COMP5212: Machine Learning

Lecture 0

Minhao Cheng

Course information

Basic

- Website: https://cse.hkust.edu.hk/~minhaocheng/teaching/comp5212f22.html
- My Office: Room 2542
- Office Hours: Tuesday 13:00-14:30 @ Room 2542
- TA: Mingxuan Fan, Shuowei Cai
- Reference:
    - “Deep Learning” (by Goodfellow, Bengio, Courville)

Course information

Syllabus (tentative)

- Part I
    - Math basics
    - Linear models(regression, classification, clustering)
    - Optimization
    - Learning theory
- Part II
    - Kernel methods
    - Tree-based methods
    - Neural network
- Part III
    - Advanced topic in machine learning
        - AutoML
        - Trustworthy machine learning
        - …

Course information

Grading policy

- Homework (40%)
  - 3 Written
  - 2 Programming
- Term project (35%)
- Final exam (25%)

Course information

Term project

- Group of at most 4 students
- Open research projects:
  - Solve an interesting problem
  - Develop a new algorithm
  - Compare state-or-the-art algorithms on some problems
  - …
- Feel free to discuss with me either by email or in the office hour

Machine Learning: Overview

Machine learning overview

From learning to machine learning

- What is learning?
  - Observation ⟶ Learning ⟶ Skill
- Skill: how to make decision (action)
  - Classify an image
  - Translate a sentence from one language to another
  - Learn to play a game
  - …

Machine learning overview

From learning to machine learning

- Human learning

Machine learning overview

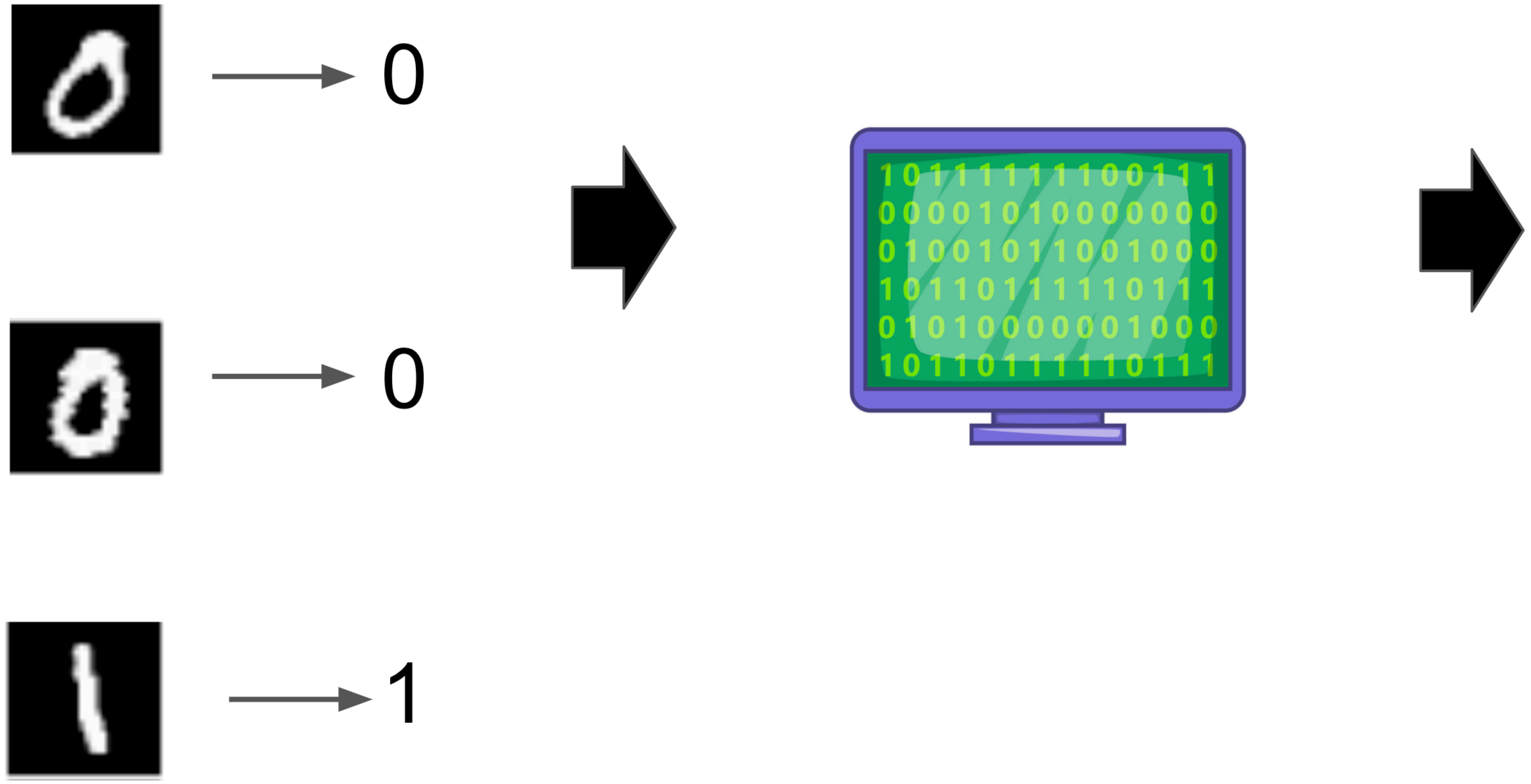
From learning to machine learning

- What is learning?
    - Observation $\longrightarrow$ Learning $\longrightarrow$ Skill
- Skill: how to make decision (action)
    - Classify an image
    - Translate a sentence from one language to another
    - Learn to play a game
    - …
- Machine learning: (Automatic the learning process)
    - Data $\longrightarrow$ Machine Learning $\longrightarrow$ Skill (decision rules)

Machine learning overview

Machine learning

Training Data **Machine Learning** **Decision rule**



Machine learning overview

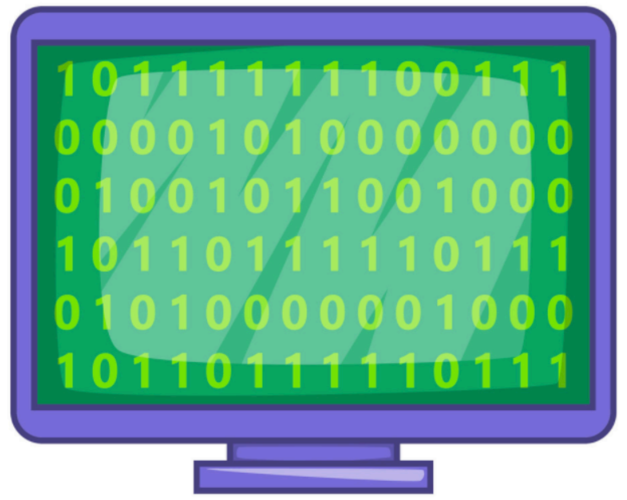
Machine learning

Training Data | Machine Learning | Decision rule

$x_1 \rightarrow 0$ (y_1)

