Machine Learning for Vision

Gary Overett

What is Machine Learning

A science that tries to learn meaningful statistics from data

Real World Problems

- Based on content of emails: SPAM or Not SPAM?

Real World Problems

- Based on image region : is there a cat?

Real World Problems

- Based on many images - how many distinct types of animals?

Example : Salmon vs Sea Bass

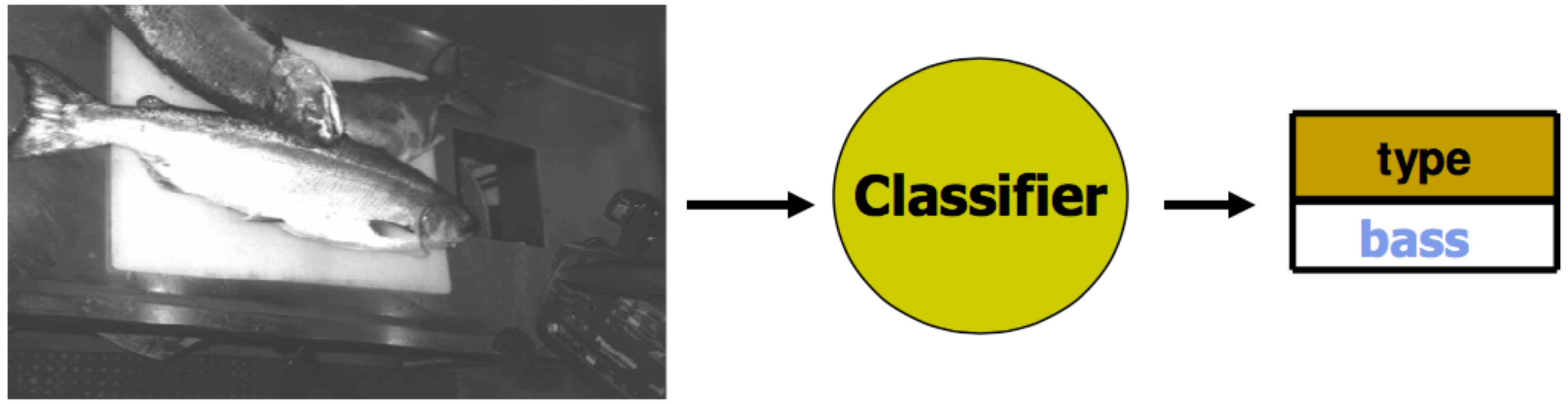
- Shamelessly lifted from https://webdocs.cs.ualberta.ca/~greiner/C-466/SLIDES/1-Introduction.pdf

