CS 540 Introduction to Artificial Intelligence

Machine Learning Overview

University of Wisconsin-Madison

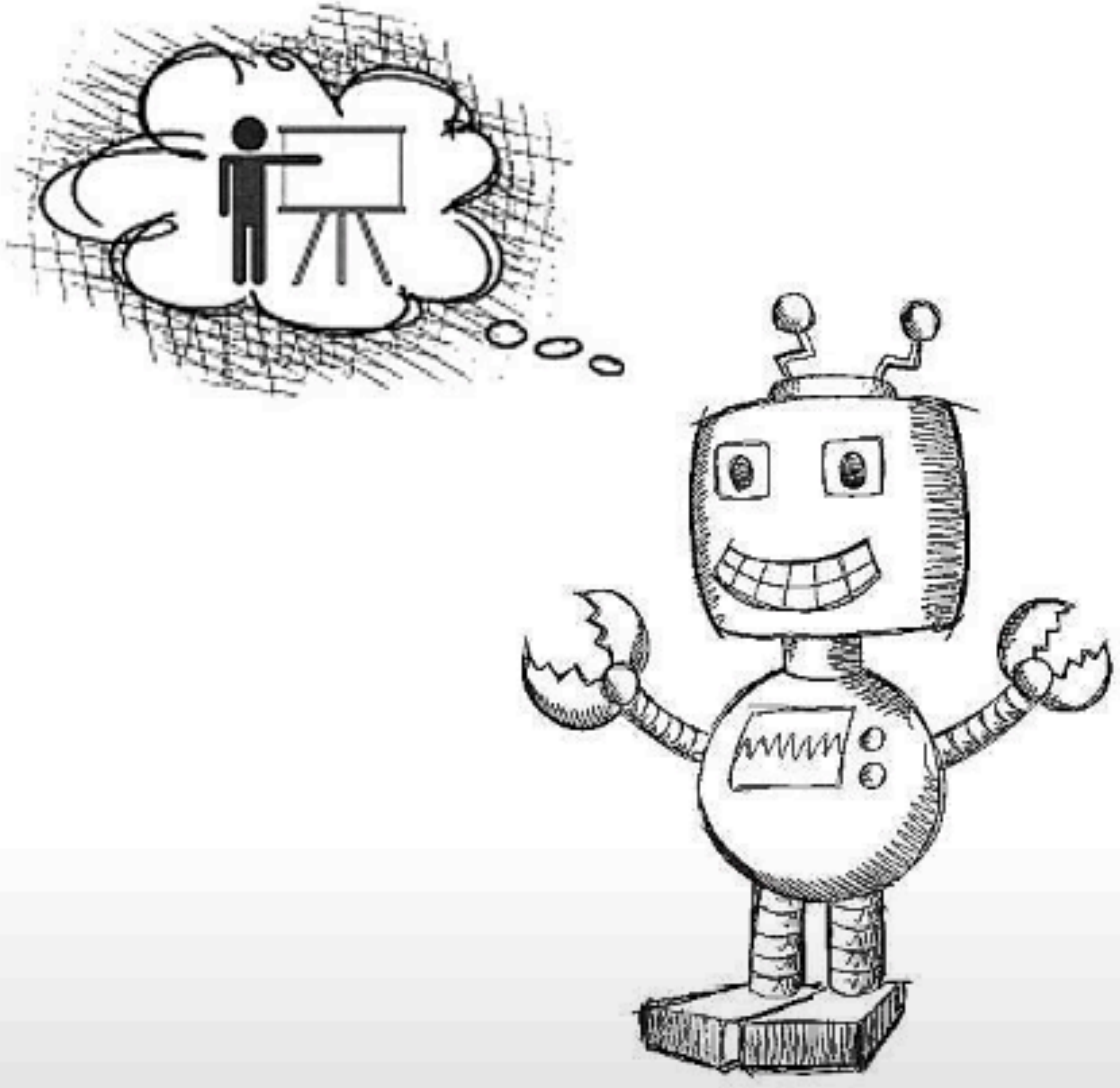
Spring 2022

Today's learning goal

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

Part I: What is machine learning?

HUMANS LEARN FROM PAST EXPERIENCES

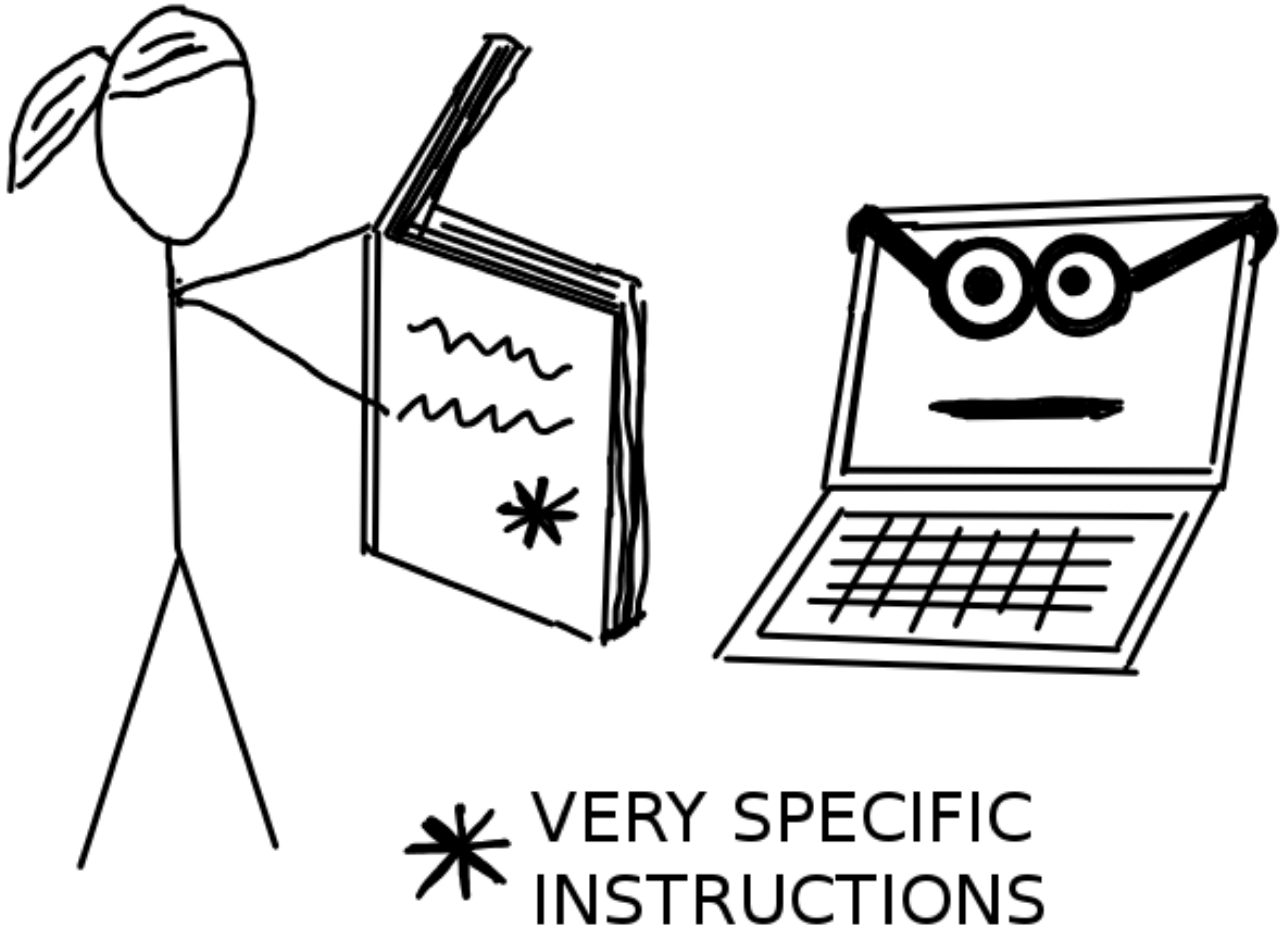


MACHINES FOLLOW INSTRUCTIONS GIVEN BY HUMANS

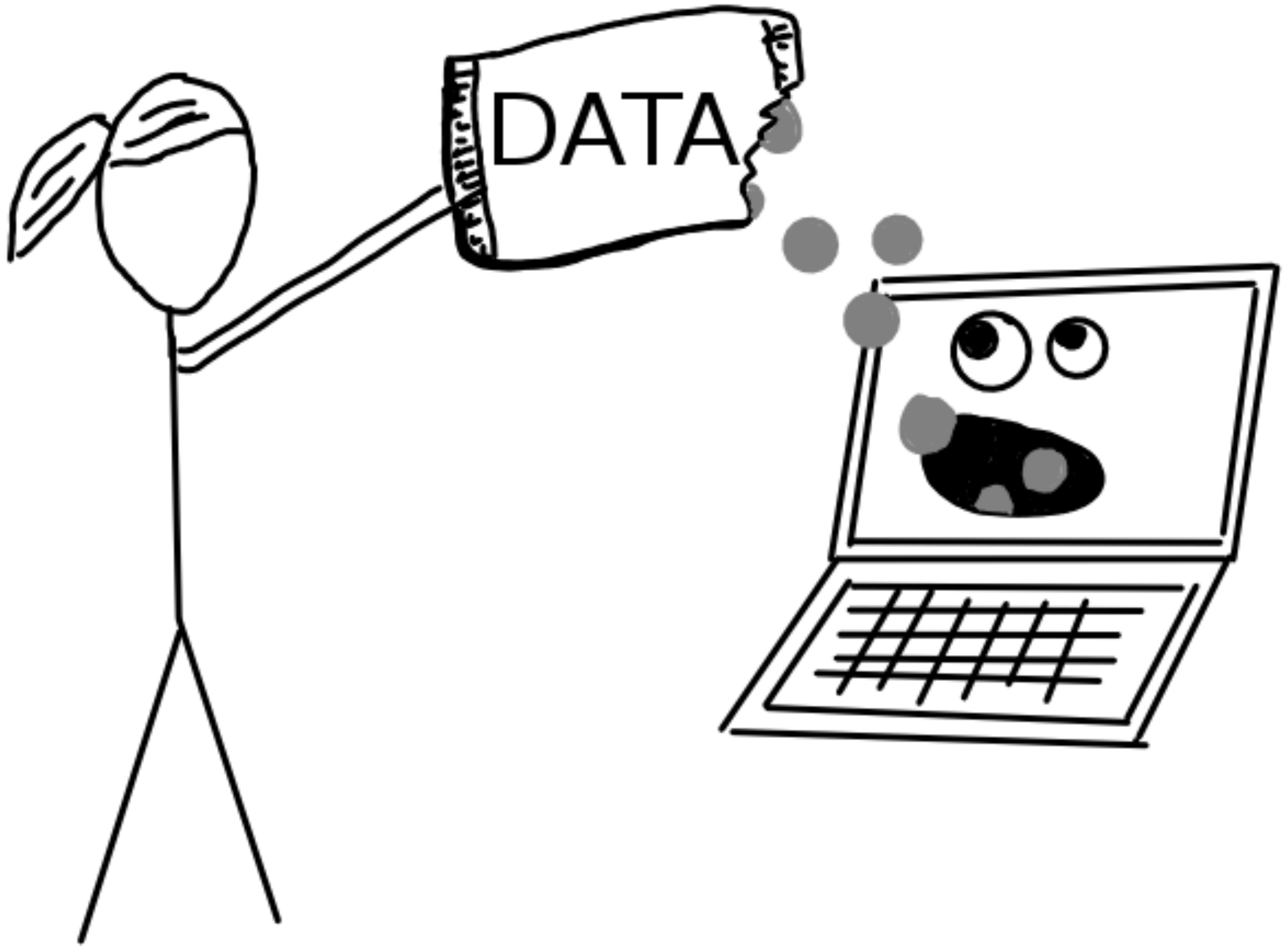
What is machine learning?

- 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



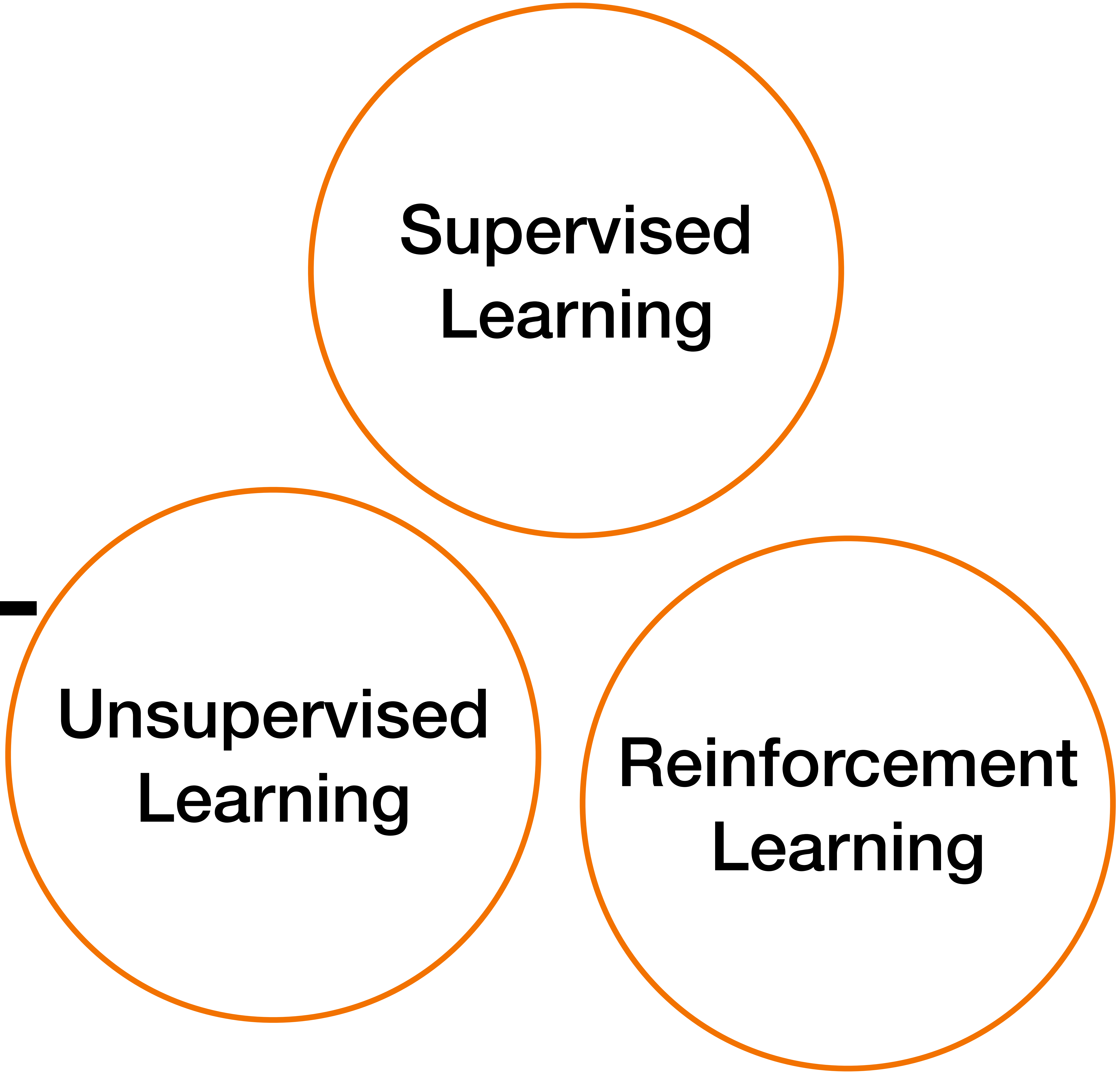
https://tung-dn.github.io/programming.html

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**.

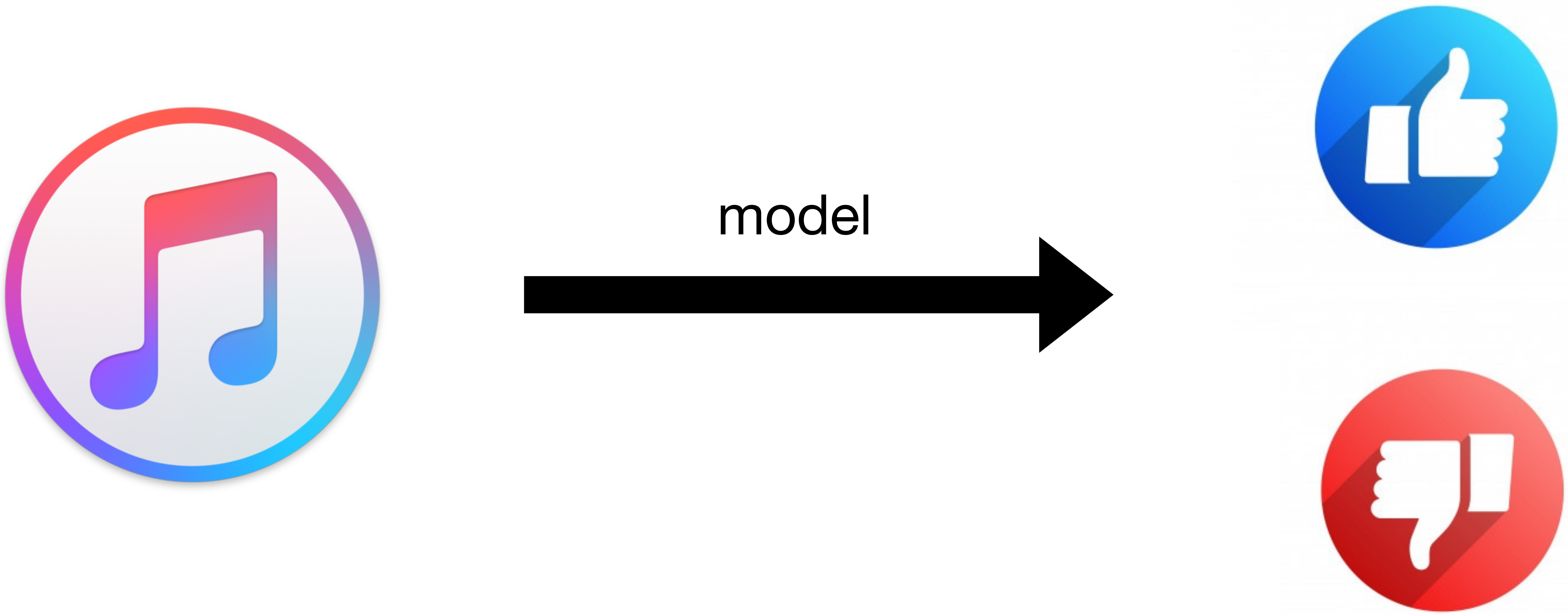
- Tom Mitchell (1997): A computer program is said to learn from **experience E** with respect to some class of **tasks T** and **performance measure P**, if its performance at tasks in T as measured by P, improves with experience E.