$x_2 \rightarrow 0$ (y_2)

$x_3 \rightarrow 1$ (y_3)

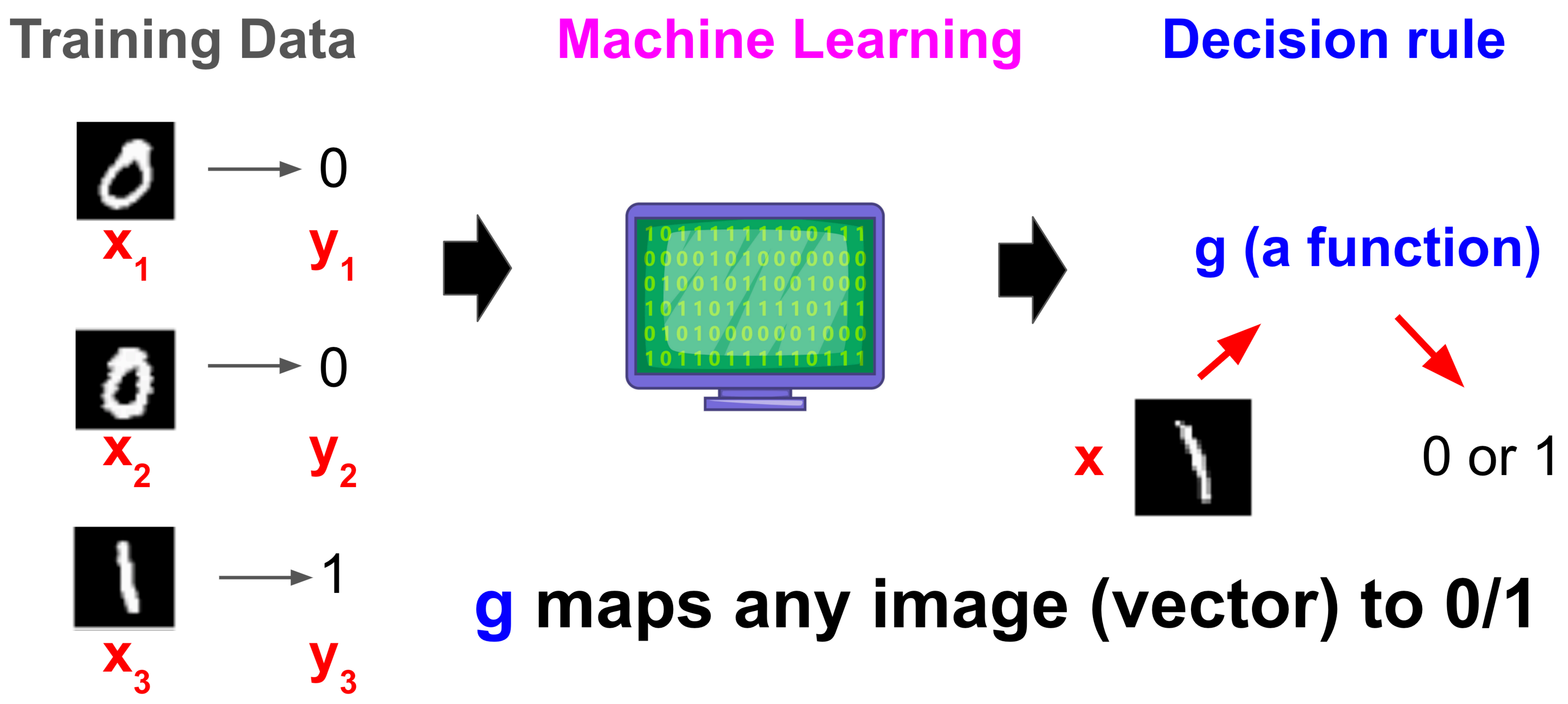


x_1: vector of pixel values [0, 24, 128, ...]

