



CS 540 Introduction to Artificial Intelligence

Machine Learning Overview

Yudong Chen
University of Wisconsin-Madison

Sep 30, 2021

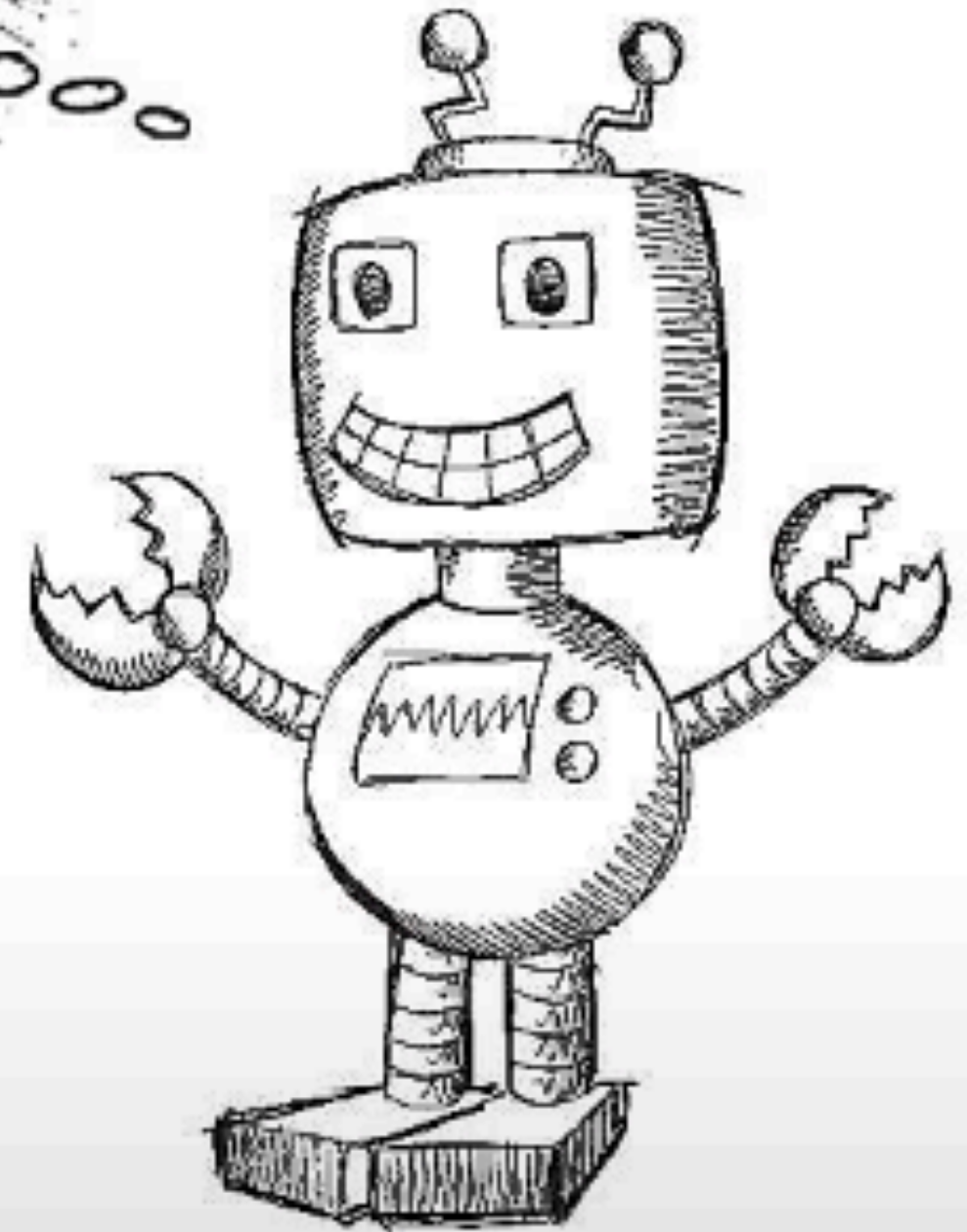
Slides created by Sharon Li [modified by Yudong Chen]

Today's outline

- What is machine learning?
- Supervised Learning
 - Classification
 - Regression
- Unsupervised Learning
 - Clustering
- Reinforcement Learning



Part I: What is machine learning?



**HUMANS LEARN FROM
PAST EXPERIENCES**

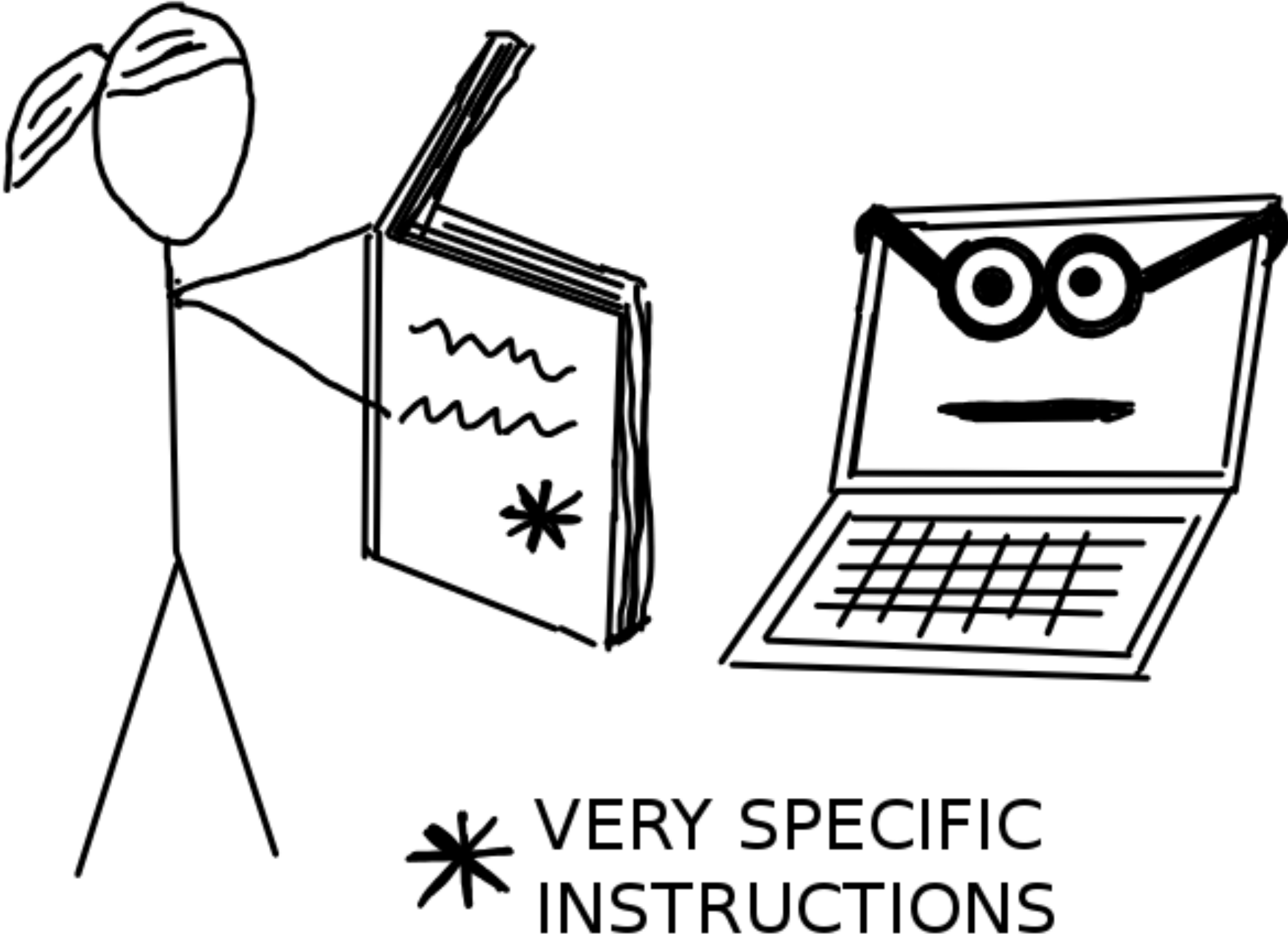
**MACHINES FOLLOW INSTRUCTIONS
GIVEN BY HUMANS**

What is **machine learning**?

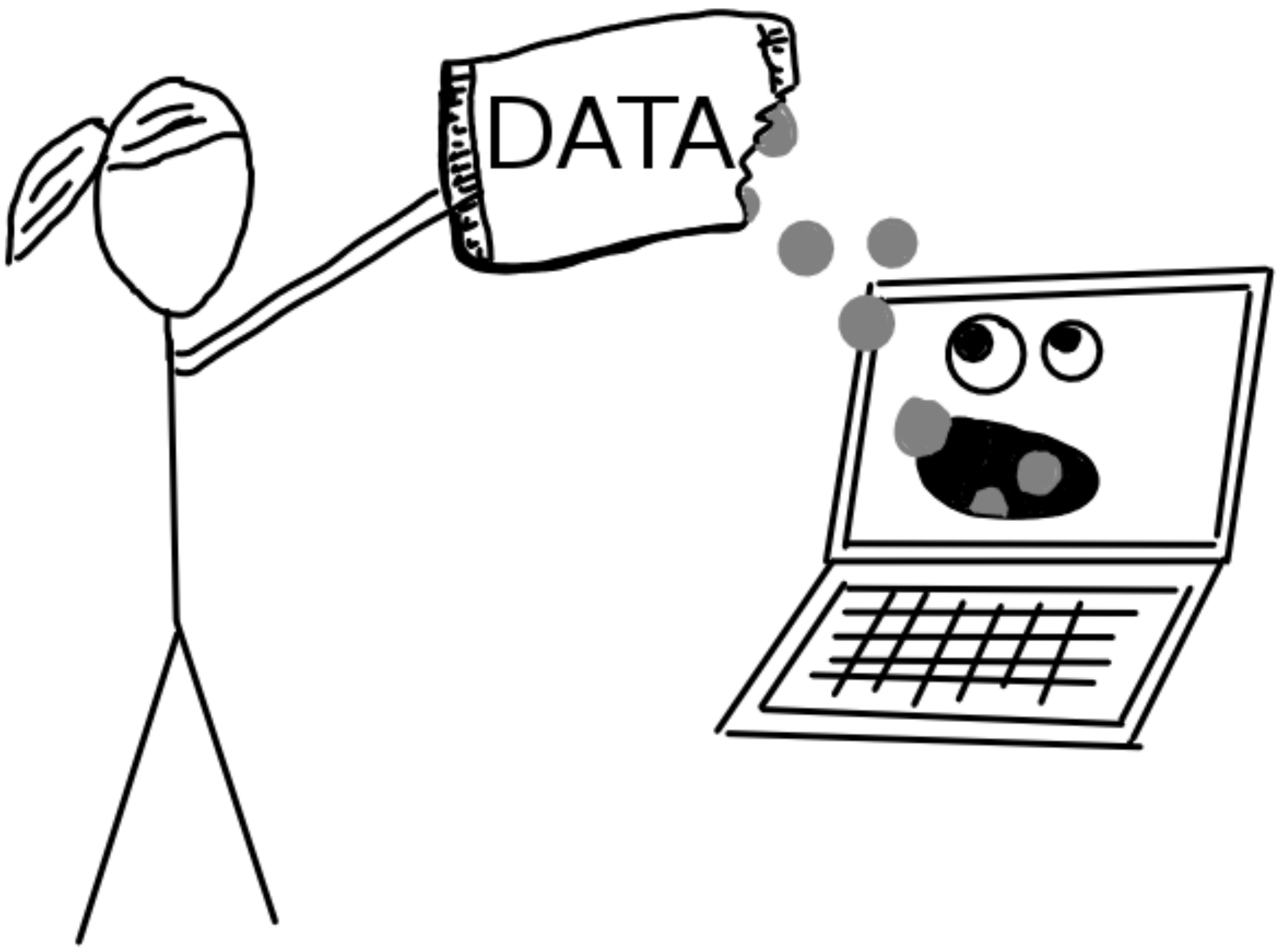
- Arthur Samuel (1959): Machine learning is the field of study that gives the computer the ability to learn **without being explicitly programmed**.



Without Machine Learning



With Machine Learning



What is **machine learning**?

- Arthur Samuel (1959): Machine learning is the field of study that gives the computer the ability to learn **without being explicitly programmed**.
- Tom Mitchell (1997): 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.



Taxonomy of ML

**Supervised
Learning**

**Unsupervised
Learning**

**Reinforcement
Learning**



Part II: Supervised Learning

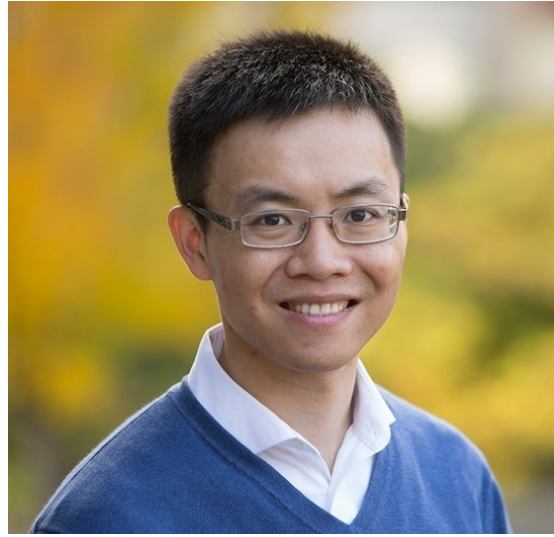
Example 1: Predict whether a user likes a song or not



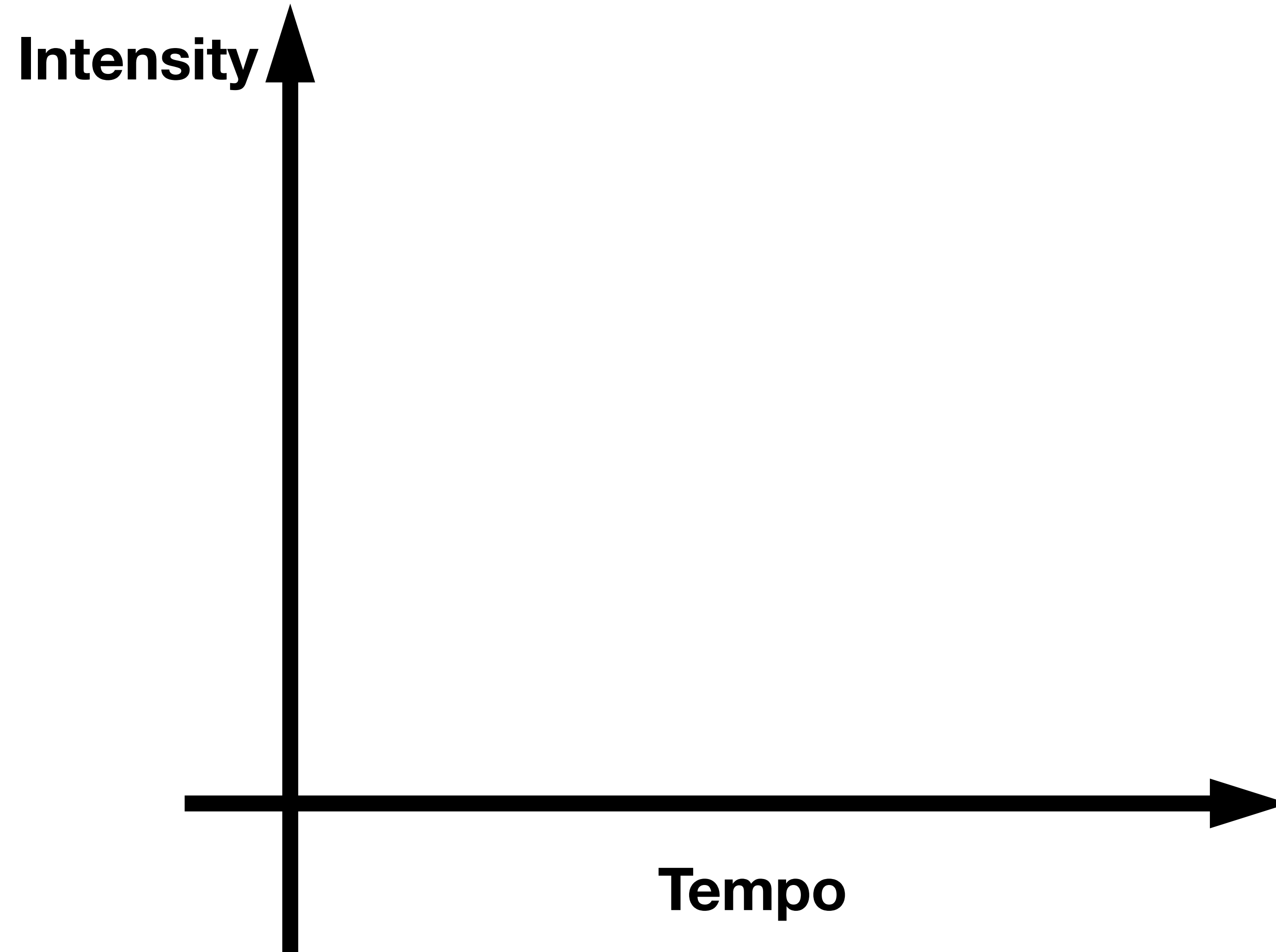
model



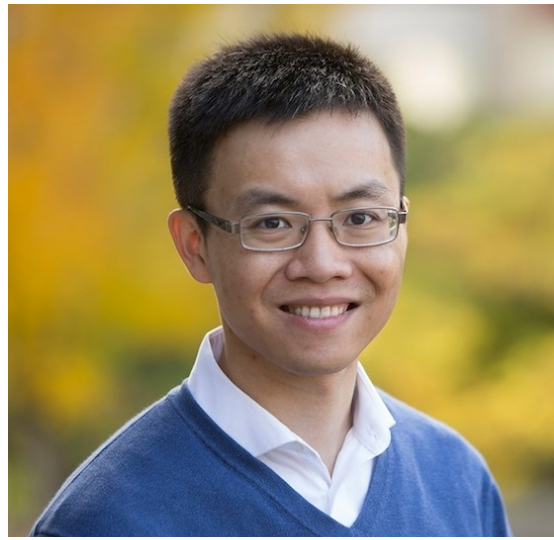
Example 1: Predict whether a user likes a song or not