Taxonomy of ML

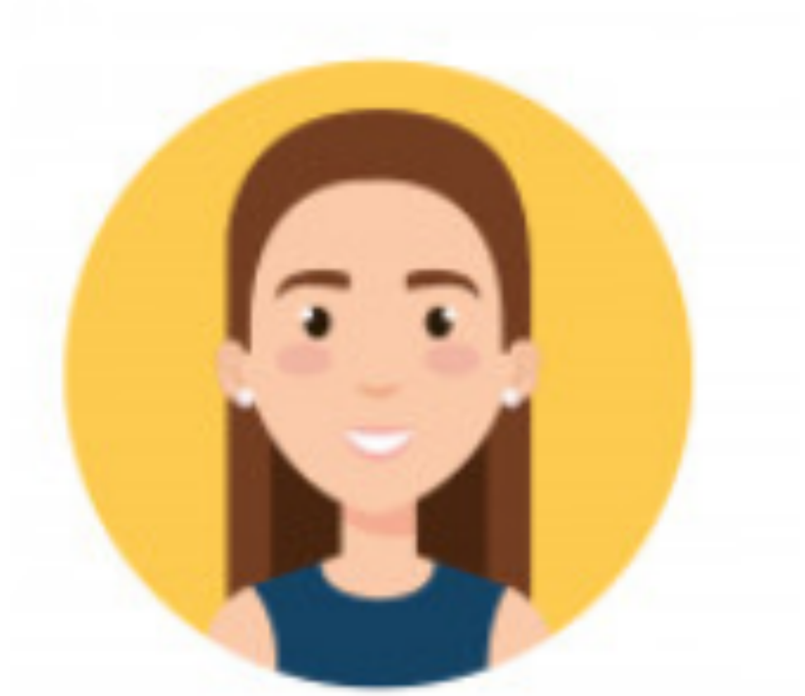


Part II: Supervised Learning

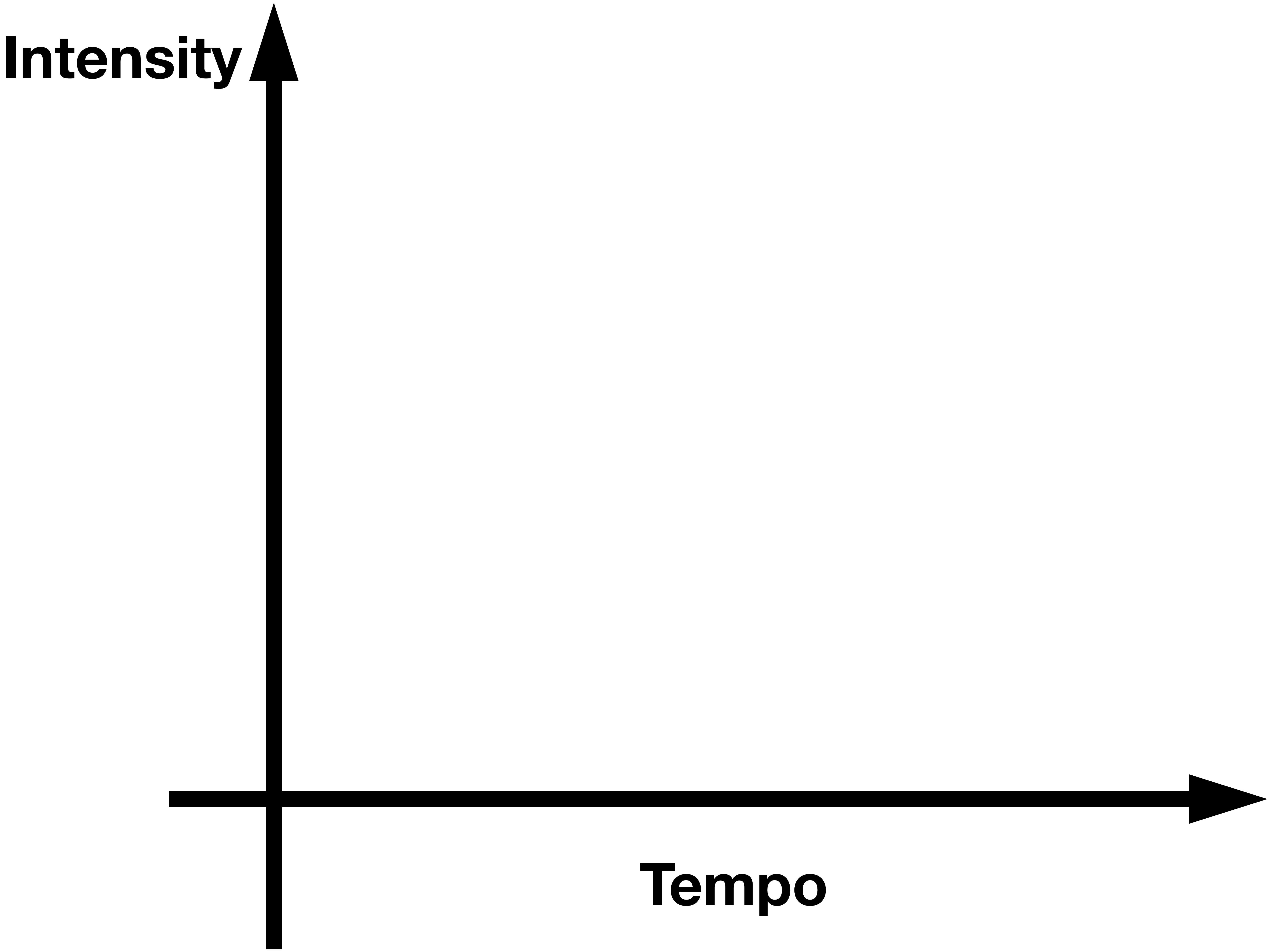
Example 1: Predict whether a user likes a song or not