y_1: 0 or 1

Machine learning overview

Machine learning



Machine Learning

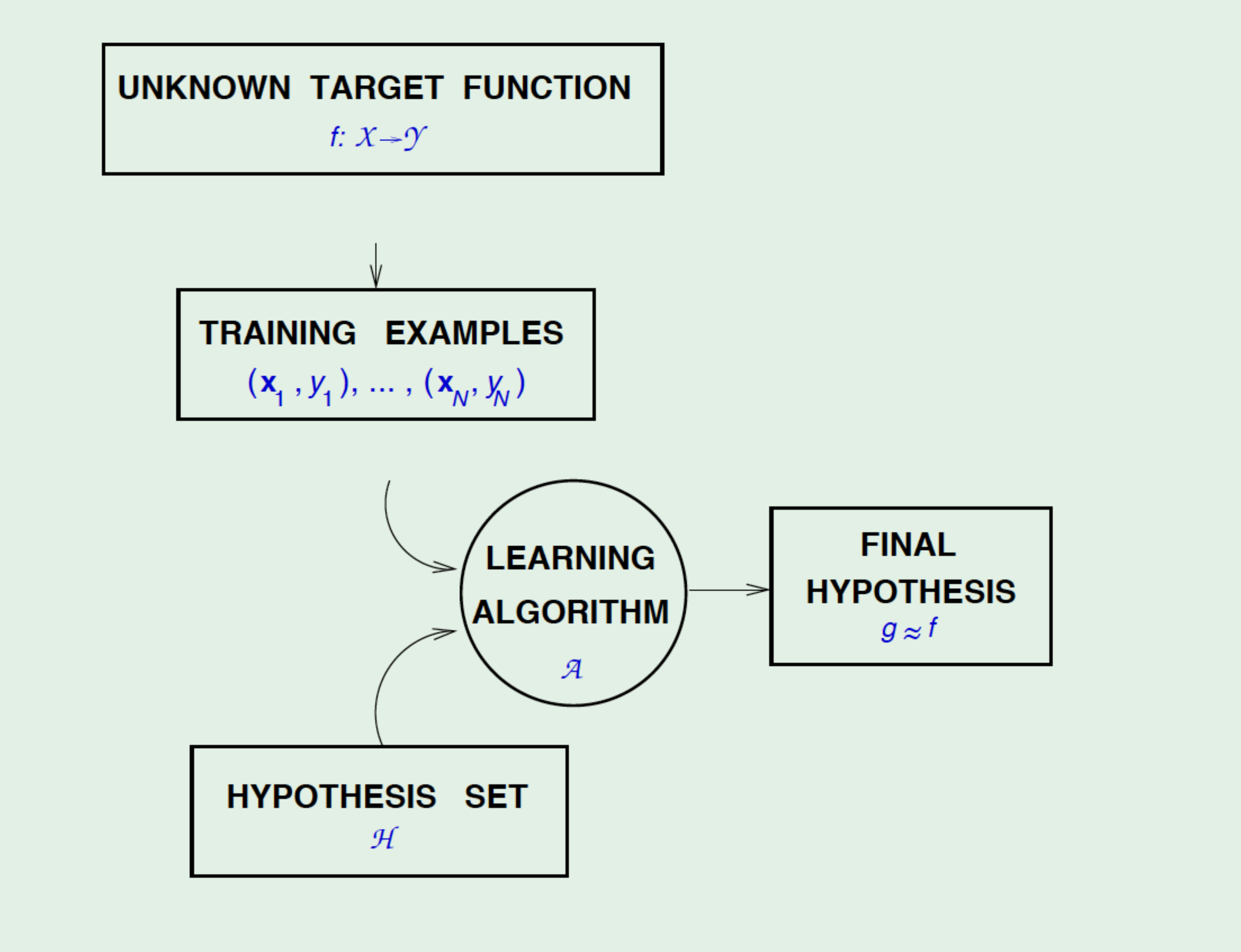
Formalization

- Input: $x \in \mathcal{X}$
- Output: $y \in \mathcal{Y}$
- Target function to be learned:
  - $f : \mathcal{X} \to \mathcal{Y}$ (ideal image classification function)
- Data:
  - $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \ldots, (X_N, y_N)\}$
- Hypothesis (model)
  - $g : \mathcal{X} \to \mathcal{Y}$ (Learned formula to be used)

$\{(\mathbf{x}_n, y_n)\}$ from $f \longrightarrow \boxed{\text{ML}} \longrightarrow g$

Machine Learning

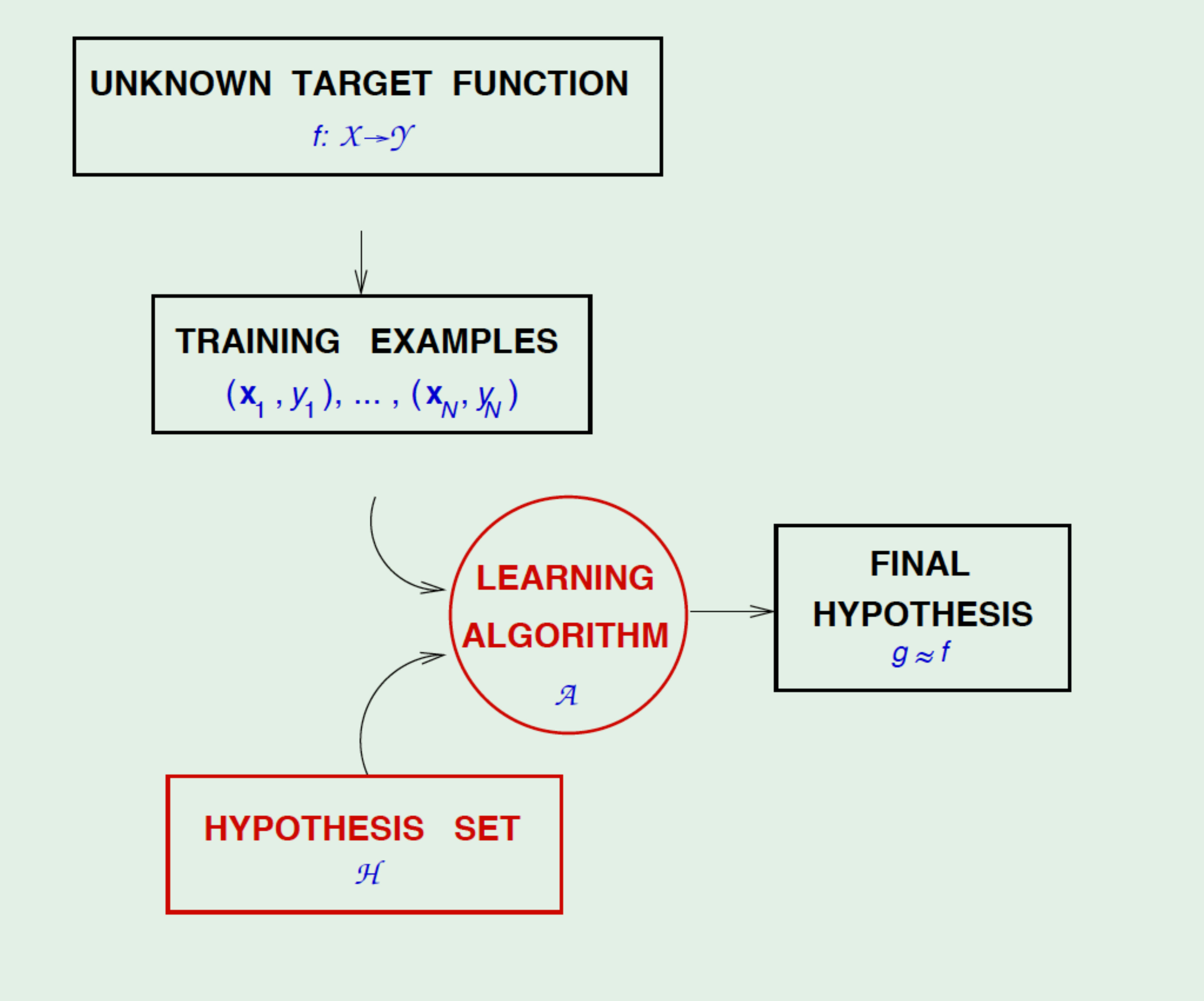
Basic setup of learning problem



Machine Learning

Learning model

- A learning model has two components:
  - The hypothesis set $\mathscr{H}$:
    - Set of candidate hypothesis (functions)
  - The learning algorithm:
    - To pick a hypothesis (function) from the $\mathscr{H}$
    - Usually optimization algorithm (choose the best function to minimize the training error)



Machine learning

Binary classification

- Data:
  - Feature for each training example: $\{x_n\}_{n=1}^{N}$, each $x_n \in \mathbb{R}^d$
  - Labels for each training example: $y_n \in \{+1, -1\}$
- Goal: learn a function
- Examples:
  - Credit: approve/disapprove
  - Email: spam/not spam
  - Patient: sick/not sick
  - …