User Yudong



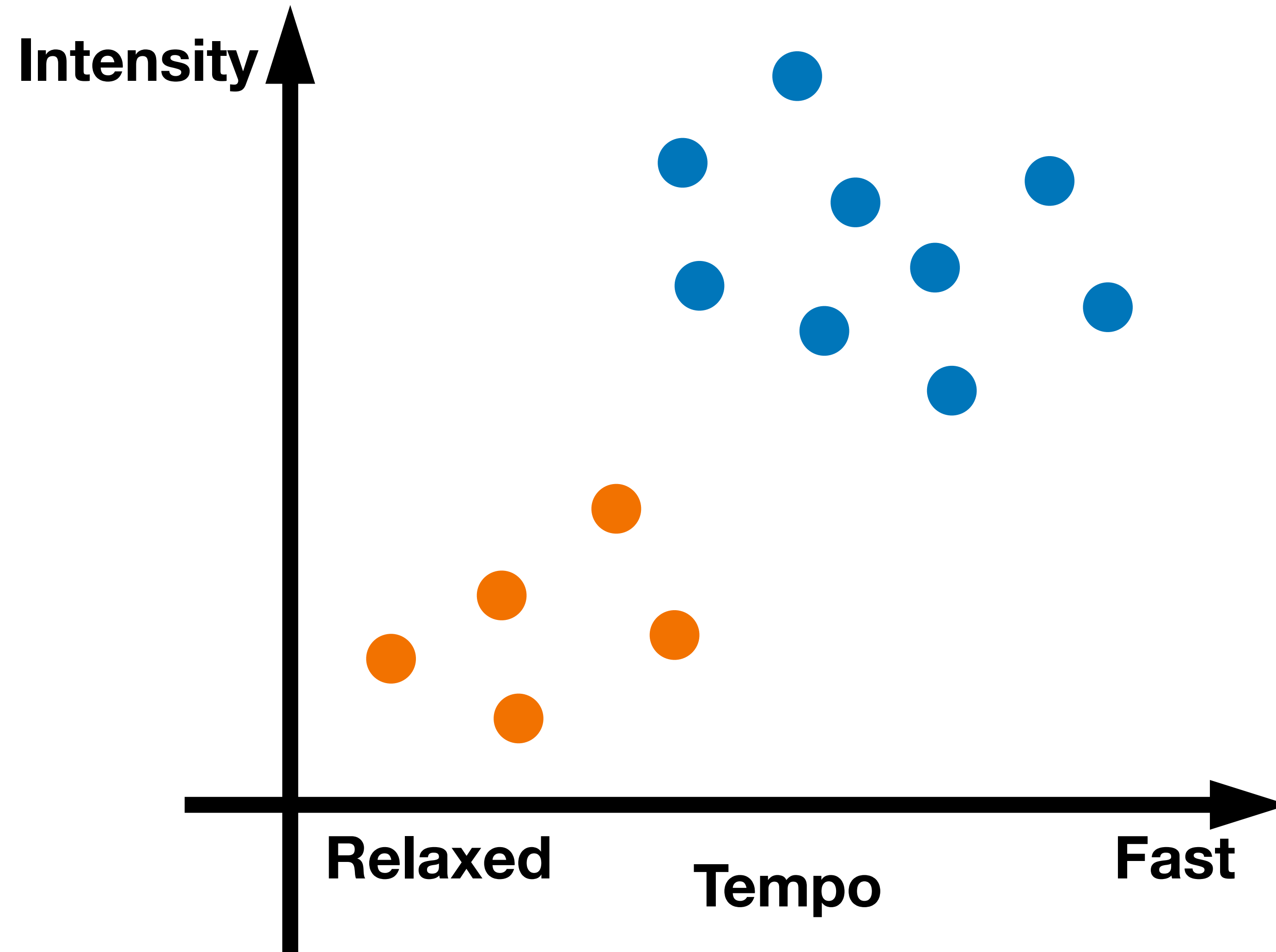
Example 1: Predict whether a user likes a song or not



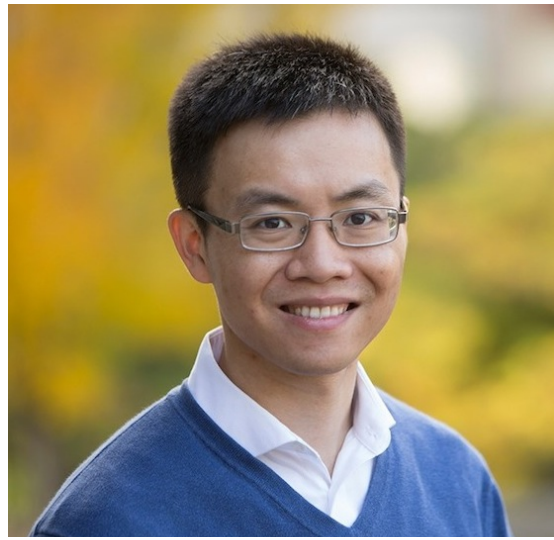
User Yudong

● DisLike

● Like



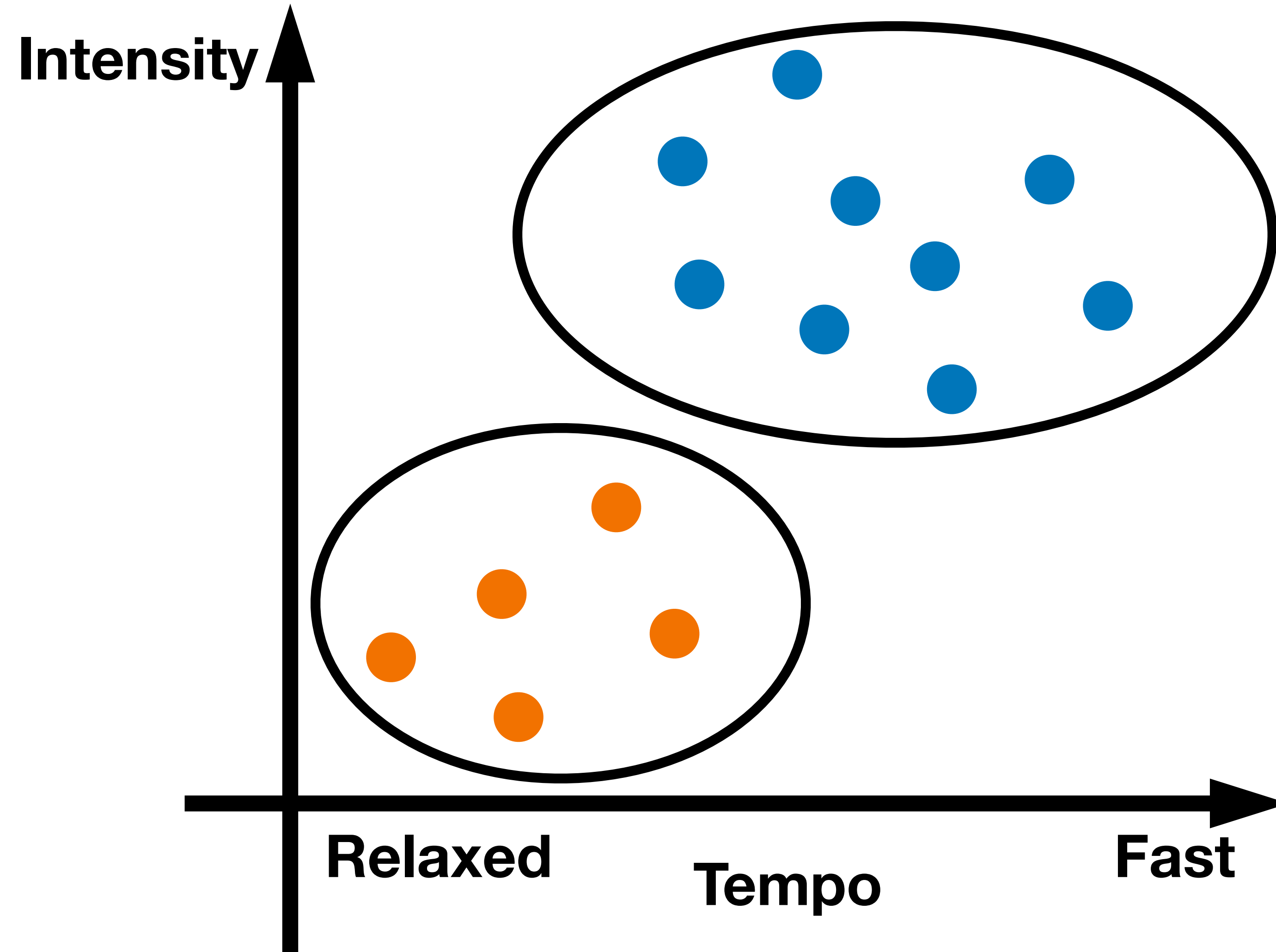
Example 1: Predict whether a user likes a song or not



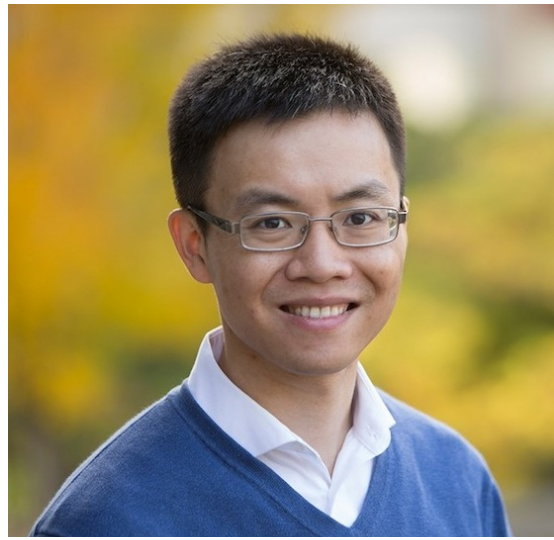
User Yudong

● DisLike

● Like



Example 1: Predict whether a user likes a song or not

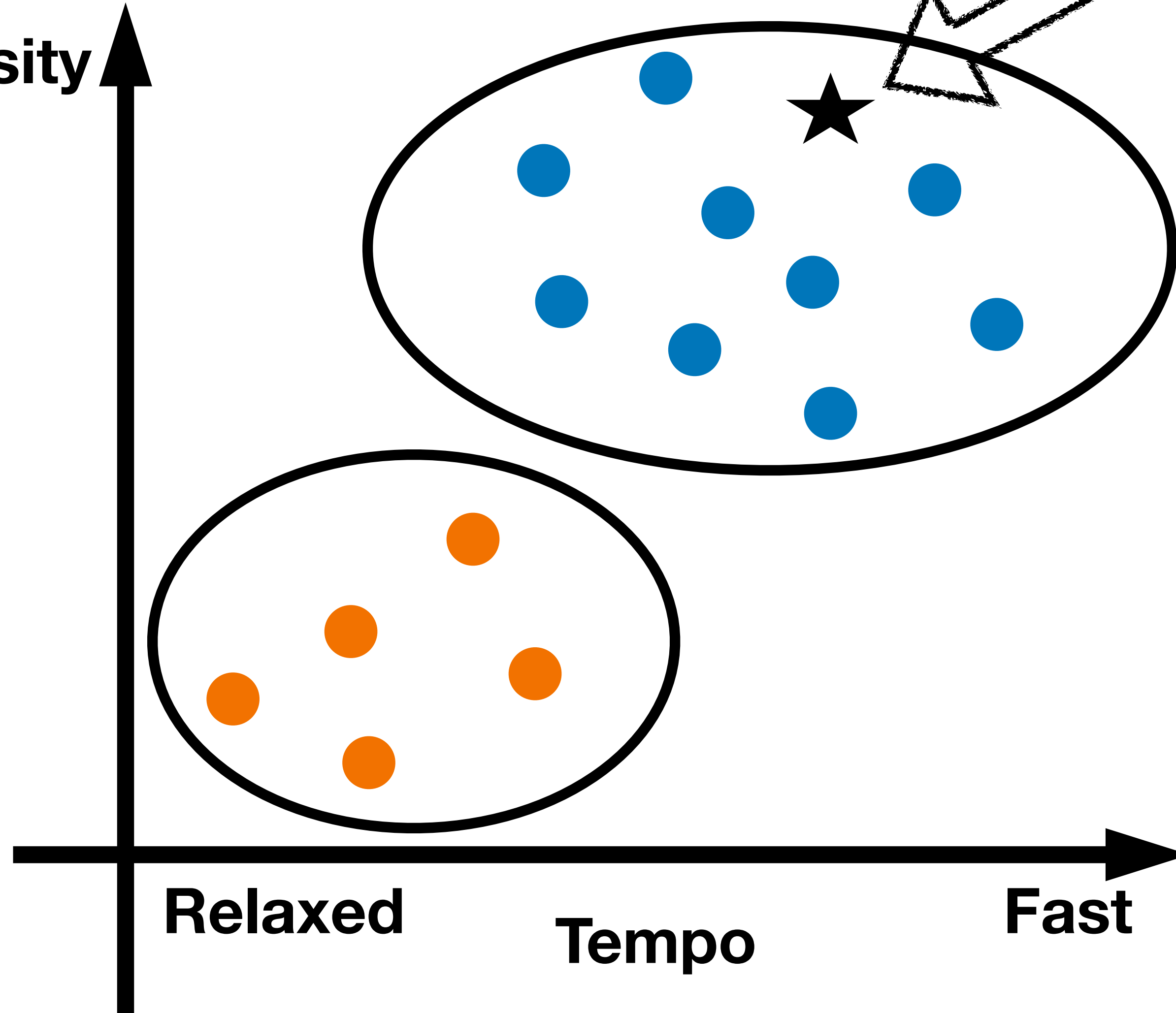


User Yudong

● DisLike

● Like

Intensity



New data

?