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

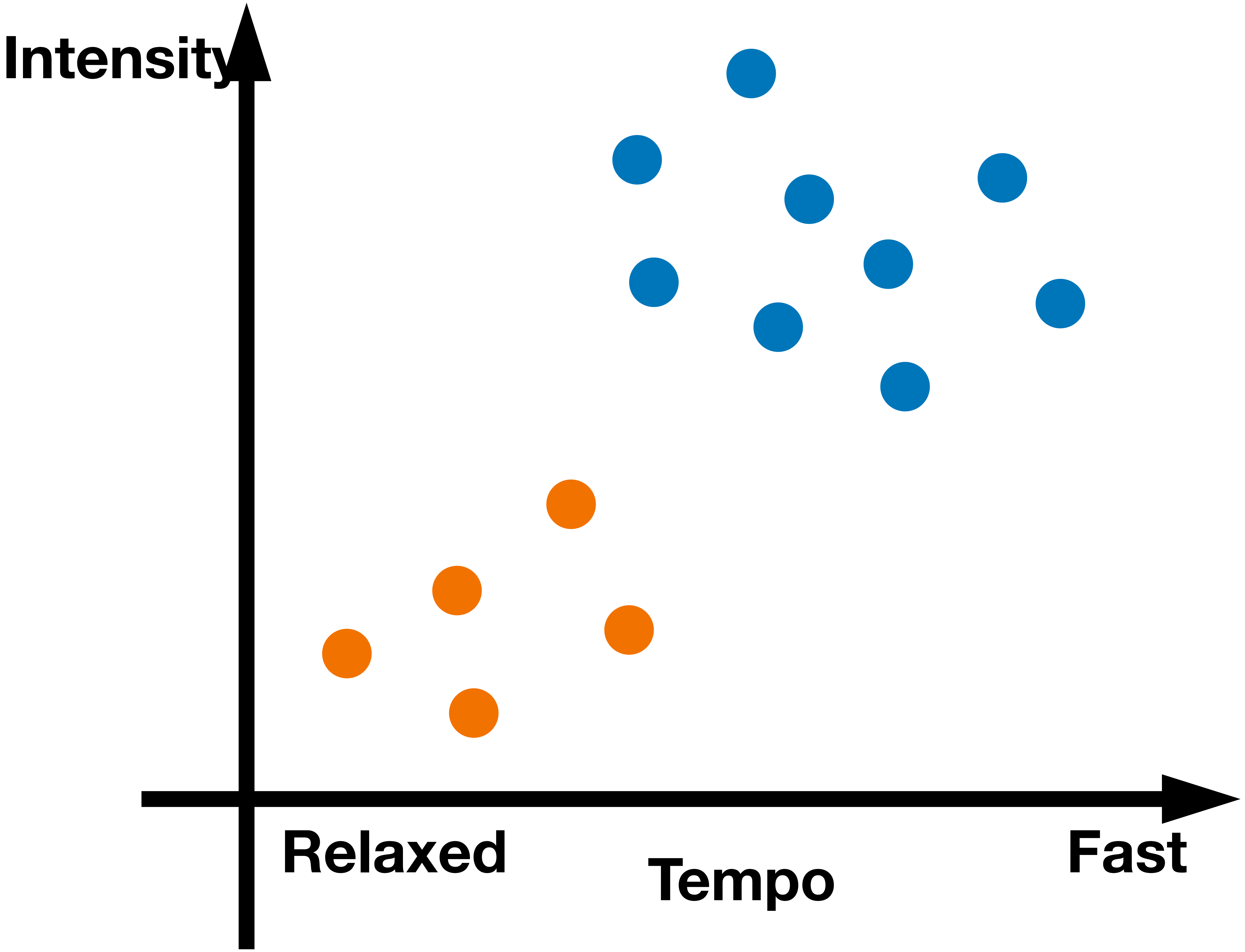


User Sharon

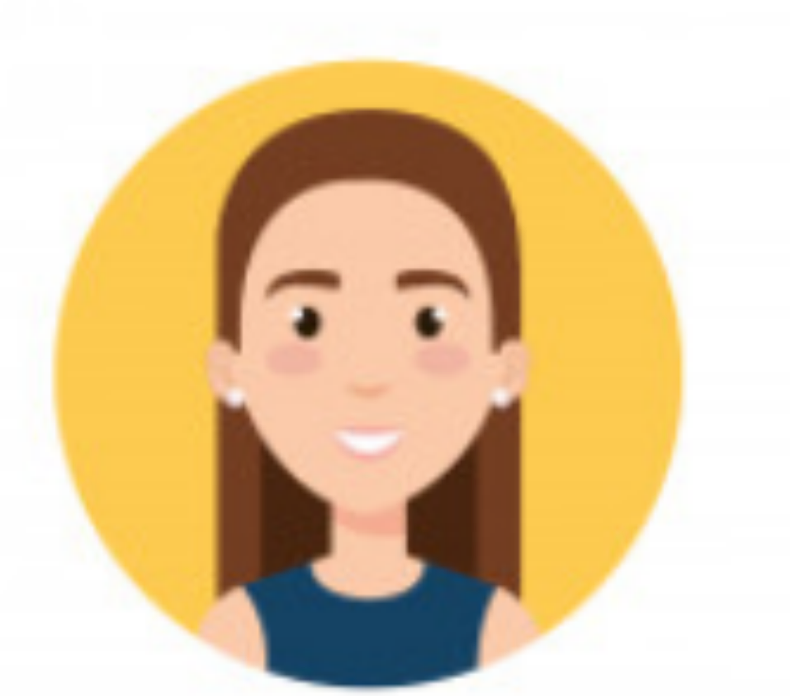


Example 1: Predict whether a user likes a song or not

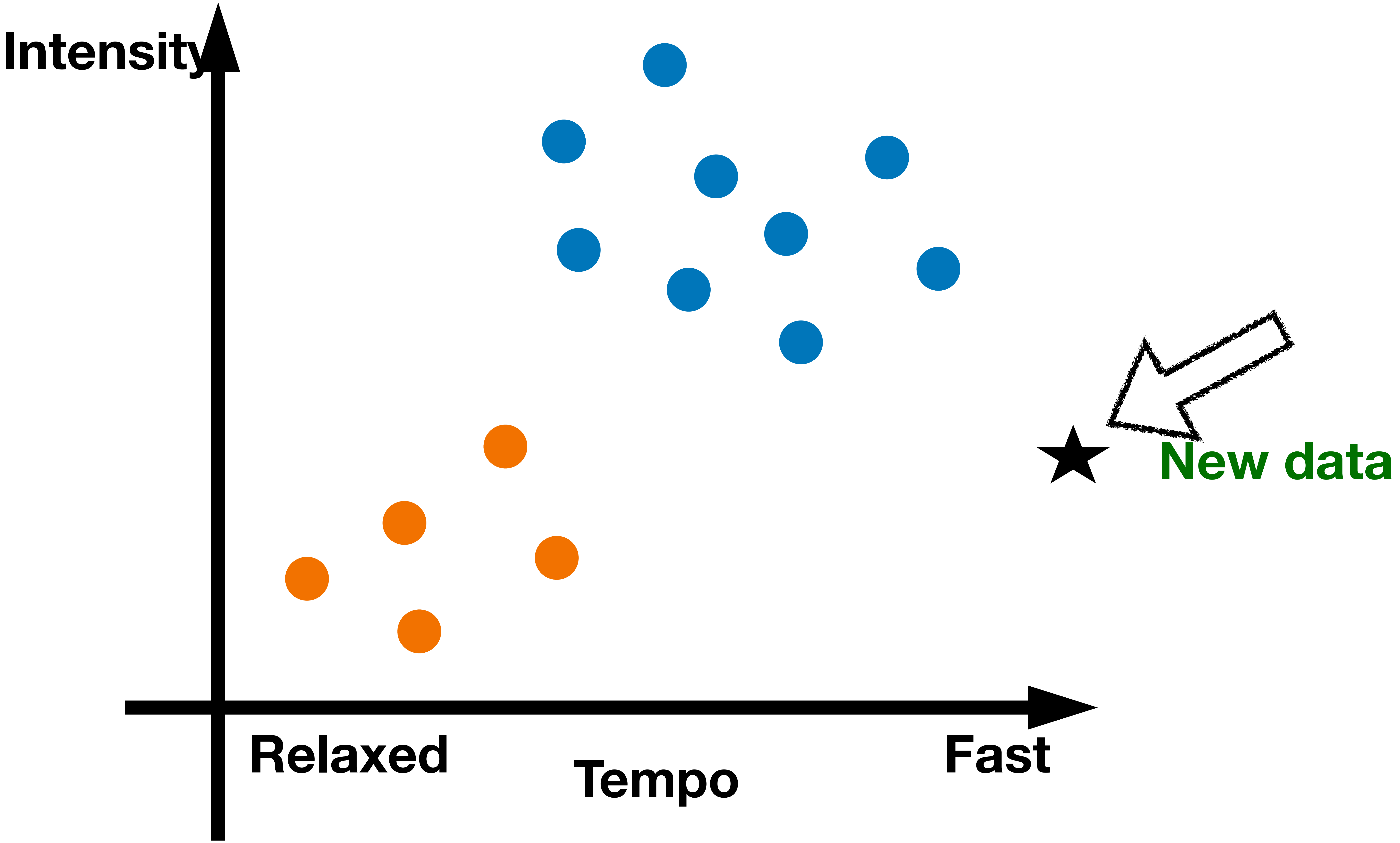
User Sharon



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



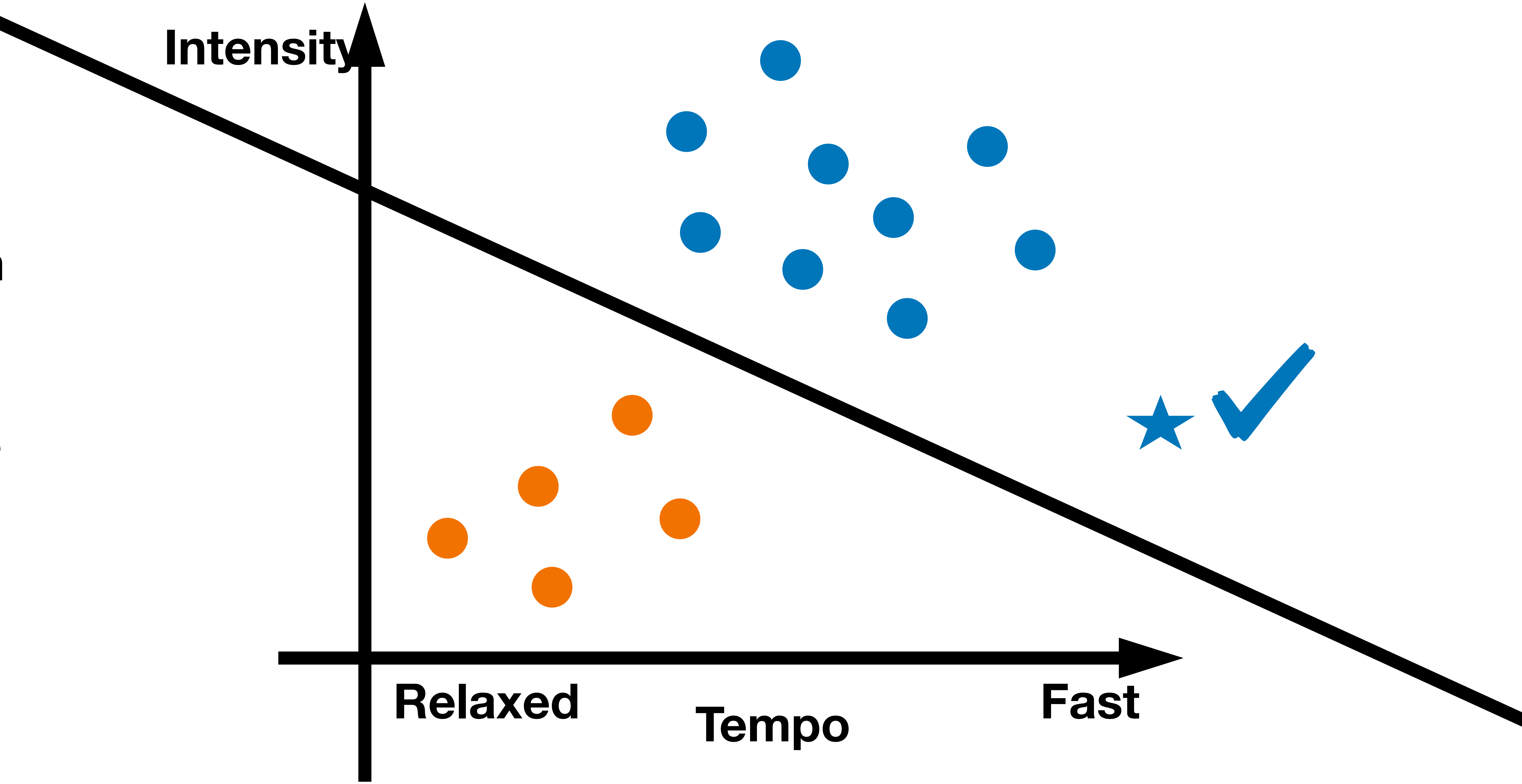
User Sharon



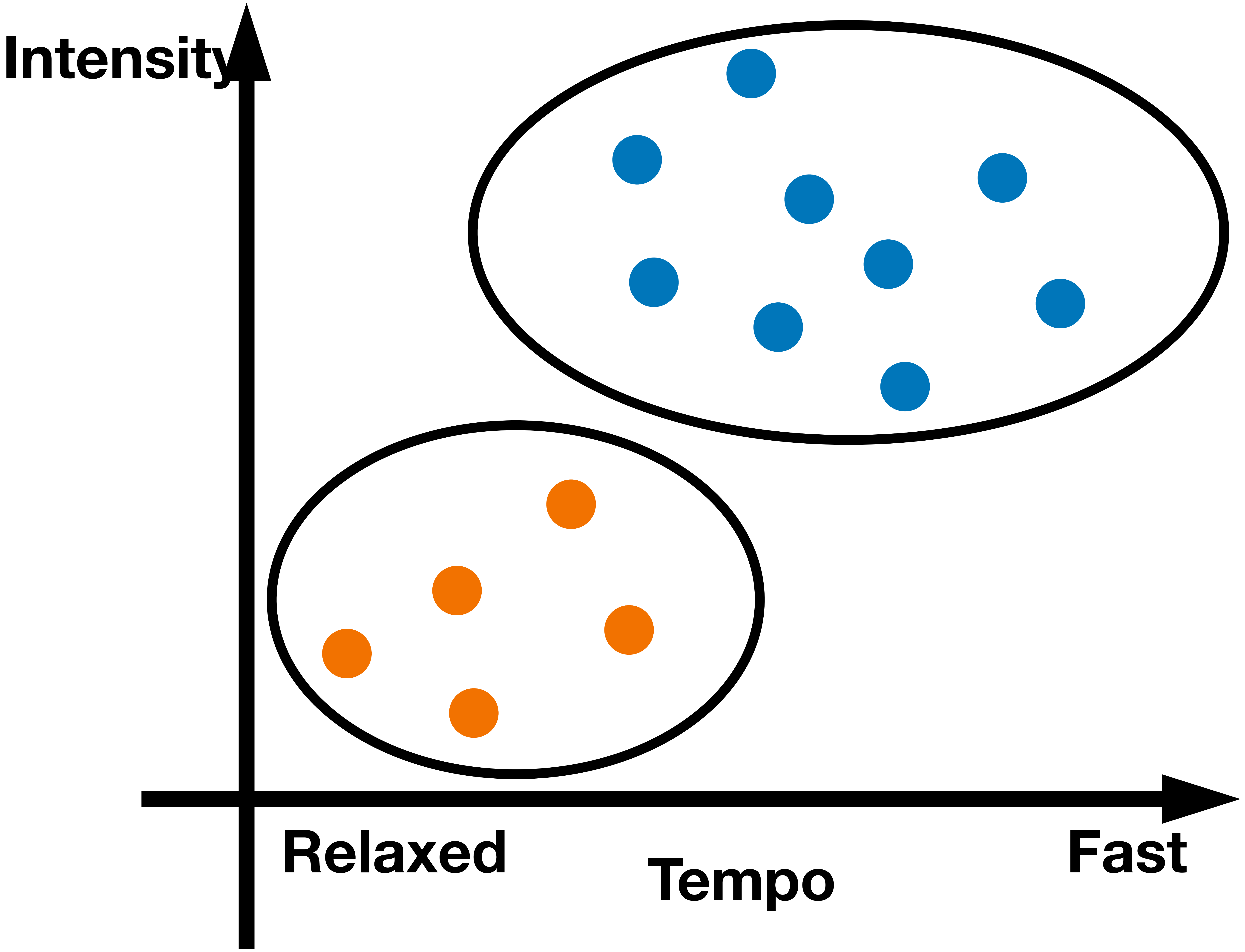
Example 1: Predict whether a user likes a song or not

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

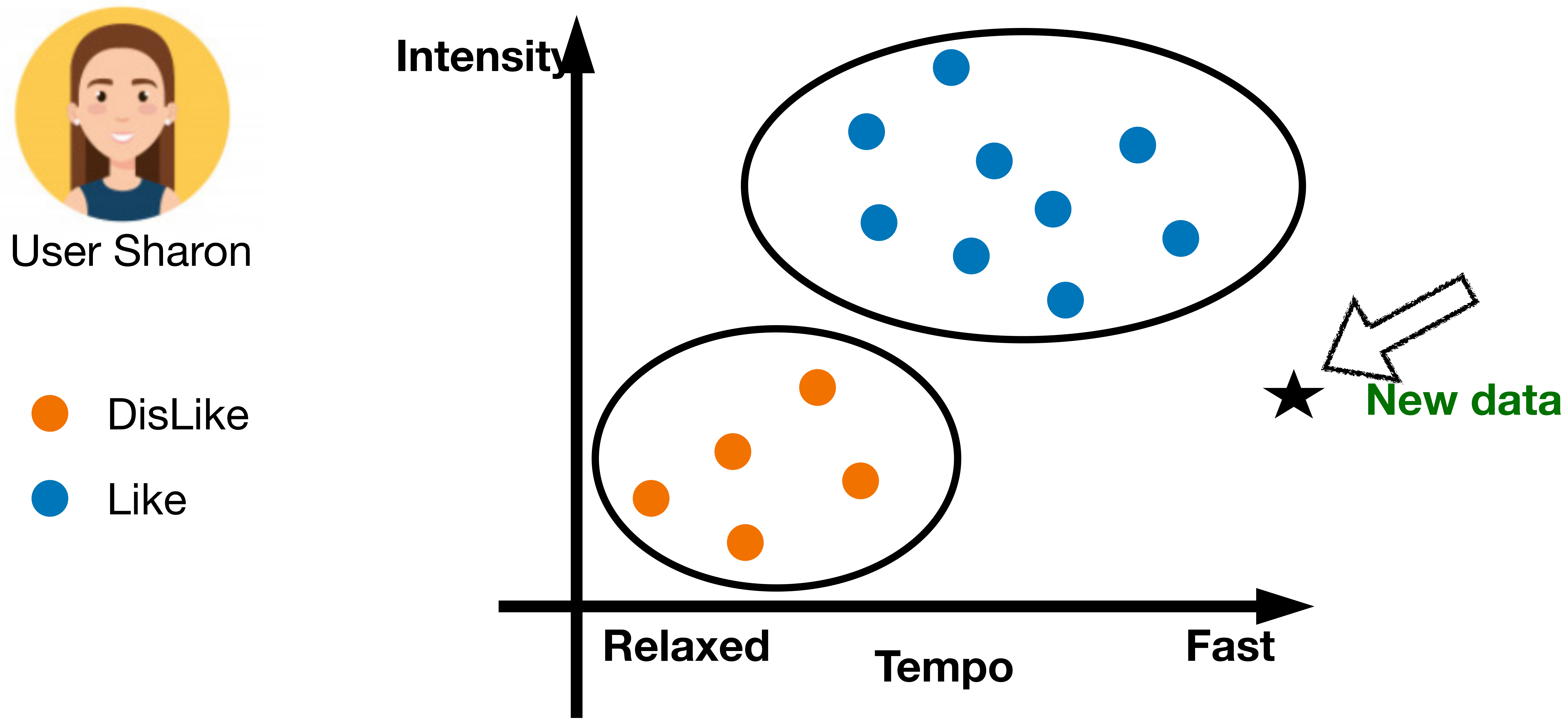
User Sharon



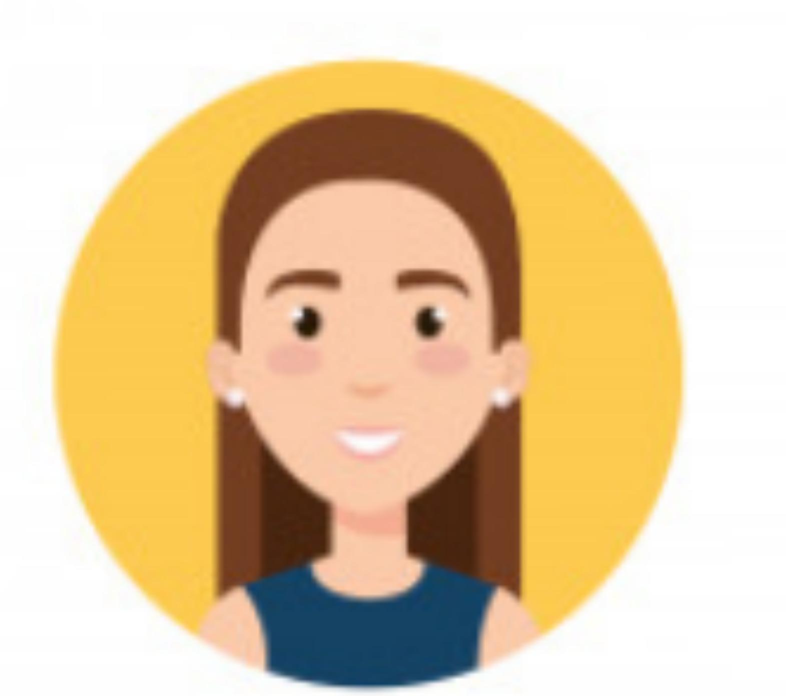
Or should the machine do this?



Or should the machine do this?



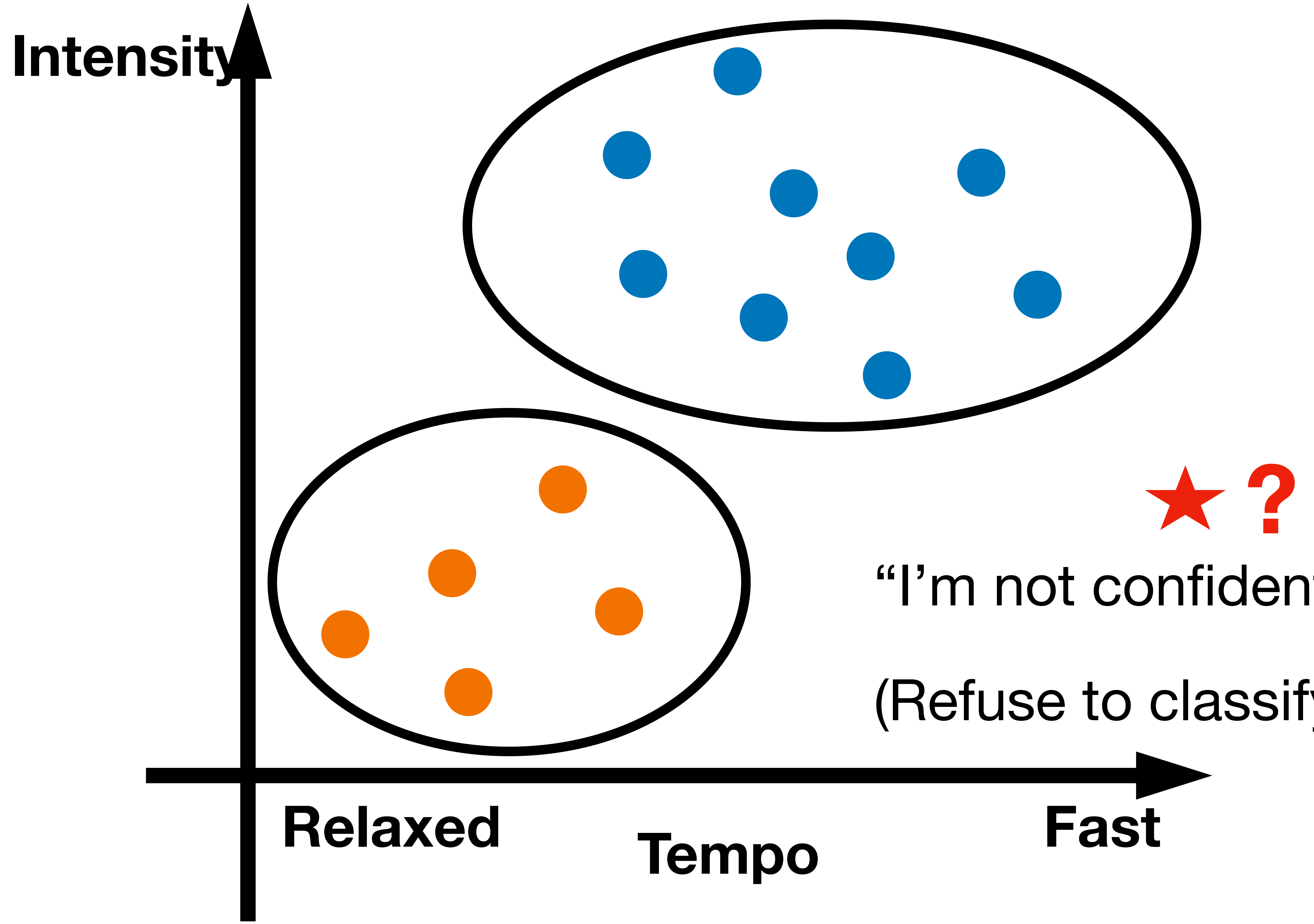
Or should the machine do this?