Fish Classifier

Sort Fish

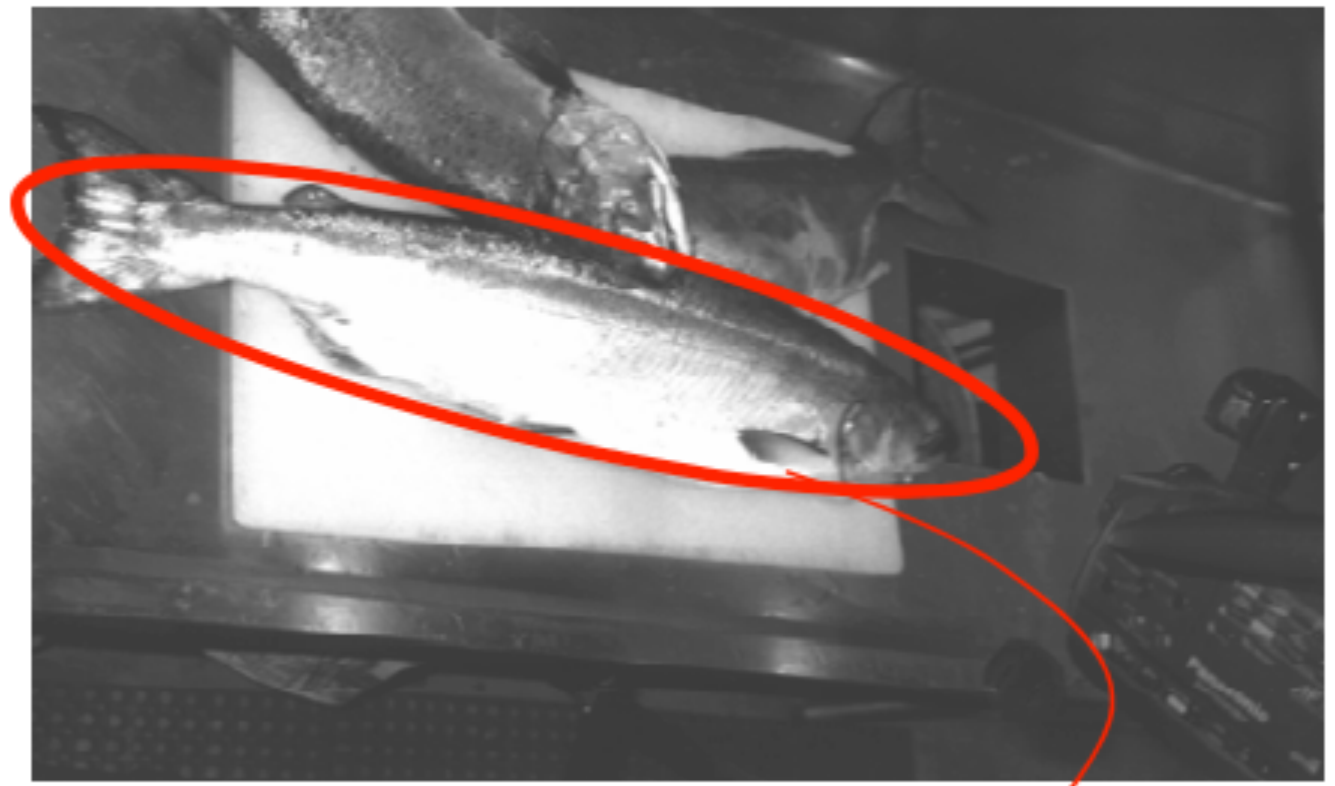
into Species → **Sea bass** / **Salmon**

using optical sensing



Problem Analysis

- Extract *features* from sample images:
  - Length
  - Width
  - Average pixel brightness
  - Number and shape of fins
  - Position of mouth
  - …



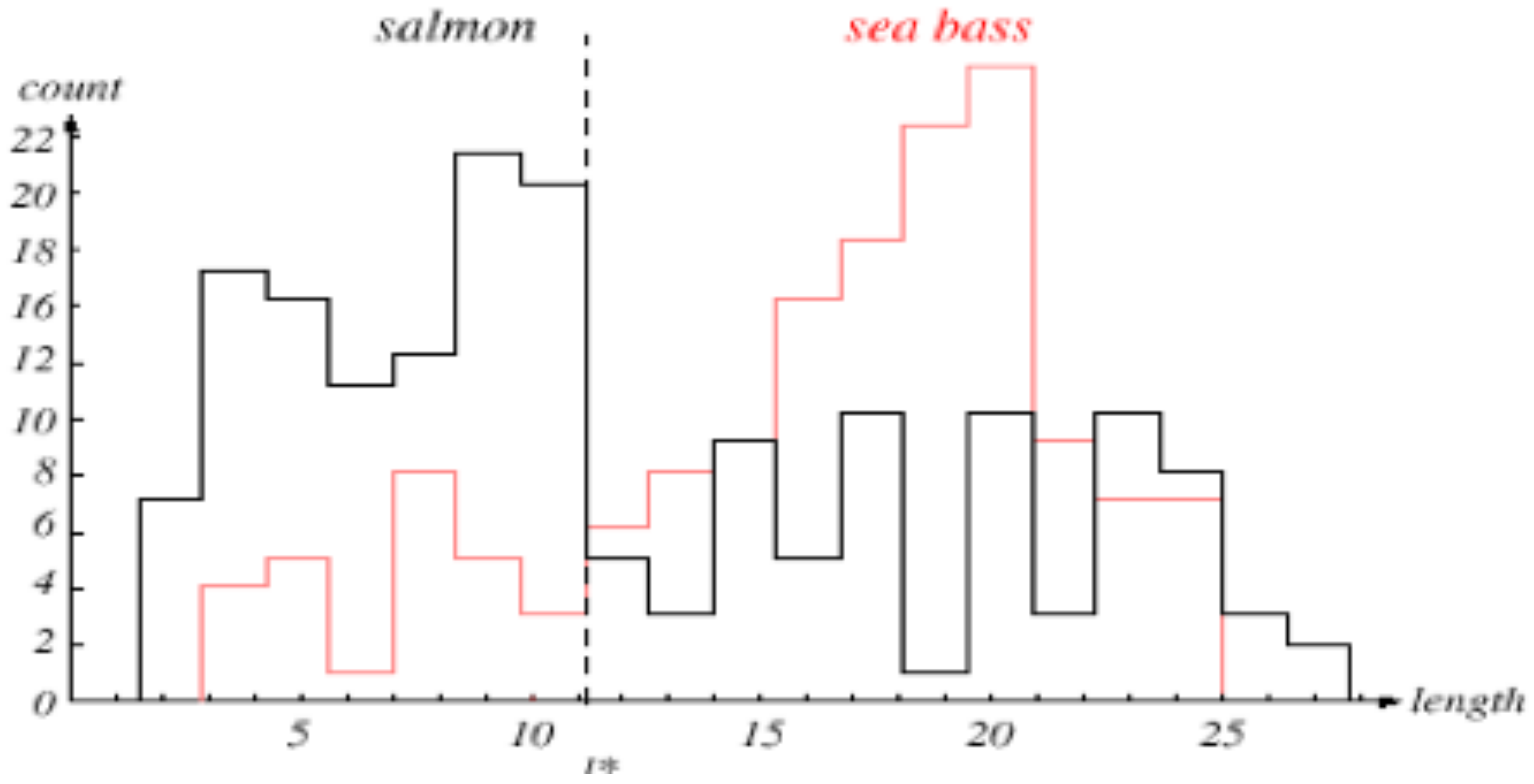
[L=50, W=10, PB=2.8, #fins=4, MP=(5,53), …]