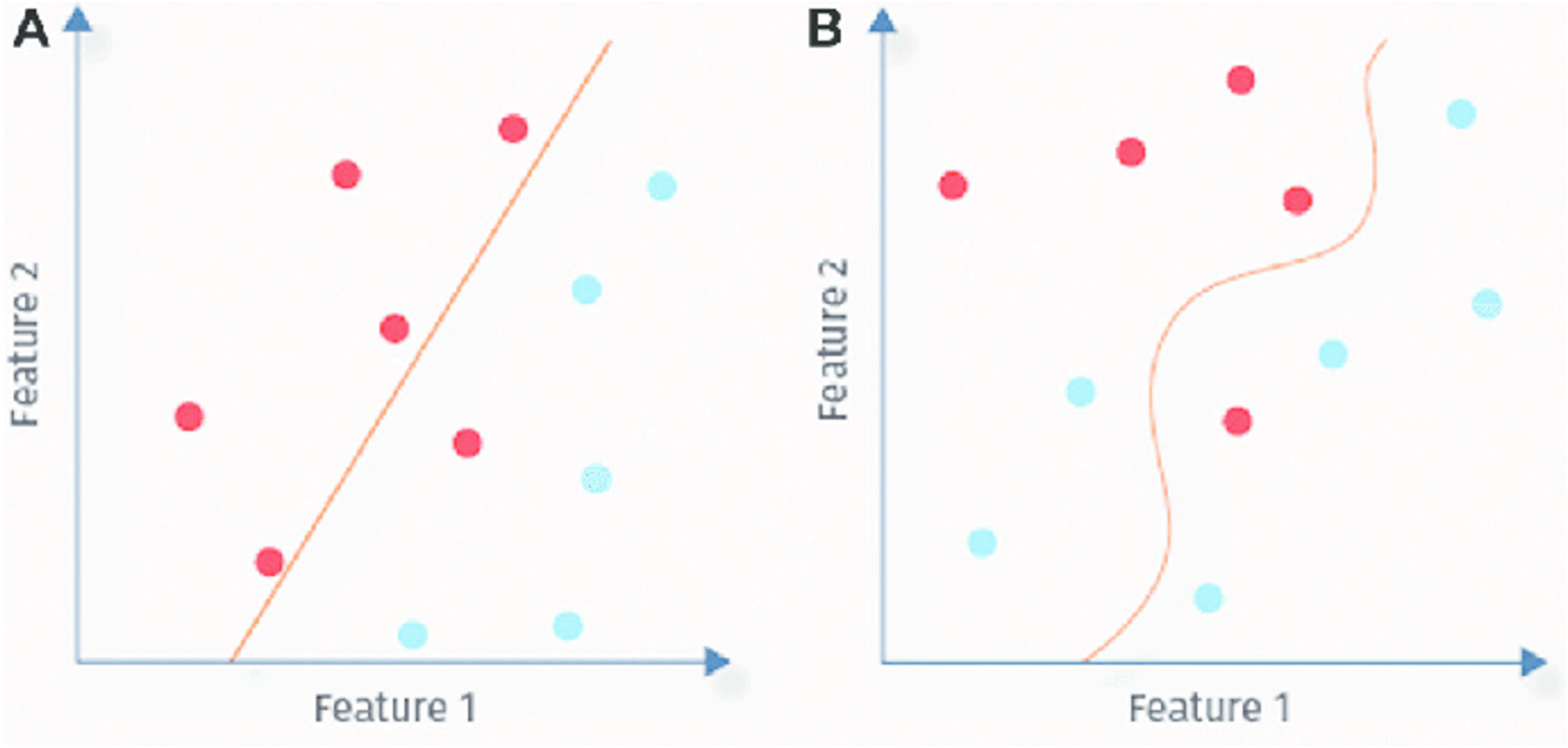
Machine learning

Types of hypothesis

- Linear hypothesis space
  - $h(x) = \text{sign}(\sum_{i=1}^{d} \boxed{w_i} x_i - \boxed{\text{threshold}})$
- Feed forward (fully connected) network:
  - $h(x) = \text{sign}(\boxed{W_L} \ldots \sigma(\boxed{W_2} \sigma(\boxed{W_1} x + \boxed{b_1}) + \boxed{b_2}) + \boxed{b_L})$
- Tree-based models
- …