User Sharon

DisLike

Like



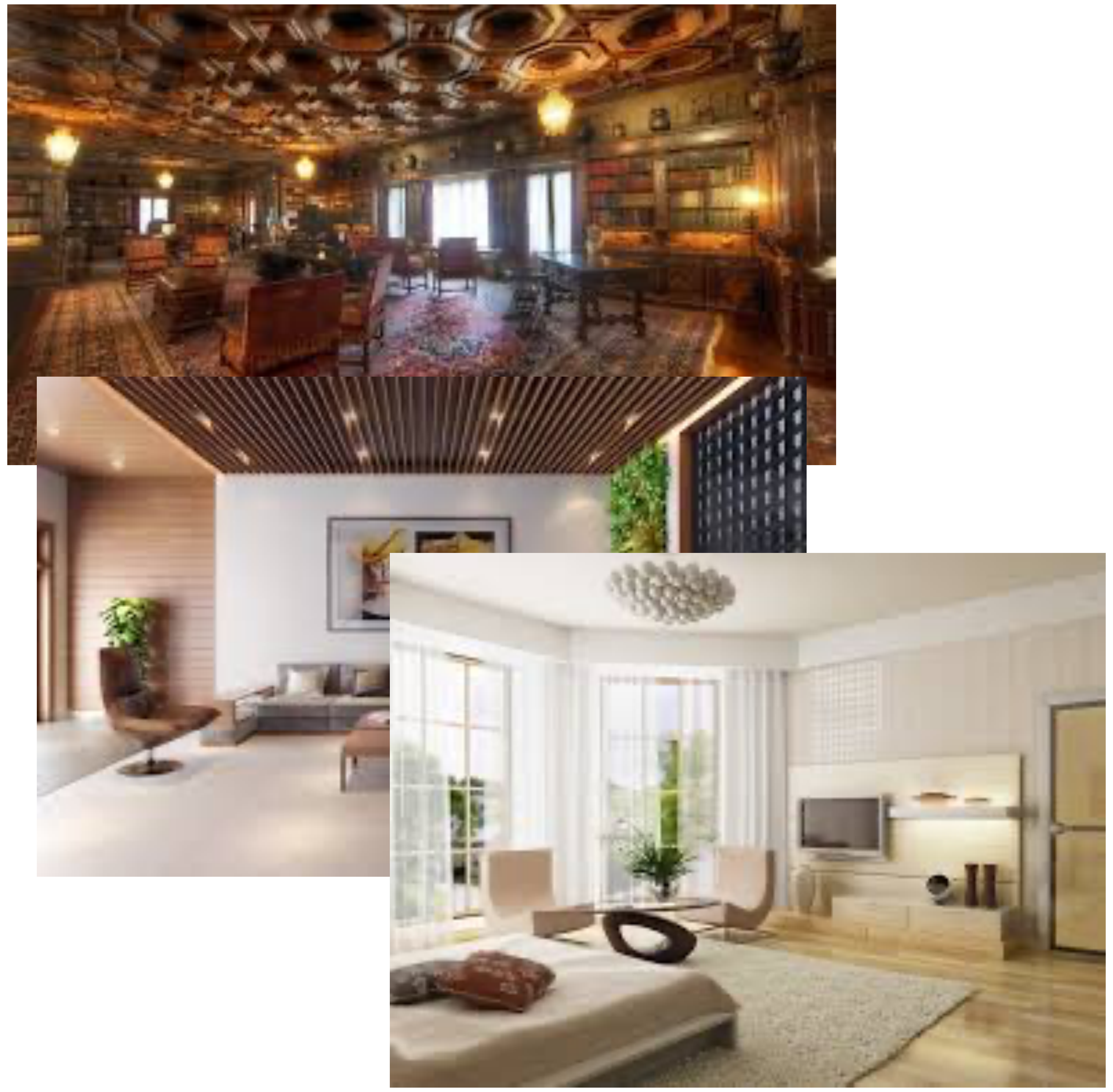
“I’m not confident enough”

(Refuse to classify)

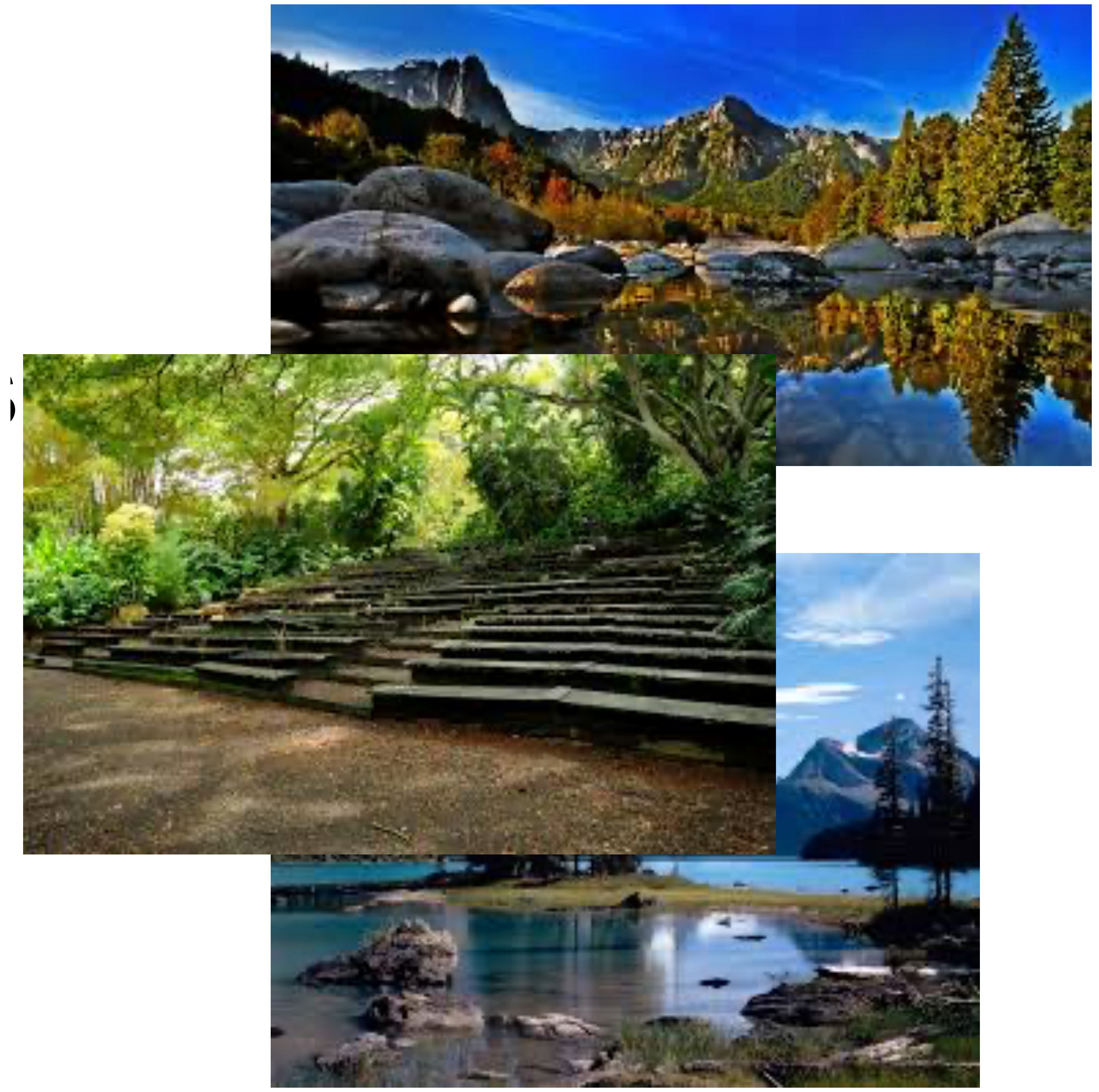
Example 2: Classify Images

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

Example 2: Classify Images

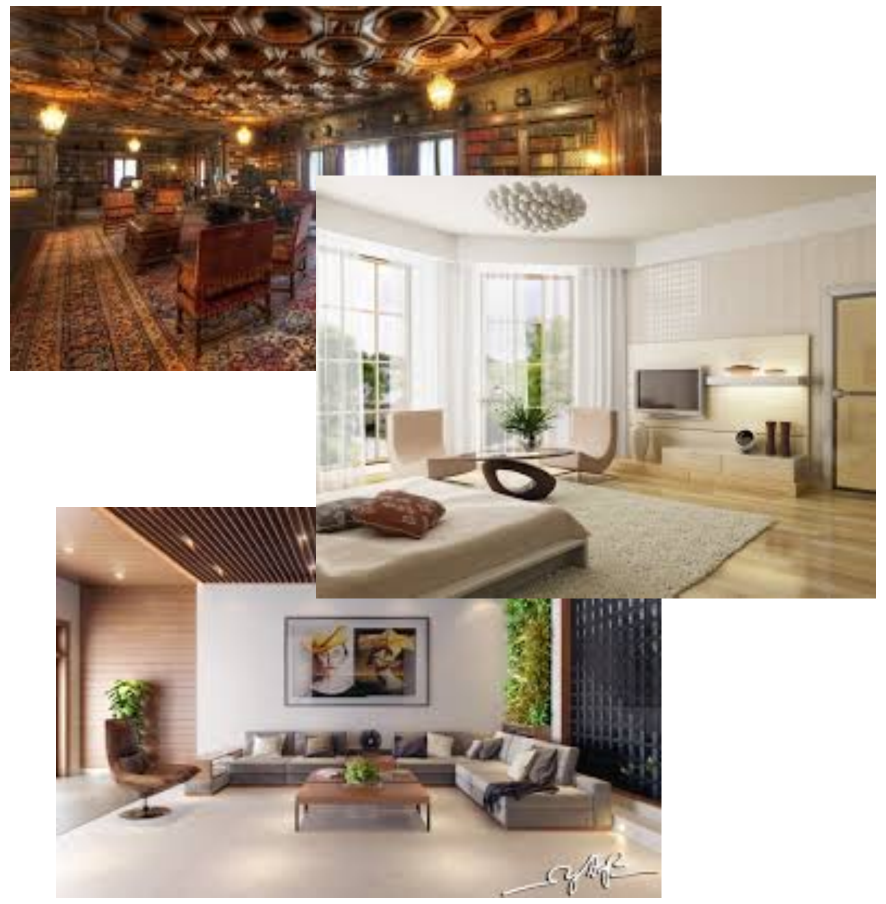


indoor

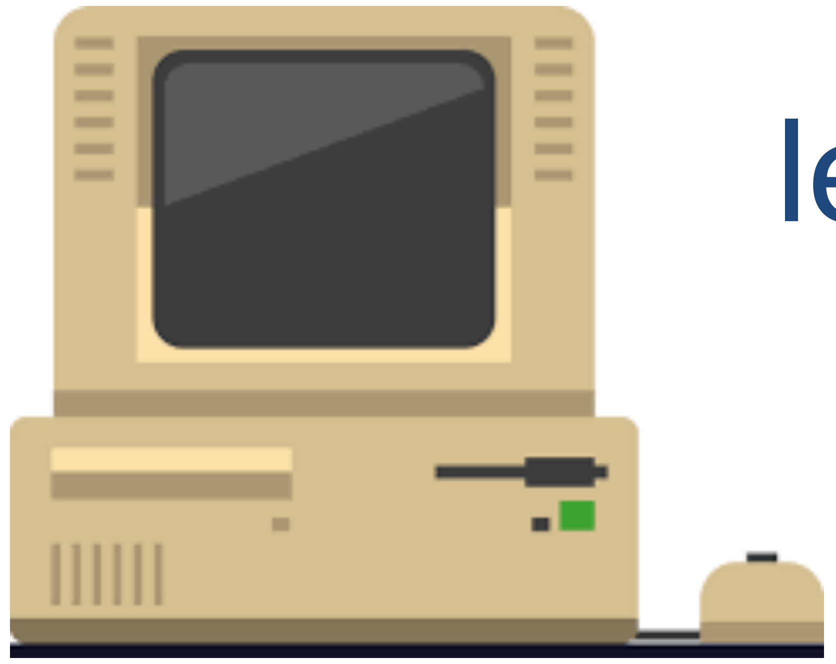


outdoor

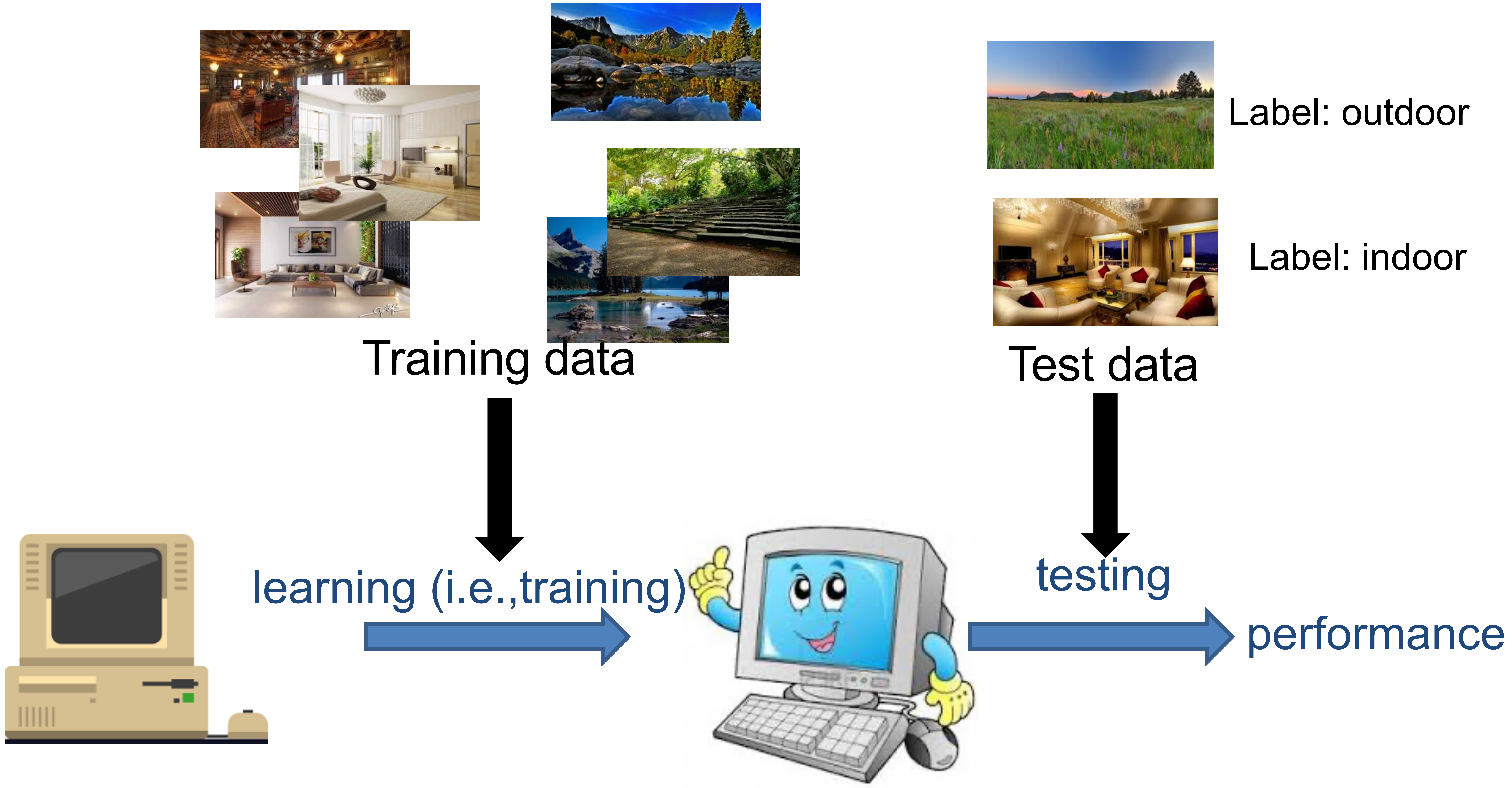
Example 2: Classify Images



Training data



learning (i.e.,training)



Training data
Label: outdoor
Label: indoor
Test data
learning (i.e.,training)
testing
performance

How to represent data?

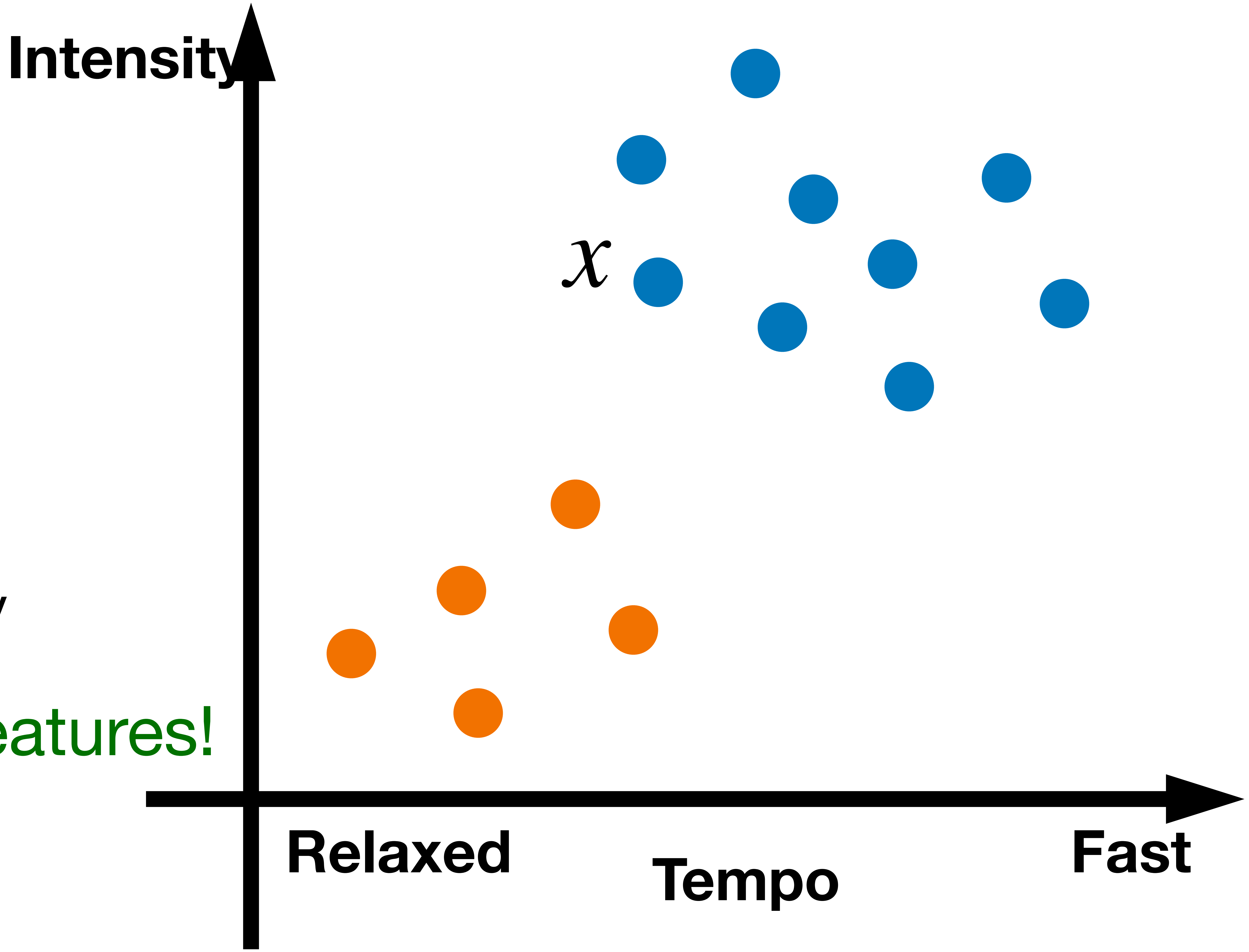
input data

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

d: feature dimension

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

There can be many features!