Length	Wtdth	Pixel Bright	…	Light
50	10	2.8	…	Pale

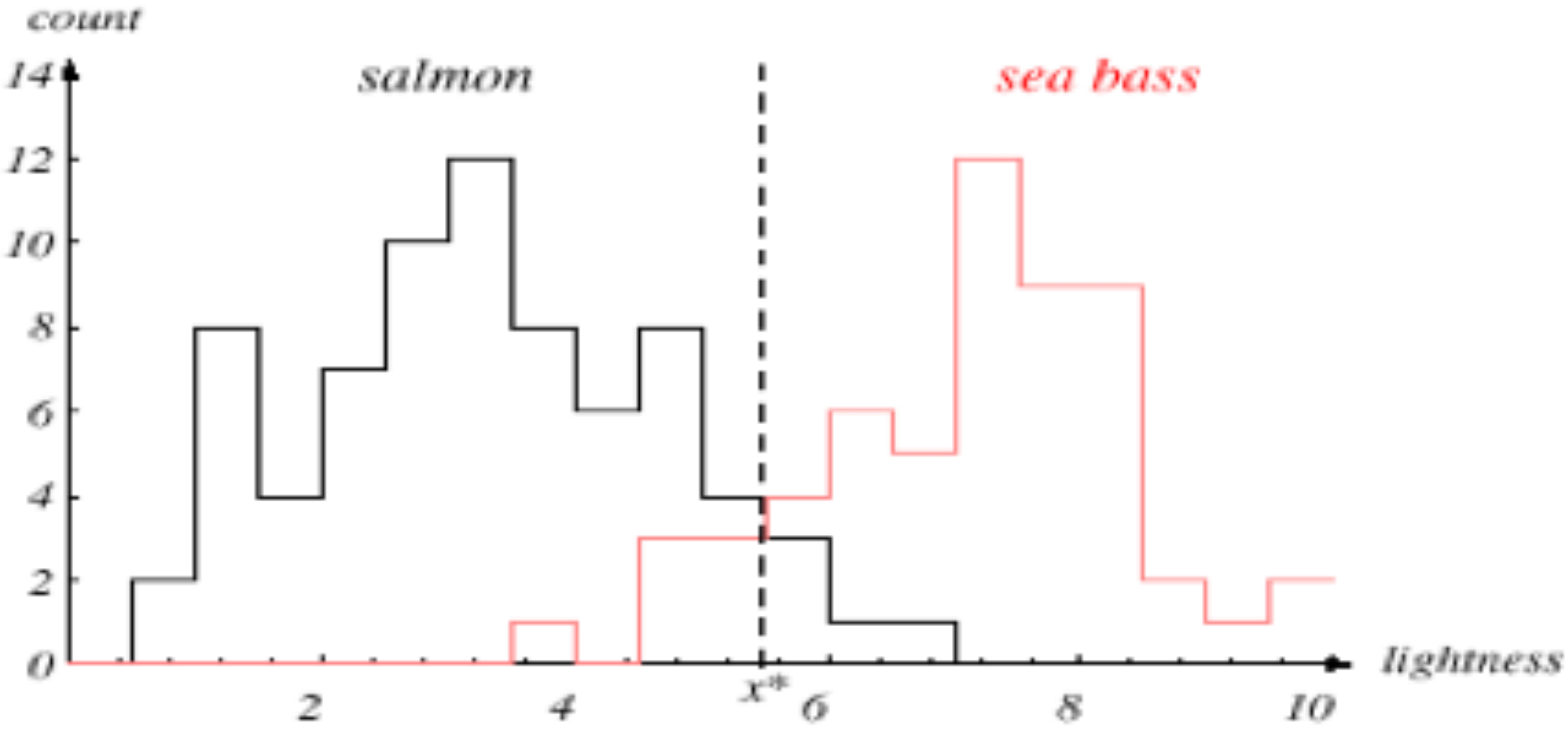