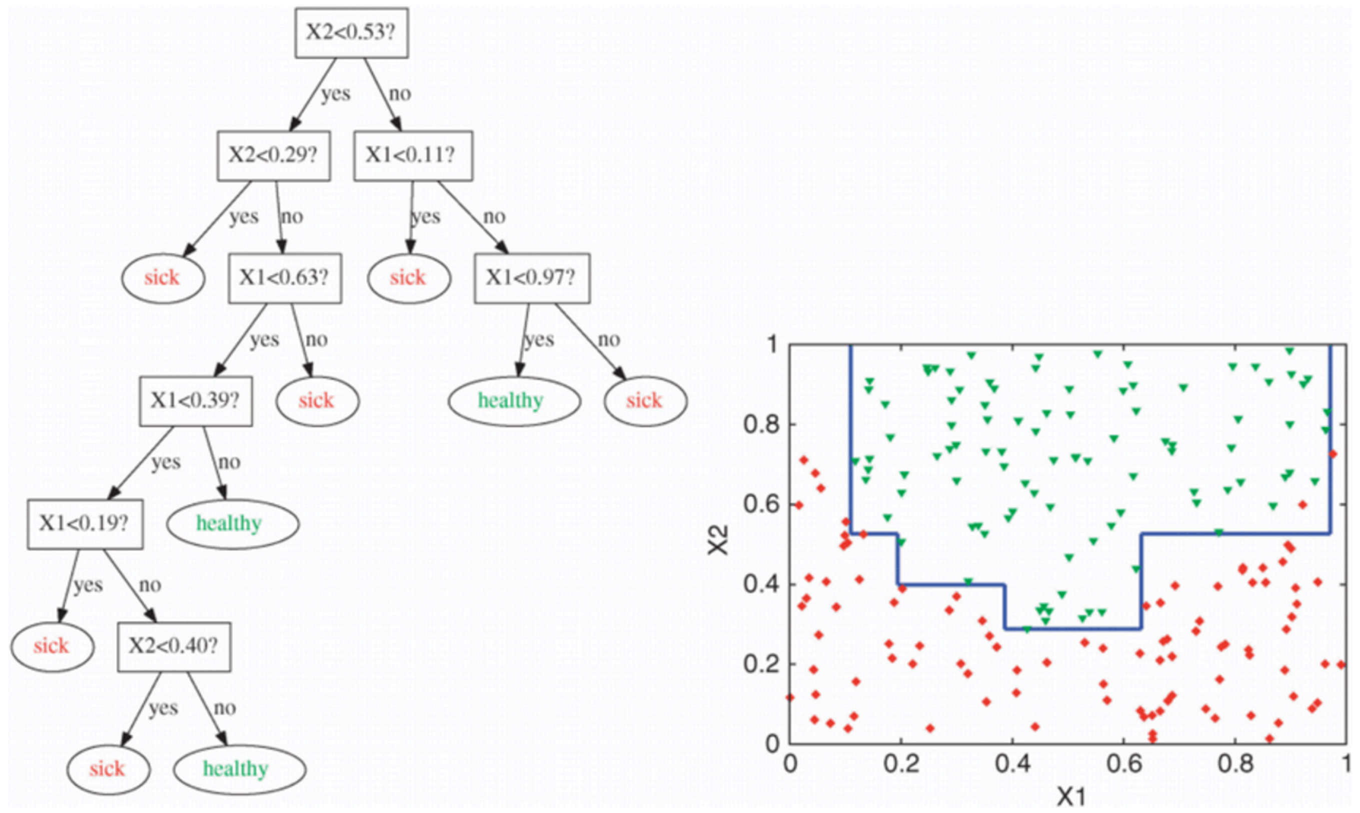
Machine learning

Types of hypothesis



Linear classification

Nonlinear classification



Tree-based classification

Machine learning

Other types of output space

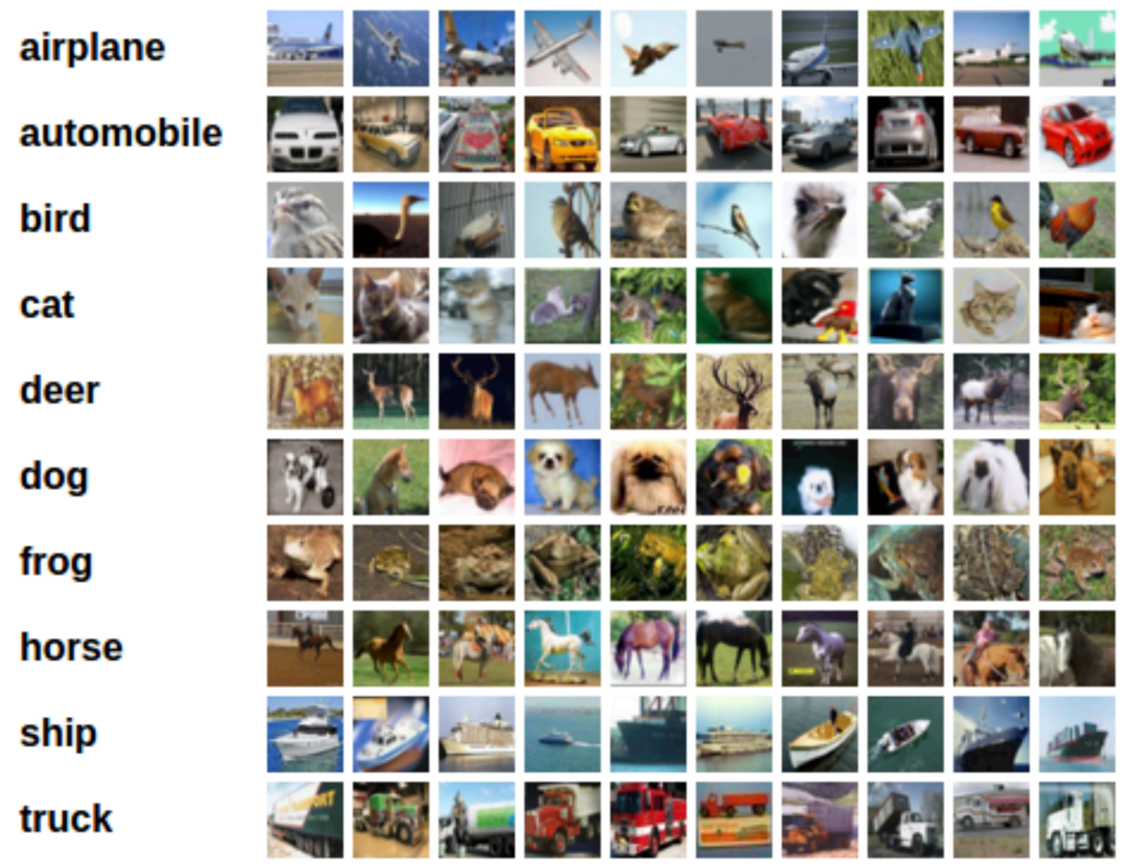
- Regression: $y_n \in \mathbb{R}$ (output is a real number)
- Example:
  - Stock price prediction
  - Movie rating prediction
  - …

Machine learning

Other types of output space

- Multi-class classification
  - $y_n \in \{1, \ldots, C\}$ (C-way)
- Examples: object classification

MNIST



CIFAR

Machine learning

Other types of output space

- Multi-label prediction
  - Multi-class problem: Each sample only has one label
  - Multi-label problem: Each sample can have multiple labels
- Example:
  - Document categorization (news/sports/economy/…)
  - Document/image tagging
  - …
- Extreme classification (large output space problems):
  - Millions of billions of labels (but usually each sample only has few labels)
  - Recommendation systems: Predict a subset of preferred items for each user
  - Document retrieval or search: Predict a subset of related articles for a query

Machine learning

Other types of output space

- Structural prediction
  - I love ML

    pronoun verb noun
- Multiclass classification for each word (word ⟶ word class)
  - (not using information of the whole sentence)
- Structure prediction problem:
  - sentence ⟶ structure (class of each word)
- Other examples: speech recognition, image captioning, machine translation, …

1. A red stop sign sitting on the side of a road.
2. A stop sign on the corner of a street.
3. A red stop sign sitting on the side of a street.