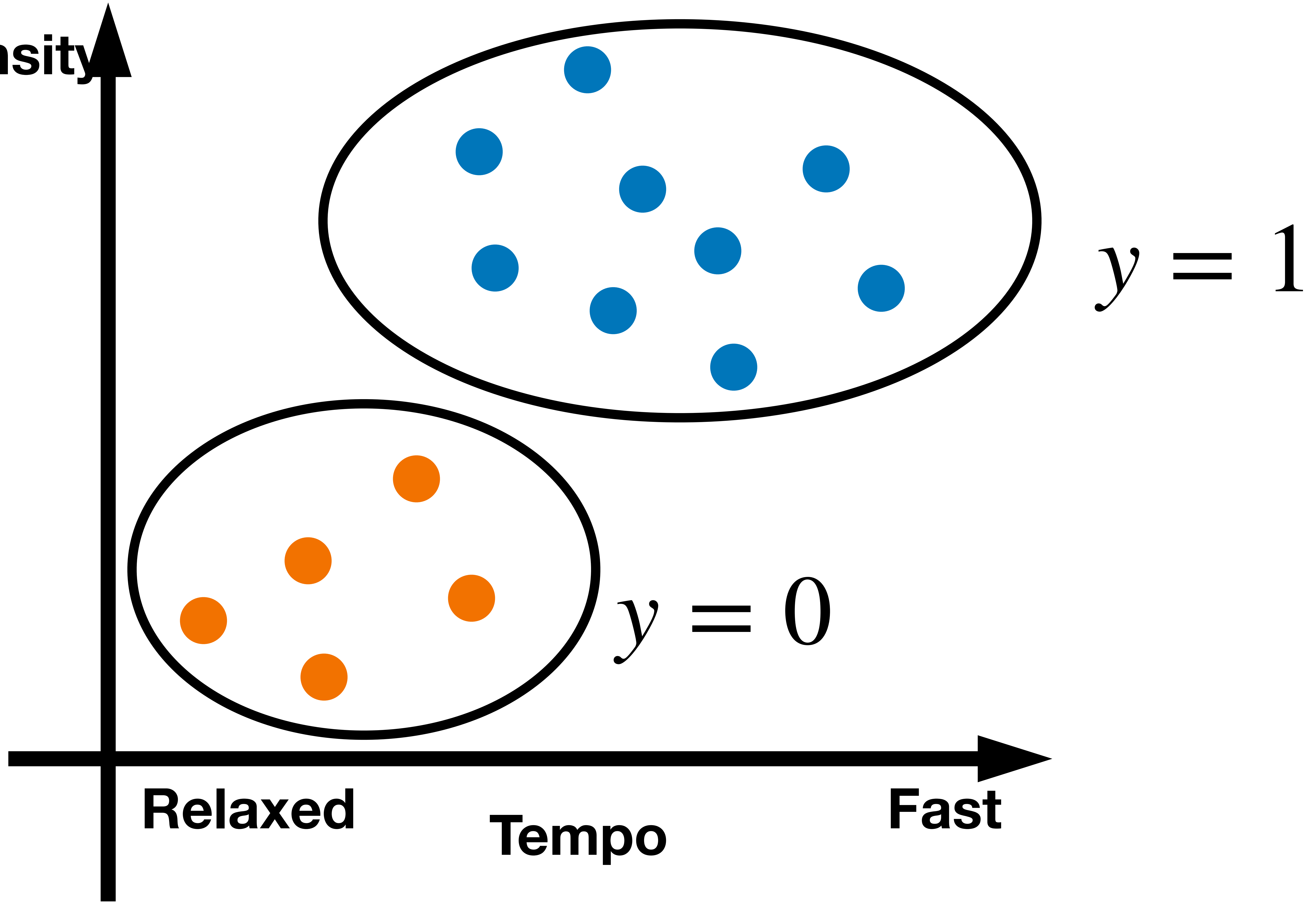


How to represent data?

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

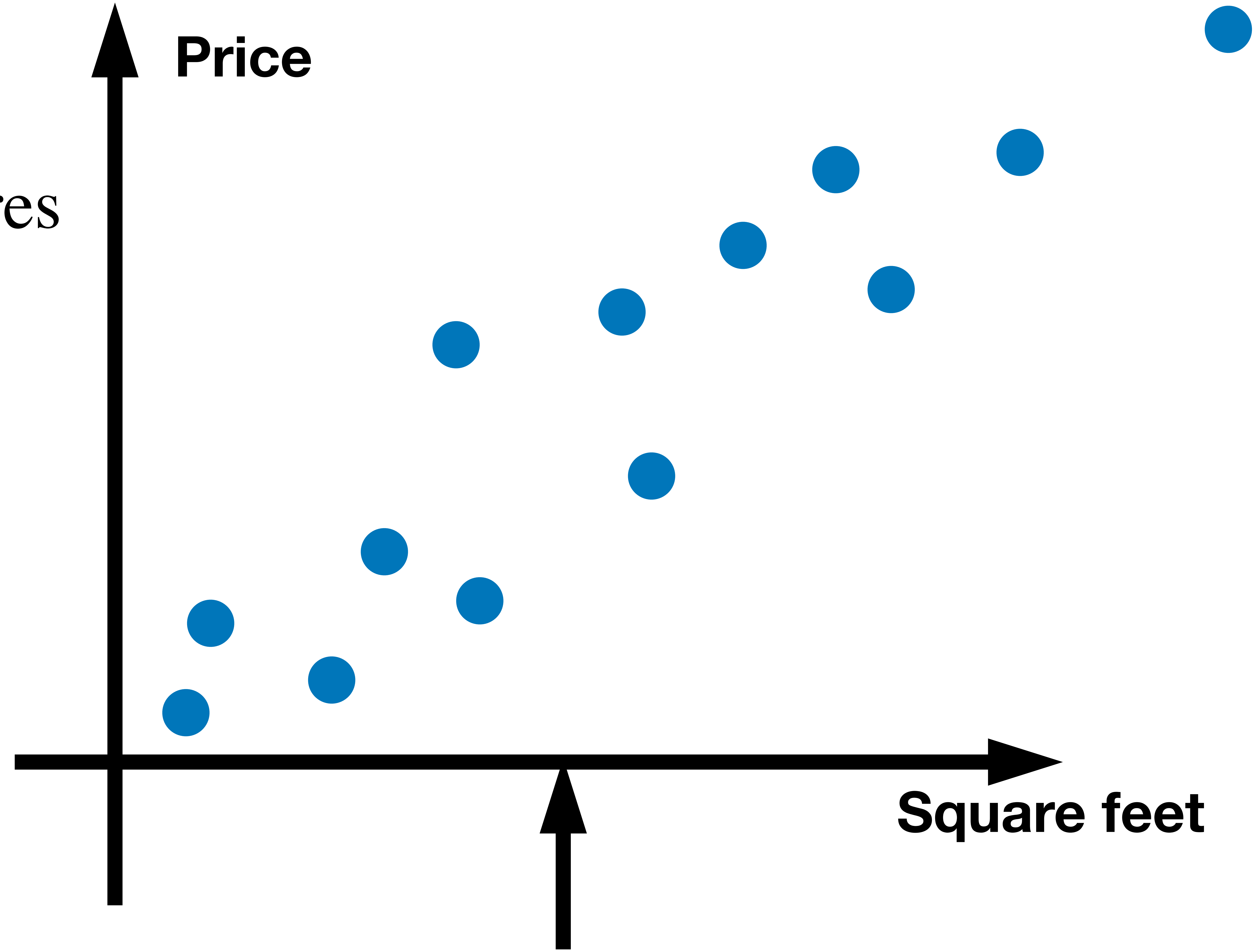
Regression

Example of regression: housing price prediction

Given: a dataset that contains n samples

$$(x_1, y_2), (x_2, y_2), \ldots, (x_n, y_n)$$

Task: if a residence has $\mathbf{x}$ squares feet, predict the price?



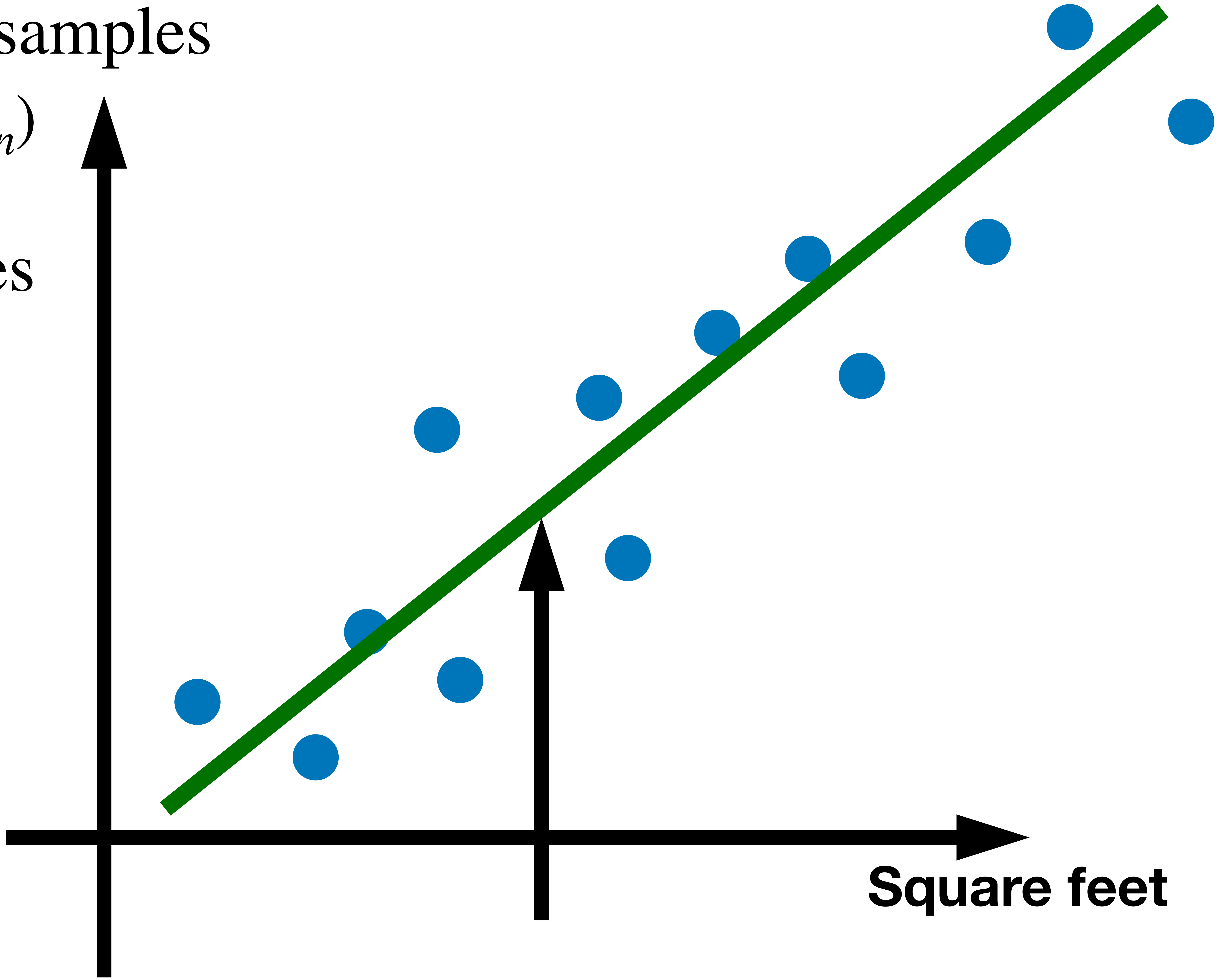
Example of regression: housing price prediction

Given: a dataset that contains n samples

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

Task: if a residence has $\mathbf{x}$ squares feet, predict the price?

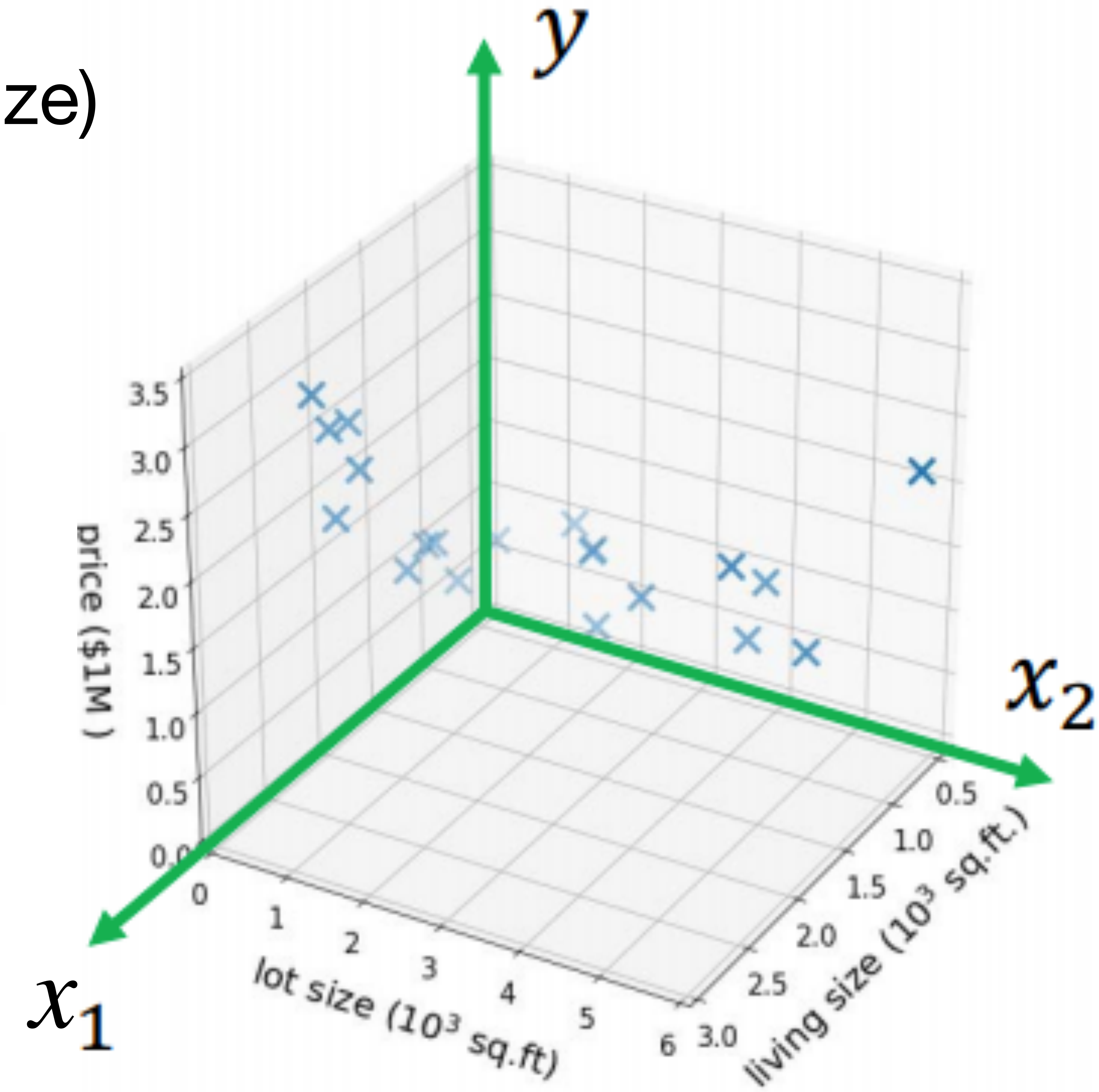
$$y \in \mathbb{R}$$



Example of regression: housing price prediction

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

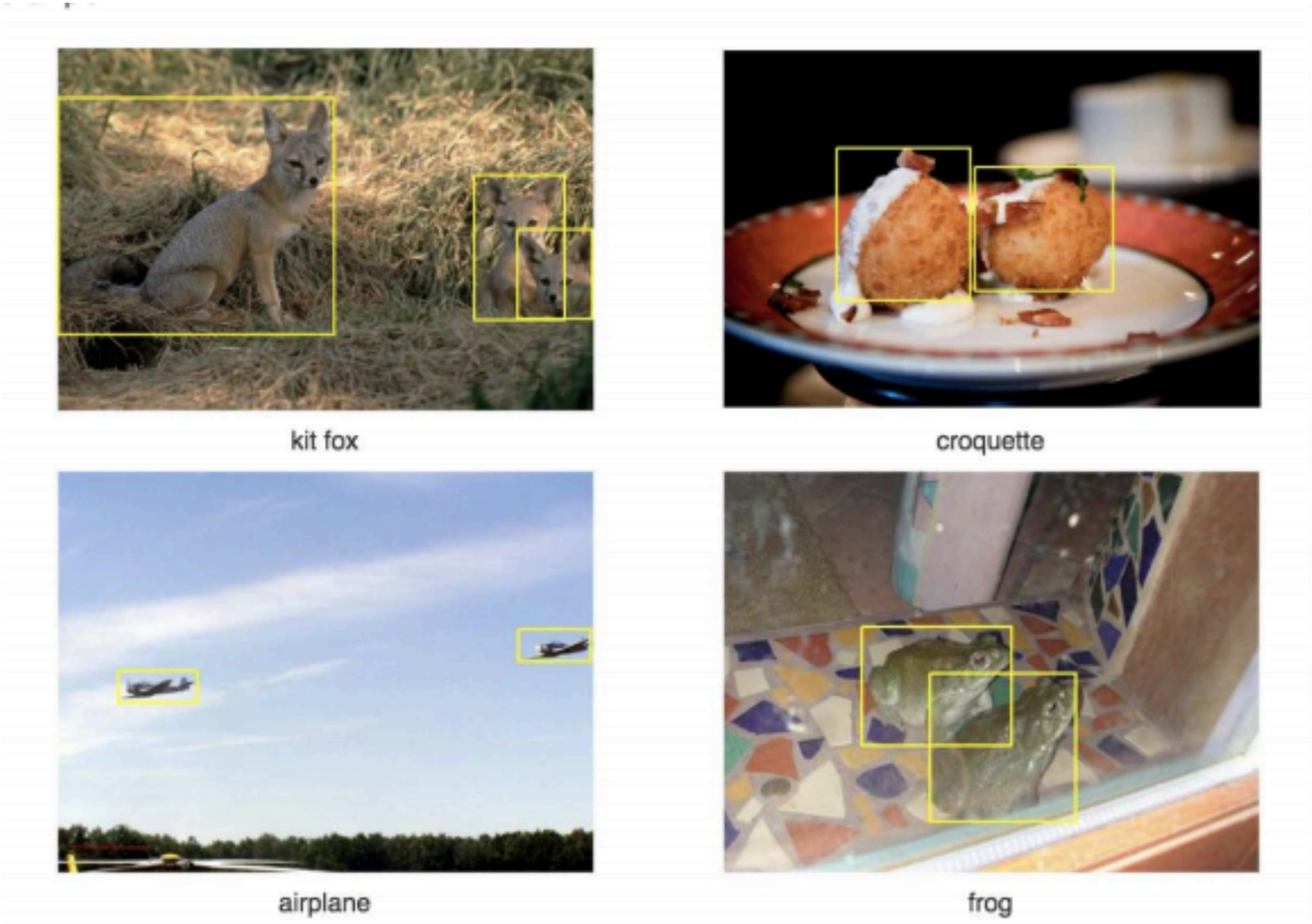
$$\underbrace{\text{(size, lot size)}}_{\substack{\text{features/input} \\ x \in \mathbb{R}^2}} \rightarrow \underbrace{\text{price}}_{\substack{\text{label/output} \\ y \in \mathbb{R}}}$$