Classifier

type
bass

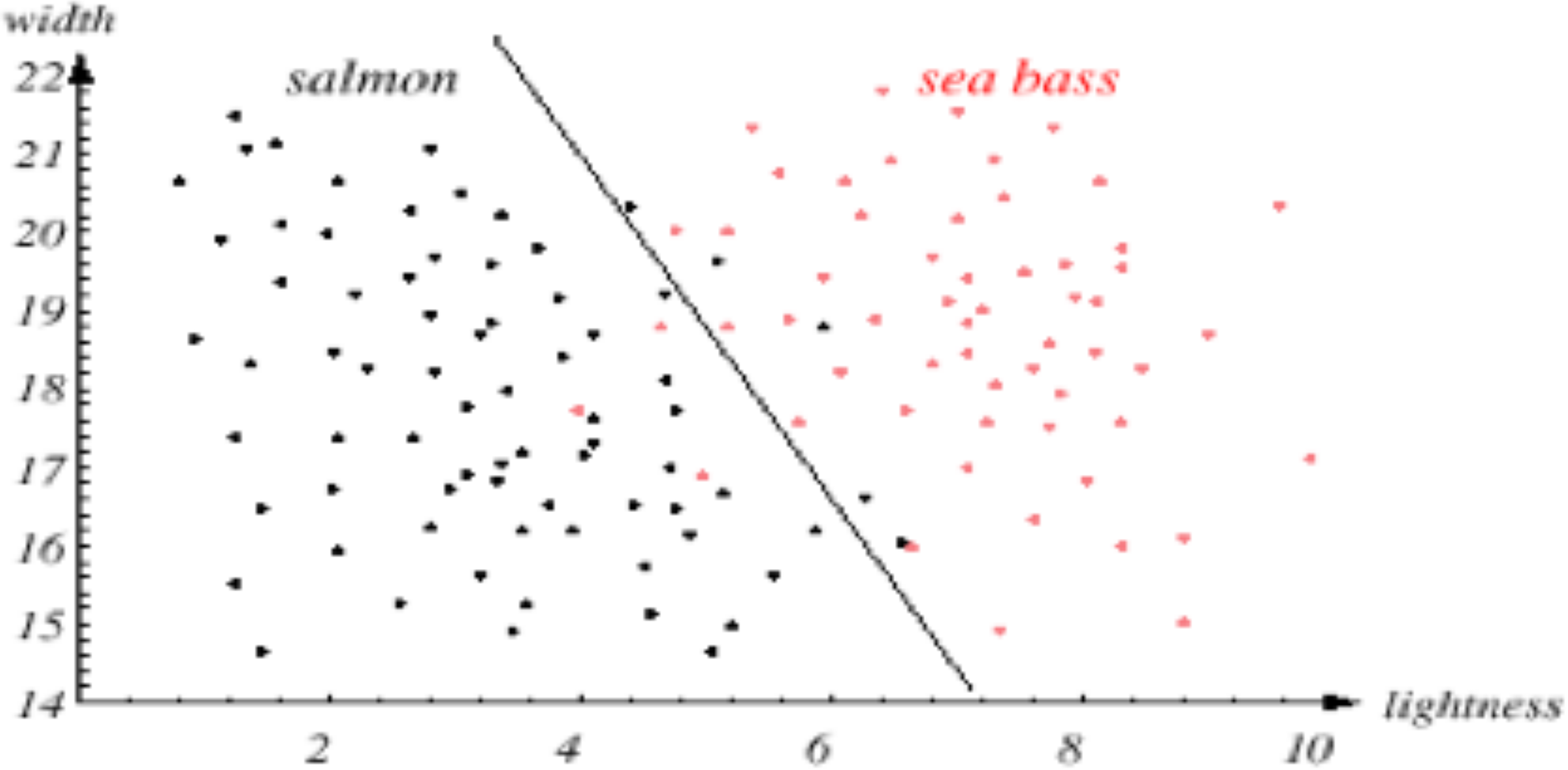
Use Length?