Machine learning overview

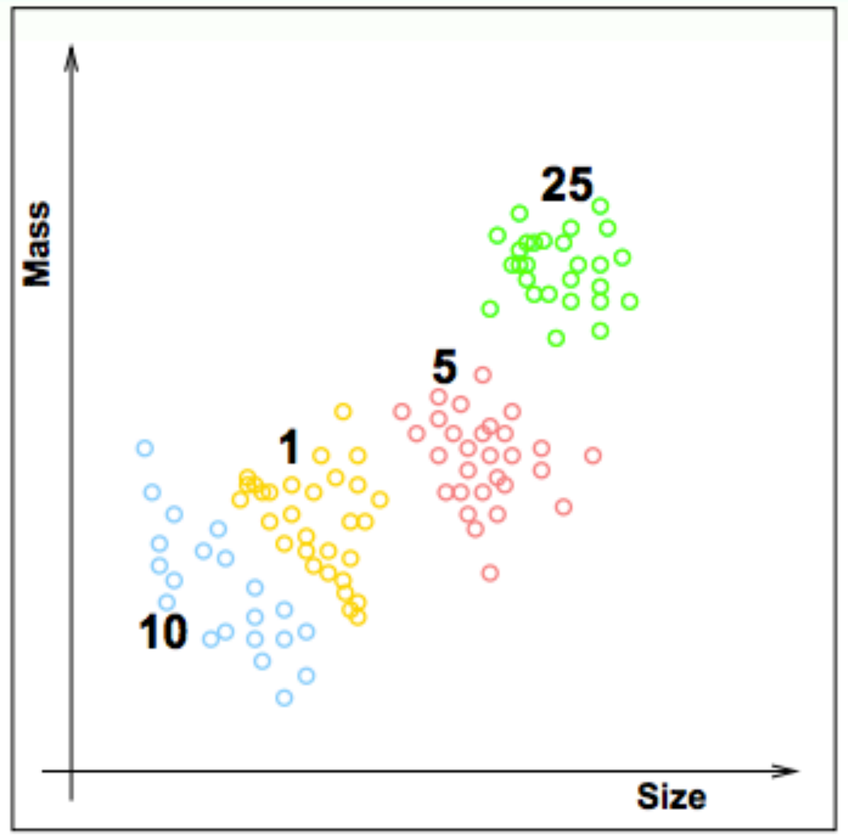
Machine Learning Problems

- Supervised learning: every x_n comes with y_n (label)
- Unsupervised learning: only x_n, no y_n
- Semi-supervised learning: Some labeled data and some unlabeled data
- Transfer learning: Transfer knowledge from source datasets to a target dataset

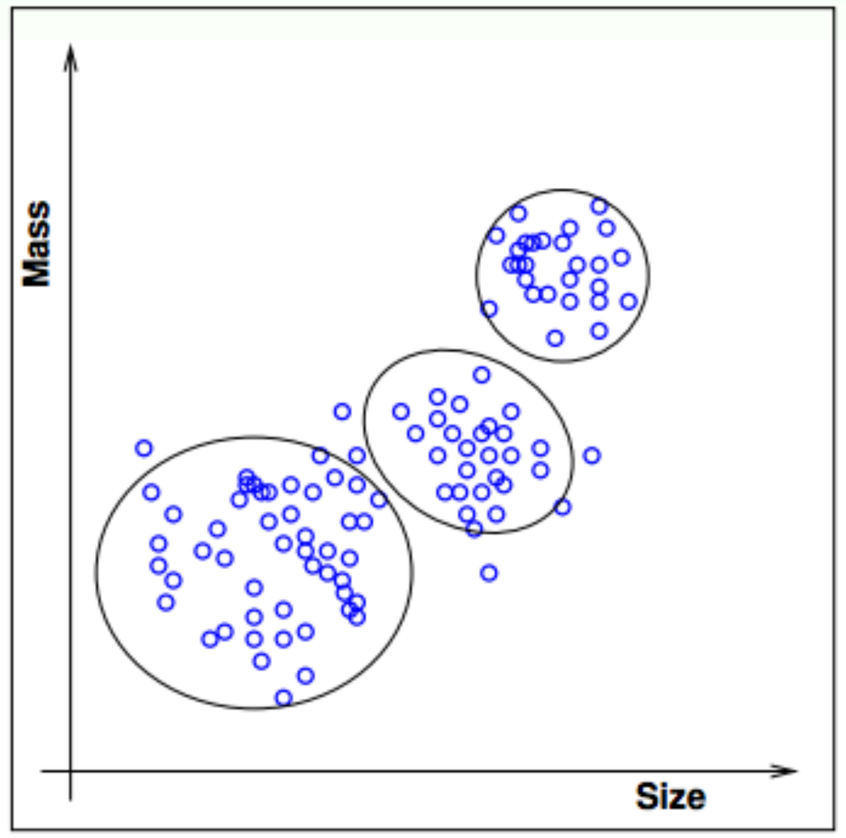
Machine learning

Unsupervised Learning (no y_n)

- Example: clustering
  - Given examples $x_1, \ldots, x_N$, classify them into K classes
- Other unsupervised learning:
  - Outlier detection: $\{x_n\} \Rightarrow$ unusual(x)
  - Dimensional reduction
  - …