(credit: stanford CS229)

Supervised Learning: More examples

x = raw pixels of the image   y = bounding boxes



Russakovsky et al. 2015

Two Types of Supervised Learning Algorithms

Classification

- the label is a **discrete** variable

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

Regression

- the label is a **continuous** variable

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

Training Set for Supervised Learning

A training set is a multiset of (instances, label) pairs to the learning algorithm:

$$\{(\boxed{x_1}, \boxed{y_2}), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)\}$$

input label

- the "**experience**" given to a learning algorithm
- multiset: can have duplicate (x,y) pairs
- Independent and identically distributed (i.i.d.) assumption:

$$(x_i, y_i) \sim p_{XY}$$

Goal of Supervised Learning

Given training set

$$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)\}$$

Learn a function mapping $f : X \to Y$, such that $f(x)$ predicts the label y on **future** data x (not in training set, but also drawn iid)

$$(x, y) \sim p_{XY}$$

Loss, empirical risk (training set error)

Loss function

- 0-1 loss for classification $\ell(f, \mathbf{x}, y) = \mathbf{1}_{[f(\mathbf{x}) \neq y]}$
- Squared loss for regression: $\ell(f, \mathbf{x}, y) = \big(f(\mathbf{x}) - y\big)^2$

Empirical risk = training set error

$$\hat{R}(f) = \frac{1}{n} \sum_{i=1}^{n} \ell(f, \mathbf{x}_i, y_i)$$

The machine learning dilemma

Can only learn from the training set (may also regularize)

$$\hat{f} = \arg\min_{f \in \mathcal{F}} \hat{R}(f) = \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \ell(f, x_i, y_i)$$

But really wants to find f^*: do well on distribution (test set, deploy, future data)

$$f^* = \arg\min_{f \in \mathcal{F}} R(f) = \arg\min_{f \in \mathcal{F}} \mathbb{E}_{(x,y)\sim p_{XY}} \ell(f, x, y)$$

Also limited by the richness of model family $\mathcal{F}$

Details in upcoming lectures :)

Quiz Break

Q1-1: Which is true about feature vectors?

A. Feature vectors can have at most 10 dimensions
B. Feature vectors have only numeric values
C. The raw image can also be used as the feature vector
D. Text data don’t have feature vectors

Quiz Break

Q1-1: Which is true about feature vectors?

A. Feature vectors can have at most 10 dimensions
B. Feature vectors have only numeric values
C. The raw image can also be used as the feature vector
D. Text data don't have feature vectors

A. Feature vectors can be in high dimen.
B. Some feature vectors can have other types of values like strings
D. Bag-of-words is a type of feature vector for text

Quiz Break

Q1-2: Which of the following is not a common task of supervised learning?

A. Object detection (predicting bounding box from raw images)
B. Classification
C. Regression
D. Dimensionality reduction

Quiz Break

Q1-2: Which of the following is not a common task of supervised learning?

A. Object detection (predicting bounding box from raw images)
B. Classification
C. Regression
D. Dimensionality reduction (PCA)

Dimensionality reduction does not require label y

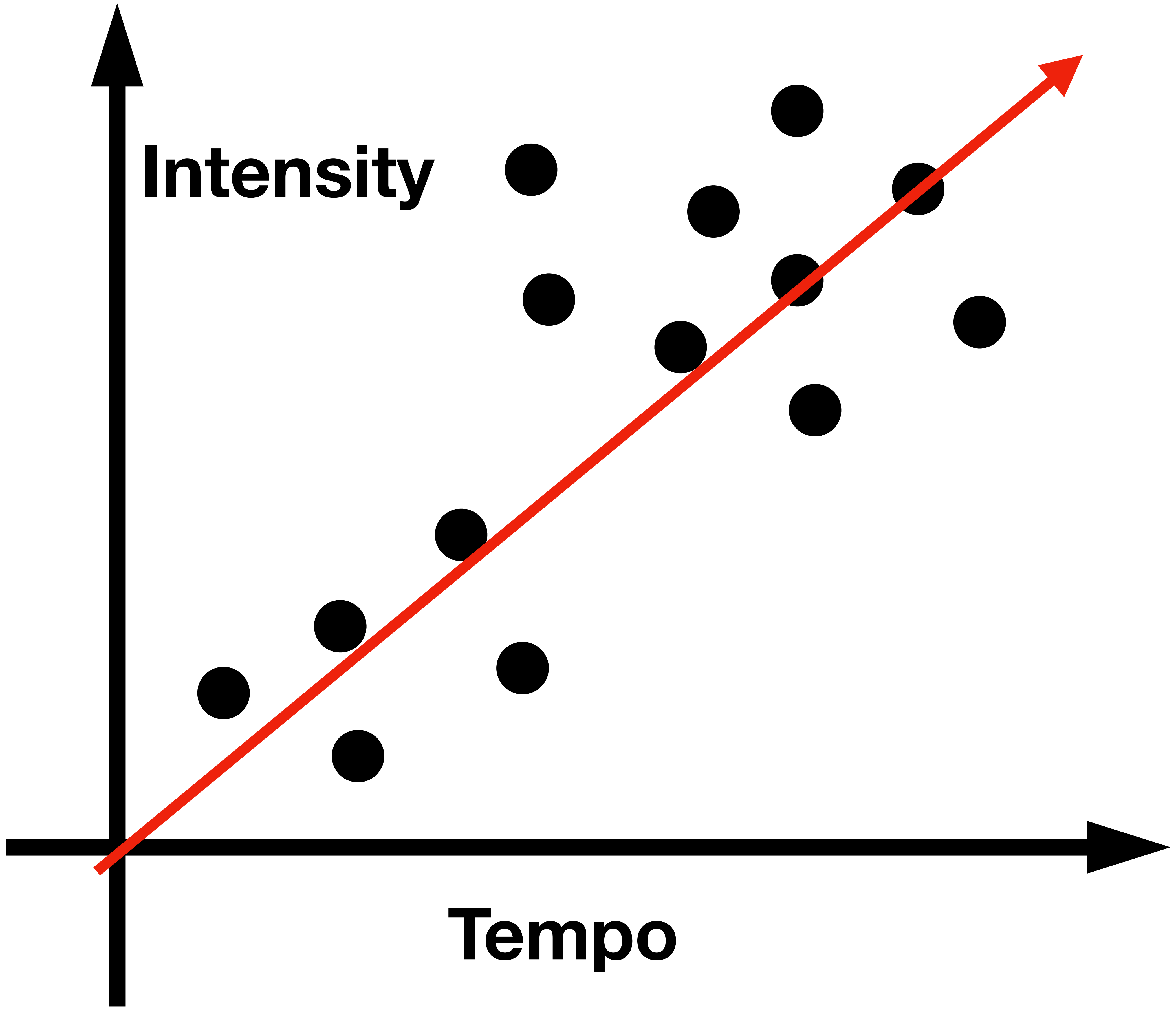
Part II: Unsupervised Learning (no labels)

Unsupervised Learning

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

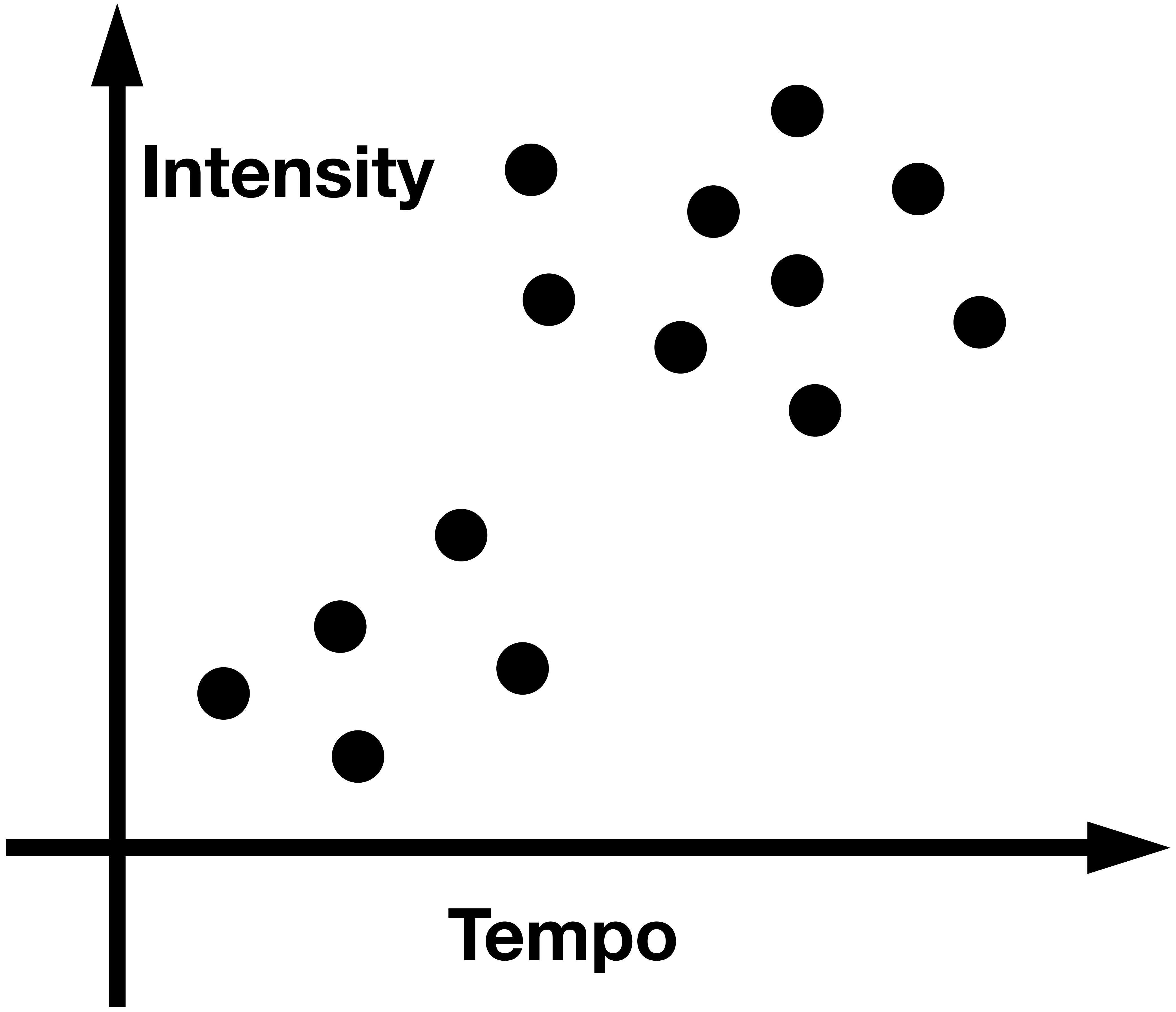
Unsupervised Learning

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



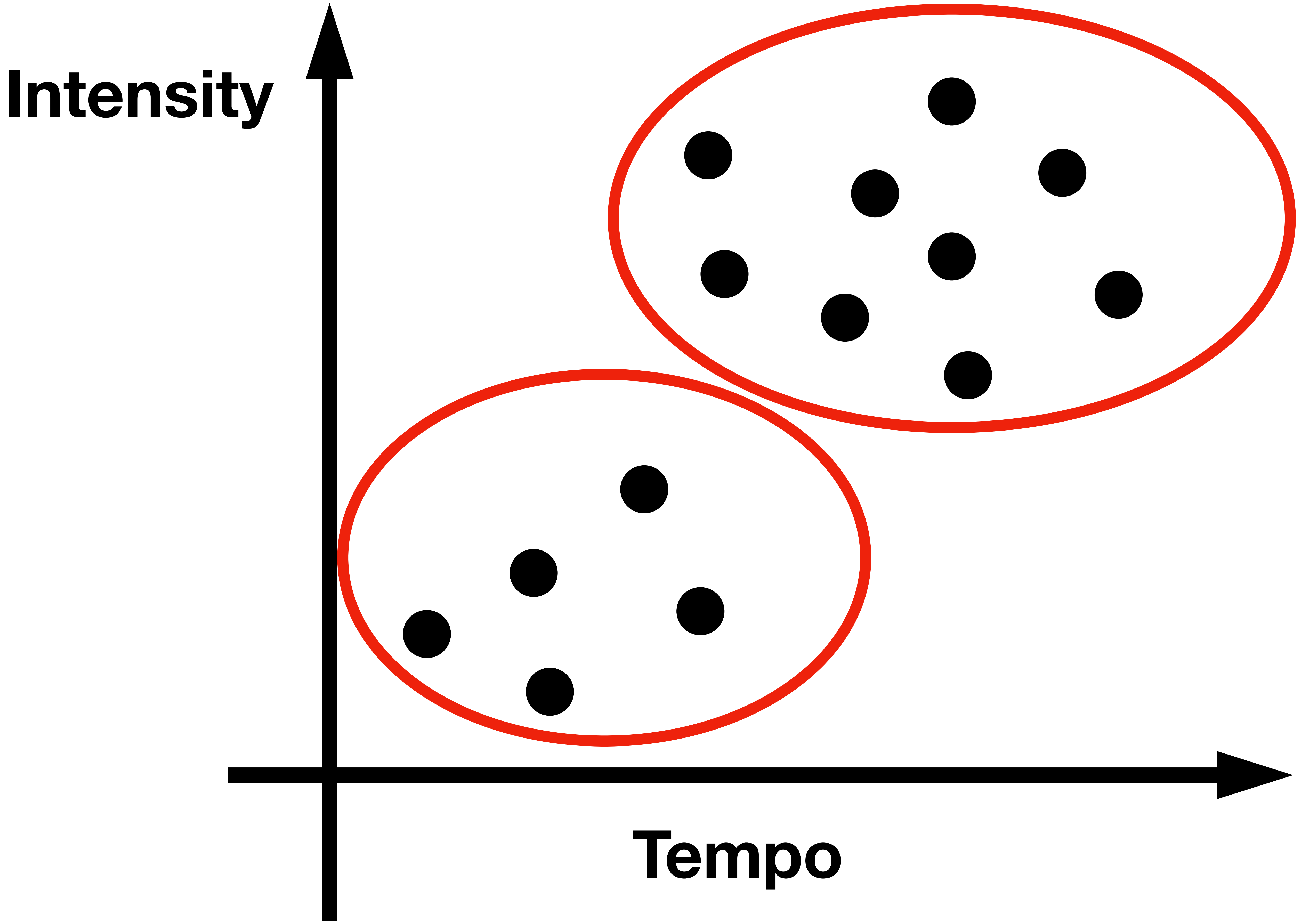
Unsupervised Learning

- Given: dataset contains **no label** $x_1, x_2, \ldots, x_n$
- **Goal:** discover interesting patterns and structures in the data
  - Dimension reduction
  - Clustering