Relaxed

Tempo

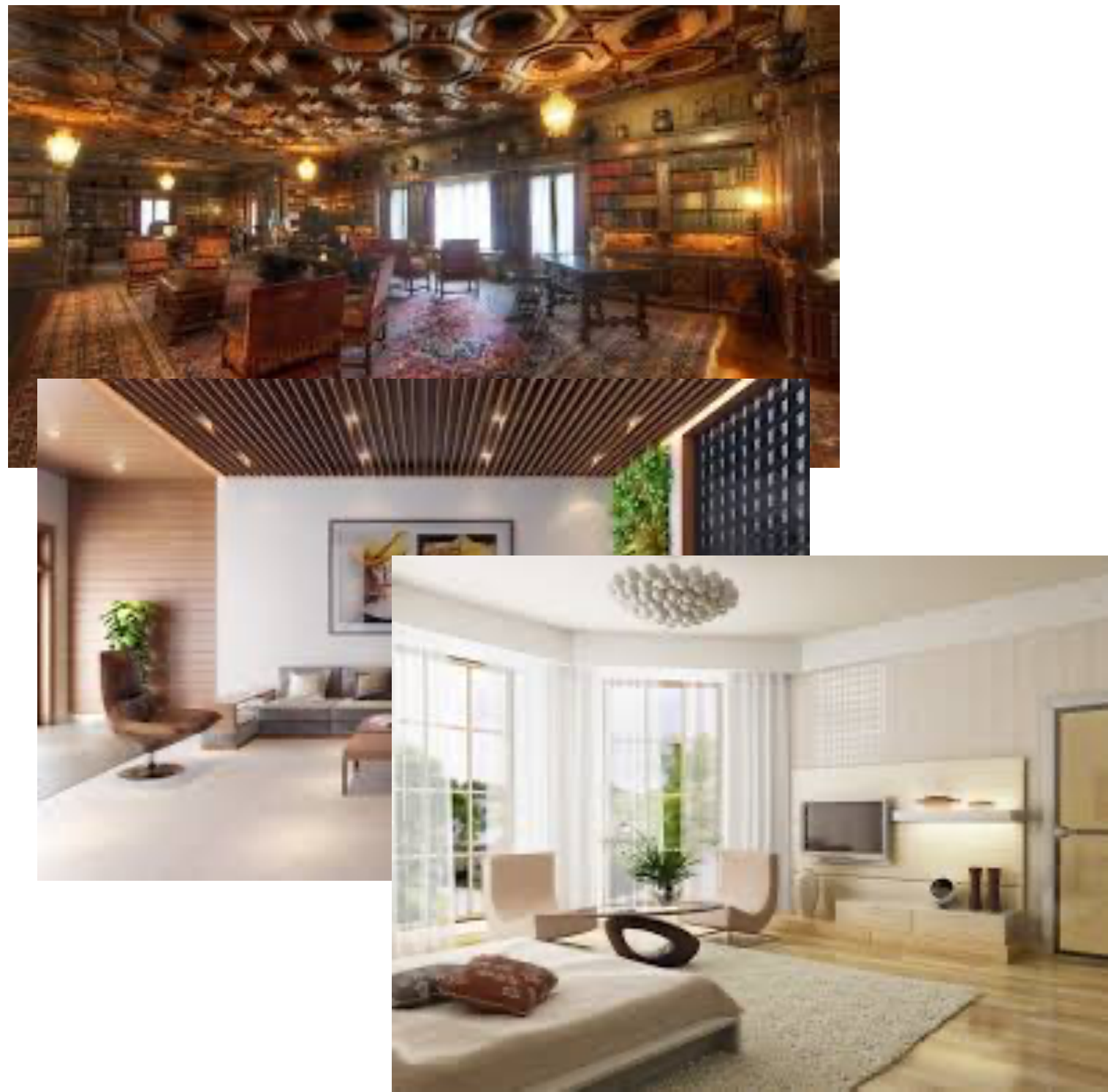
Fast

Example 2: Classify Images

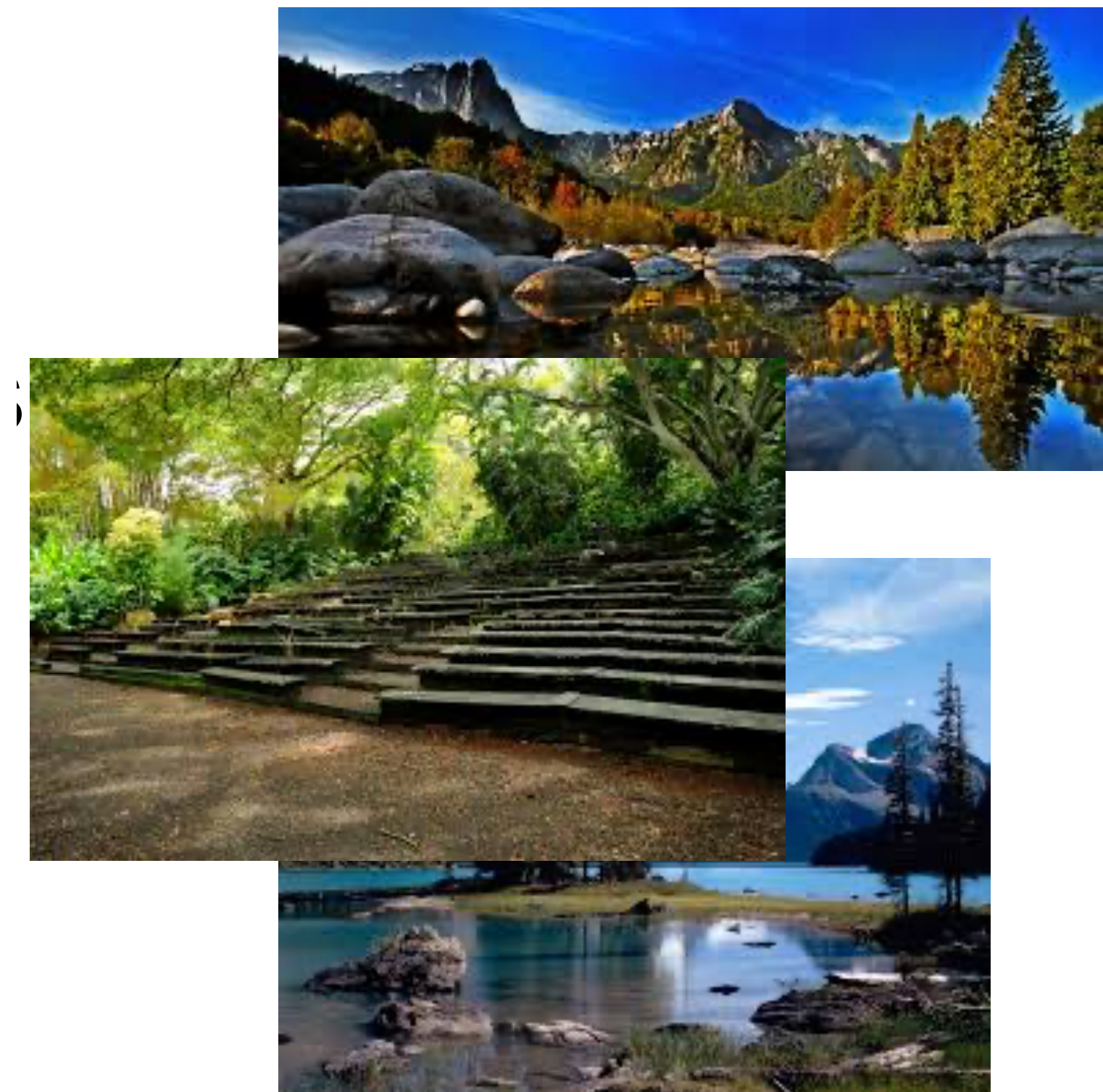
<http://www.image-net.org/>



Example 2: Classify Images

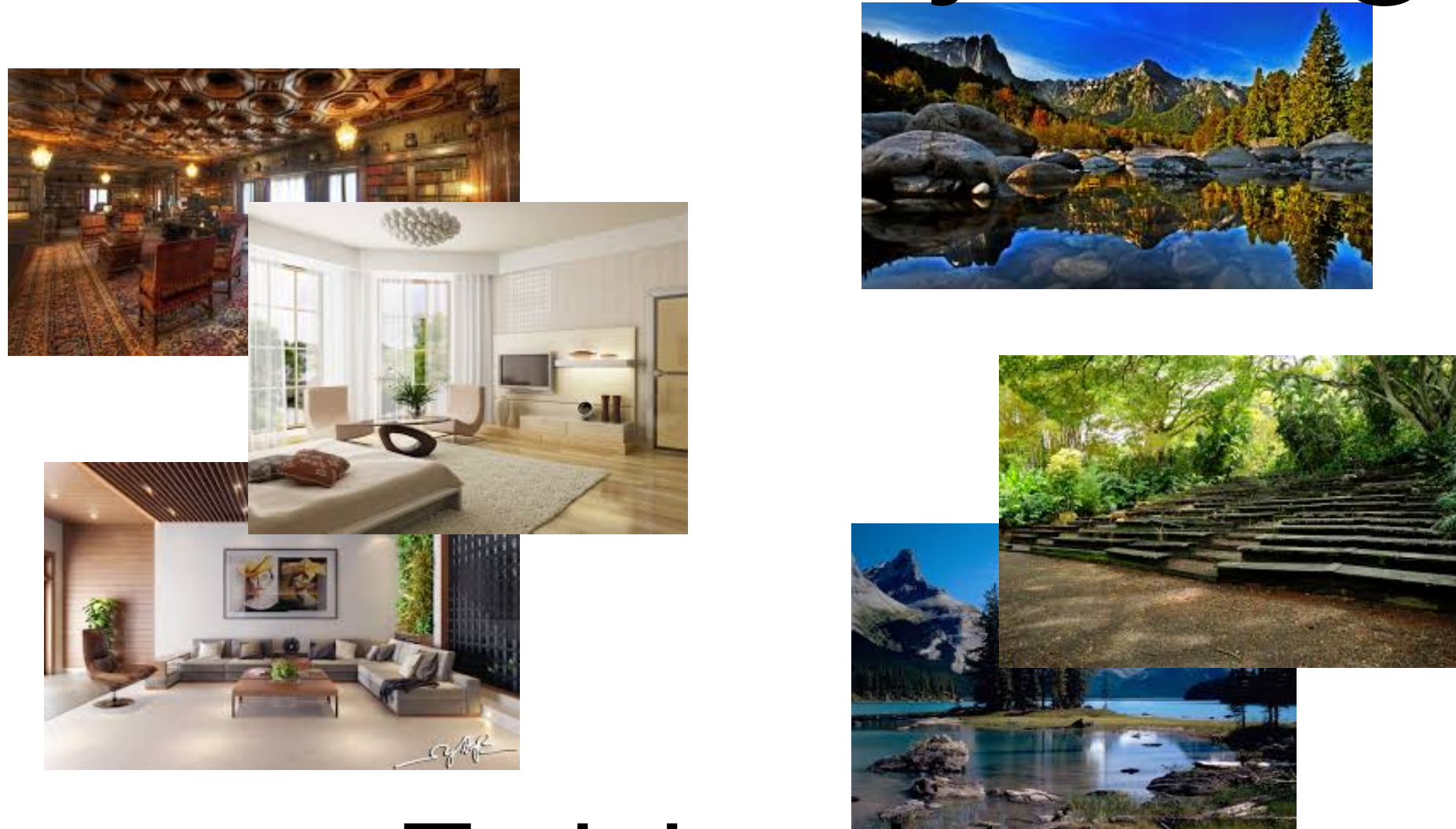


indoor



outdoor

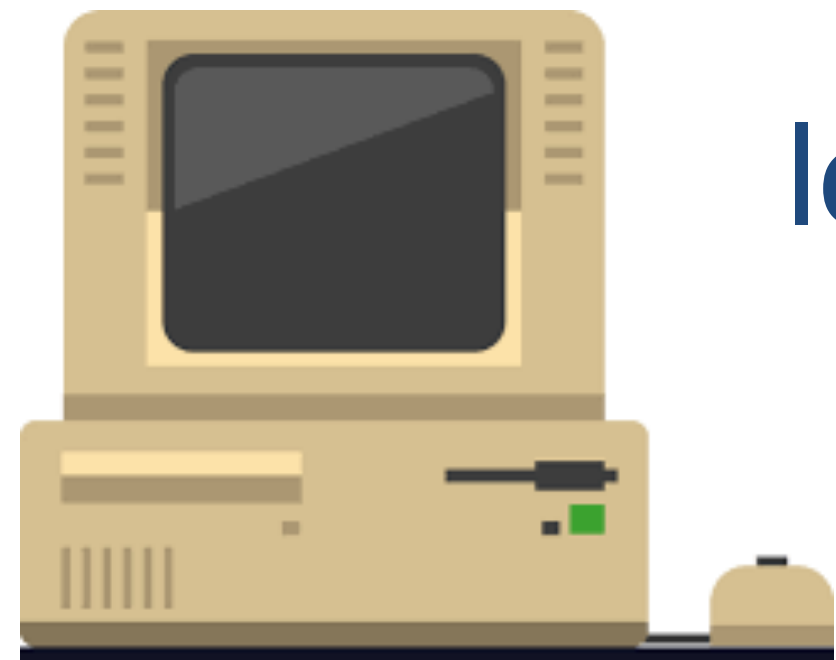
Example 2: Classify Images

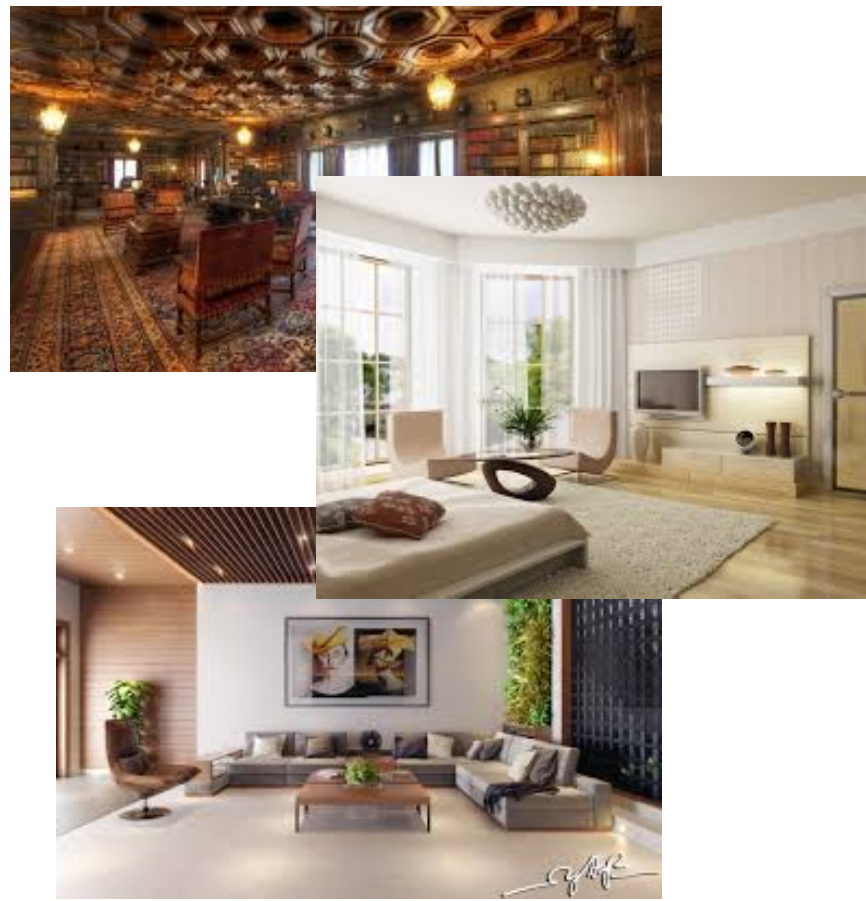


Training data



learning (i.e., training)





Label: outdoor

Label: indoor

Training data

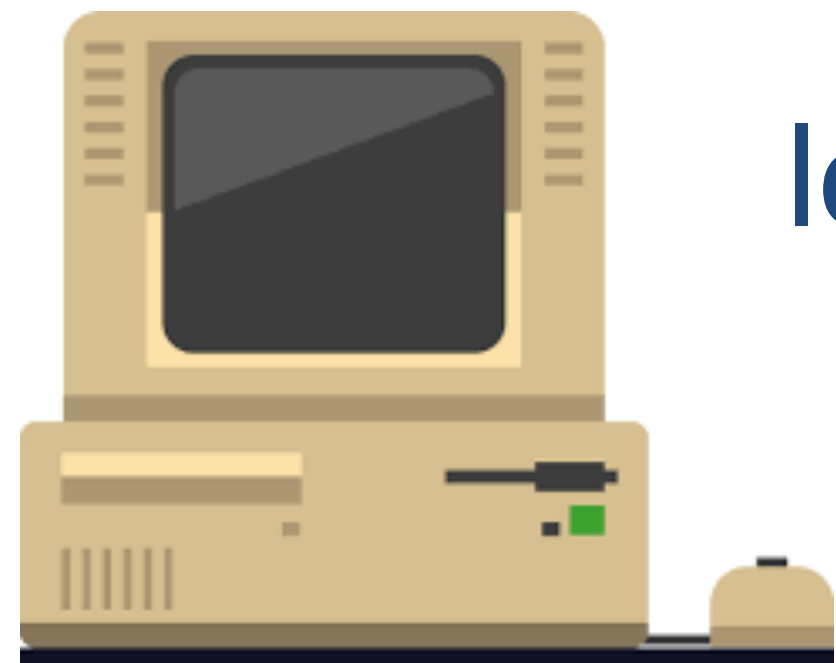
Test data



learning (i.e., training)

testing

performance



How to represent data?