supervised multiclass classification

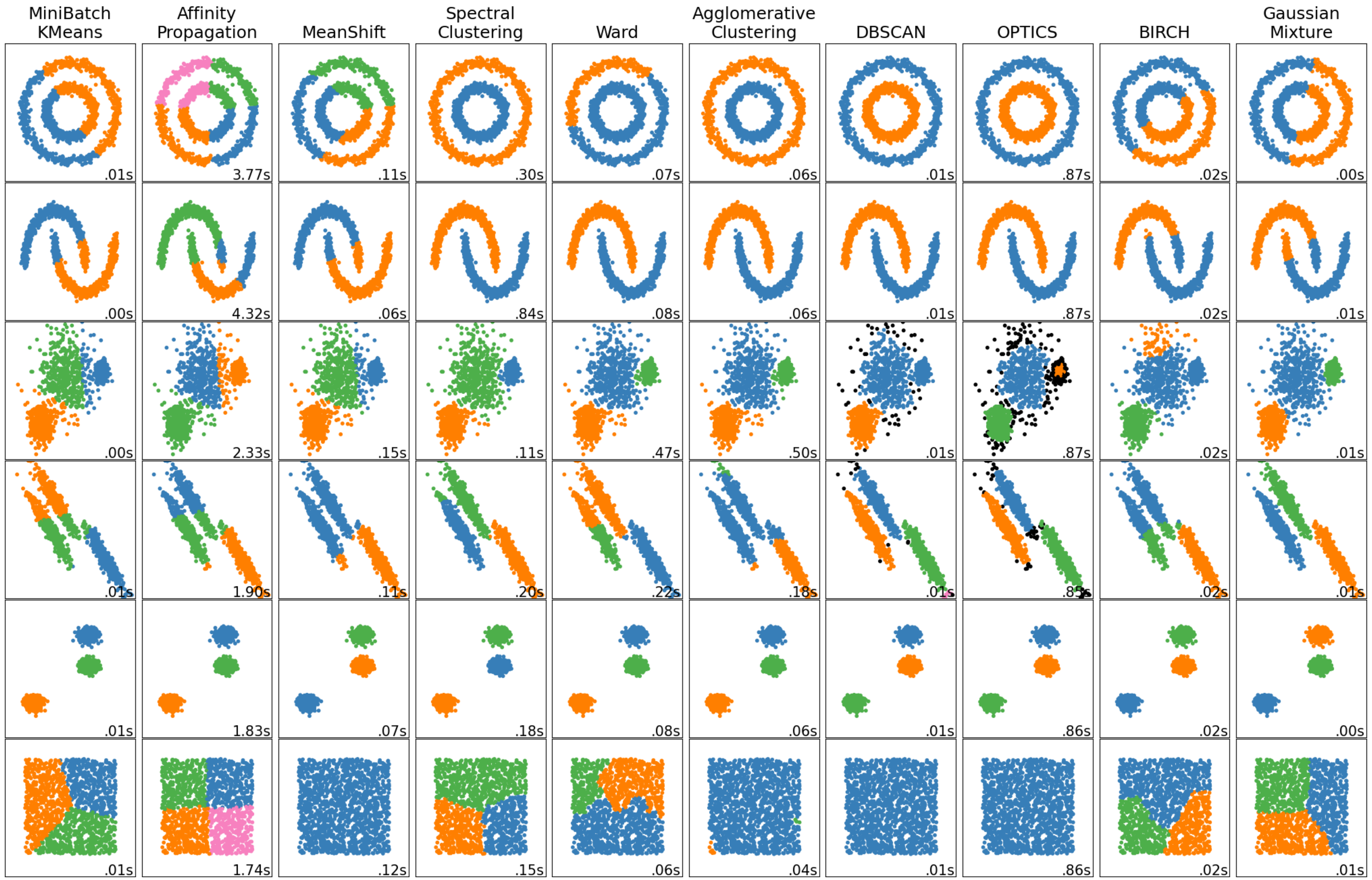


unsupervised multiclass classification
$\Longleftrightarrow$ 'clustering'

