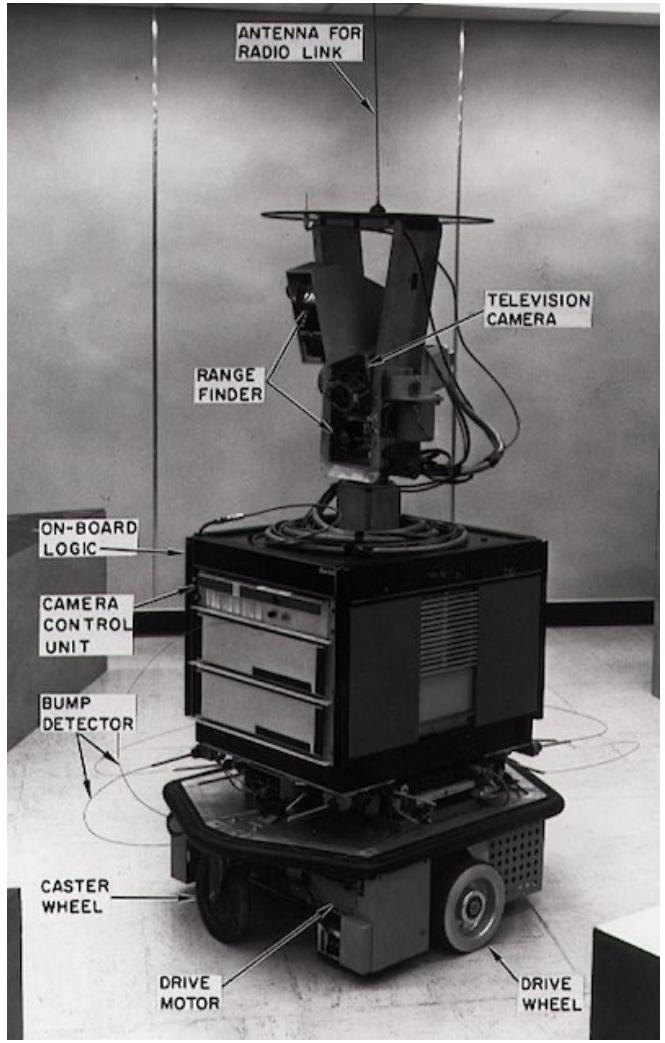


Image source: https://en.wikipedia.org/wiki/Shakey_the_robot

Machine Learning & Artificial Intelligence & Data Science

- Artificial Intelligence (AI)
  - Expert systems in 1980s
    - Rules engine defining knowledge / expertise
    - Inference engine leveraging rules to deduce new facts
    - Too many rules are needed to achieve satisfactory intelligence
    - Too specialized and centralized
    - Not taking advantage of the modern explosion of big data

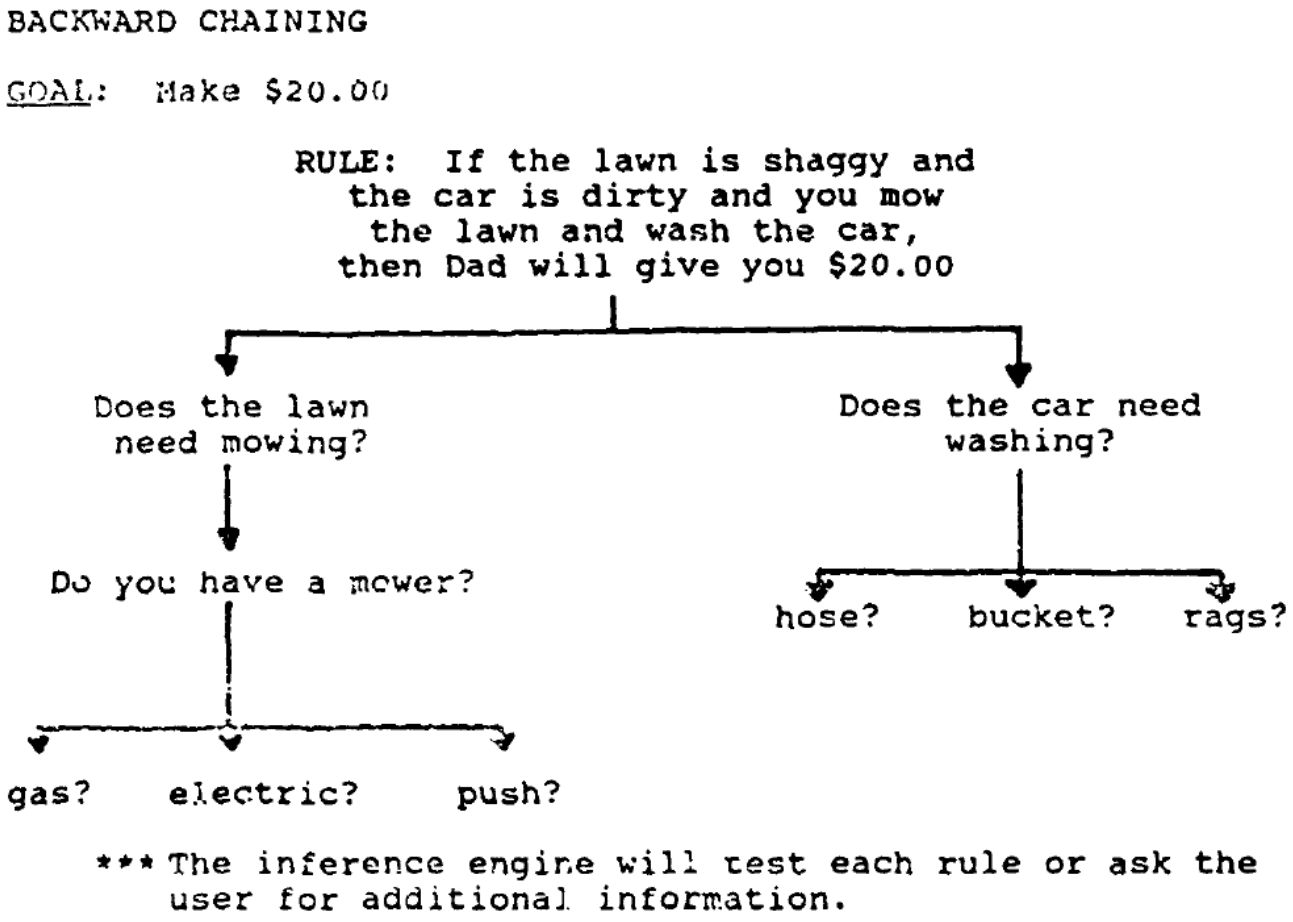


Image source: https://en.wikipedia.org/wiki/Expert_system

Machine Learning & Artificial Intelligence & Data Science

- Artificial Intelligence (AI)
  - Machine learning is the most successful and popular approach to AI these days.
- Machine Learning (ML)
  - ML is a subset of AI.
  - ML uses specialized algorithms to make decisions by learning from data.
  - Deep learning is a subset of ML that uses deep artificial neural networks for learning.

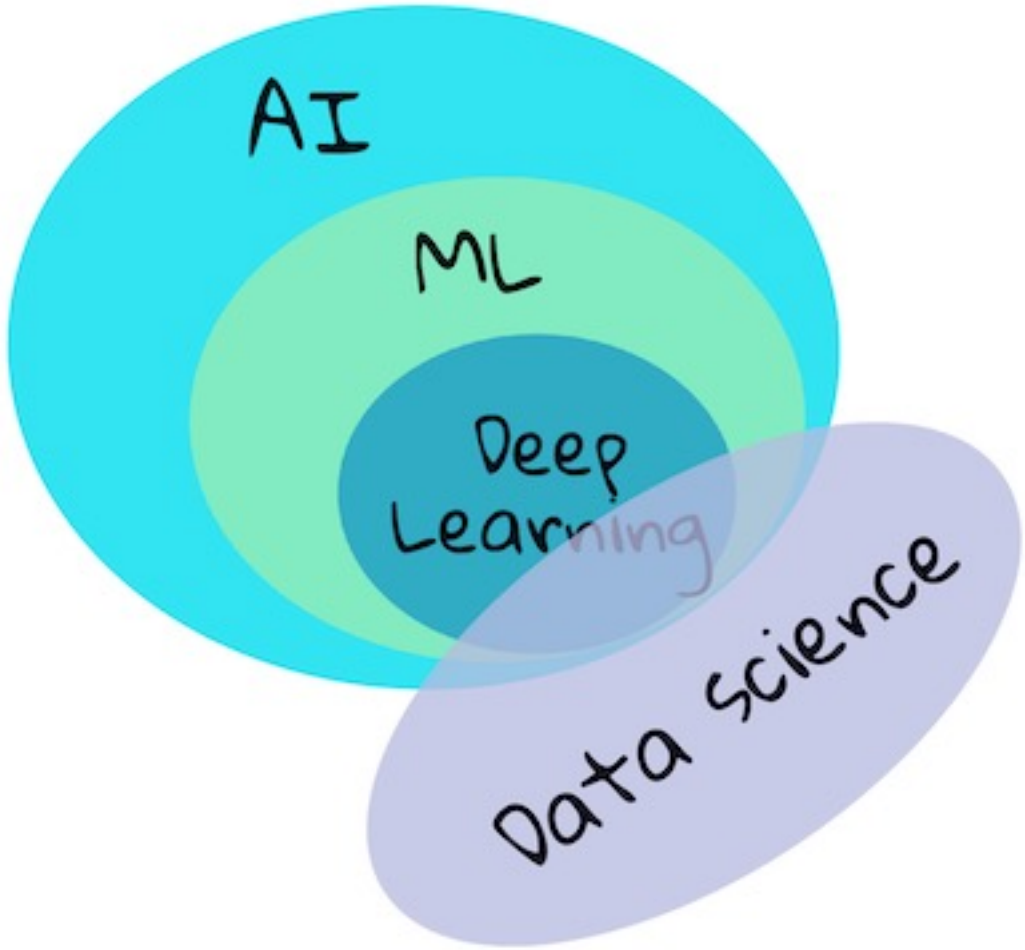


Image source: https://github.com/microsoft/ML-For-Beginners/blob/main/1-Introduction/1-intro-to-ML/README.md