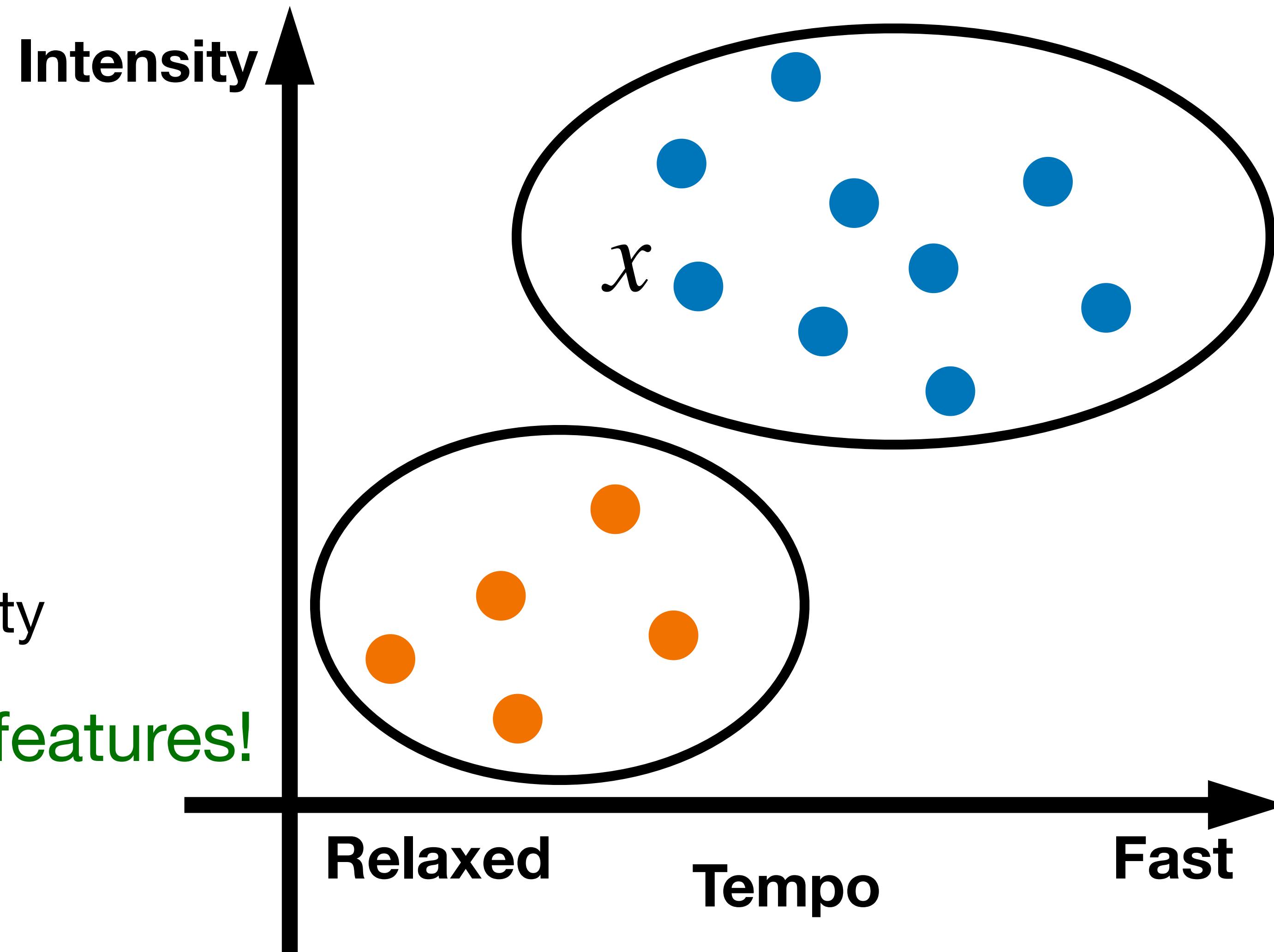
input data

$$x \in \mathbb{R}^d$$

d : feature dimension

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{array}{l} \text{Tempo} \\ \text{Intensity} \end{array}$$

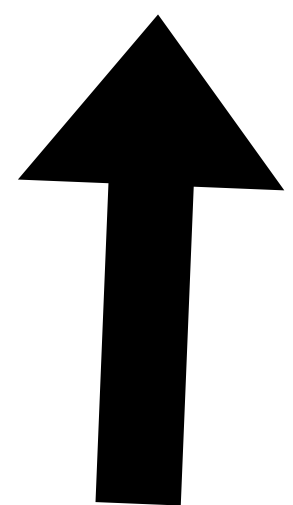
There can be many features!



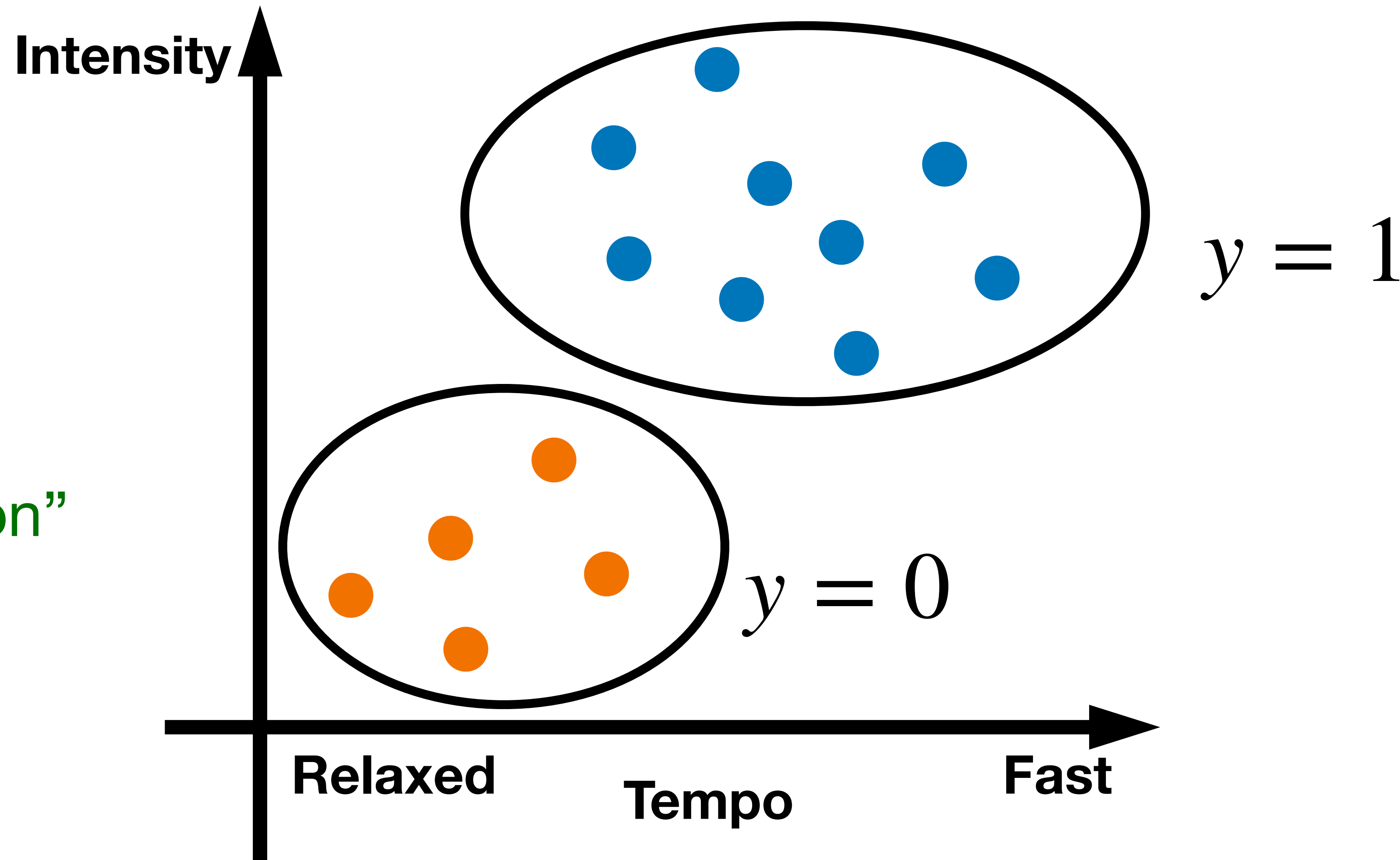
How to represent data?

Label

$$y \in \{0,1\}$$



Where “supervision”
comes from



Represent various types of data

- Image
 - Pixel values
- Bank account
 - Credit rating, balance, # deposits in last day, week, month, year, #withdrawals

Two Types of Supervised Learning Algorithms

Classification

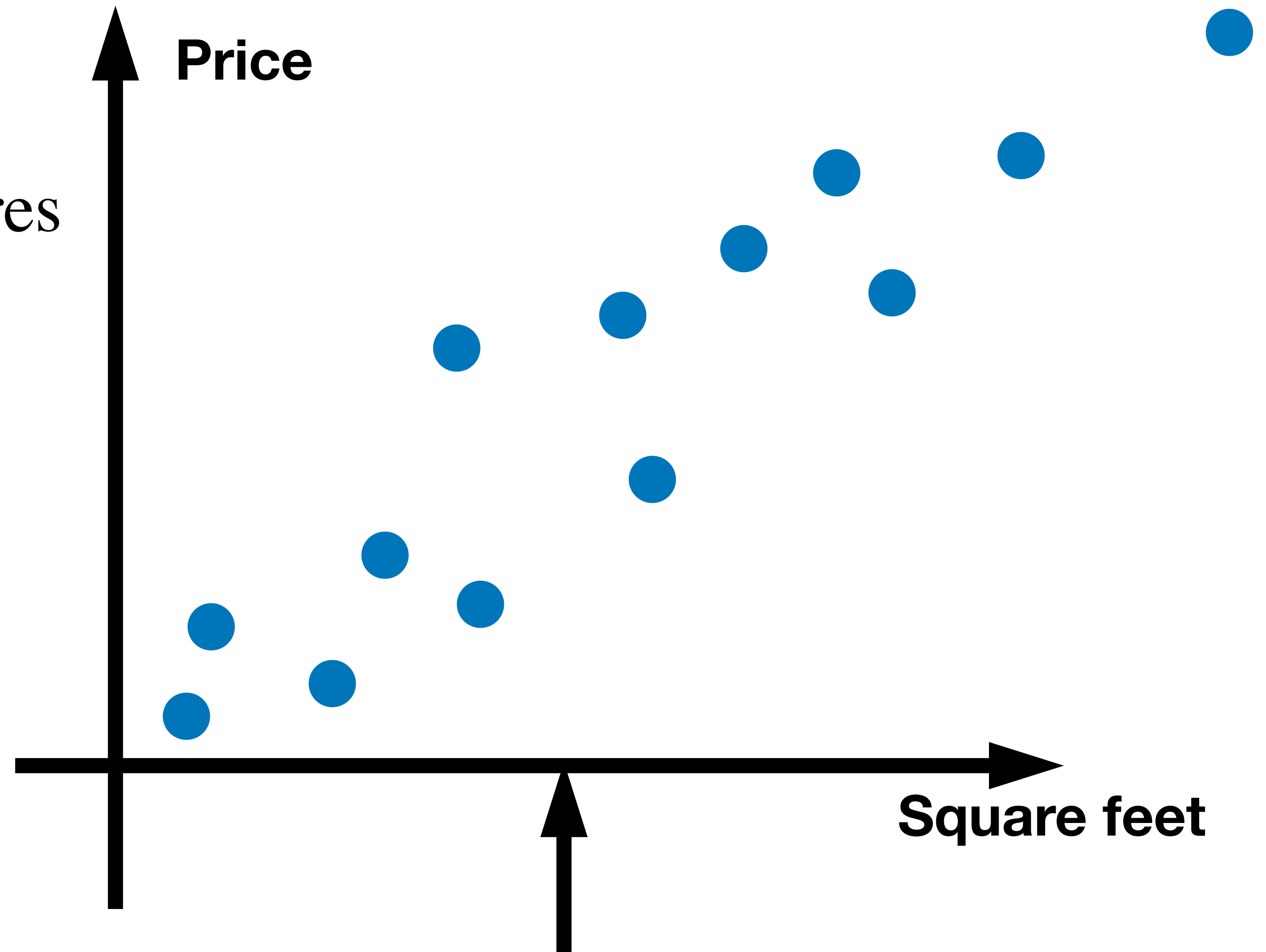
Regression

Example of regression: housing price prediction

Given: a dataset that contains n samples

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

Task: if a residence has x squares feet, predict the price?



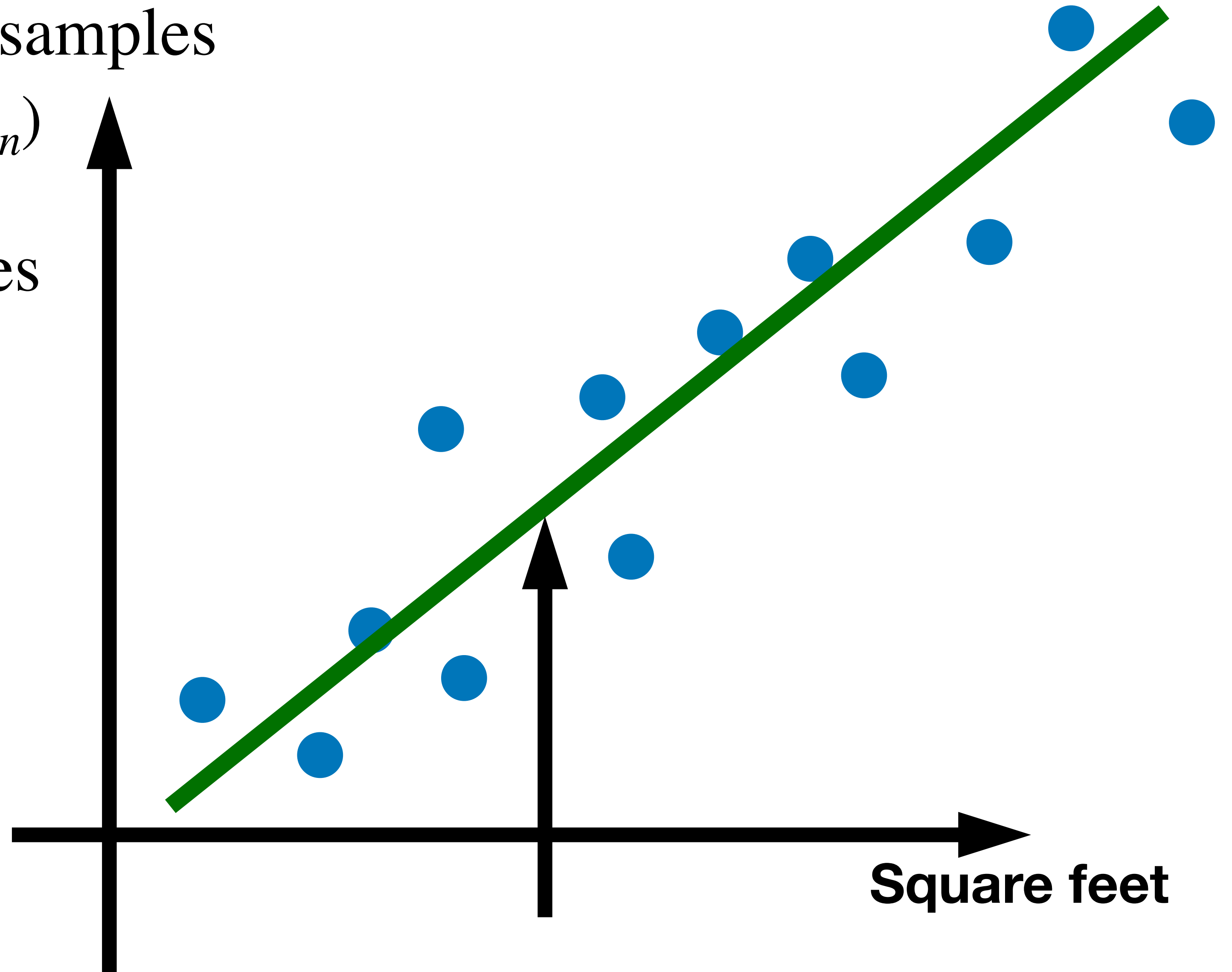
Example of regression: housing price prediction

Given: a dataset that contains n samples

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$$

Task: if a residence has x squares feet, predict the price?

$$y \in \mathbb{R}$$



Example of regression: housing price prediction

Input with more features (e.g., lot size)

(size, lot size)