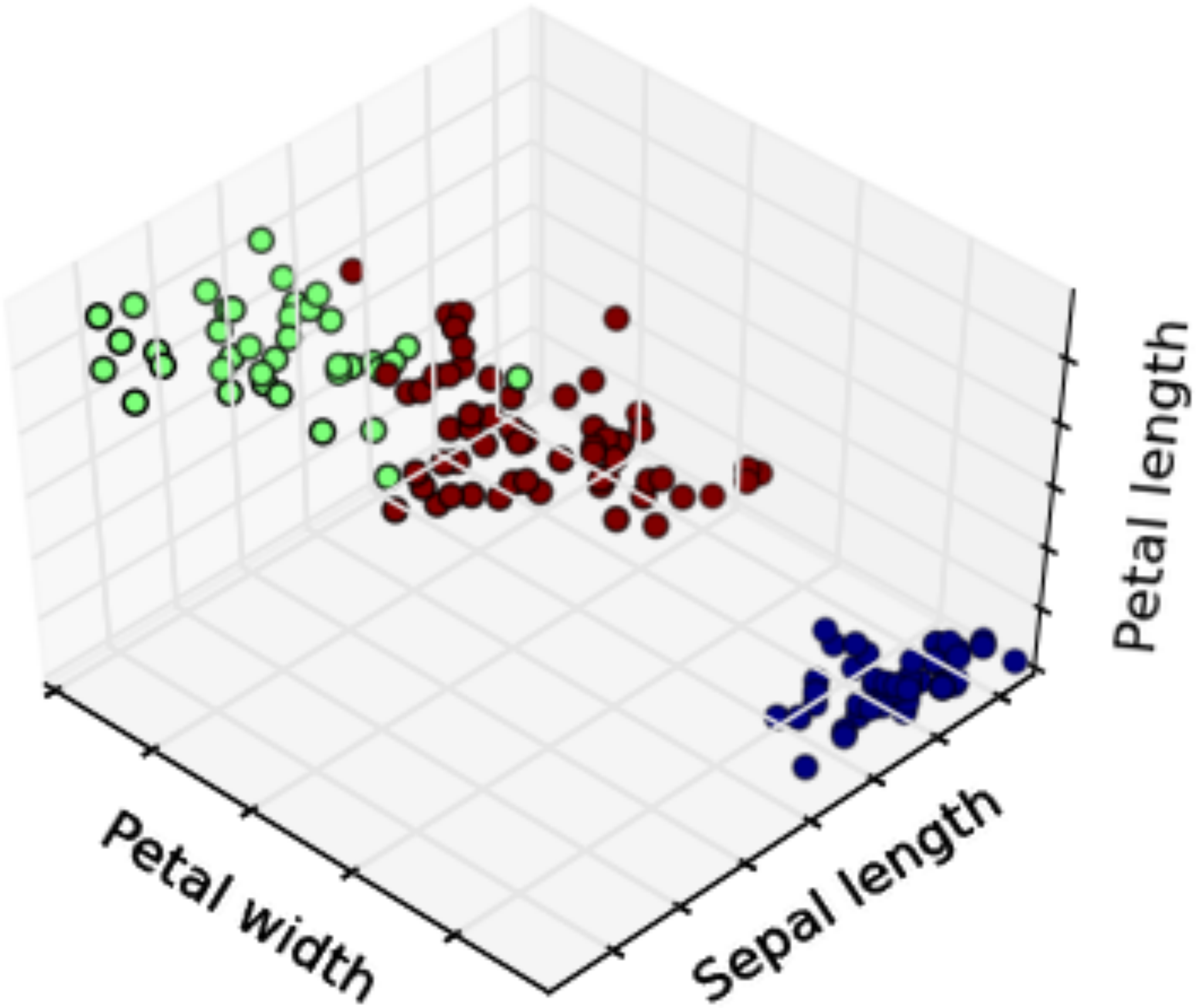


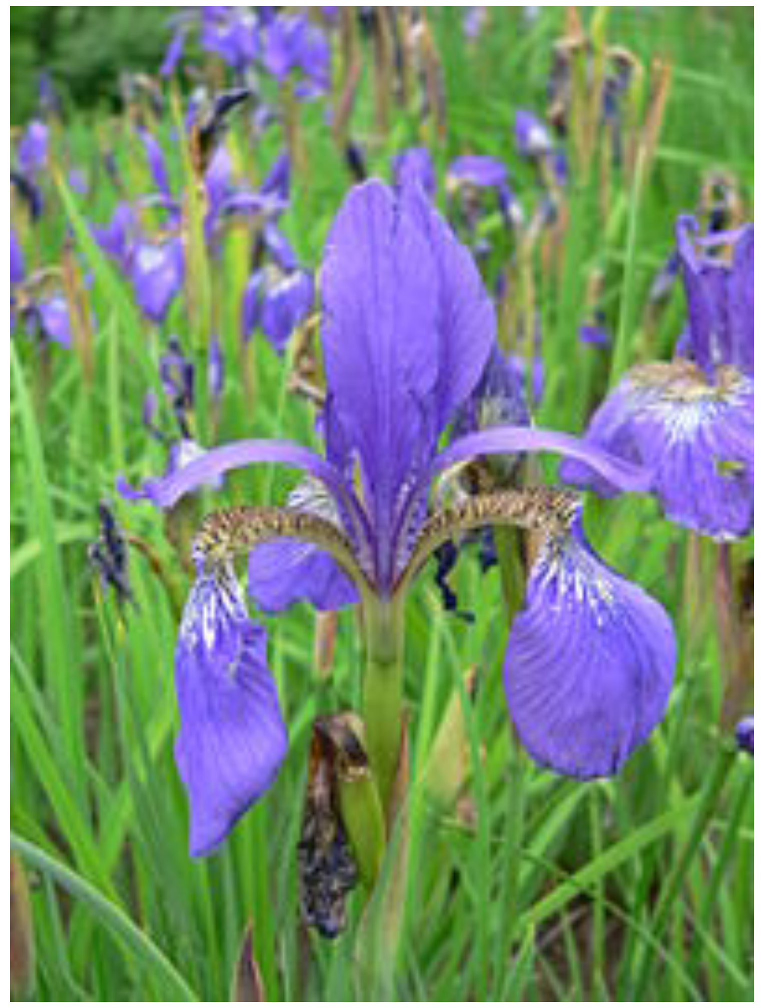
Clustering

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



Clustering



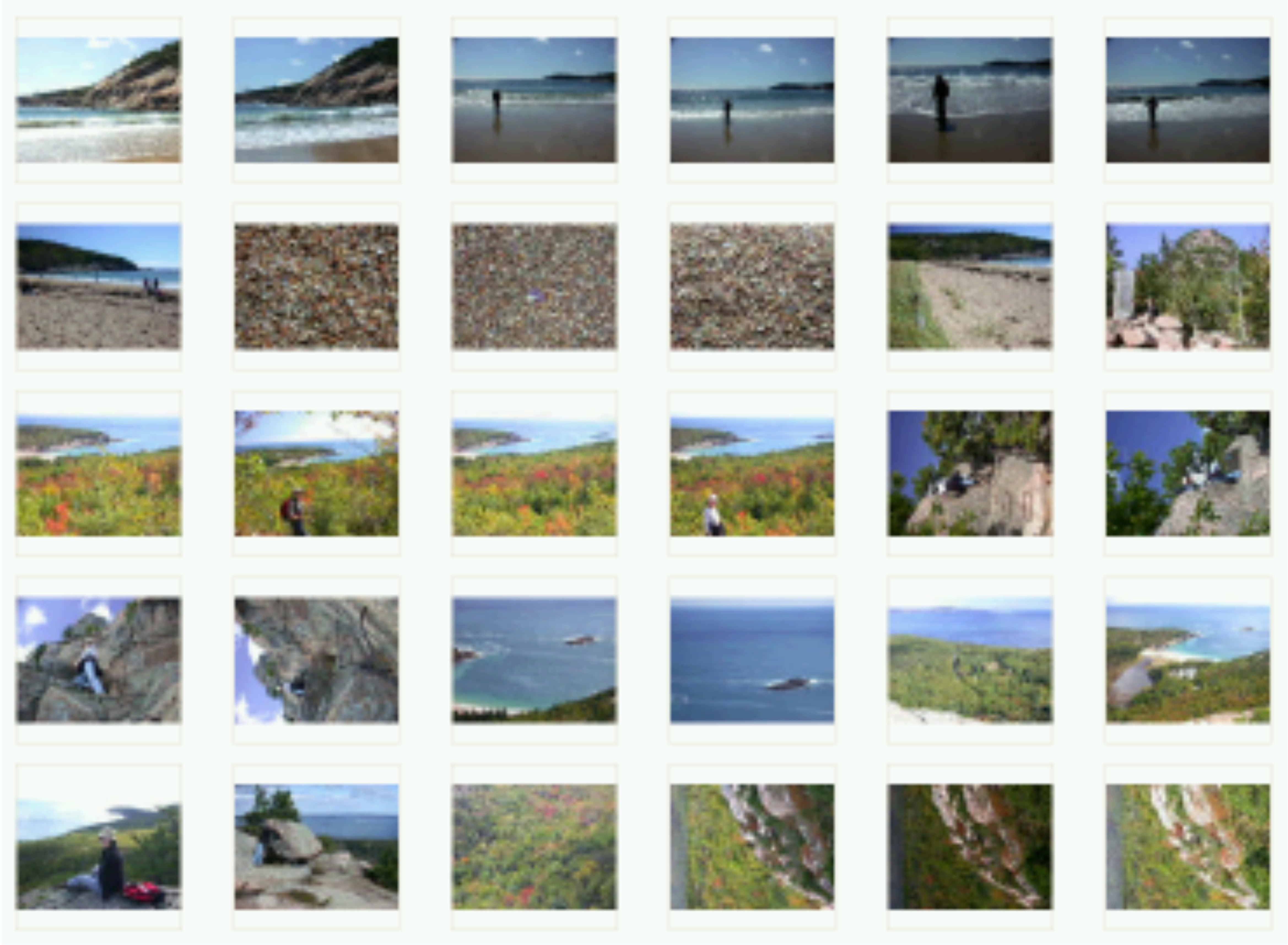


Clustering Irises using three different features

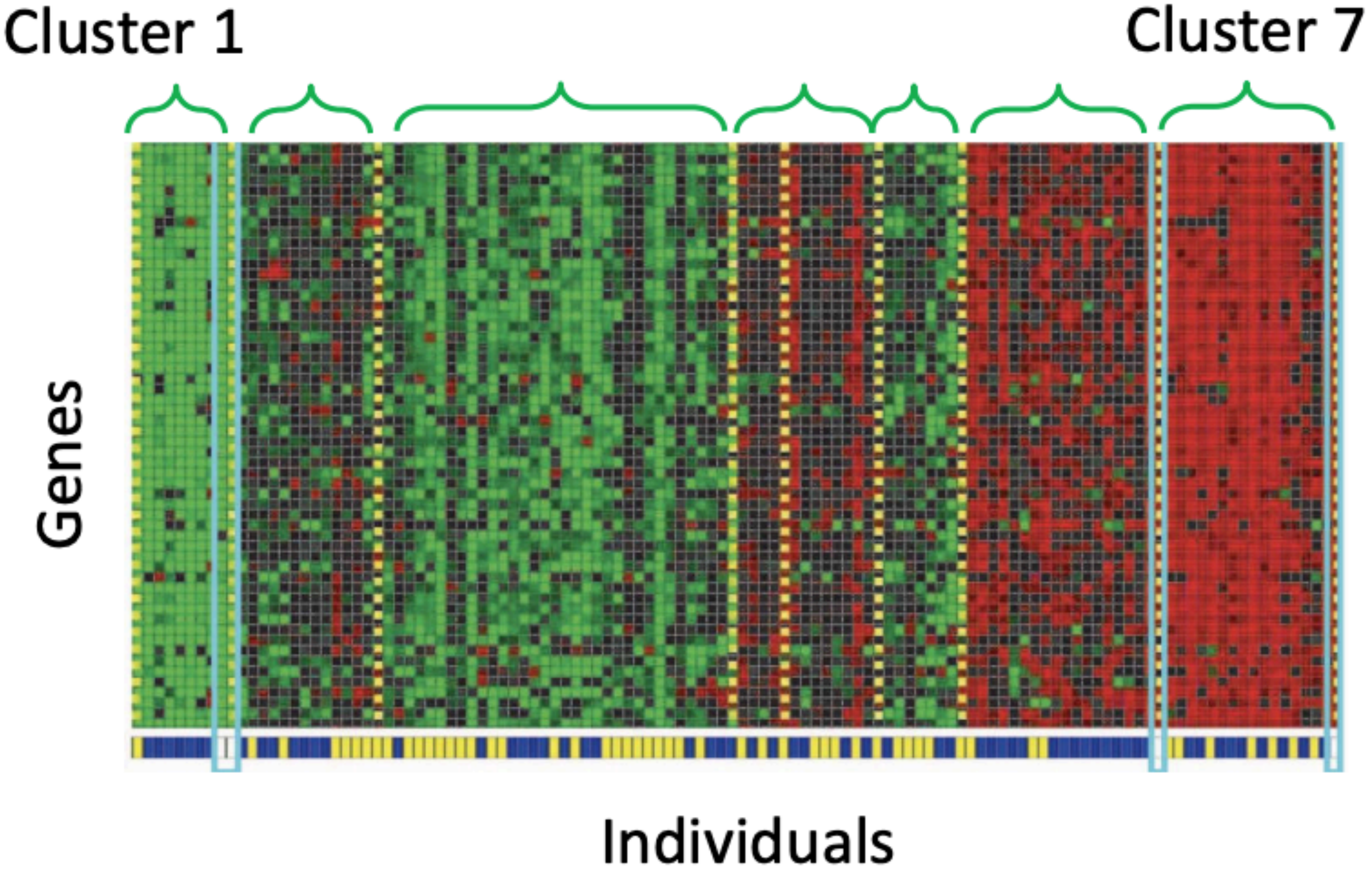
The colors represent clusters identified by the algorithm, not y’s provided as input

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)

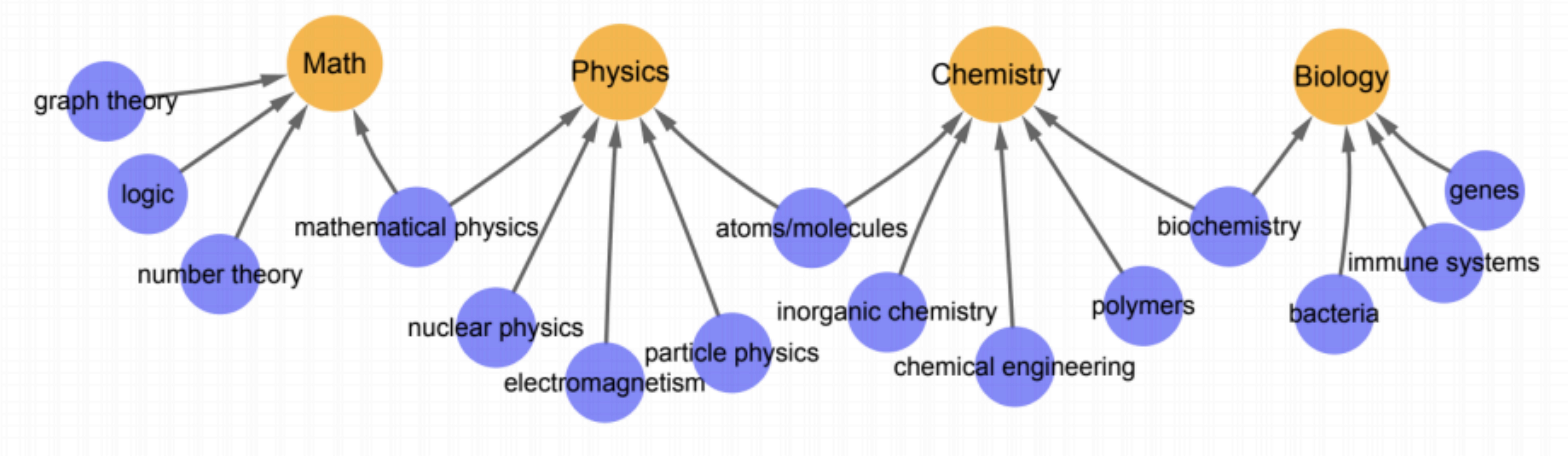


Clustering Genes



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-M.-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

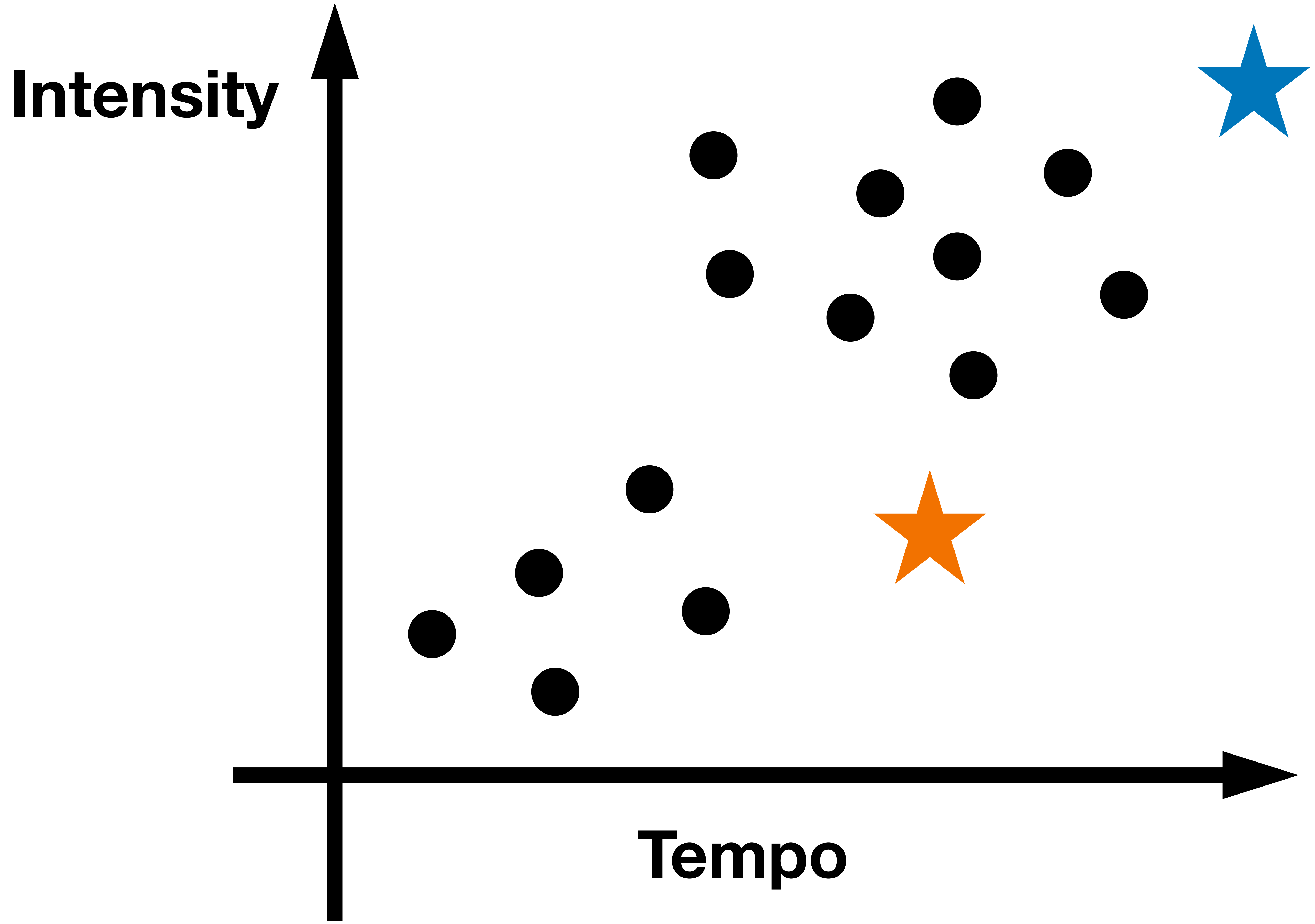
- Very popular clustering method

- Don’t confuse it with k-NN classifier

- Input: a dataset $x_1, x_2, \ldots, x_n$, and assume the number of clusters **k** is given