Machine Learning & Artificial Intelligence & Data Science

- Machine Learning (ML)
  - Learn from data
  - Create models to make predictions
  - Focus on the future
- Data Science (DS)
  - Analyze and interpret data
  - Extract insights from data
  - Focus on the past

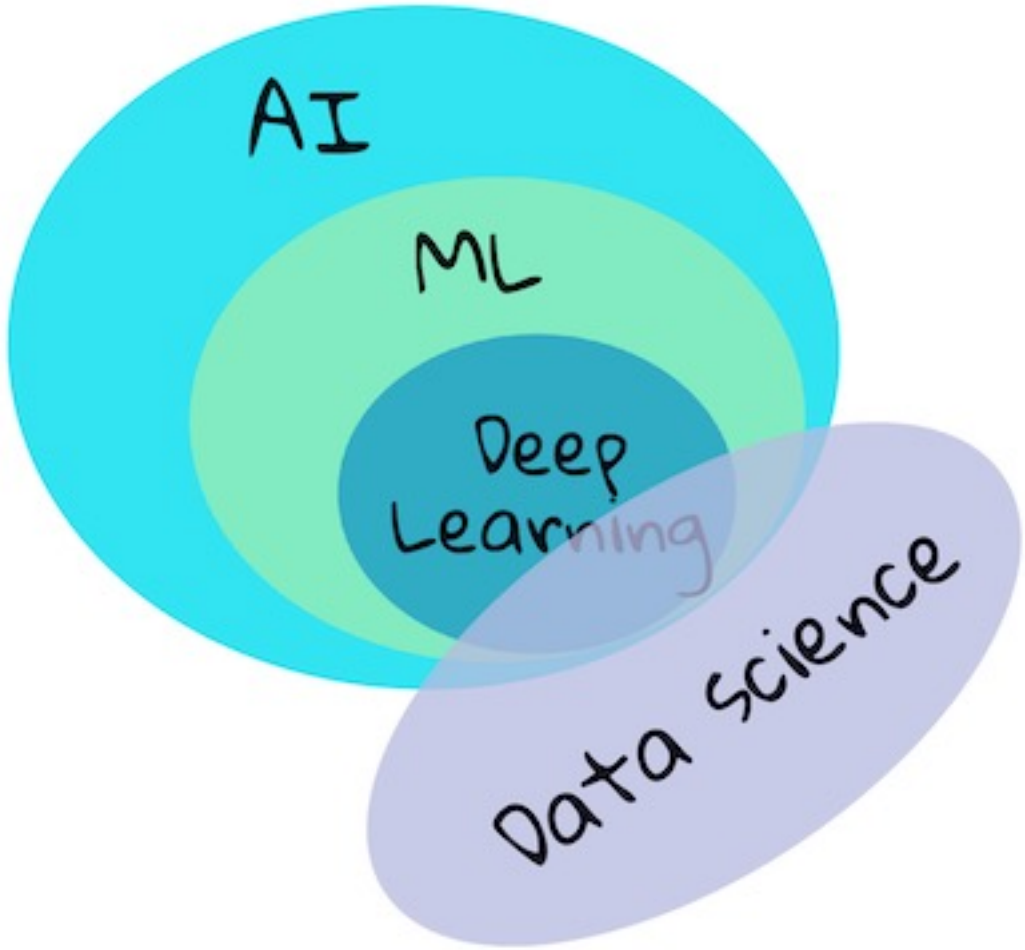


Image source: https://github.com/microsoft/ML-For-Beginners/blob/main/1-Introduction/1-intro-to-ML/README.md

Machine Learning & Artificial Intelligence & Data Science

- ML vs. DS

Size of the house (m^2)	# of bedrooms	Newly renovated	Price (1000 €)
523	1	N	115
645	1	N	150
660	2	N	210
1034	3	Y	280
2290	4	N	355
2545	4	Y	440

- ML provided an estimated price (230 **€**) for a newly renovated house with 2 bedrooms, whose size is 630 m^2.
- DS found that a house with more bedrooms is more expensive than the one with less bedrooms but similar size.

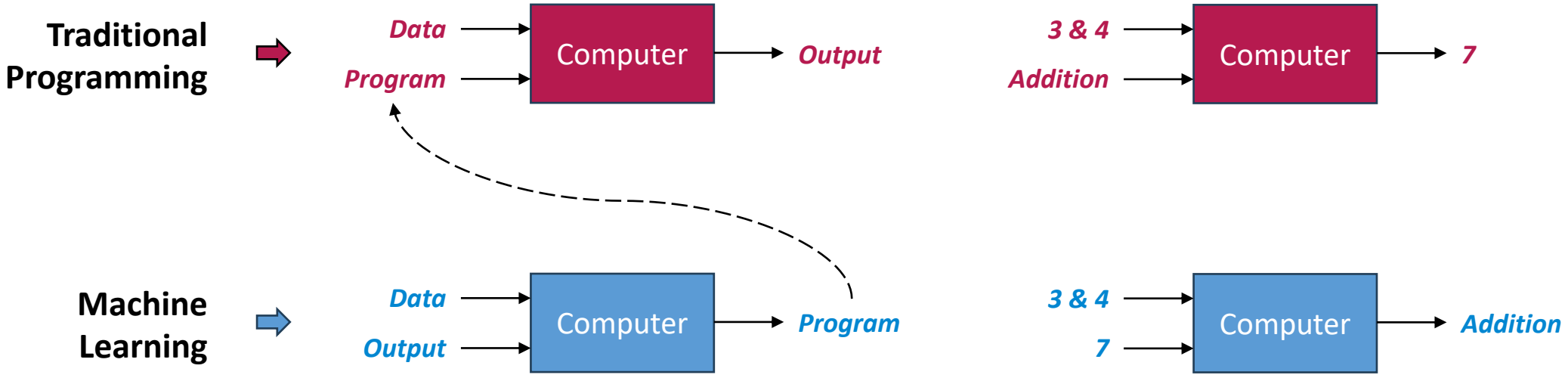
Outline

- Machine Learning & Artificial Intelligence & Data Science
- **What is Machine Learning?**
- Basic terms & Foundational concepts
- Types of Machine Learning
- Workflow of Machine Learning

What is Machine Learning?

What is Machine Learning?

- Early definition of Machine Learning
  - "The field of study that gives computers the ability to learn without being explicitly programmed." from Arthur Samuel (1959)



What is Machine Learning?

- Inspiration of Machine Learning
  - How the human brain learns certain things based on the observations it perceives from the outside world?
    - Children perceive the facts of their surroundings and gradually learn the hidden patterns of life which help them to craft logical rules to identify learned patterns.
      - Perceived fact: Cry loudly → Parents' attention
      - Learned rule: I need my parent's attention → I could cry loudly
    - Adults perceive things from the real world, process the perceived information, make rational decisions, and perform certain actions based on circumstances.

What is Machine Learning?

- Similarly,
  - ML uses specialized algorithms to uncover meaningful information and find hidden patterns from perceived data to support the rational decision-making process.
  - From a system perspective, ML is defined as the creation of automated systems that can learn hidden patterns from data to aid in making intelligent decisions.

What is Machine Learning?

- Application examples:

Historical weather data + ML ➡  Weather forecast app

Bilingual text data + ML ➡  Translation app

Conversation data + ML ➡  Chatbot

What is Machine Learning?

- ML changes the philosophy of solving problems:
  - From thinking logically and mathematically
  - To making observations about the real world, collecting data, running experiments with specific algorithms to build models, analyzing the performance of models using statistics.

- Why a business world want to use ML instead of creating hand-coded rule-based programs?

What is Machine Learning?

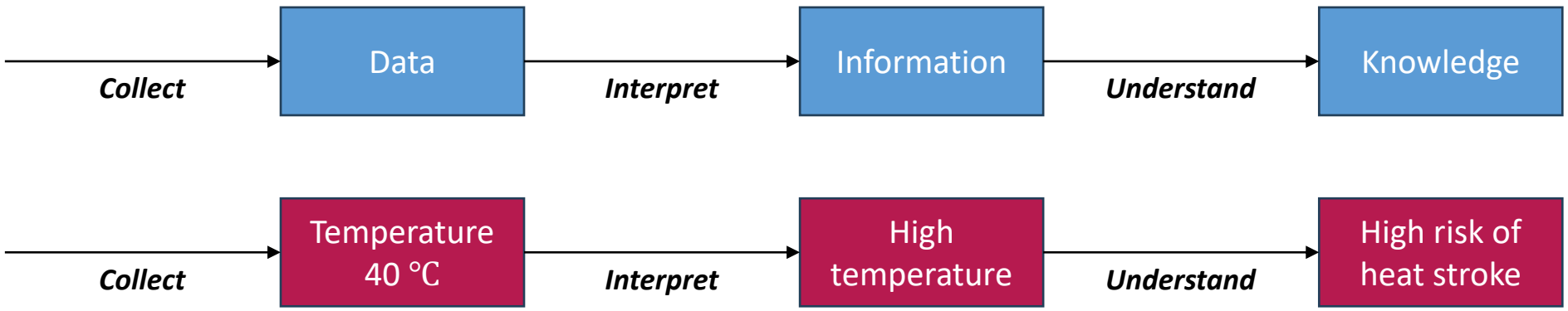
- Some practical benefits:
  - ML gives you a tool to reduce the time you spend programming.
  - ML allows you to customize your products, making them better for specific groups of people.
  - ML lets you solve problems that you, as a programmer, have no idea how to do by hand.