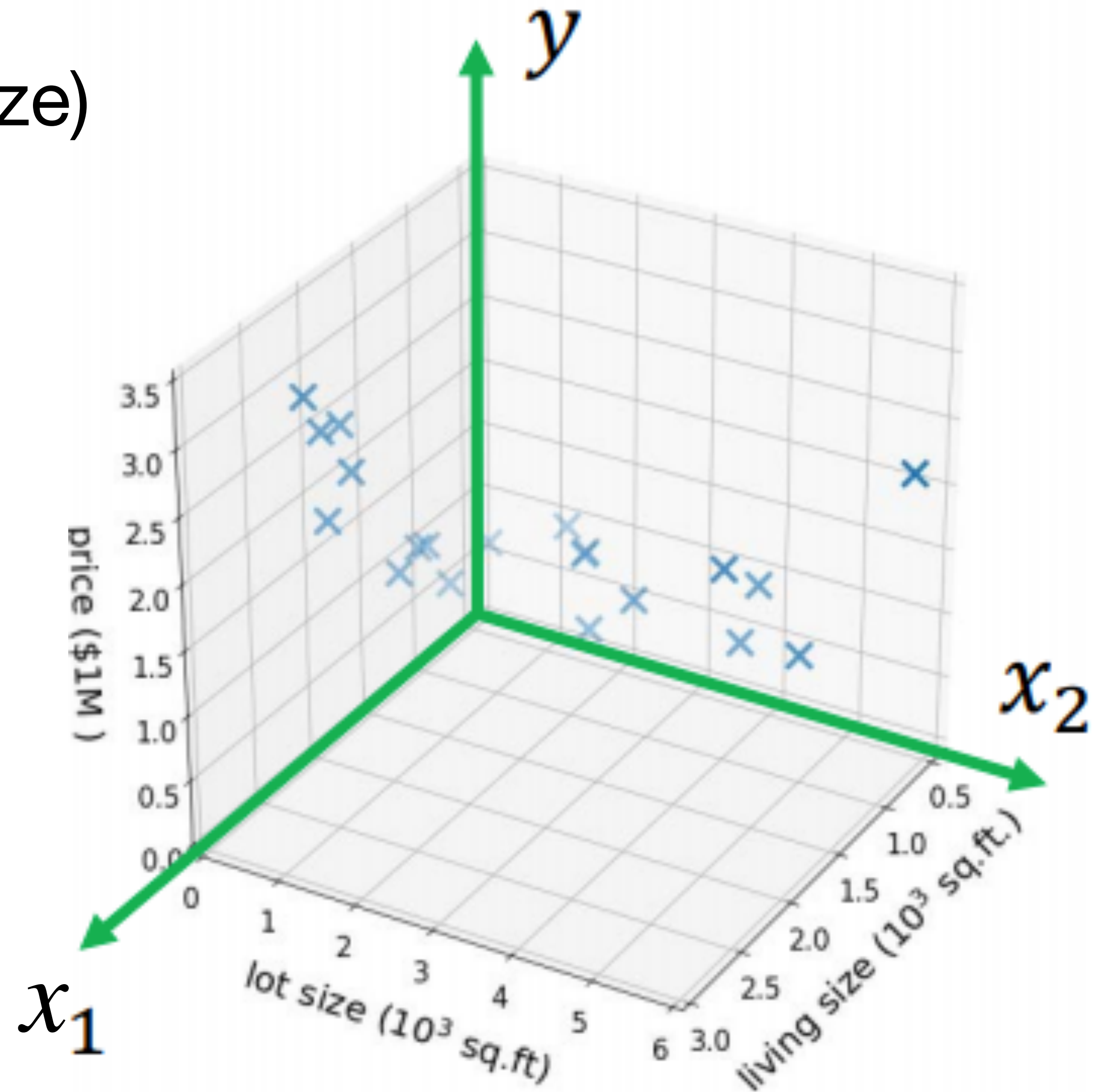
→ price

features/input

label/output

$$x \in \mathbb{R}^2$$

$$y \in \mathbb{R}$$

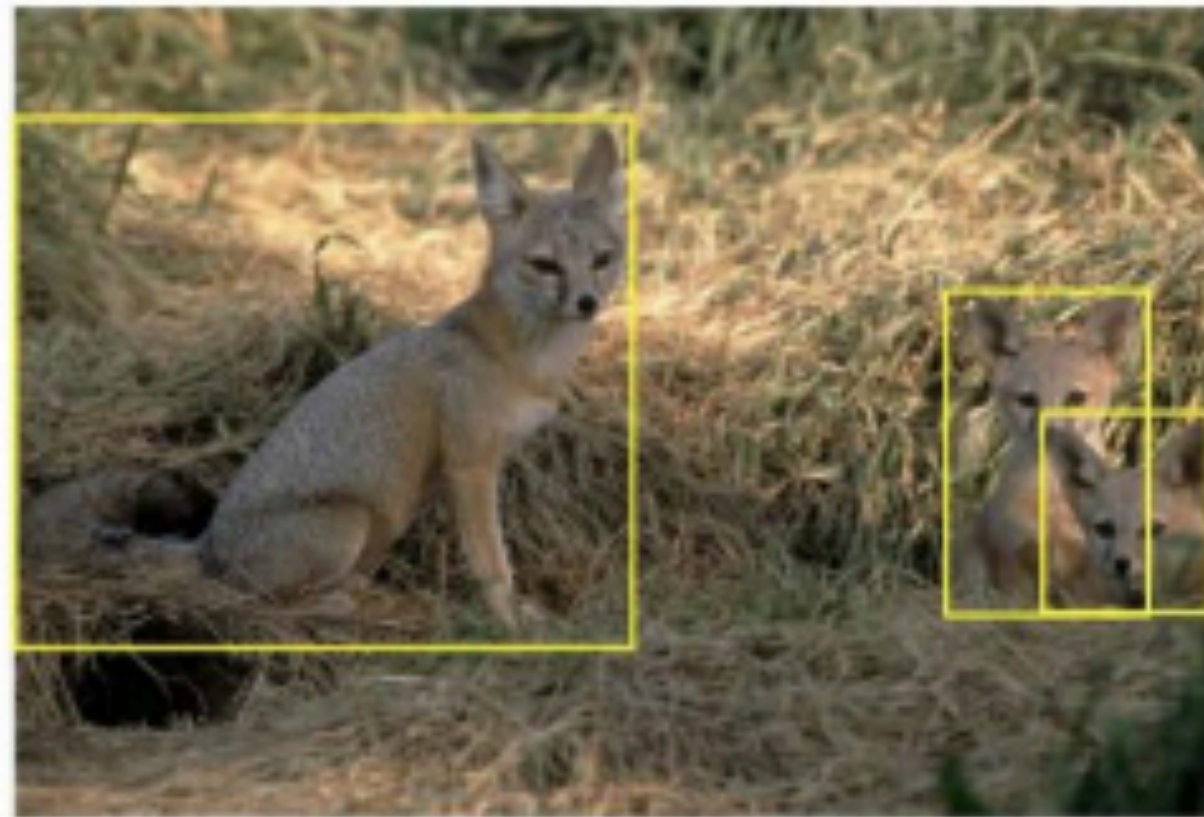


(credit: stanford CS229)

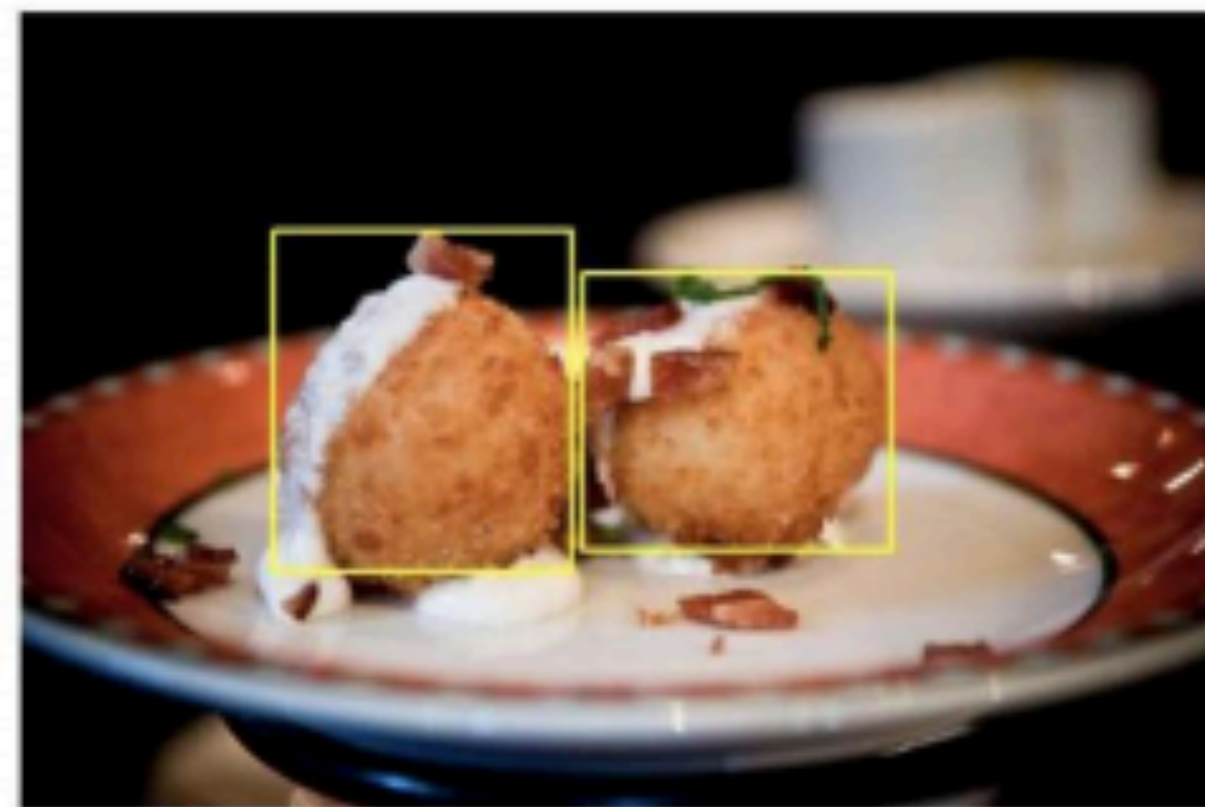
Supervised Learning: More examples

x = raw pixels of the image

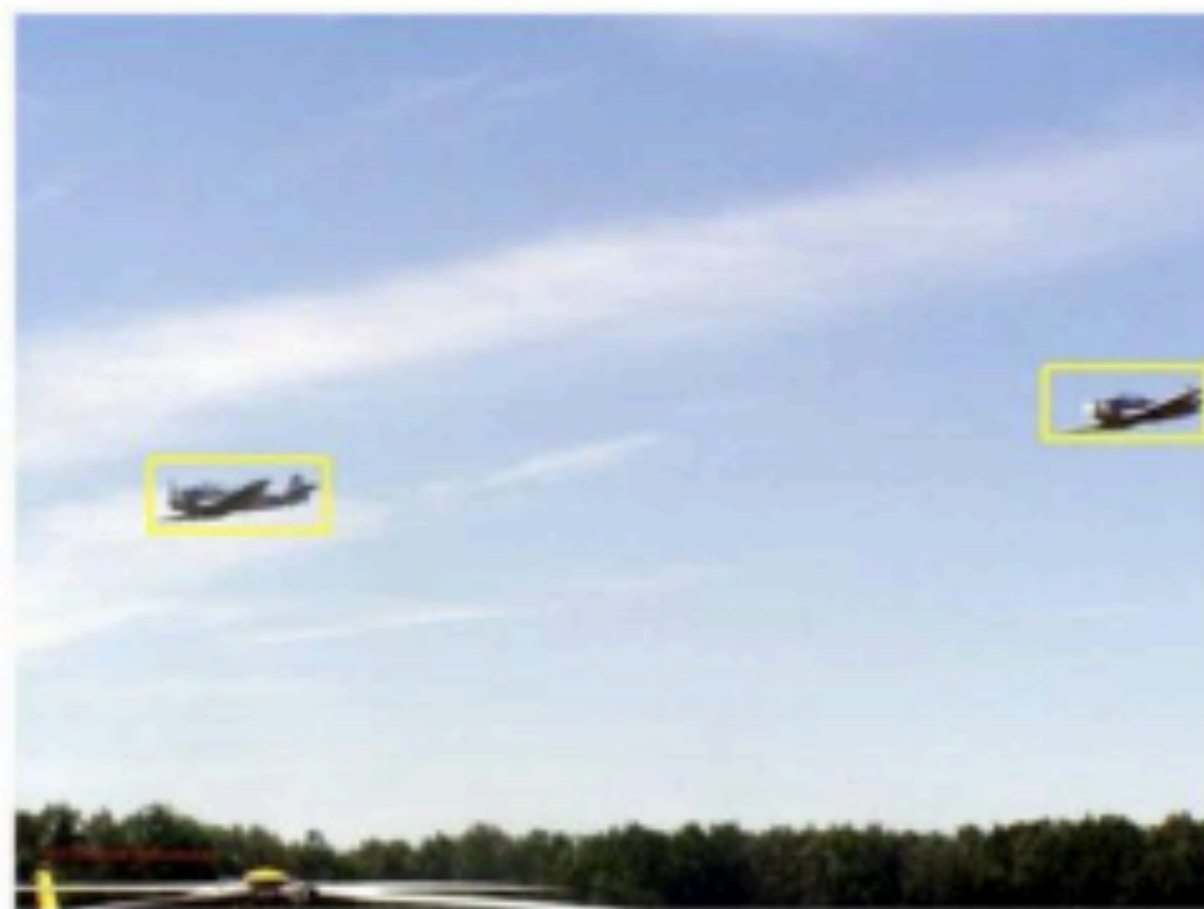
y = bounding boxes



kit fox



croquette



airplane



frog

Two Types of Supervised Learning Algorithms

Classification

- the label is a **discrete** variable

$$y \in \{1, 2, 3, \dots, K\}$$

Regression

- the label is a **continuous** variable

$$y \in \mathbb{R}$$

Training Data for Supervised Learning

Training data is a collection of input instances to the learning algorithm:

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$$

input label

A training data is the “**experience**” given to a learning algorithm

Goal of Supervised Learning

Given training data

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$$

Learn a function mapping $f : X \rightarrow Y$, such that $f(x)$ predicts the label y on **future** data x (not in training data)

Goal of Supervised Learning

Training set error

- 0-1 loss for classification $\ell = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(f(\mathbf{x}_i) \neq y_i)$
- Squared loss for regression: $\ell = \frac{1}{n} \sum_{i=1}^n (f(\mathbf{x}_i) - y_i)^2$

A learning algorithm optimizes the training objective

$$f^* = \arg \min \mathbb{E}_{(x,y)} \ell(f(x), y)$$

Details in upcoming lectures :)

Break & Quiz

Q 1.1: Which is true about feature vectors?

- A. Feature vectors can have at most 10 dimensions
- B. Feature vectors have only integer values
- C. Raw images can be used as feature vectors
- D. Text data cannot be represented as feature vectors

Break & Quiz

Q 1.2: Which of the following is not typically supervised learning?

- A. Object detection (identifying bounding boxes on objects)
- B. Classification
- C. Regression
- D. Dimensionality Reduction (e.g., PCA)



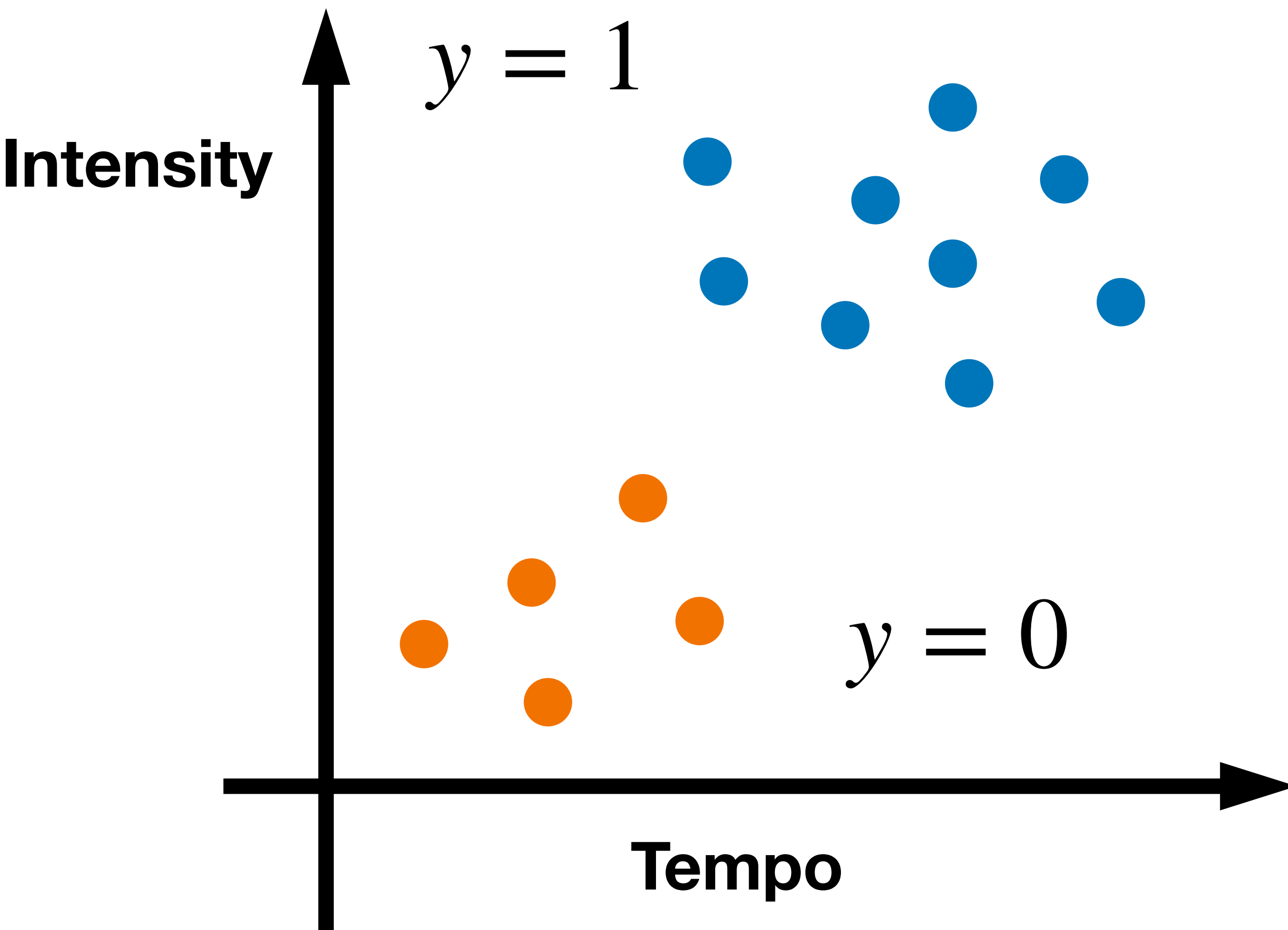
Part II: Unsupervised Learning

Unsupervised Learning

- Given: dataset contains **no label**: x_1, x_2, \dots, x_n
- **Goal**: discover interesting patterns and structures in the data