Machine learning

Unsupervised Learning

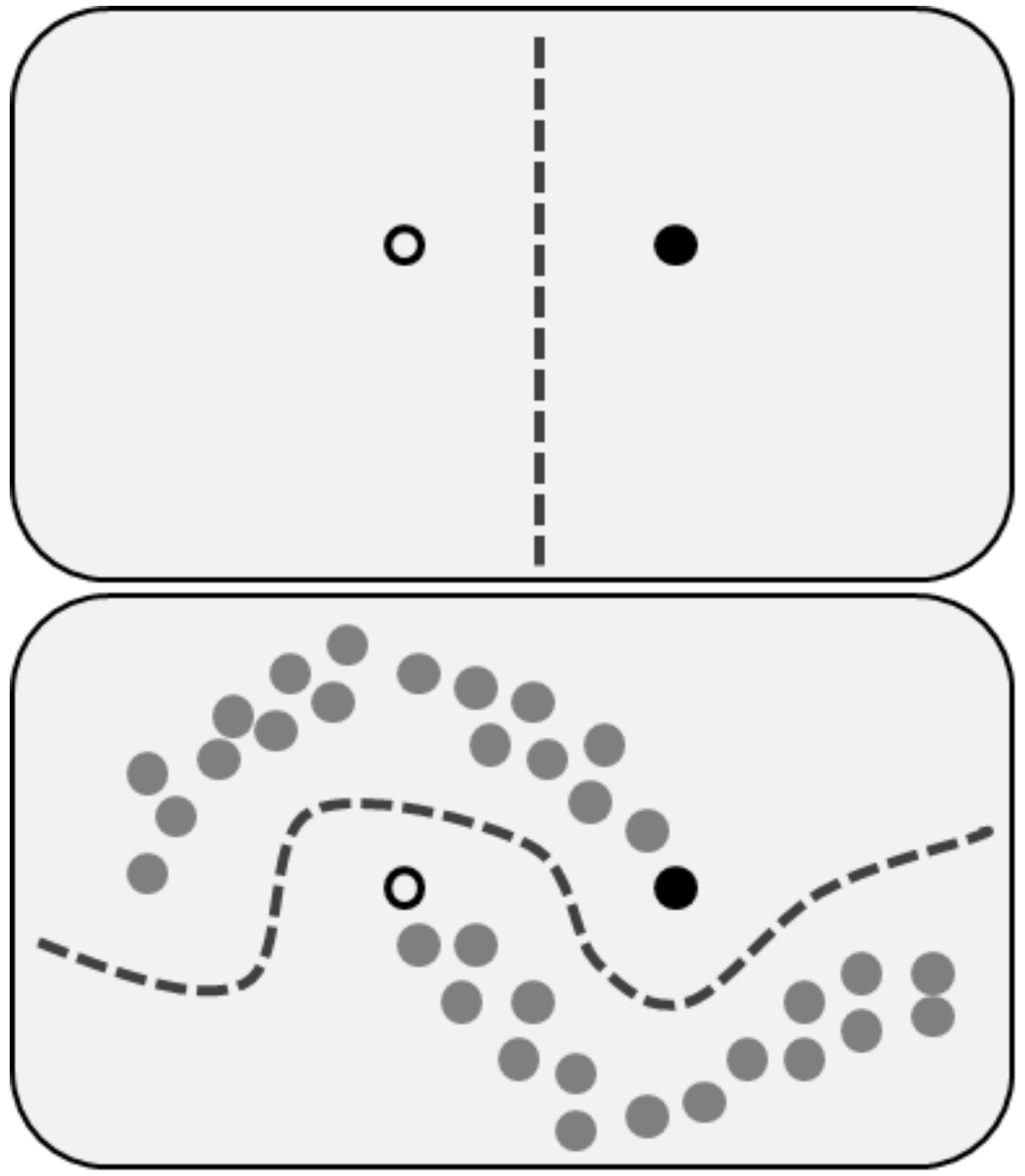
- Example: clustering



Machine learning

Semi-supervised learning

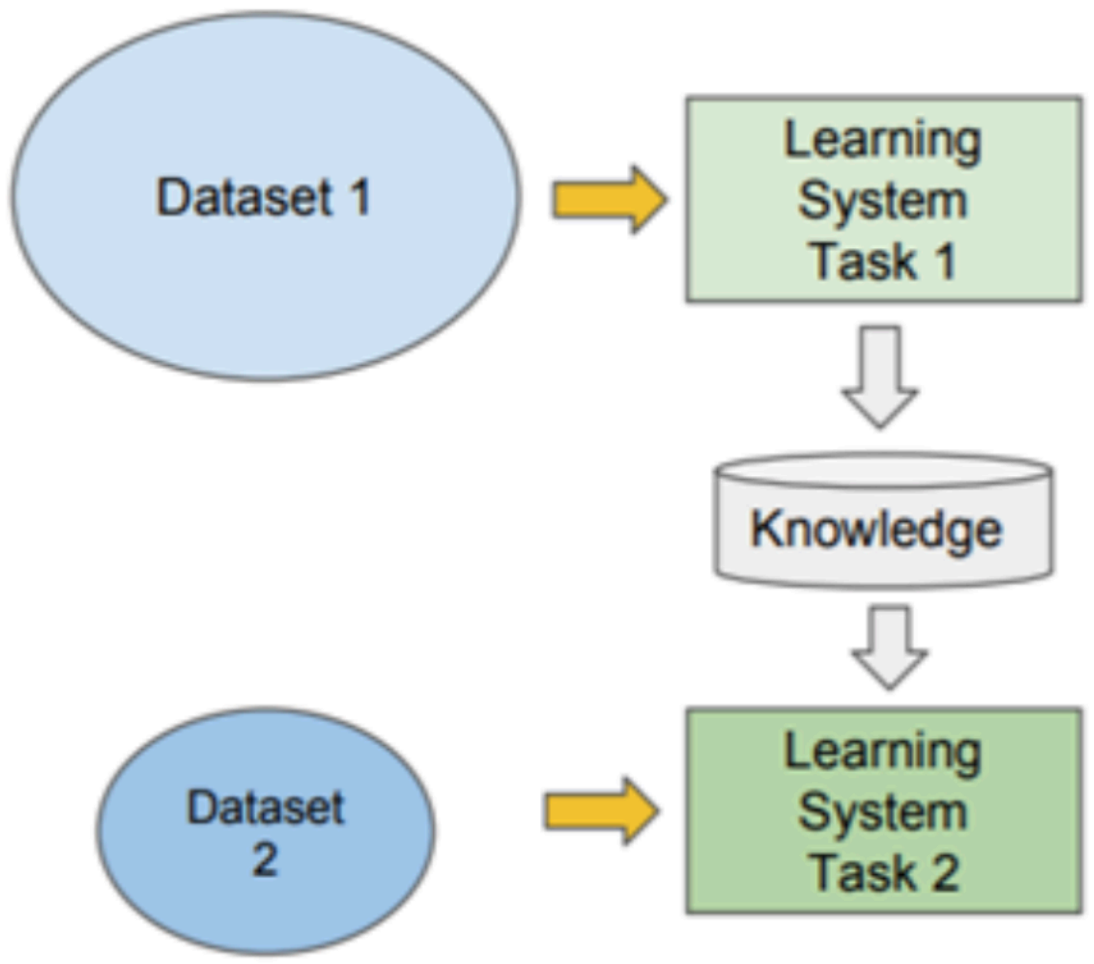
- Only some (few) x_n has y_n
  - Labeled data is much more expensive than unlabeled data

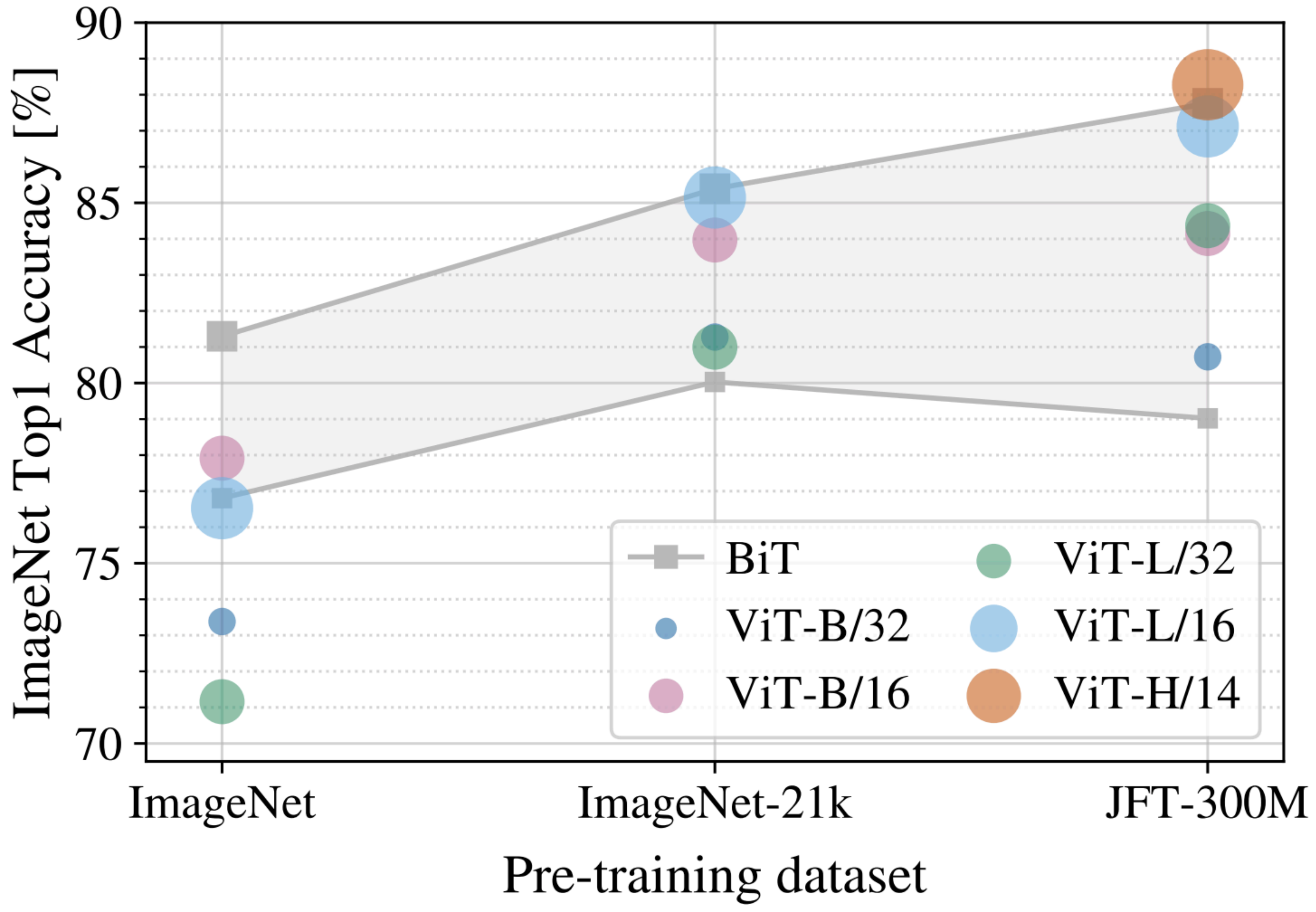


Machine learning

Transfer learning

- Source dataset D_{source} and target dataset D_{target}
- How to leverage the information of D_{source} to improve the performance of target task?
- Useful when source data has much richer information than target data
- (Pre)train the neural network based on the source data
- Fine-tune some parts or the entire network on target data





Machine Learning

Self-supervised learning

- The pretraining can be done with unlabeled data (easy to collect gigantic unlabeled data)
    - Example: We can get almost unlimited unlabeled text from Internet
- Define the training task based on unlabeled data
    - Example: predict a word in a sentence
- Transfer the model to end task

Original sentence:

In Autumn the leaves fall from the trees.

Masked sentence:

In Autumn the [   ] fall from the trees.

leaves
apples
raindrops
branches

Predicted words by the model

Masked language modeling (pretraining for text model)

Contrastive learning (pretraining for text model)