Outline

- Machine Learning & Artificial Intelligence & Data Science
- What is Machine Learning?
- **Basic terms & Foundational concepts**
- Types of Machine Learning
- Workflow of Machine Learning

Basic terms & Foundational concepts

- Data:
  - Data is the driving force of ML.
  - Data is a collection of values that convey information, describing the raw, unprocessed facts that may be further interpreted formally.



Basic terms & Foundational concepts

- Data:

Structured data

City	Long	Lat
Paris	2°21’E	48°51’N
London	0°7’E	51°31’N
Berlin	13°23’E	52°31’N
Beijing	116°24’E	39°54’N
Sydney	151°12’E	33°51’S

Unstructured data

Text

Image

Audio

Video

… …

Semi-structured data

… …

Basic terms & Foundational concepts

- Data:
  - Data is a collection of values which can be discrete or continuous.
    - Continuous value
      - Temperature: $-2℃$, $0℃$, $37.6℃$, …
    - Discrete value
      - Nominal value
        - Post code: 33001 (Bordeaux), 75001 (Paris), …
        - Boolean value: 0 (represent False), 1 (represent True)
      - Ordinal value
        - Academic grade: A, B, C, D, …
        - Clothes size: XXL, XL, L, M, S, XS, …

- ***The values don't need to be numerical.***
- ***To facilitate computer processing, we usually convert these values into numbers.***

Basic terms & Foundational concepts

- Data:
  - Feature
    - A feature is a measurable property of your data.
    - In ML, a feature is an ***input*** variable → x
    - A data set may contain multiple features → $x_1, x_2, \dots, x_N$
  - Label
    - A label is the ***target*** we try to predict.
    - In ML, a label is the ***output*** variable → y
  - Example
    - An example is a particular instance of data
      - Labeled example → $(x_1, x_2, \dots, x_N, y)$
      - Unlabeled example → $(x_1, x_2, \dots, x_N)$

Basic terms & Foundational concepts

- Data:
  - For a tabular data set

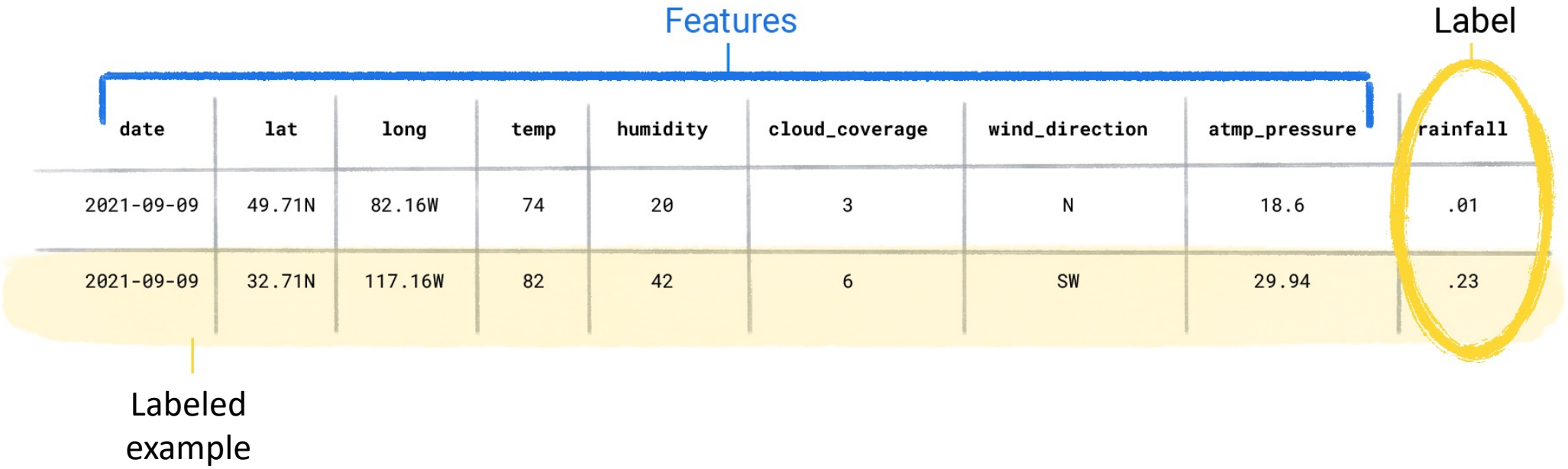


Image source: https://developers.google.com/machine-learning/intro-to-ml/supervised

Basic terms & Foundational concepts

- Data:
  - For a tabular data set

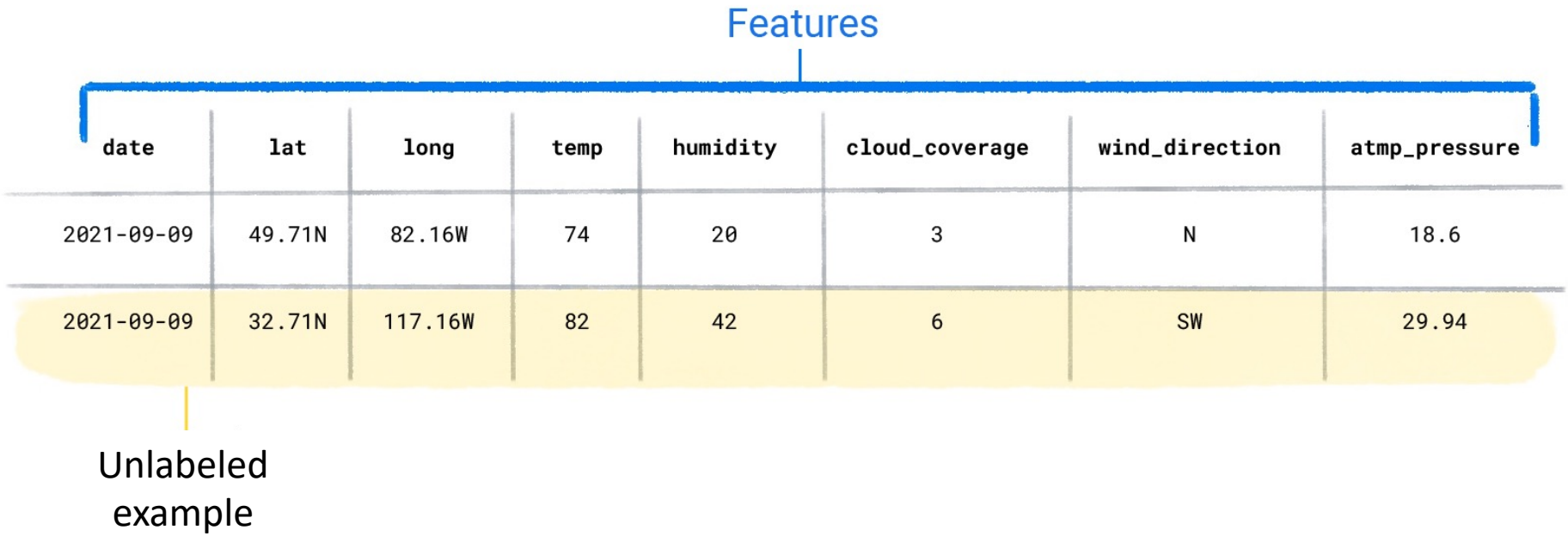


Image source: https://developers.google.com/machine-learning/intro-to-ml/supervised

Basic terms & Foundational concepts

- Data:
  - Feature
    - A data set may contain multiple features → $x_1, x_2, \dots, x_N$
  - Label
    - In ML, a label is the ***output*** variable → y
- Model
  - In ML, a model is the mathematical representation of the relationship from specific input feature patterns to specific output label values.
  - The model defines the hidden pattern learned from the data set, which can be used to make predictions on never-before-seen data.
  - A model can be represented as $\hat{y} = \mathrm{F}(x_1, x_2, \dots, x_N)$