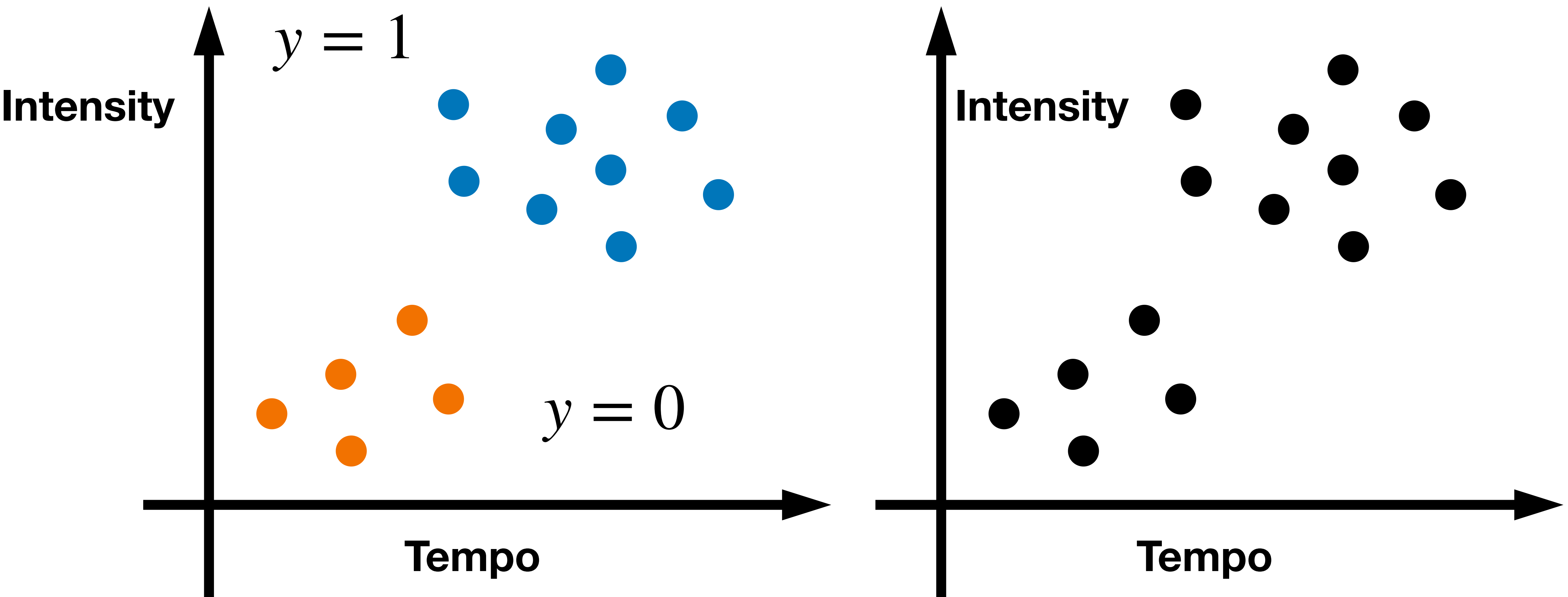
Unsupervised Learning

- Given: dataset contains **no label**: x_1, x_2, \dots, x_n
- **Goal**: discover interesting patterns and structures in the data



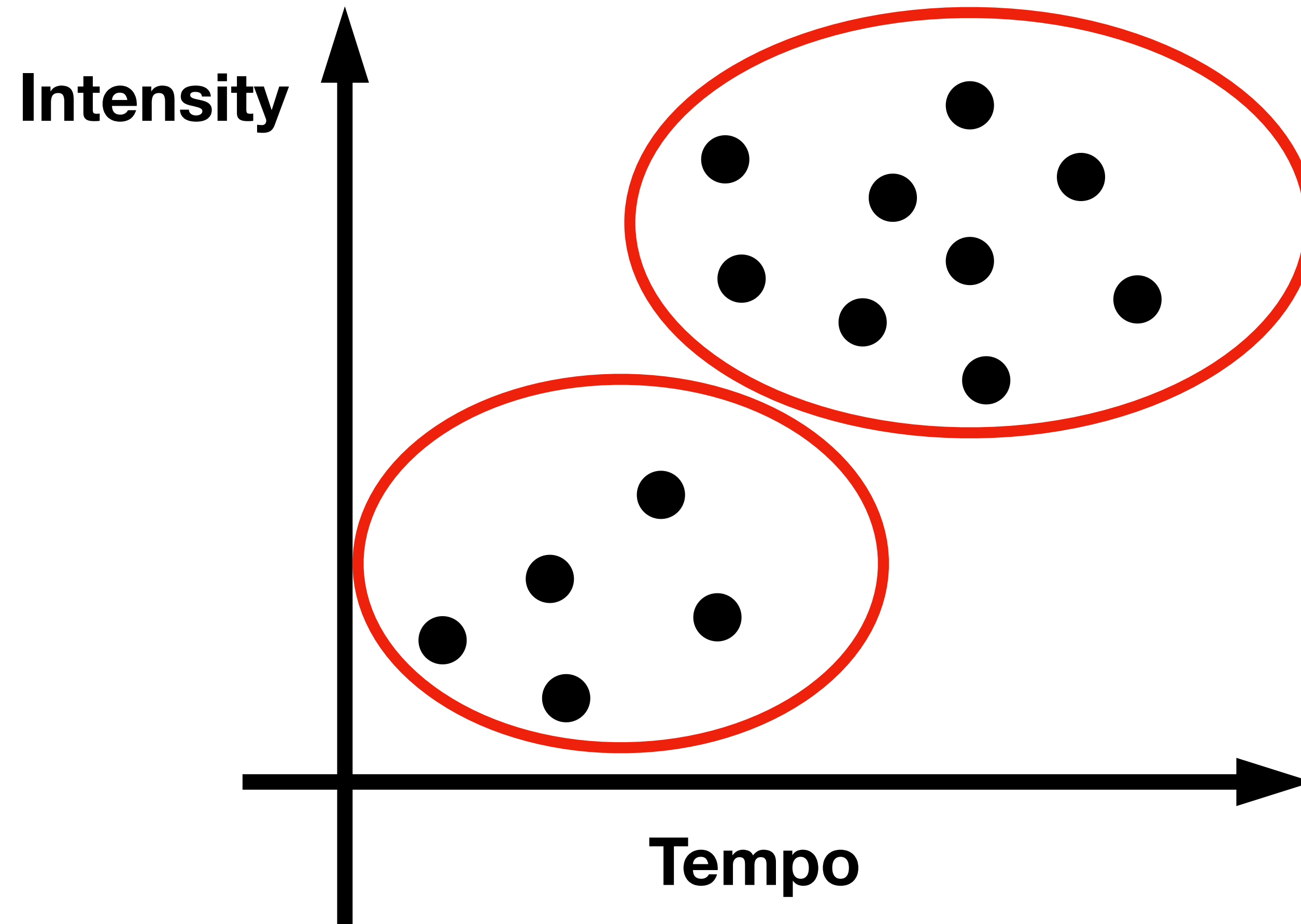
Unsupervised Learning

- Given: dataset contains **no label**: x_1, x_2, \dots, x_n
- **Goal**: discover interesting patterns and structures in the data

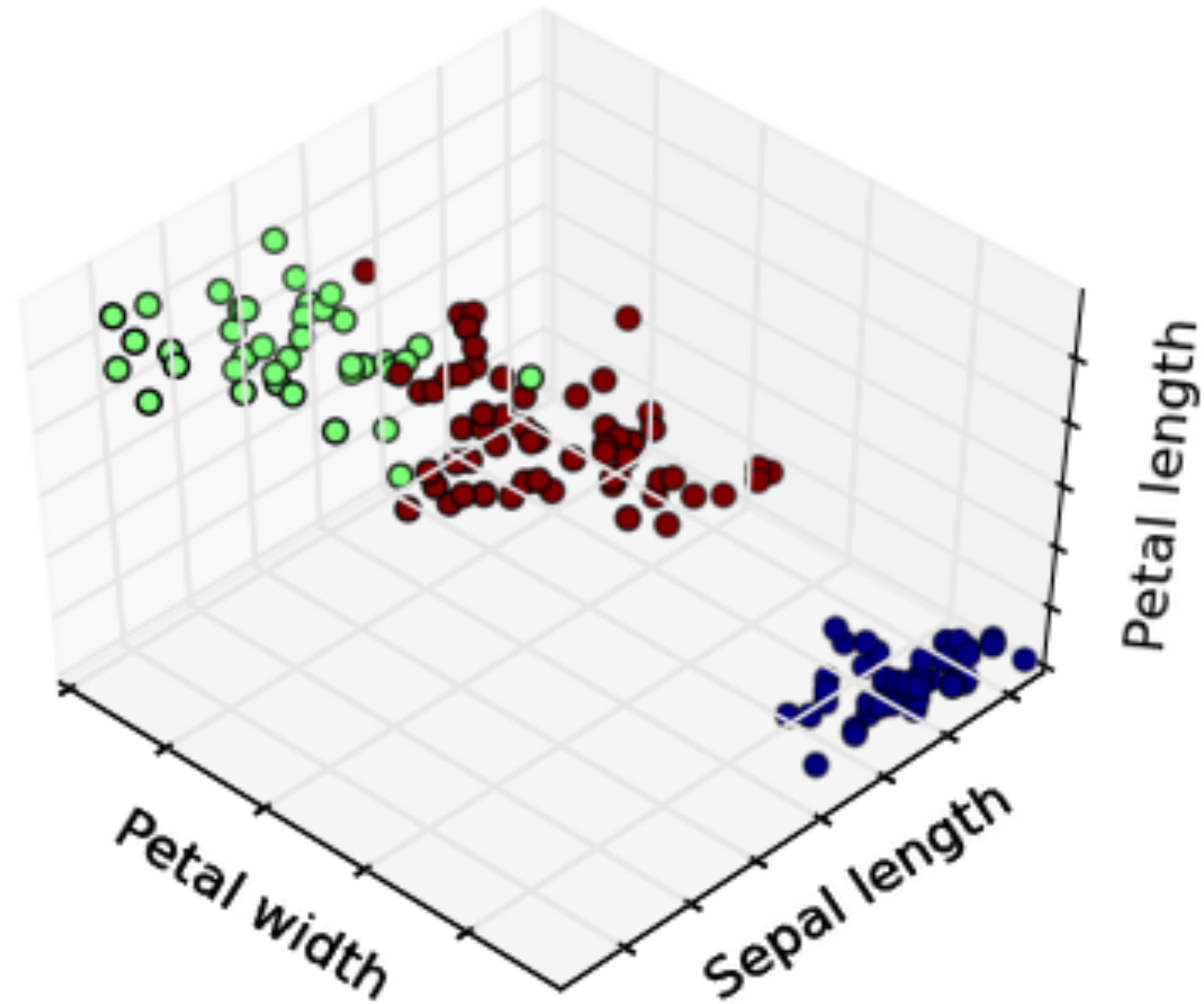


Clustering

- Given: dataset contains **no label**: x_1, x_2, \dots, x_n
- **Output**: divides the data into clusters such that there are intra-cluster similarity and inter-cluster dissimilarity