salmon
sea bass
count
22
20
18
16
12
10
8
6
4
2
0
length
5
10
15
20
25
l*

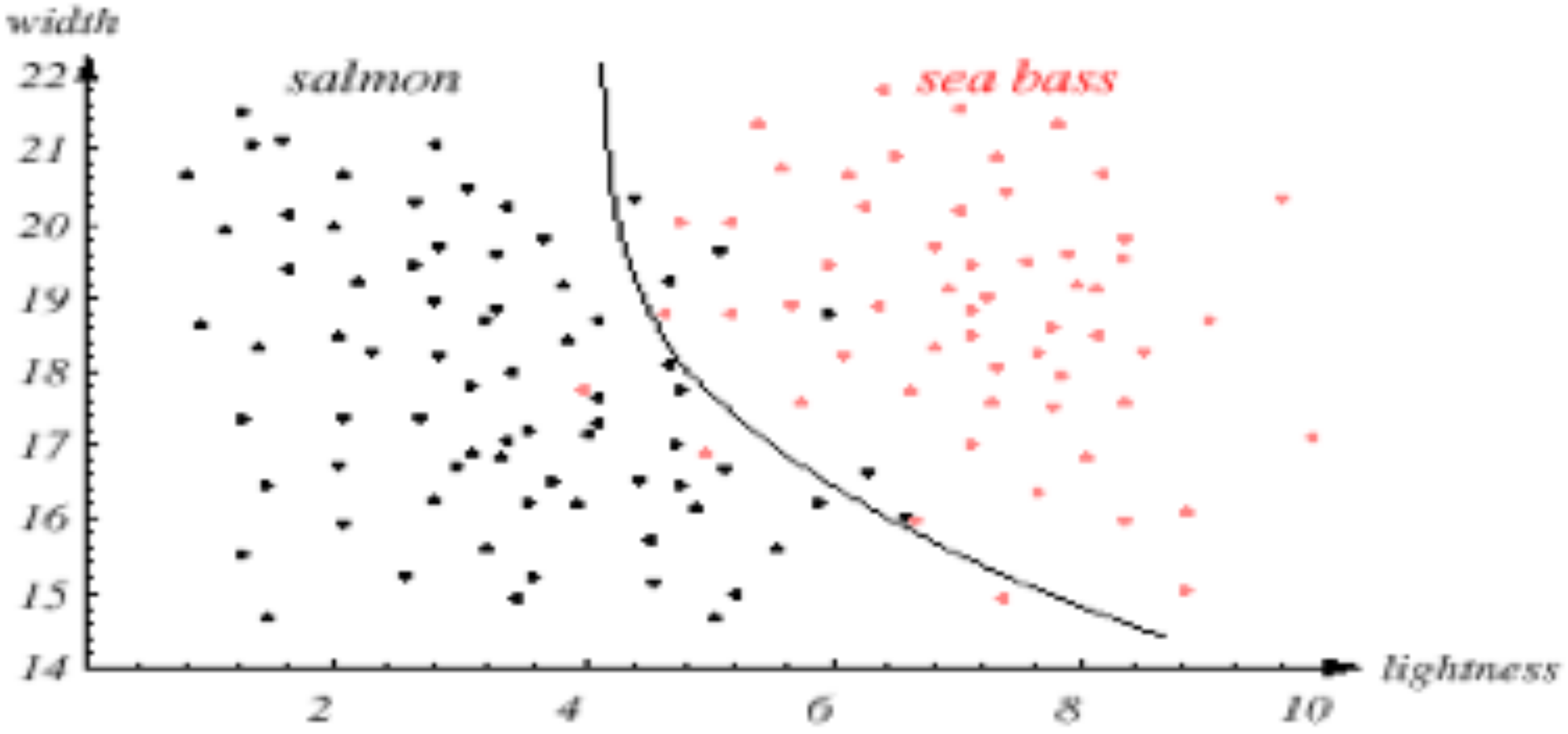
Use Lightness?