Basic terms & Foundational concepts

- Model
  - Building
    - ***Training***
    - Evaluating
  - Inference

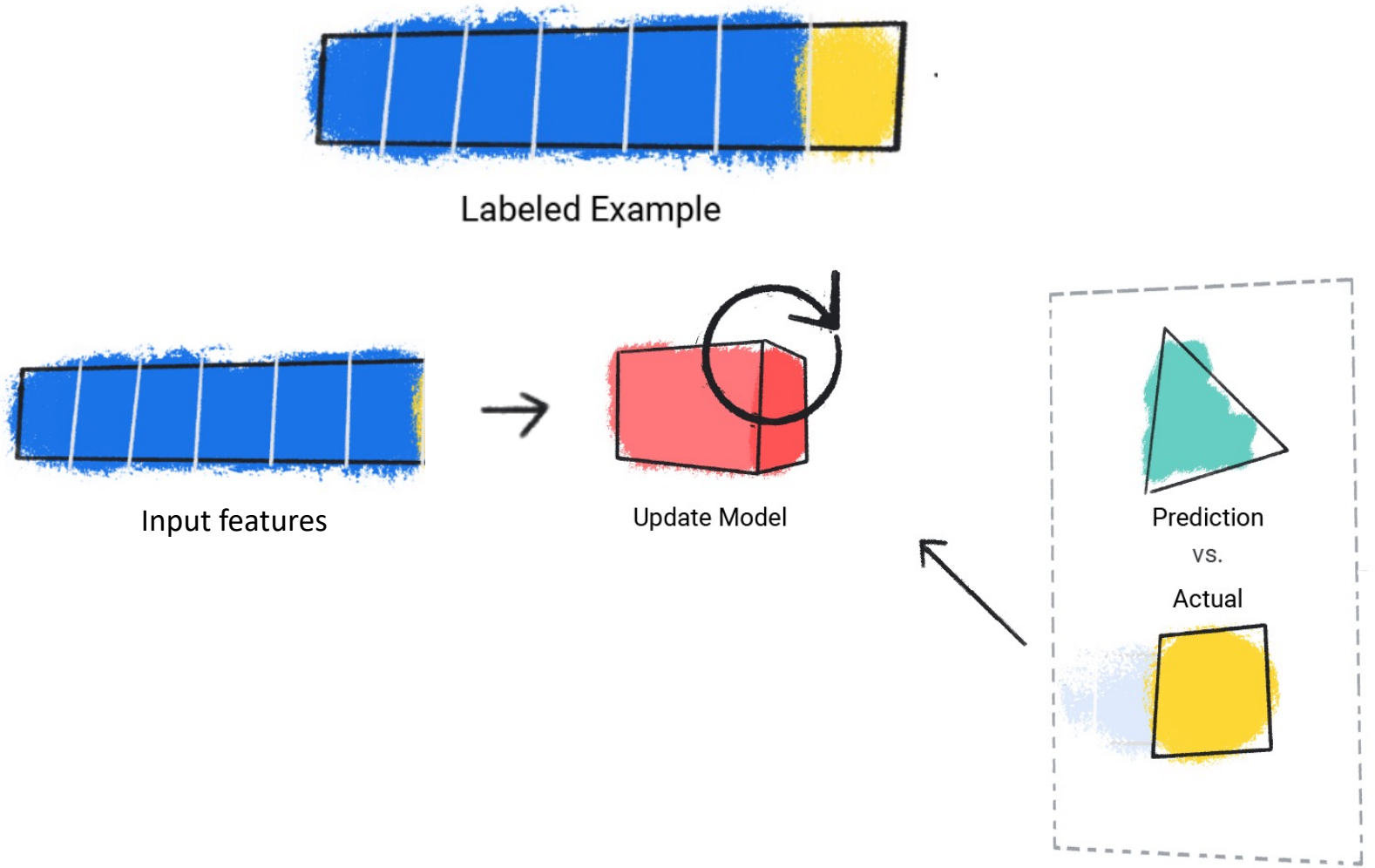
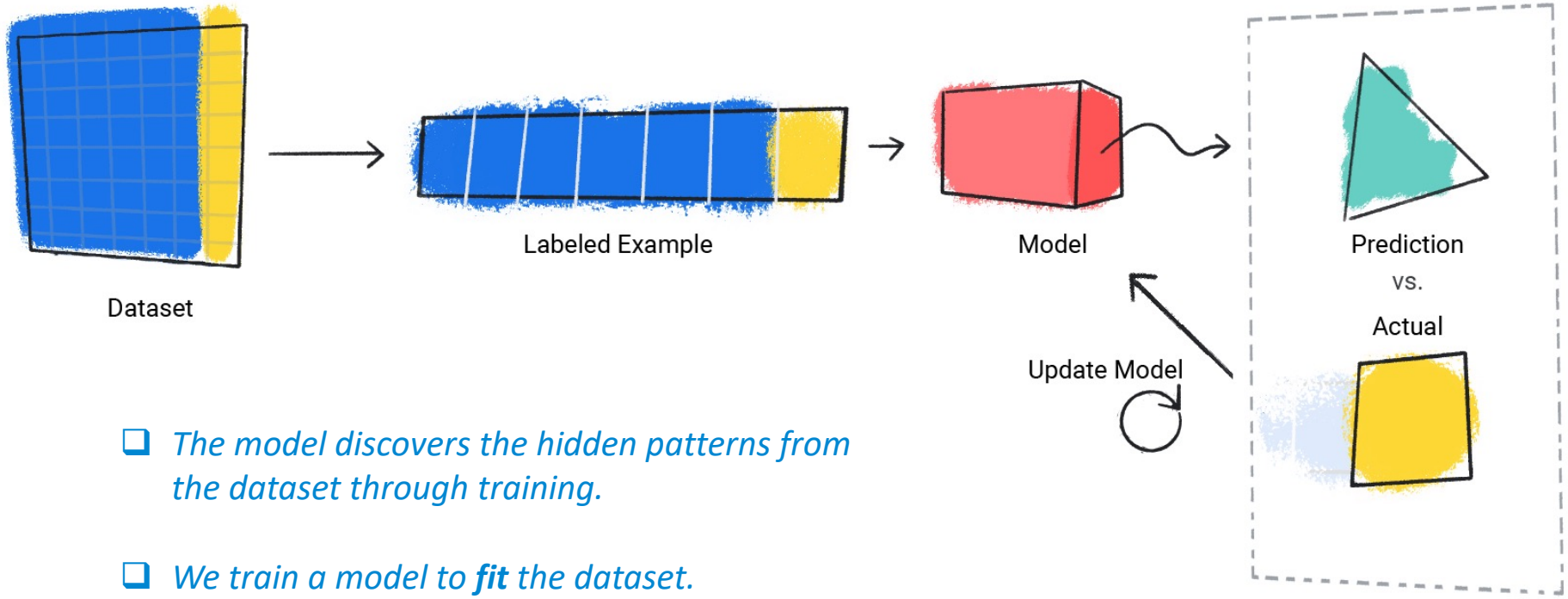


Image source: https://developers.google.com/machine-learning/intro-to-ml/supervised

Basic terms & Foundational concepts

- Model
  - Building
    - ***Training***
    - Evaluating
  - Inference



- *The model discovers the hidden patterns from the dataset through training.*
- *We train a model to **fit** the dataset.*
  - *model.fit()*

Image source: https://developers.google.com/machine-learning/intro-to-ml/supervised

Basic terms & Foundational concepts

- Model
  - Building
    - Training
    - ***Evaluating***
  - Inference

- *We evaluated a trained model to determine how well it learned.*
- *First, use a trained model to predict a value from a given dataset with labeled examples.*
- *Then, compare predicted values with actual label values to see how well they match each other.*

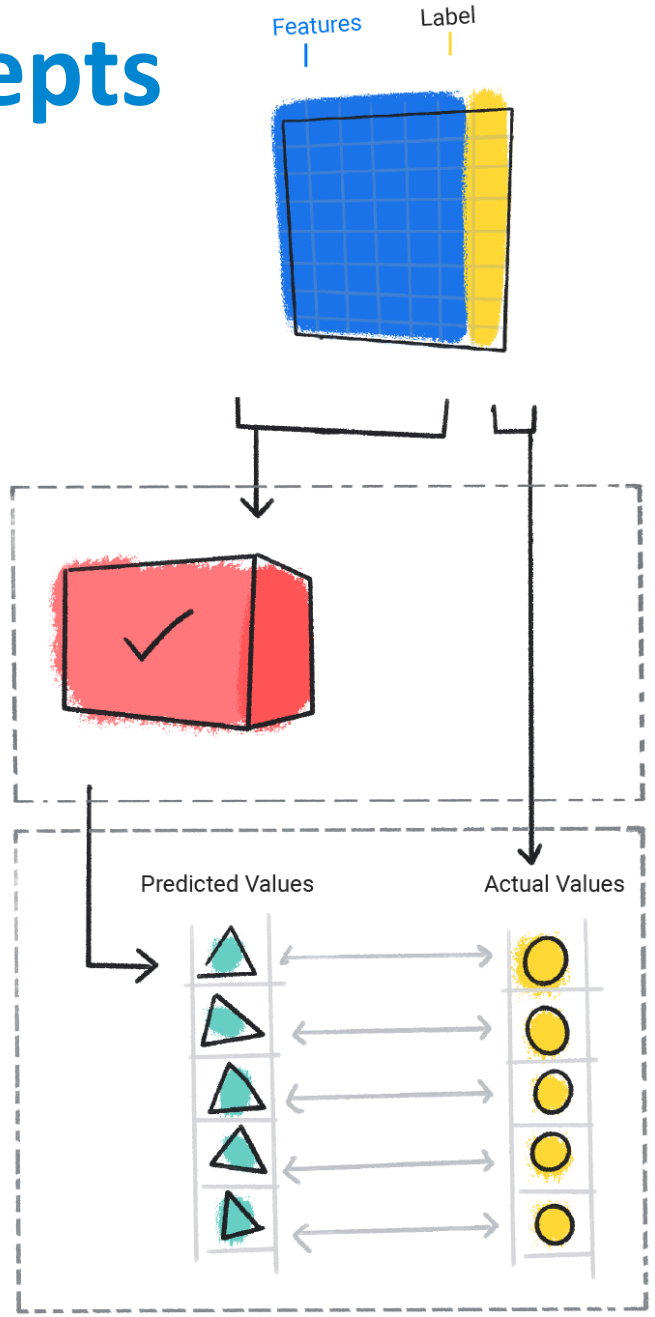


Image source: https://developers.google.com/machine-learning/intro-to-ml/supervised

Basic terms & Foundational concepts

- Model
  - Building
    - Training
    - Evaluating
  - ***Inference***
    - Once we are satisfied with the results from evaluating the model, we can use the model to make predictions, called inference, on unlabeled examples.