Clustering

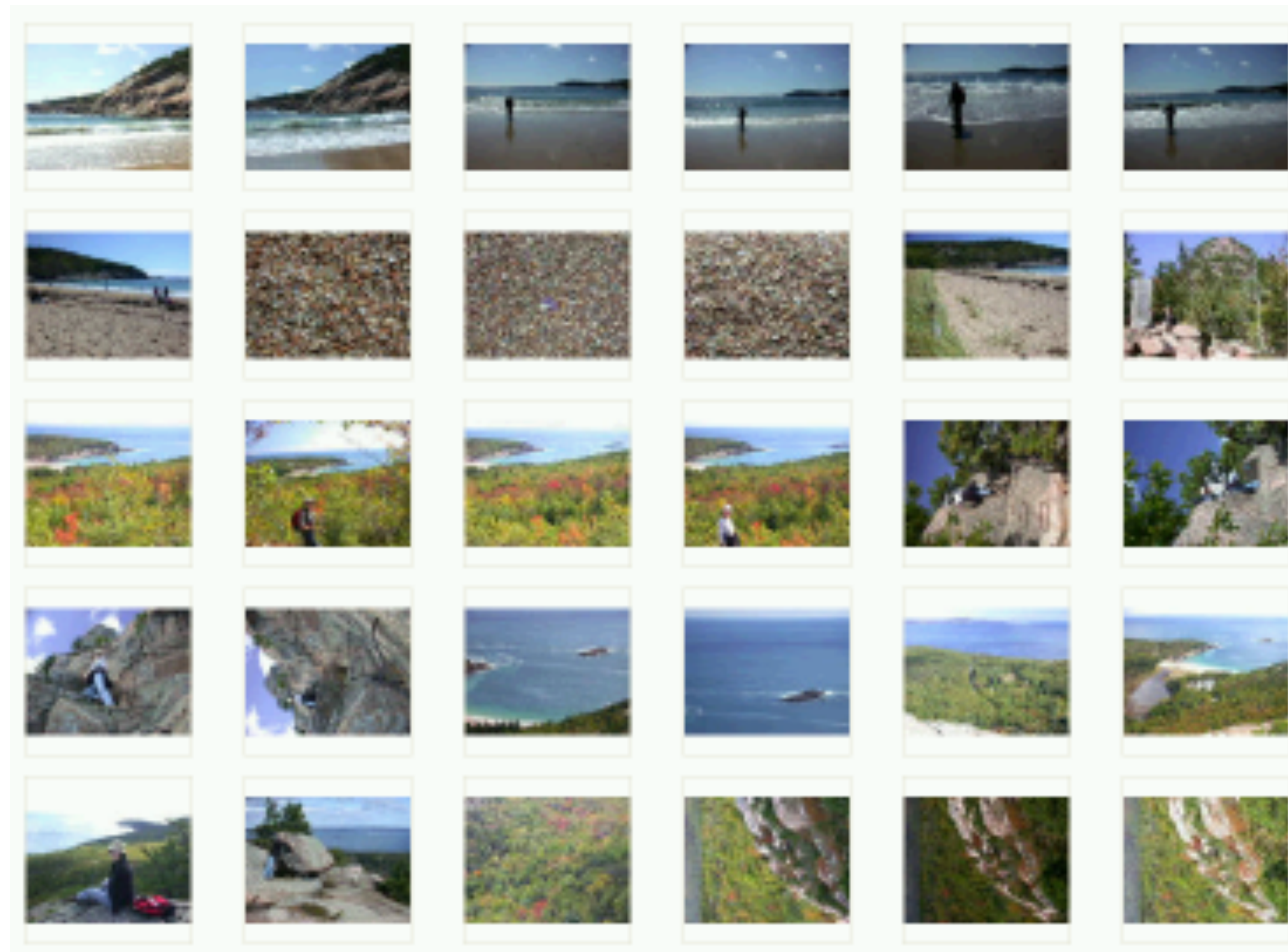


Clustering Irises using three different features

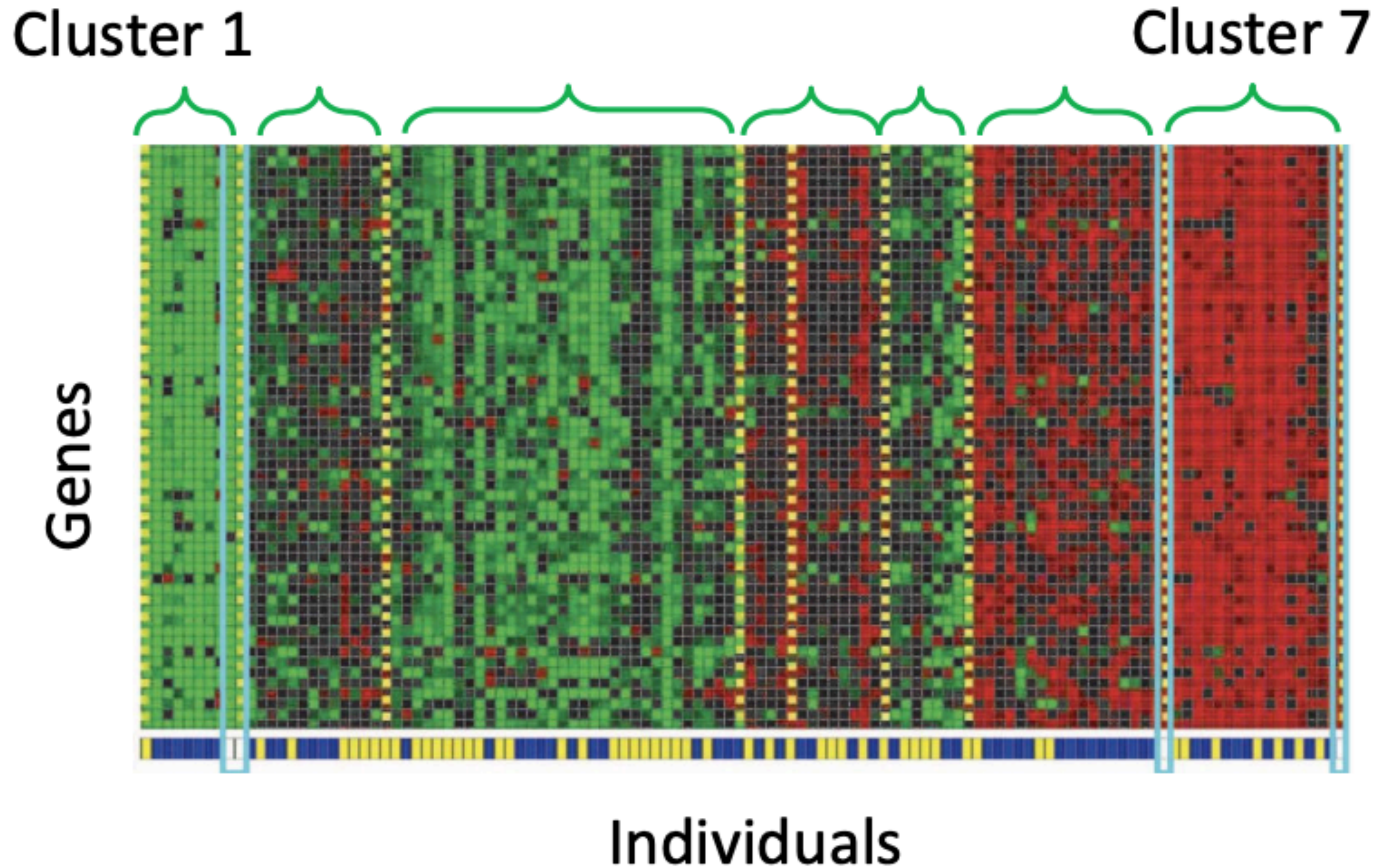
The colors represent clusters identified by the algorithm, not provided as input

Clustering

- You probably have >1000 digital photos stored on your phone
- After this class you will be able to organize them better (based on visual similarity)

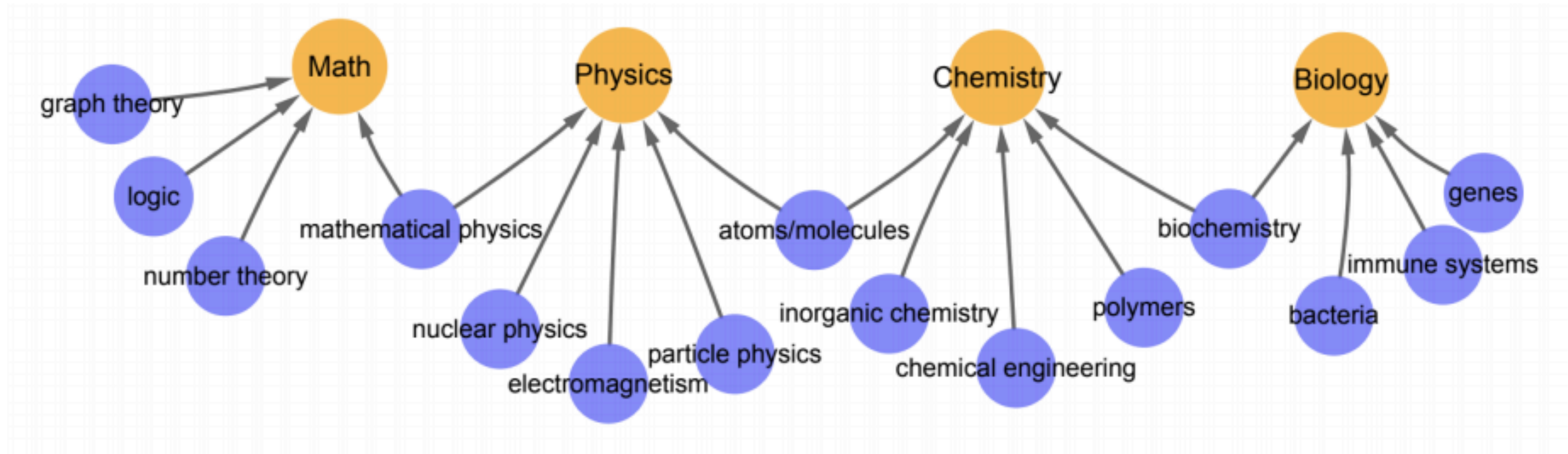


Clustering Genes



Identifying Regulatory Mechanisms using Individual Variation Reveals Key Role for Chromatin Modification. [Su-In Lee, Dana Pe'er, Aimee M. Dudley, George M. Church and Daphne Koller. '06]

Clustering Words with Similar Meanings



[Arora-Ge-Liang-Ma-Risteski, TACL'17,18]

How do we perform clustering?

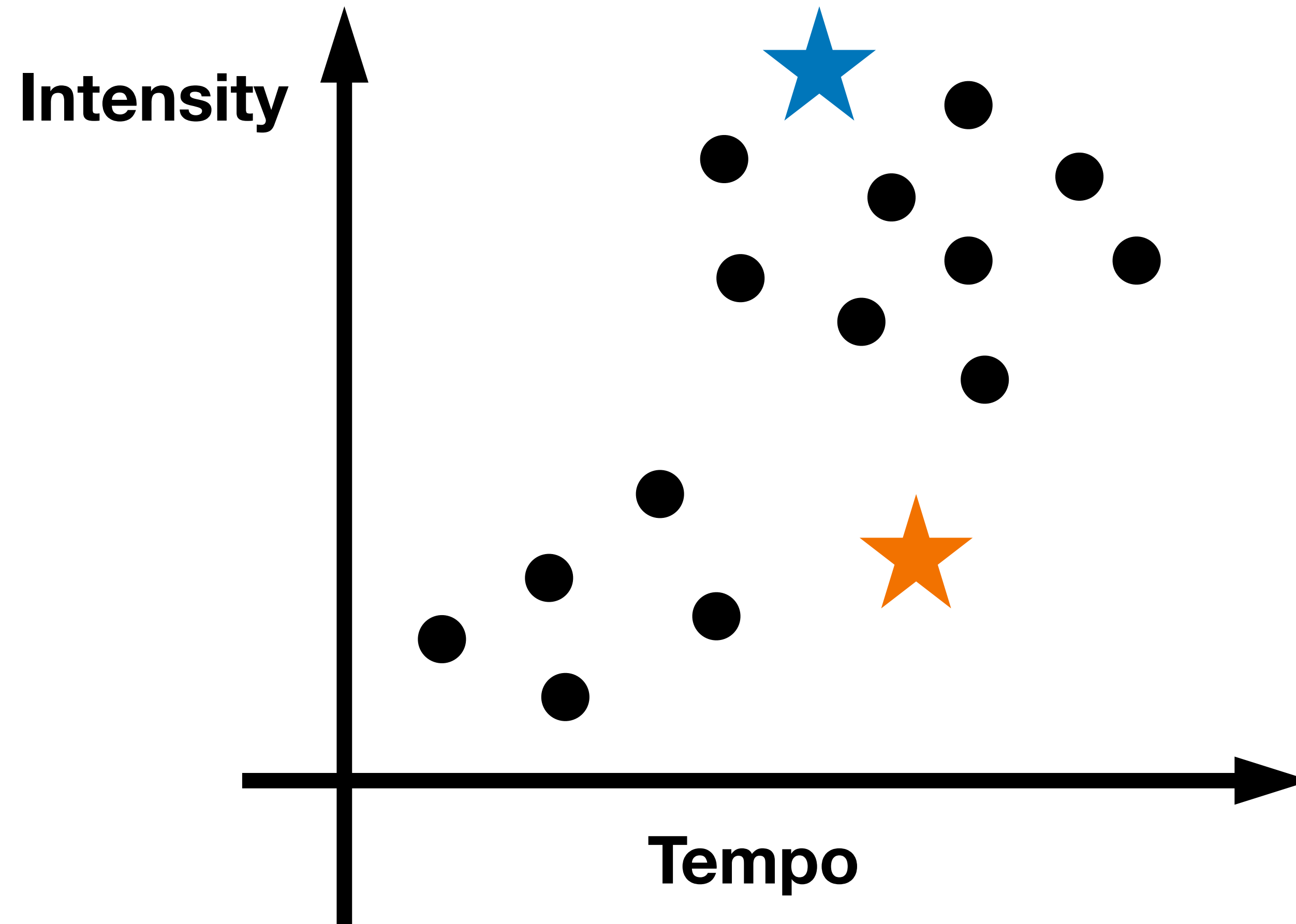
- Many clustering algorithms. We will look at the two most frequently used ones:
 - *K-means clustering*: we specify the desired number of clusters, and use an iterative algorithm to find them
 - *Hierarchical clustering*: we build a binary tree over the dataset

K-means clustering

- Very popular clustering method
- Don't confuse it with k-NN classifier
- Input: a dataset x_1, x_2, \dots, x_n , and assume the number of clusters **k** is given

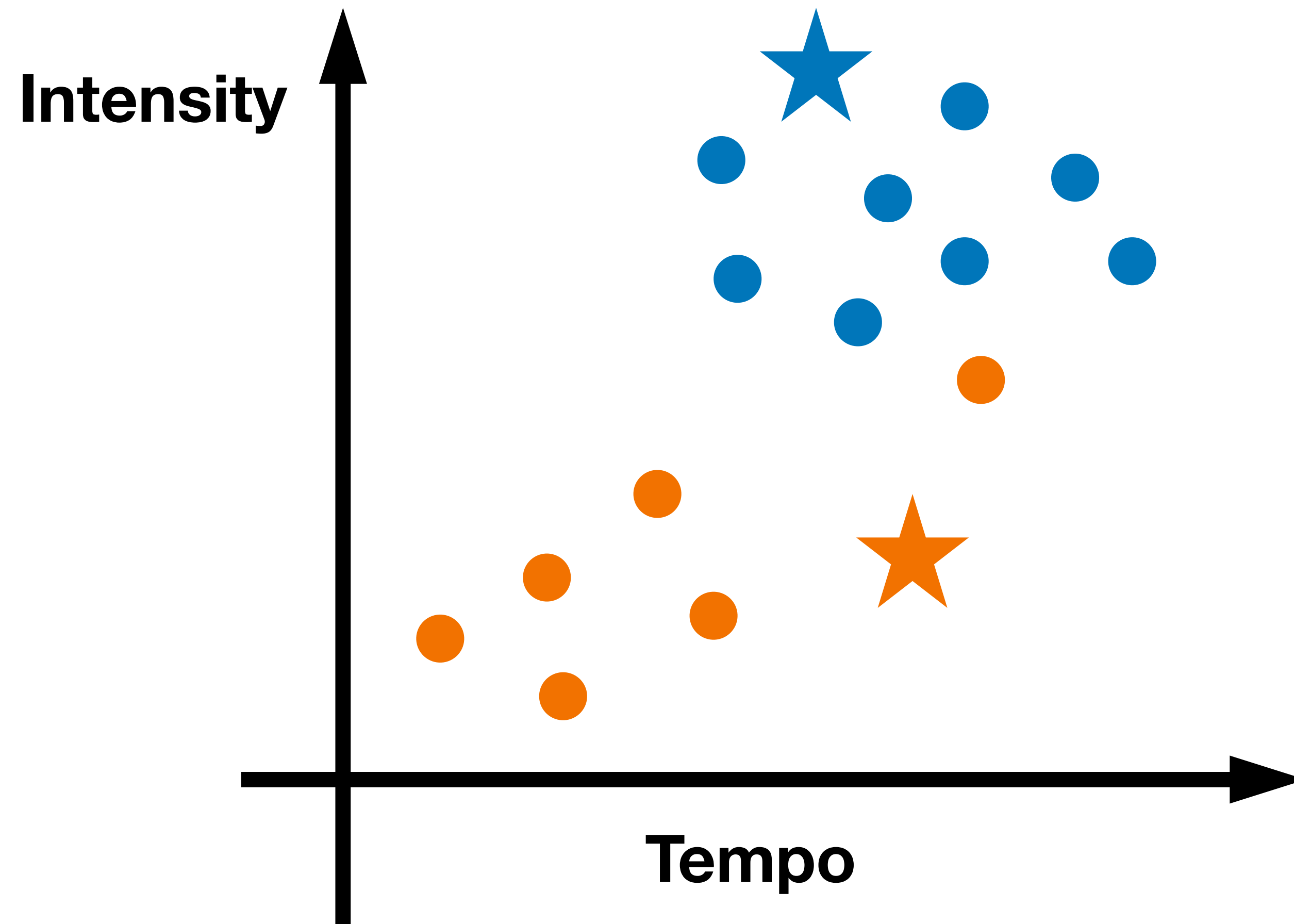
K-means clustering

Step 1: **Randomly** picking 2 positions as initial cluster centers (not necessarily a data point)



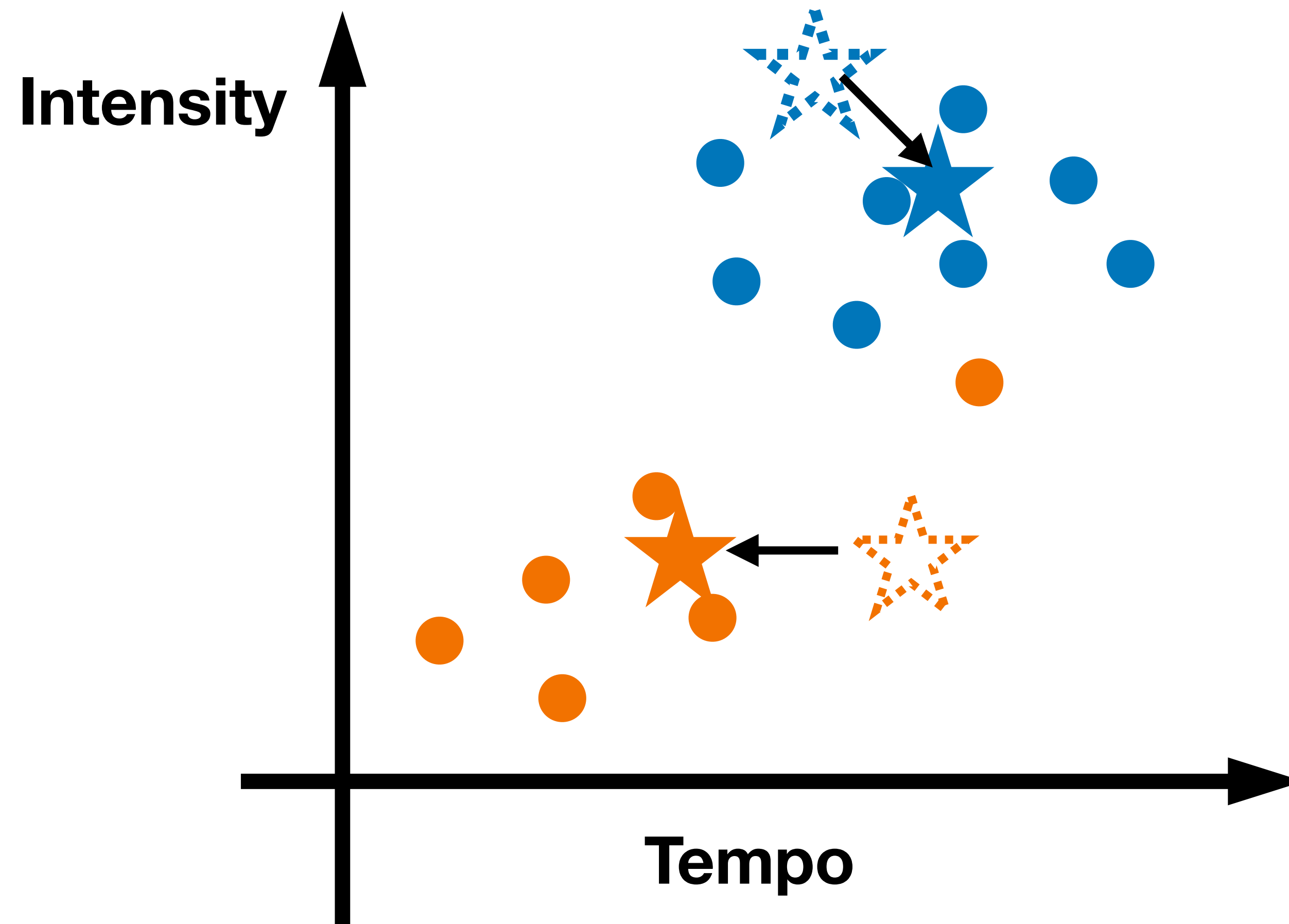
K-means clustering

Step 2: for each point x , determine its cluster: find the closest center in Euclidean space