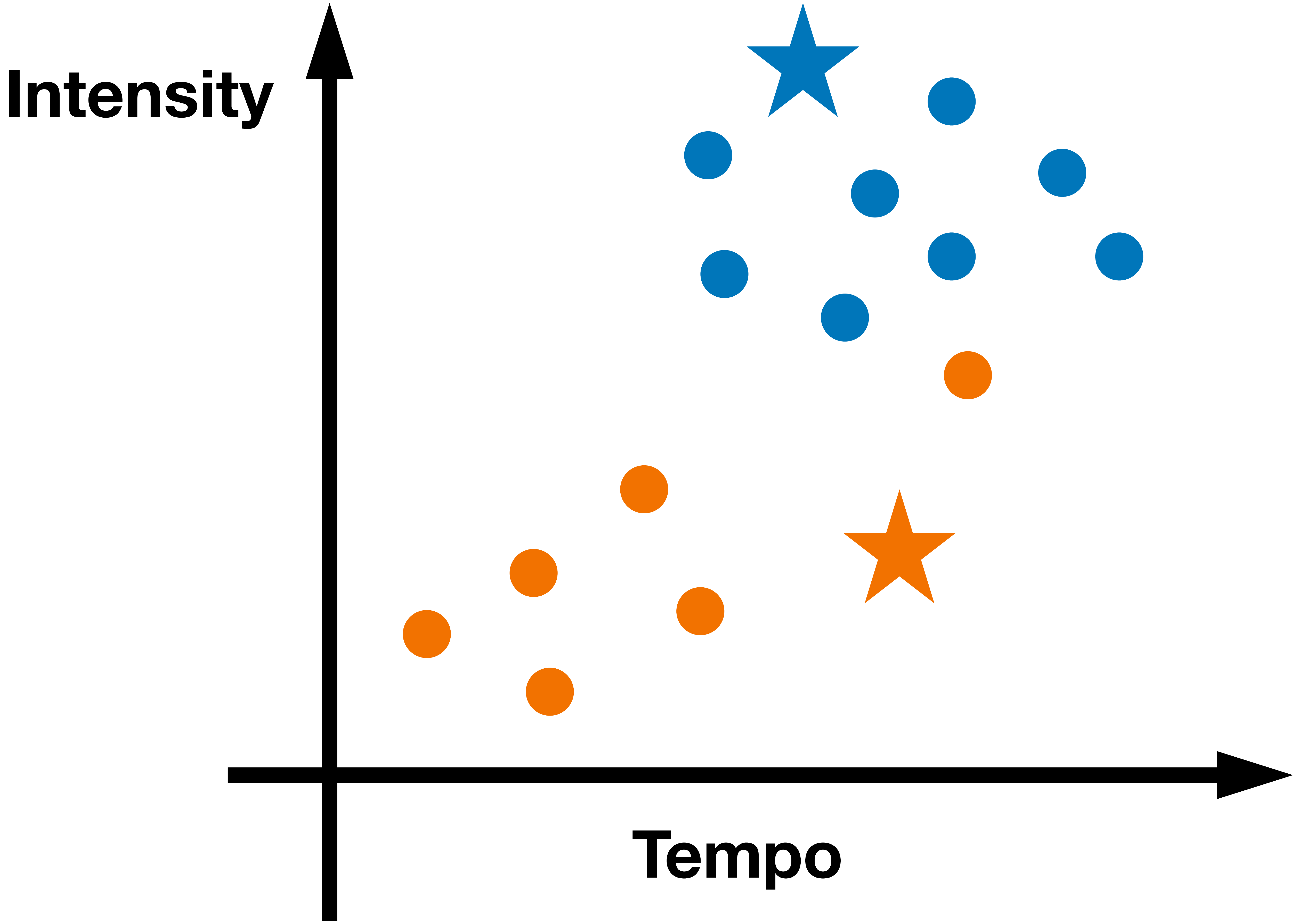
K-means clustering

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



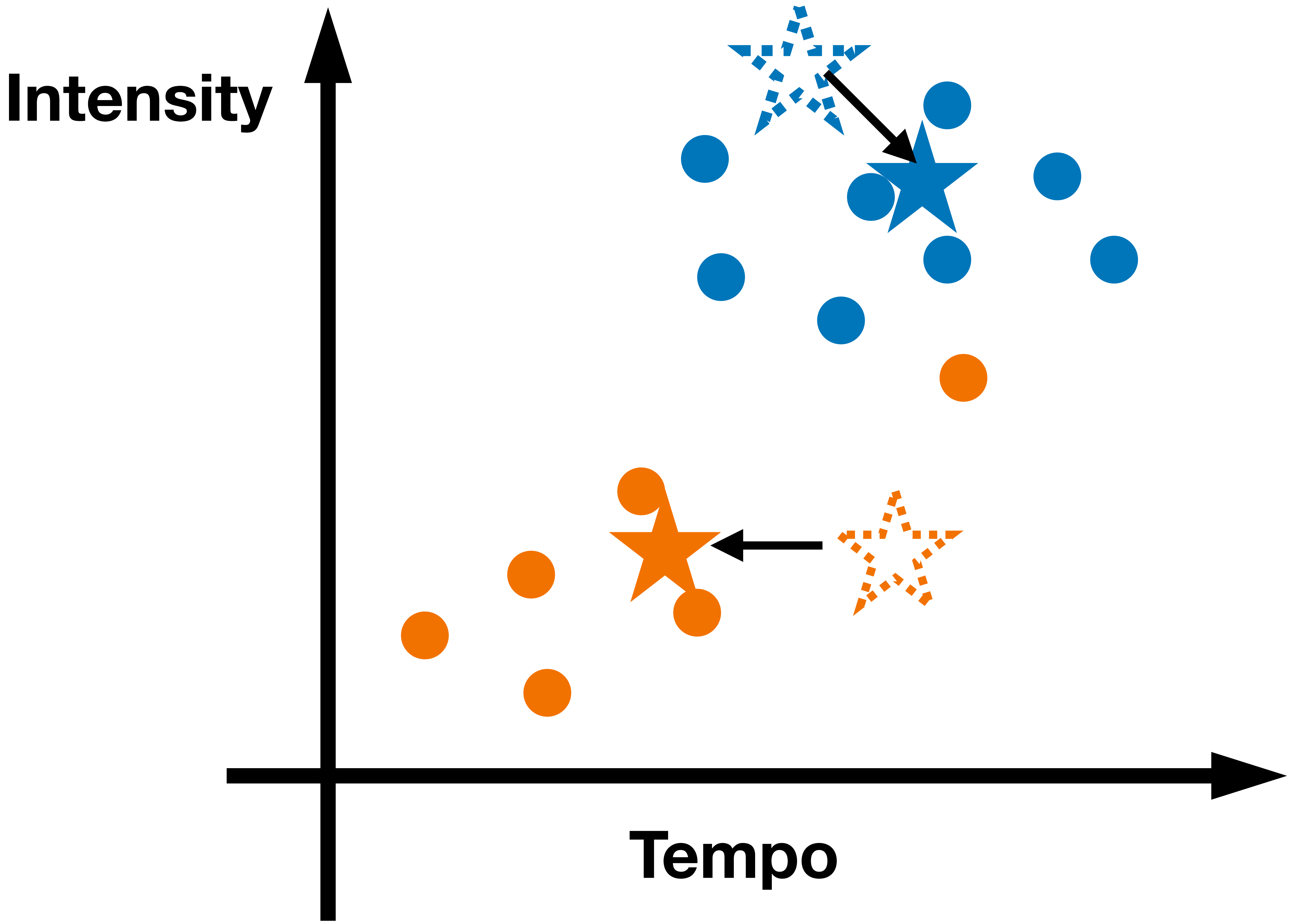
K-means clustering

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



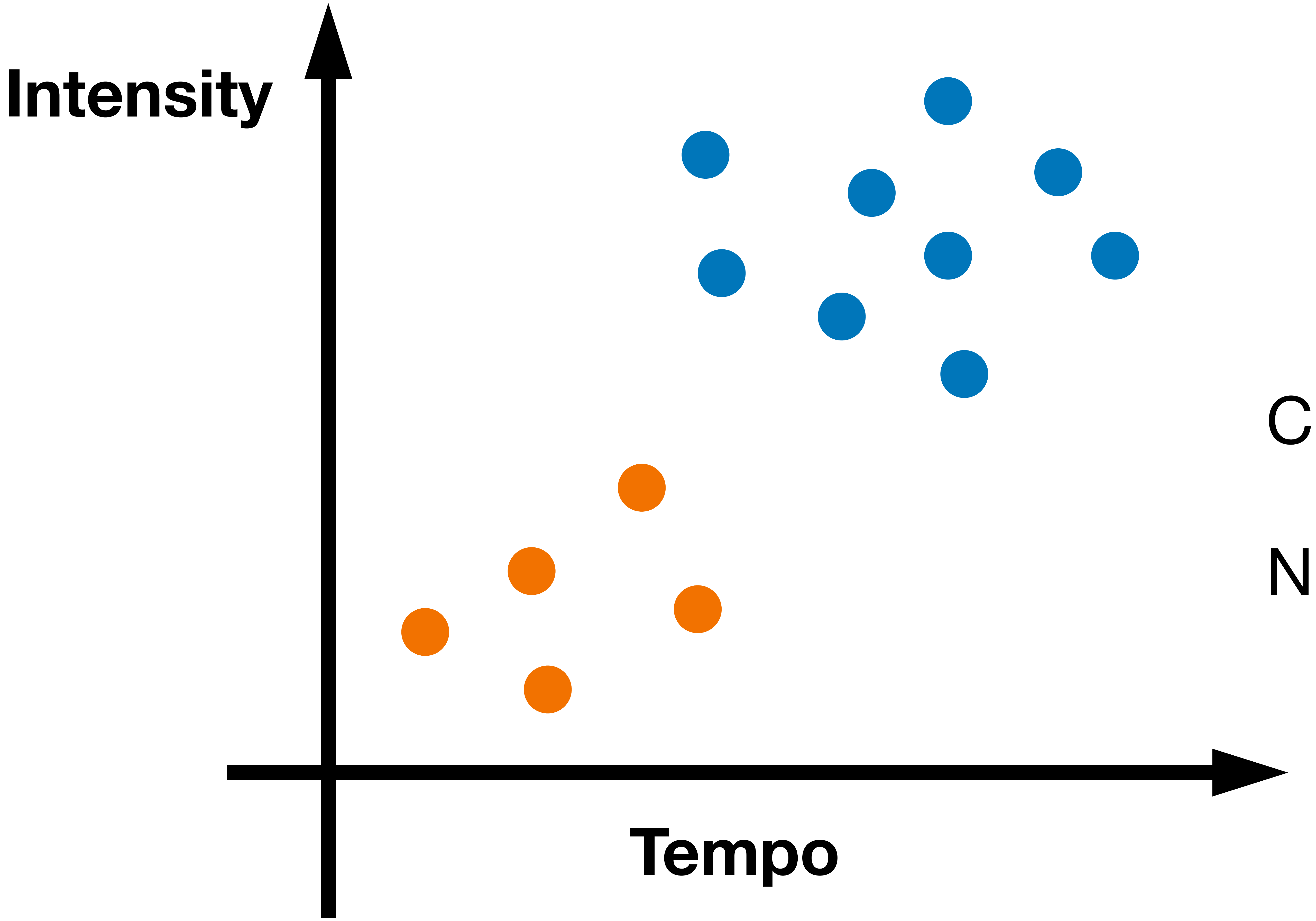
K-means clustering

Step 3: update all cluster centers as the centroids



K-means clustering

Repeat step 2 & 3 until convergence



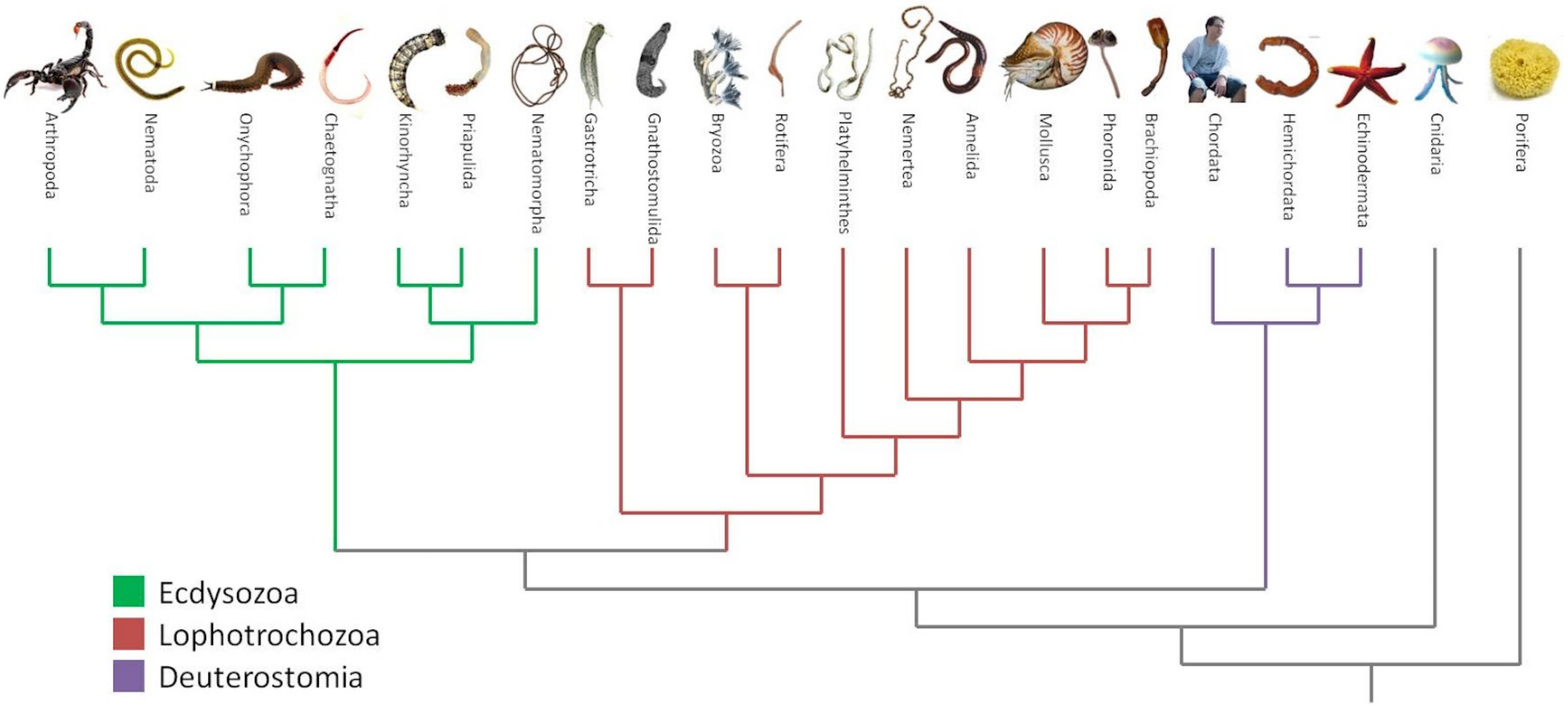
Converged solution!

No labels required!

K-means clustering: A demo

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

Hierarchical Clustering (more to follow next lecture)



Quiz Break

Q2-1: Which is true about machine learning?

A. The process doesn’t involve human inputs
B. The machine is given the training and test data for learning
C. In clustering, the training data also have labels for learning
D. Supervised learning involves labeled data

Quiz Break

Q2-1: Which is true about machine learning?

A. The process doesn’t involve human inputs
B. The machine is given the training and test data for learning
C. In clustering, the training data also have labels for learning
D. Supervised learning involves labeled data

A. The labels are human inputs
B. The machine should not have test data for learning
C. No labels available for clustering

Quiz Break

Q2-2: Which is true about unsupervised learning?

A. There are only 2 unsupervised learning algorithms
B. Kmeans clustering is a type of hierarchical clustering
C. Kmeans algorithm automatically determines the number of clusters k
D. Unsupervised learning is widely used in many applications

Quiz Break

Q2-2: Which is true about unsupervised learning?

A. There are only 2 unsupervised learning algorithms
B. Kmeans clustering is a type of hierarchical clustering
C. Kmeans algorithm automatically determines the number of clusters k
D. Unsupervised learning is widely used in many applications

What we’ve learned today…

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

Thanks!