Outline

- Machine Learning & Artificial Intelligence & Data Science
- What is Machine Learning?
- Basic terms & Foundational concepts
- **Types of Machine Learning**
- Workflow of Machine Learning

Types of Machine Learning

- Machine learning approaches are traditionally divided into three broad categories, which corresponding to **learning paradigms**.

Supervised learning

Unsupervised learning

Reinforcement learning

Types of Machine Learning

- Supervised learning
  - The computer is presented with example input features and their desire output labels, given by a “supervisor”, and the goal is to learn a general rule that maps input features to output labels.
  - Supervised learning models can make predictions after seeing lots of data with the correct answer and then discovering the connections between the features in the data that product the correct answers.
  - These ML systems are "supervised" in the sense that a human gives the ML system data with the known correct results.

Types of Machine Learning

- Supervised learning
  - According to the prediction tasks, supervised learning models can be divided into the following sub-categories:
    - Regression model:
      - Predict a numeric value
      - For example, estimate the price of a house

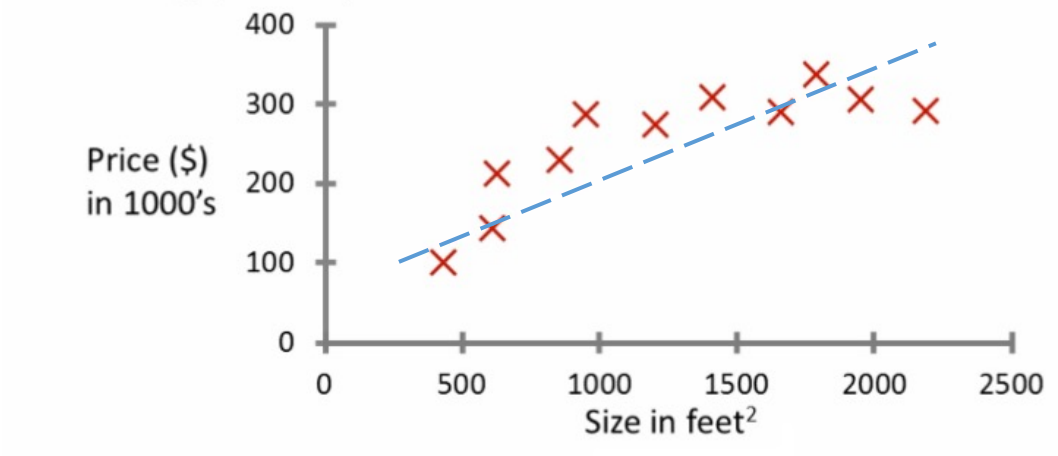


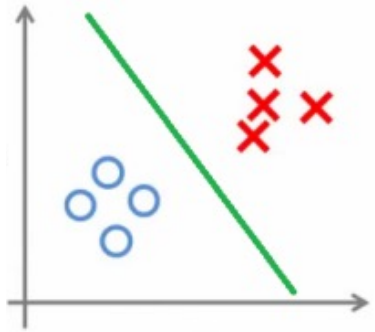
Image source: https://medium.com/@humantsattabhayya/all-about-supervised-learning-3ef9cd5efe06

Types of Machine Learning

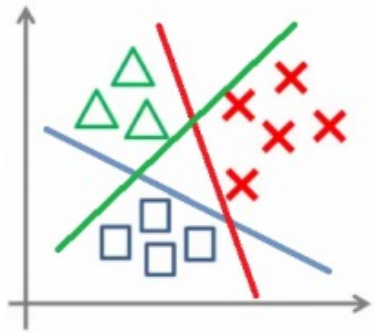
- Supervised learning
  - According to the prediction tasks, supervised learning models can be divided into the following sub-categories:
    - Regression model:
      - Predict a numeric value
      - For example, estimate the price of a house
    - Classification model:
      - Predict the likelihood that something belongs to a category.
      - The predicted value states whether or not something belongs to a particular category.
      - For example, determine if an email is spam

Types of Machine Learning

- Supervised learning
  - According to the prediction tasks, supervised learning models can be divided into the following sub-categories:
    - Regression model:
    - Classification model:
      - Binary classification model
        - There are only 2 classes.
        - *'Dog'* or '*Cat'*
        - '*Dog*' or '*Not dog*'
      - Multi-class classification model
        - There are more than 2 classes.
        - *'Dog'* or *'Cat'* or *'Horse'* or *'Sheep'*



Binary

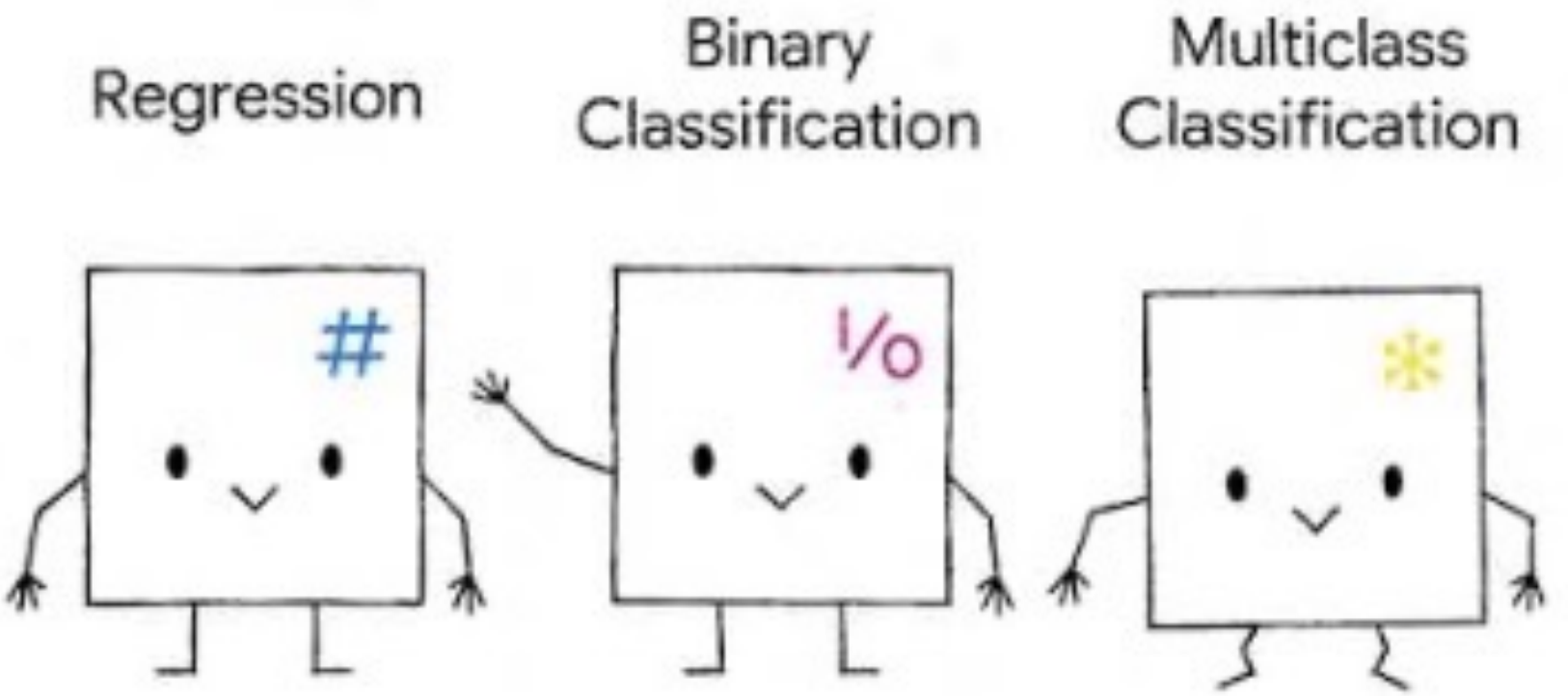


3 classes

KEDGE BUSINESS SCHOOL

Image source: https://medium.com/analytics-vidhya/common-loss-functions-in-machine-learning-for-classification-model-931cbf564d42