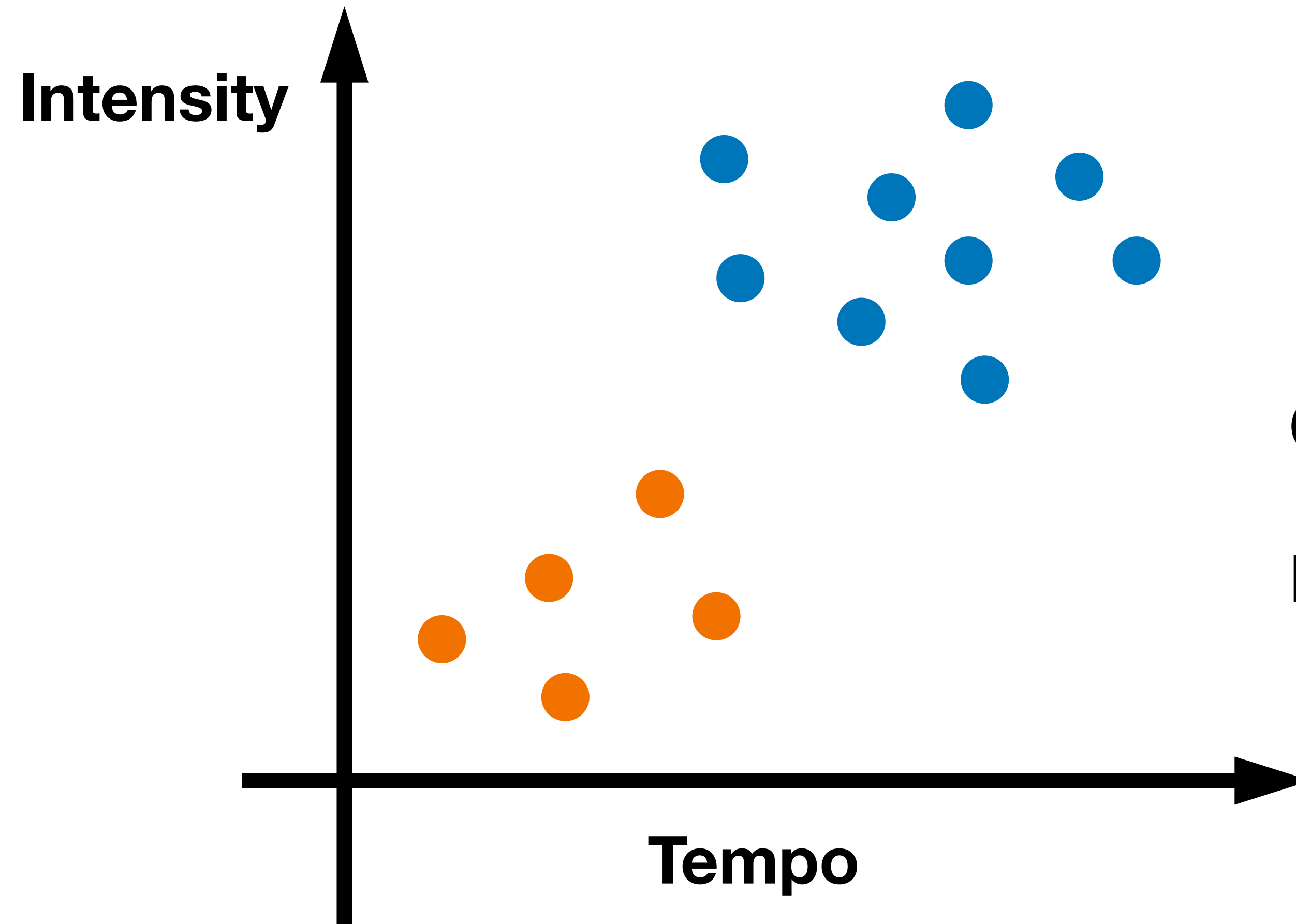
K-means clustering

Step 3: update all cluster centers as the centroids



K-means clustering

Repeat step 2 & 3 until convergence



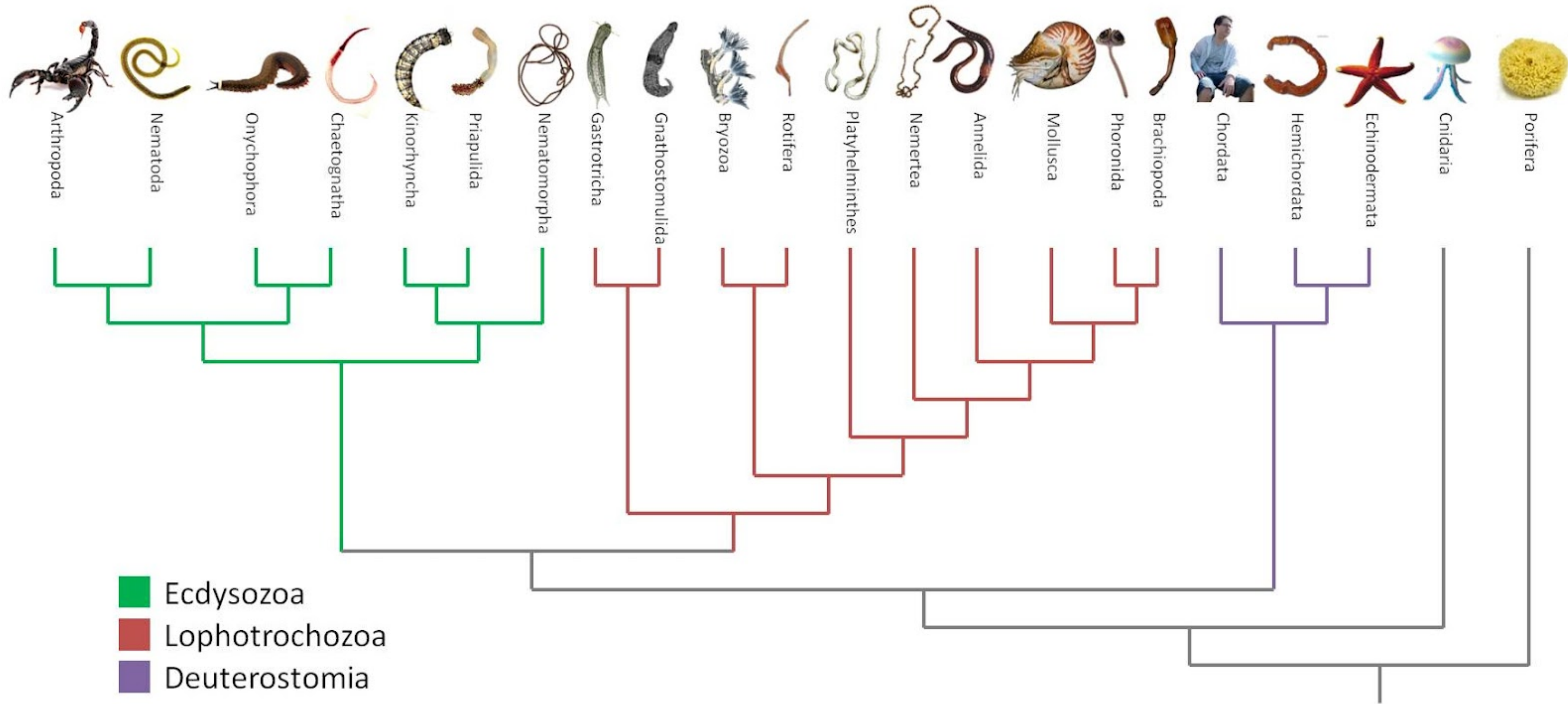
Converged solution!

No labels required!

K-means clustering: A demo

<https://www.naftaliharris.com/blog/visualizing-k-means-clustering/>

Hierarchical Clustering (more to follow next lecture)



Break & Quiz

Q 1.2: Which is true about supervised learning?

- A. The process doesn't involve human input
- B. The machine is learning from training and test data
- C. Clustering data makes use of labelled data
- D. Supervised learning requires labels

Break & Quiz

Q 1.2: Which is true about unsupervised learning?

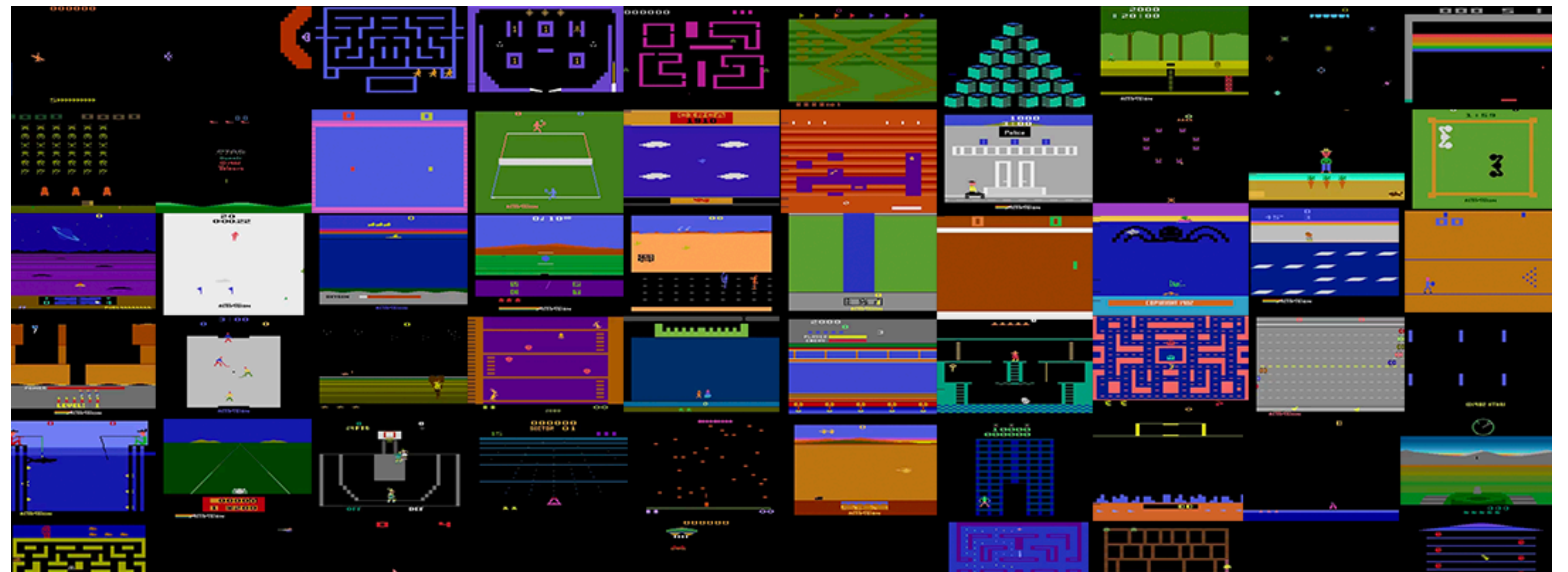
- A. There are only 2 types of unsupervised learning algorithms.
- B. K-means clustering is a type of hierarchical clustering.
- C. K-means clustering automatically determines the number of clusters.
- D. Unsupervised learning is widely used in applications.



Part III: Reinforcement Learning

Reinforcement Learning

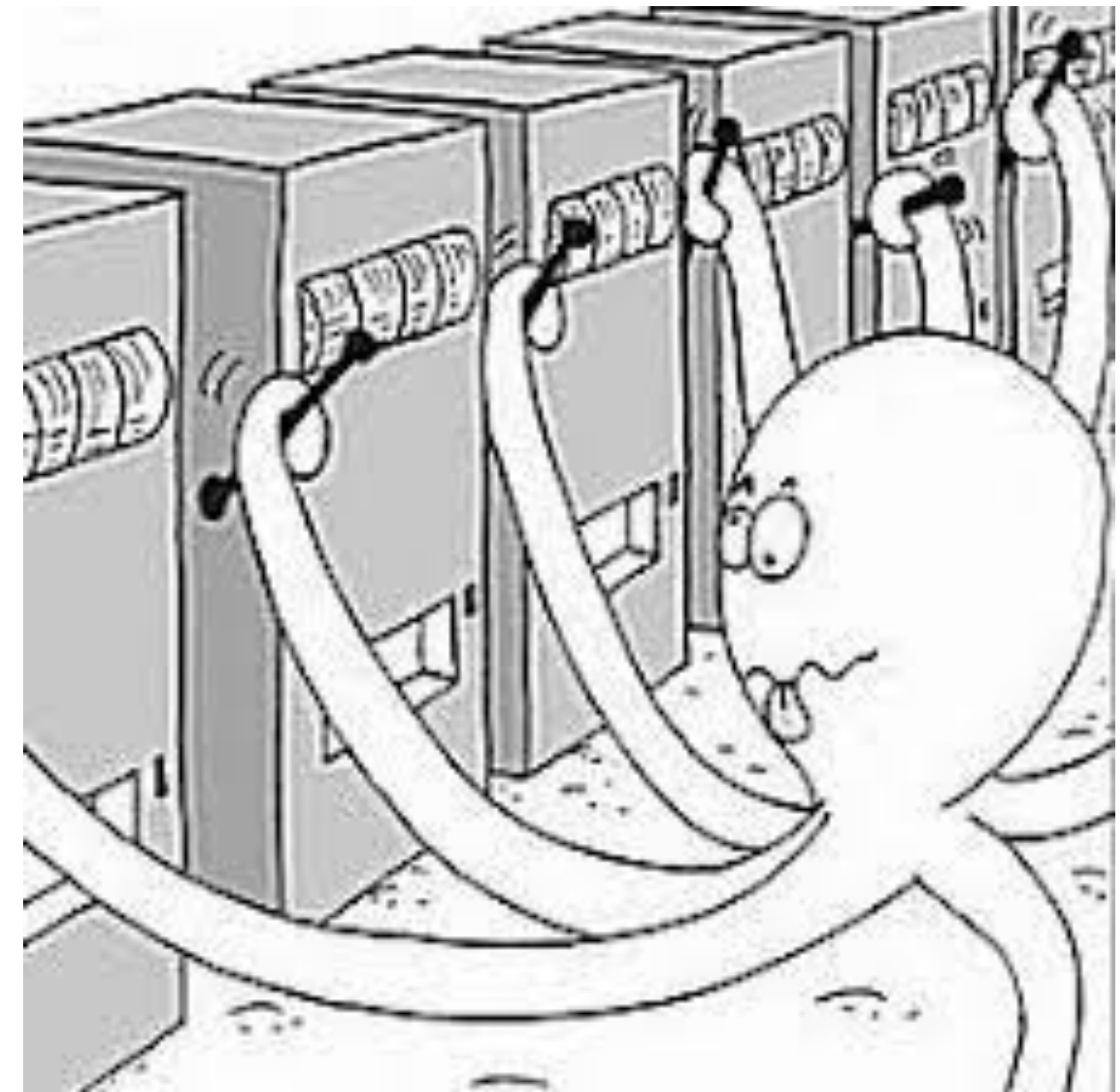
- Given: an agent that can take **actions** and a **reward** function specifying how good an action is.
- **Goal:** learn to choose actions that maximize total future reward.



Google Deepmind

Reinforcement Learning Key Problems

1. Problem: maximal reward action is unknown
 - **Exploration-exploitation** trade-off
 - Try new restaurants, or stick with known ones?
2. Problem: actions may have delayed effects.
 - Requires **credit-assignment**
 - Which moves are good in a chess game?



Multi-armed Bandit

Today's recap

- What is machine learning?
- Supervised Learning
 - Classification
 - Regression
- Unsupervised Learning
- Reinforcement Learning



Thanks!