Use Both + Simple Line?

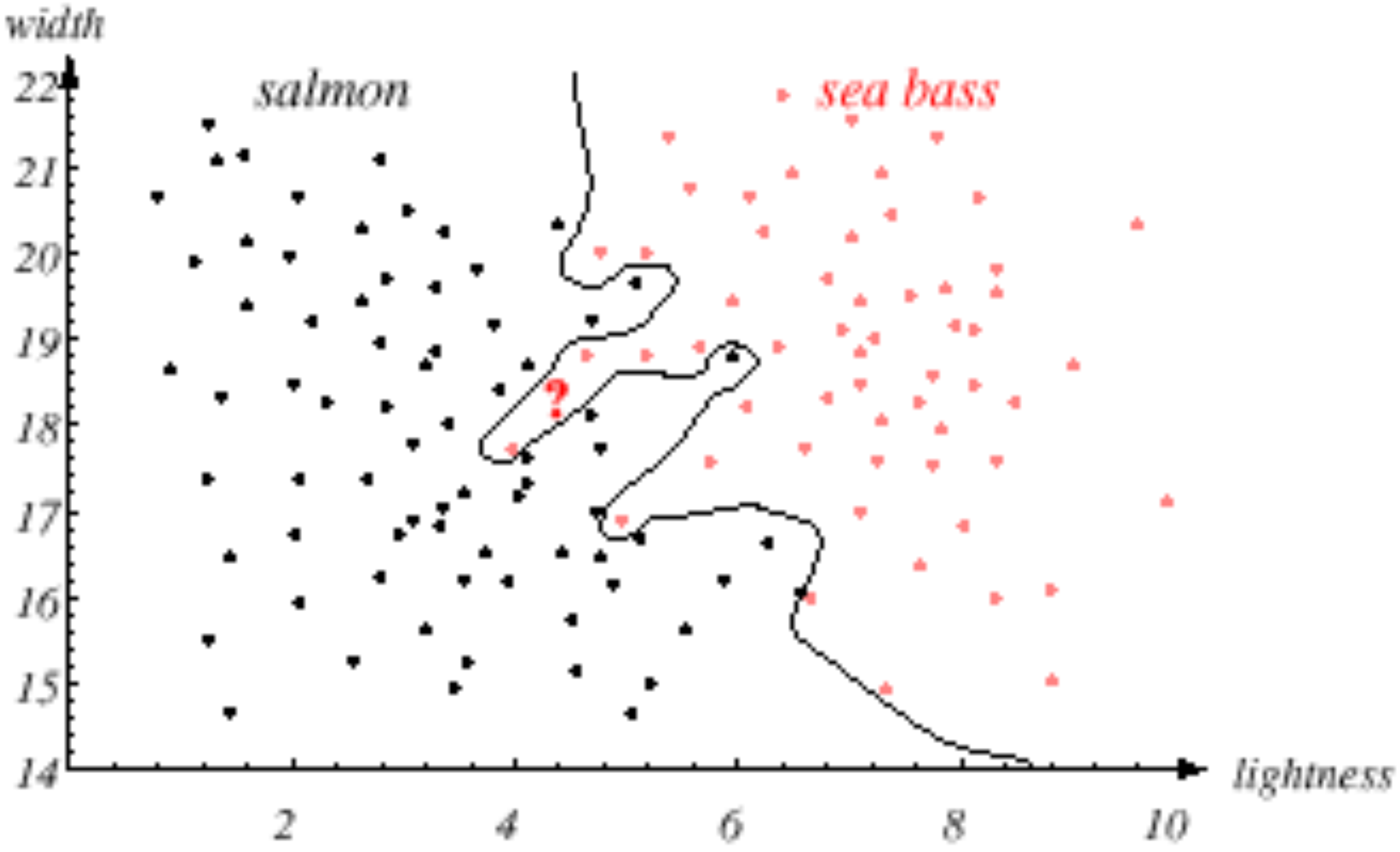


Curve?