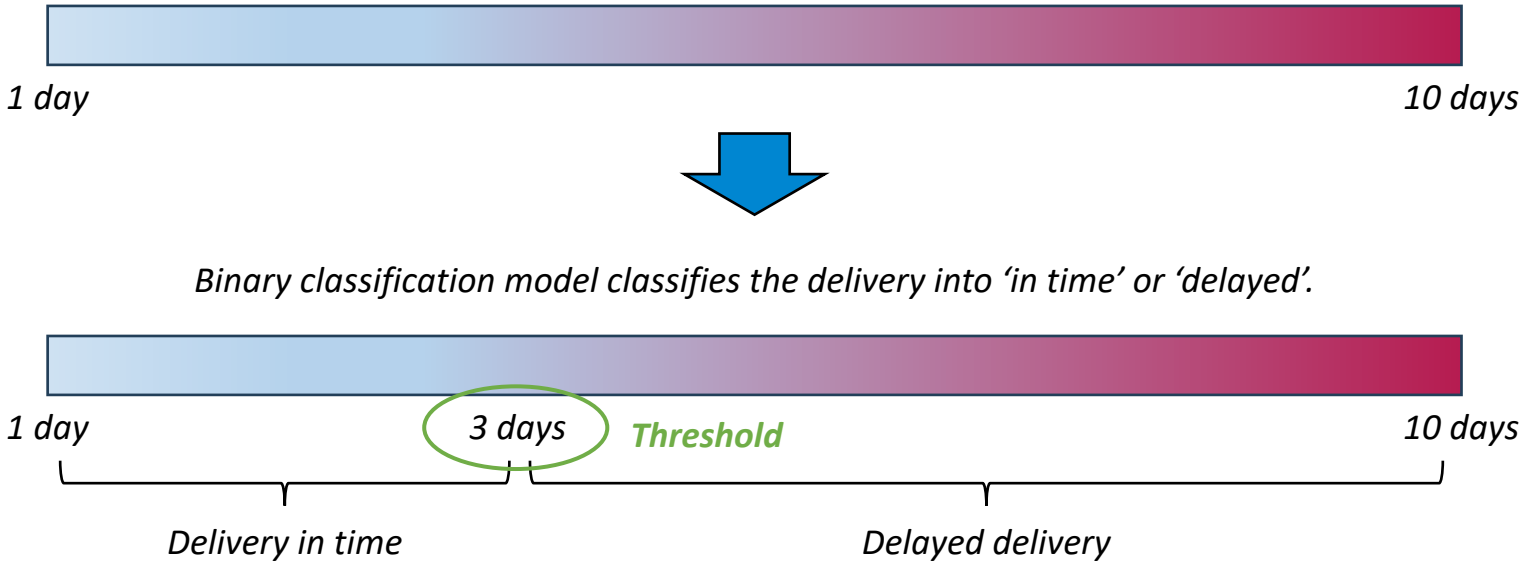
Types of Machine Learning



Types of Machine Learning

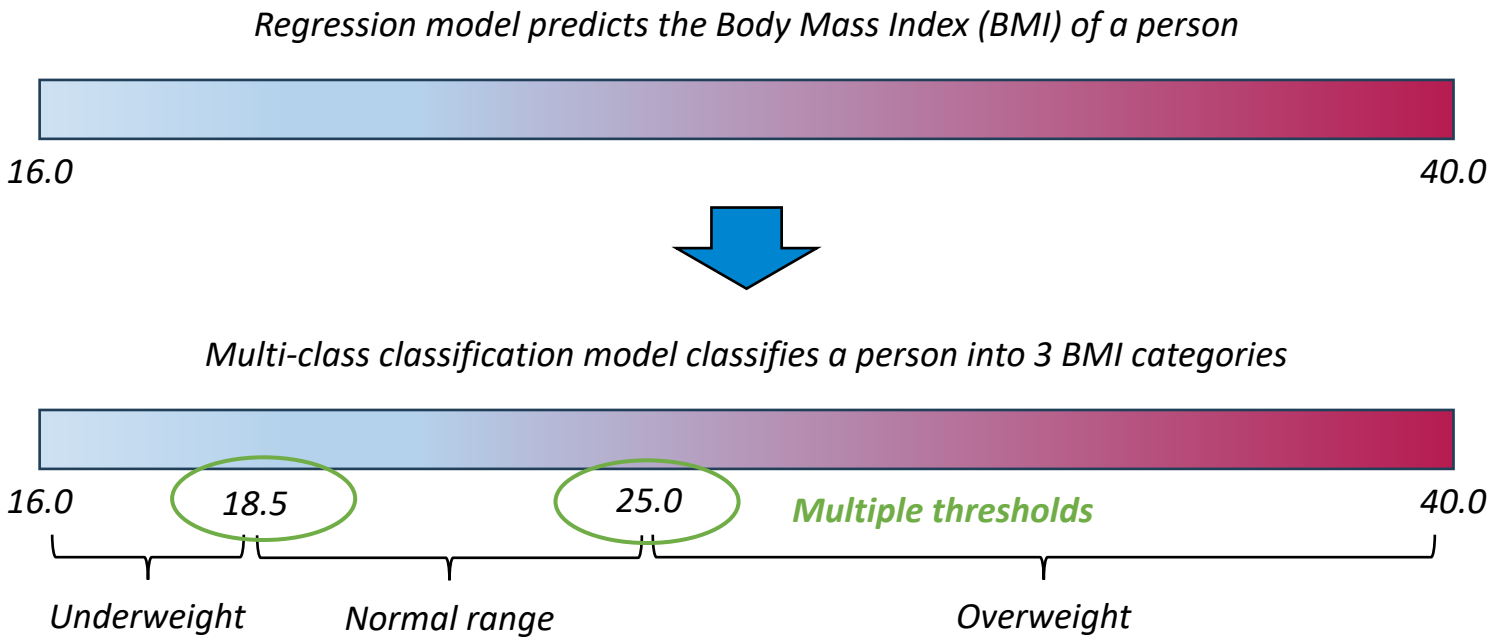
- Supervised learning
  - Some prediction tasks can be converted into others.
    - Regression → Classification

Regression model predicts the lead time (in days) for a delivery.



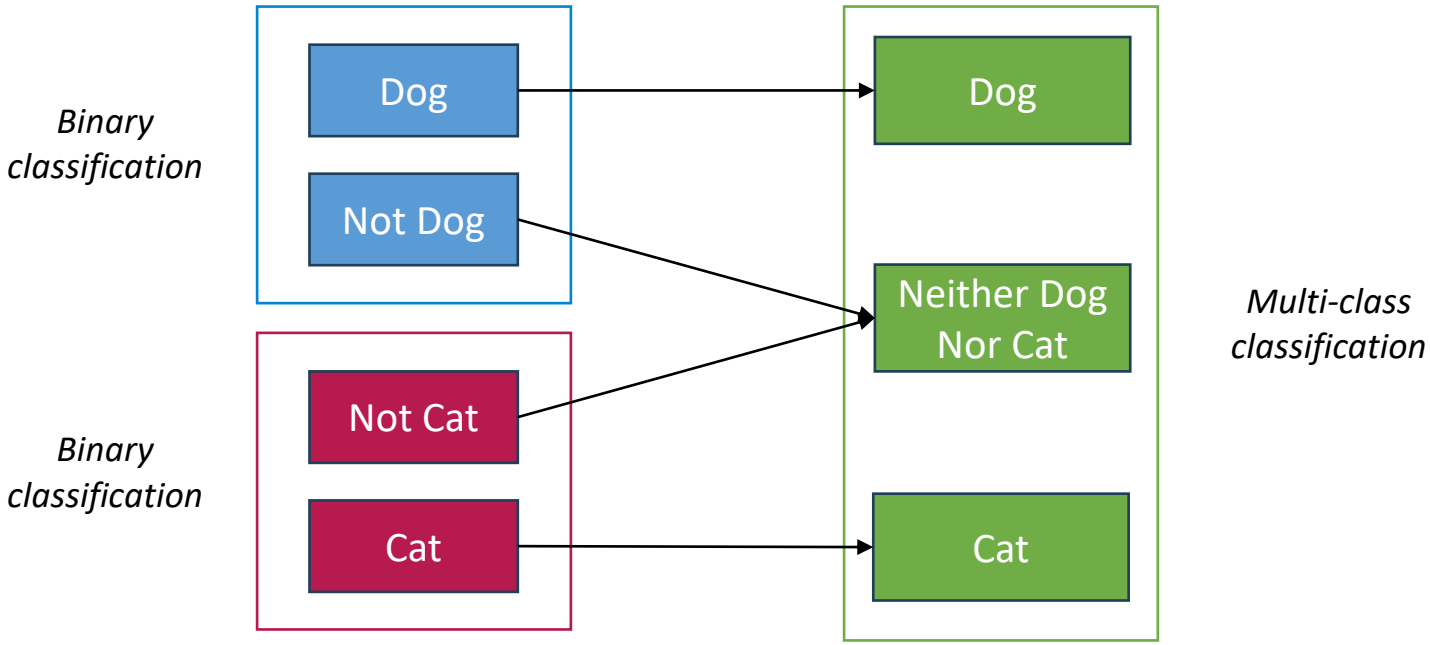
Types of Machine Learning

- Supervised learning
  - Some prediction tasks can be converted into others.
    - Regression → Classification



Types of Machine Learning

- Supervised learning
  - Some prediction tasks can be converted into others.
    - Binary classification → Multi-class classification



Types of Machine Learning

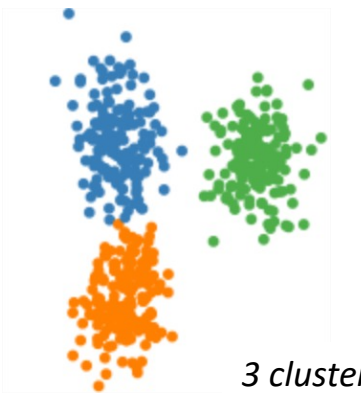
- Unsupervised learning
  - Unsupervised learning algorithms take a set of data that contains only features, and find structure in the data, like grouping or clustering of data points.
  - An unsupervised learning model's goal is to identify meaningful patterns in an unlabeled dataset, without knowing any correct answers.

Types of Machine Learning

- Unsupervised learning
  - Clustering
    - A commonly used unsupervised learning technique
    - Group data points into different natural clusters

2 clusters

2 clusters



3 clusters

3 clusters

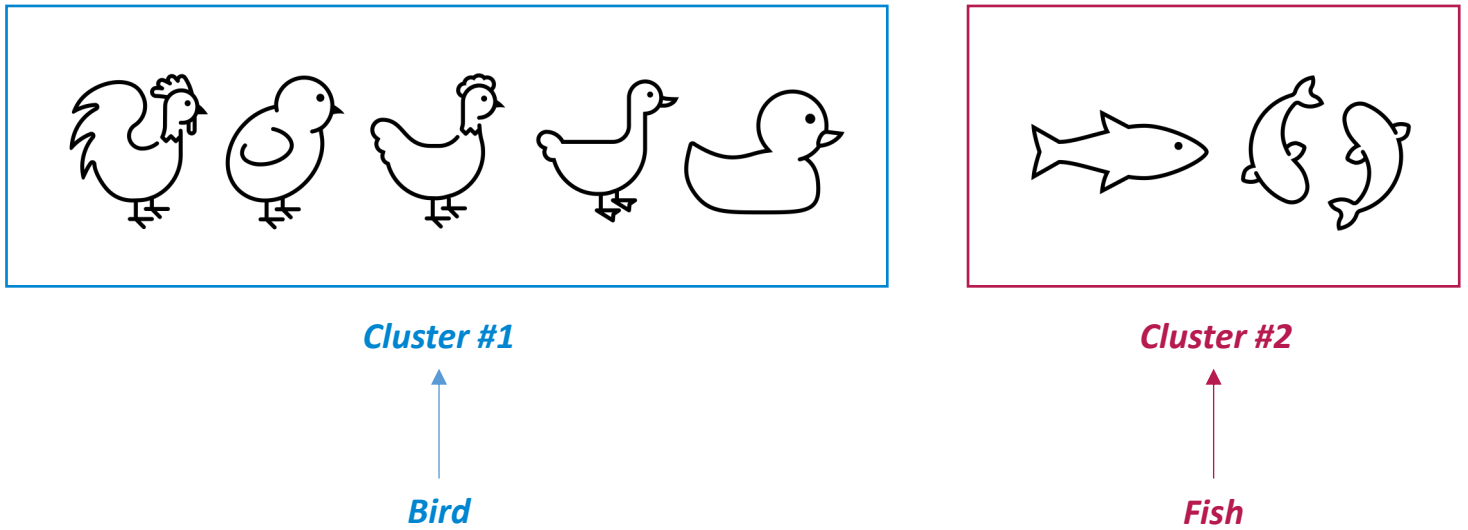
Image source: https://scikit-learn.org/stable/modules/clustering.html

Types of Machine Learning

- Unsupervised learning
  - Clustering vs. Classification
    - Clustering and classification both assign a category (cluster or class) to a data point.
    - There is a correct answer for classification, but not for clustering.
    - The classes are defined by human for classification model, but the clusters are not.
      - You may know the number of clusters in the dataset, but you don’t know what are they.
      - After the clustering algorithm outputs the resulting clusters, you may attempt to name those clusters based on your understanding.

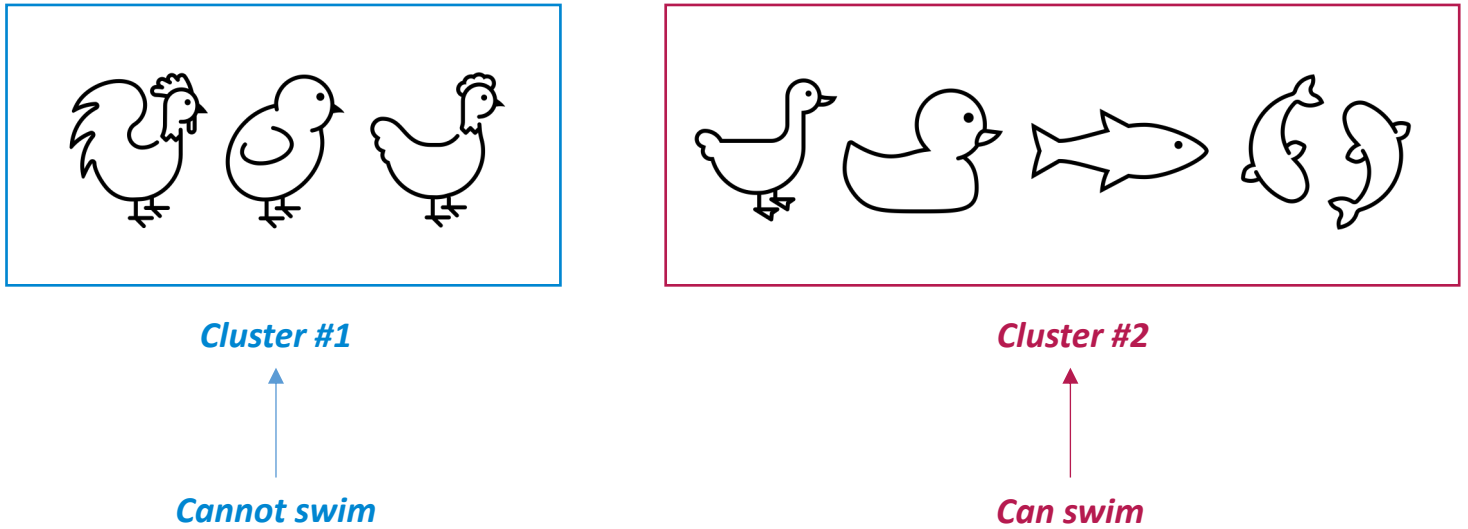
Types of Machine Learning

- Unsupervised learning
  - Clustering



Types of Machine Learning

- Unsupervised learning
  - Clustering



Types of Machine Learning

- 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.
  - The outputs of a reinforcement learning model is a series of actions to be taken in an environment to maximize reward.

Types of Machine Learning

- Reinforcement learning

RL system	Function	Output actions	Rewards
Robot in a maze	Navigate in the maze	Turn right / left	Get closer to the exit of the maze
AlphaGo	Play the game of Go	Next move in Go	Win the game of Go

Outline

- Machine Learning & Artificial Intelligence & Data Science
- What is Machine Learning?
- Basic terms & Foundational concepts
- Types of Machine Learning
- **Workflow of Machine Learning**

Workflow of Machine Learning

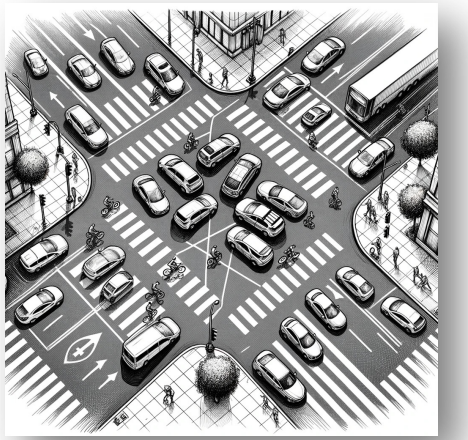
- For a classical supervised learning process:
  1) Frame problem
  2) Collect and prepare data
  3) Build the model
  4) Deploy the model

Workflow of Machine Learning

- For a classical supervised learning process:
  1) ***Frame problem***
     - First, determine the problem you want to solve
       - E.g., detect spam emails, forecast future weather, cancer diagnosis …
     - Then, decide if ML is the right approach for solving this problem
       - Is the problem too simple that can be solved by a hand-coded rule-based program?
       - Is the problem too complex that ML cannot learn the hidden pattern?
       - Is there any available data for developing ML models?
     - If so, frame the problem using ML terms
       - Regression or binary classification or multi-class classification?
       - What are the features? What is the label to predict?

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***2) Collect and prepare data***
    - Collect data
      - Be aware of the sources of data
        - Is the data source reliable?
        - Is there any inherent biases in the data?
    - Prepare data



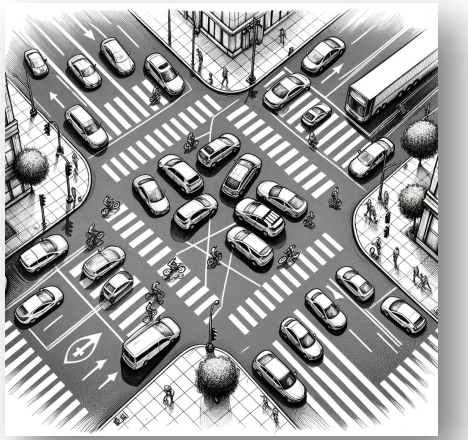
Data reflect the traffic jam

Official alert → *More reliable*

Crowdsourcing alert → *Less reliable*

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***2) Collect and prepare data***
    - Collect data
      - Be aware of the sources of data
        - Is the data source reliable?
        - Is there any inherent biases in the data?
    - Prepare data



Data reflect the traffic jam

Official alert → *Poor timeliness, less reliable*

Crowdsourcing alert → *Good timeliness, more reliable*

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***2) Collect and prepare data***
    - Collect data
      - Be aware of the sources of data
        - Is the data source reliable?
        - Is there any inherent biases in the data?
    - Prepare data

KFC customer satisfaction survey

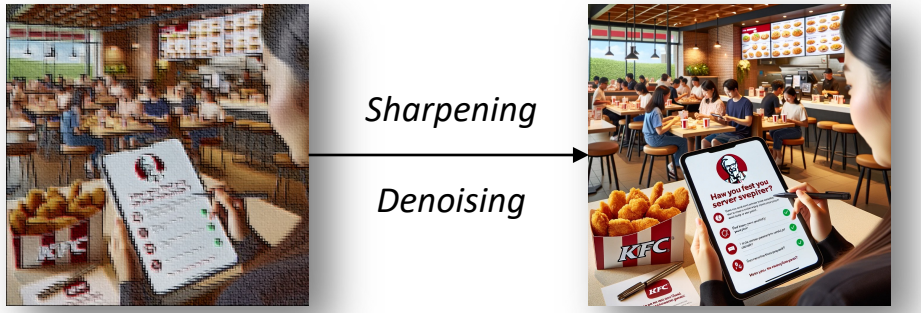
- *Only cover the on-site customers*
- *Ignore the takeout customers*

Workflow of Machine Learning

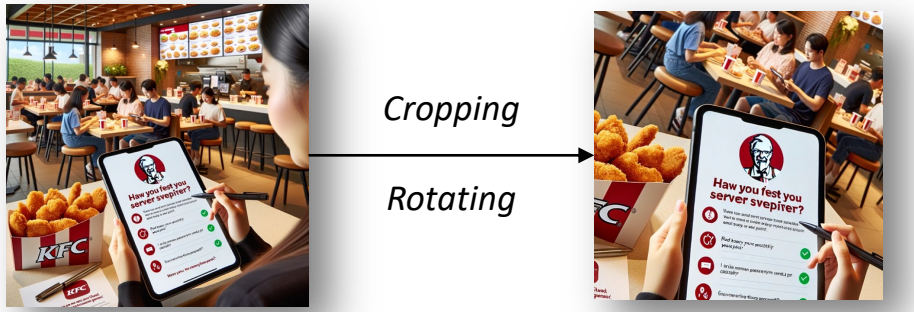
- For a classical supervised learning process:
  - ***2) Collect and prepare data***
    - Collect data
    - Prepare data
      - Normalize data from diverse sources
        - E.g., customer satisfaction survey
          - Source #1: ★★☆
          - Source #2: 1 to 10 points
          - Normalize both source to the range from -1 to 1

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***2) Collect and prepare data***
    - Collect data
    - Prepare data
      - Normalize data from diverse sources
      - Improve the data's quality & quantity
        - E.g., for an image



Improve the quality



Improve the quantity

Workflow of Machine Learning

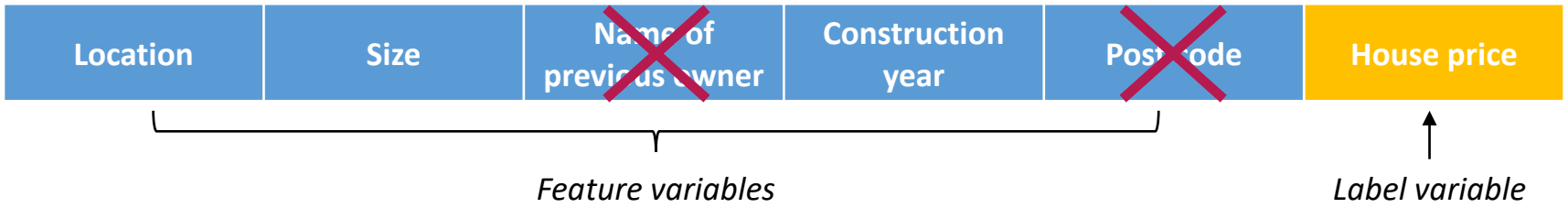
- For a classical supervised learning process:
  - ***2) Collect and prepare data***
    - Collect data
    - Prepare data
      - Normalize data from diverse sources
      - Improve the data's quality & quantity
      - Clean and edit data
        - Delete examples with missing values
        - Edit examples with wrong values
        - …

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***2) Collect and prepare data***
    - Collect data
    - Prepare data
      - Normalize data from diverse sources
      - Improve the data's quality & quantity
      - Clean and edit data
      - Randomly shuffle the data if needed

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***2) Collect and prepare data***
    - Collect data
    - Prepare data
      - Normalize data from diverse sources
      - Improve the data's quality & quantity
      - Clean and edit data
      - Randomly shuffle the data if needed
      - Decide what are the feature variables and which is the label variable
        - You may not need all the features for your model



Workflow of Machine Learning

- For a classical supervised learning process:
  - ***2) Collect and prepare data***
    - Collect data
    - Prepare data
      - Normalize data from diverse sources
      - Improve the data's quality & quantity
      - Clean and edit data
      - Randomly shuffle the data if needed
      - Decide what are the feature variables and which is the label variable
      - Split the dataset into:
        - Training dataset: Fit the model to the training dataset during training.
        - Testing dataset: Use the testing dataset to evaluate the trained model.

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***2) Collect and prepare data***
    - Collect data
    - Prepare data
      - Normalize data from diverse sources
      - Improve the data's quality & quantity
      - Clean and edit data
      - Randomly shuffle the data if needed
      - Decide what are the feature variables and which is the label variable
      - Split the dataset into training and testing dataset
      - Visualize your data always help you understand it

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***3) Build the model***
    - Choose a training method / algorithms

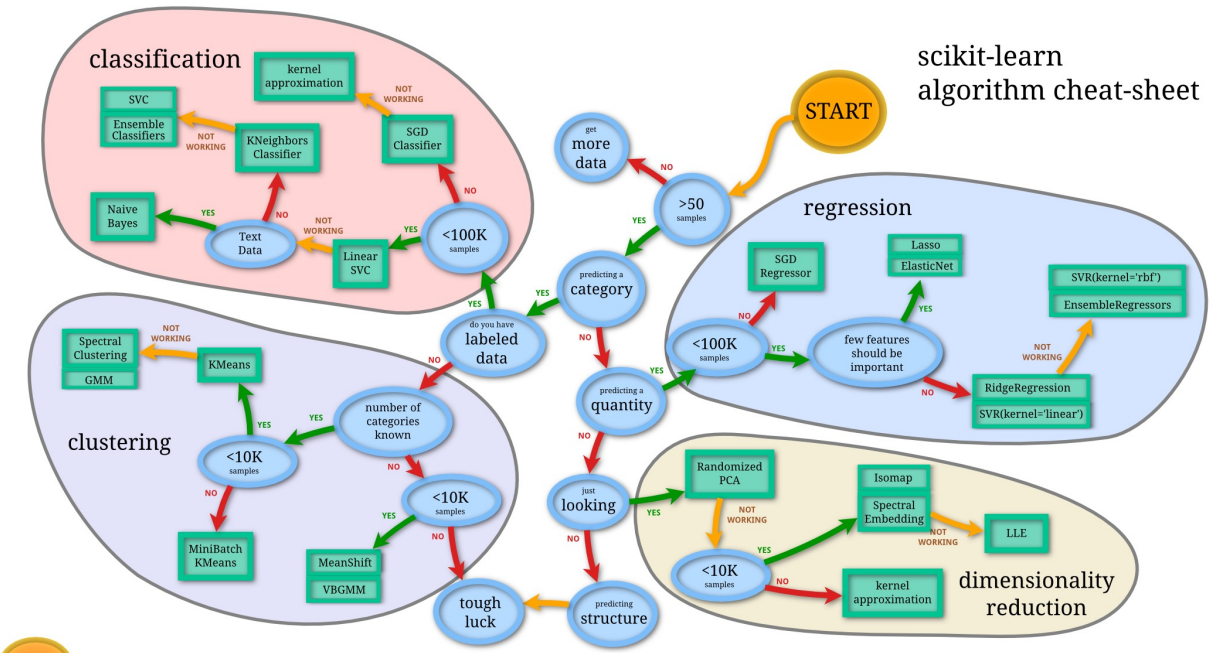


Image source: https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***3) Build the model***
    - Choose a training method / algorithms
    - Train the model
      - Use the training dataset
      - model.fit()

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***3) Build the model***
    - Choose a training method / algorithms
    - Train the model
    - Evaluate the model
      - Use the testing dataset
      - Use metrics to tell how well the model learned

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***4) Deploy the model***
    - Deploy the model when you are satisfied with the evaluation results of the model
    - Deploying a ML model to the business means using completely new data to test the model's performance
      - Model monitoring
        - Is there performance degradation?
        - What are the causes?
          - The new data might be different from the training data.
            - E.g., for a house price estimation model
              - Training data only covers the houses.
              - Now users are using it to estimate the price of apartments.

Workflow of Machine Learning

- For a classical supervised learning process:
  - ***4) Deploy the model***
    - Deploy the model when you are satisfied with the evaluation results of the model
    - Deploying a ML model to the business means using completely new data to test the model's performance
      - Model monitoring
      - Model adaptation
        - Retrain the model to adapt it to the newest data.

Workflow of Machine Learning

- For a classical supervised learning process:
  1) Frame problem
  2) Collect and prepare data
  3) Build the model
  4) Deploy the model

Hands-on Exercises

- NumPy Recap
- Pandas Recap