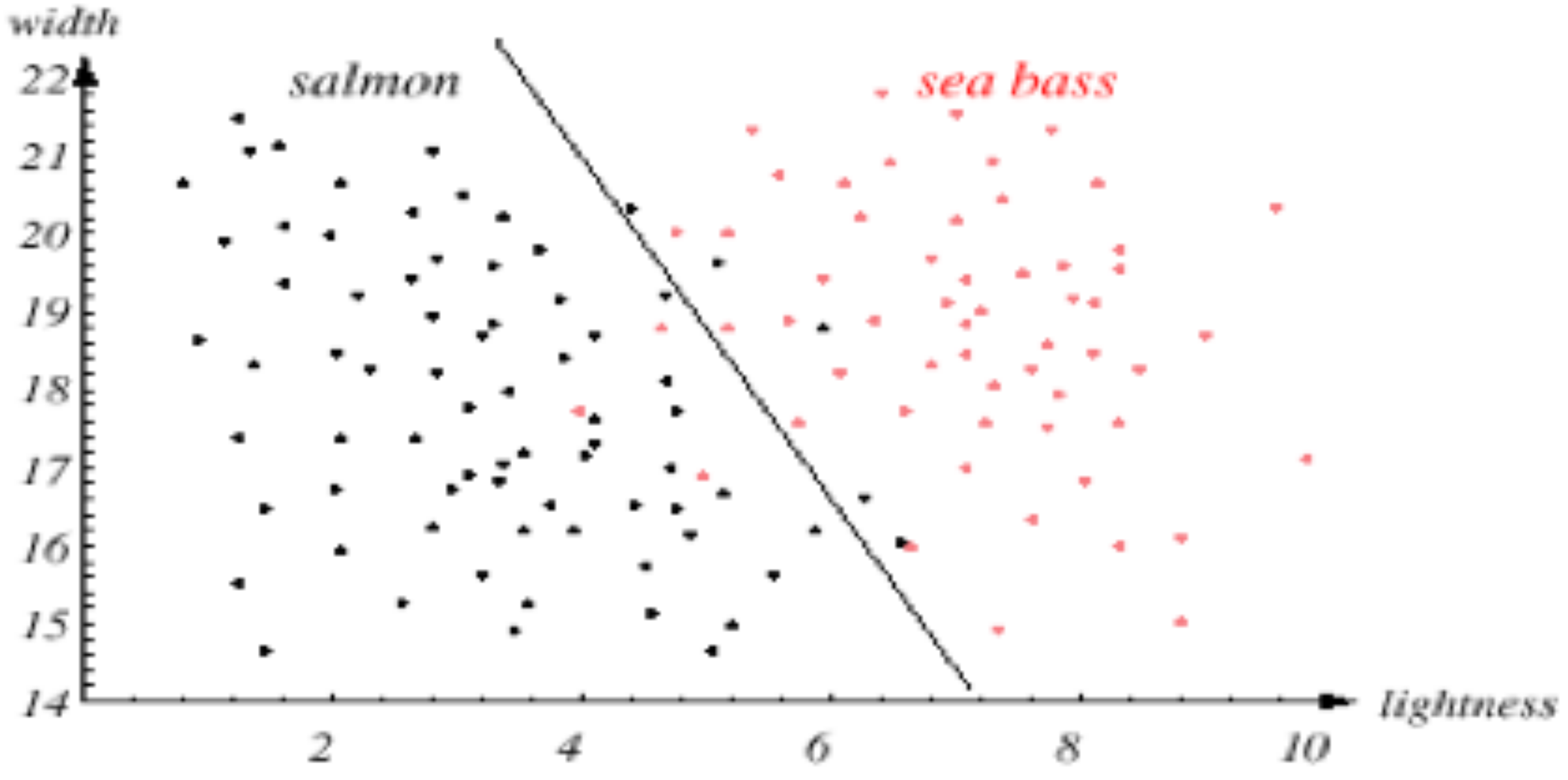


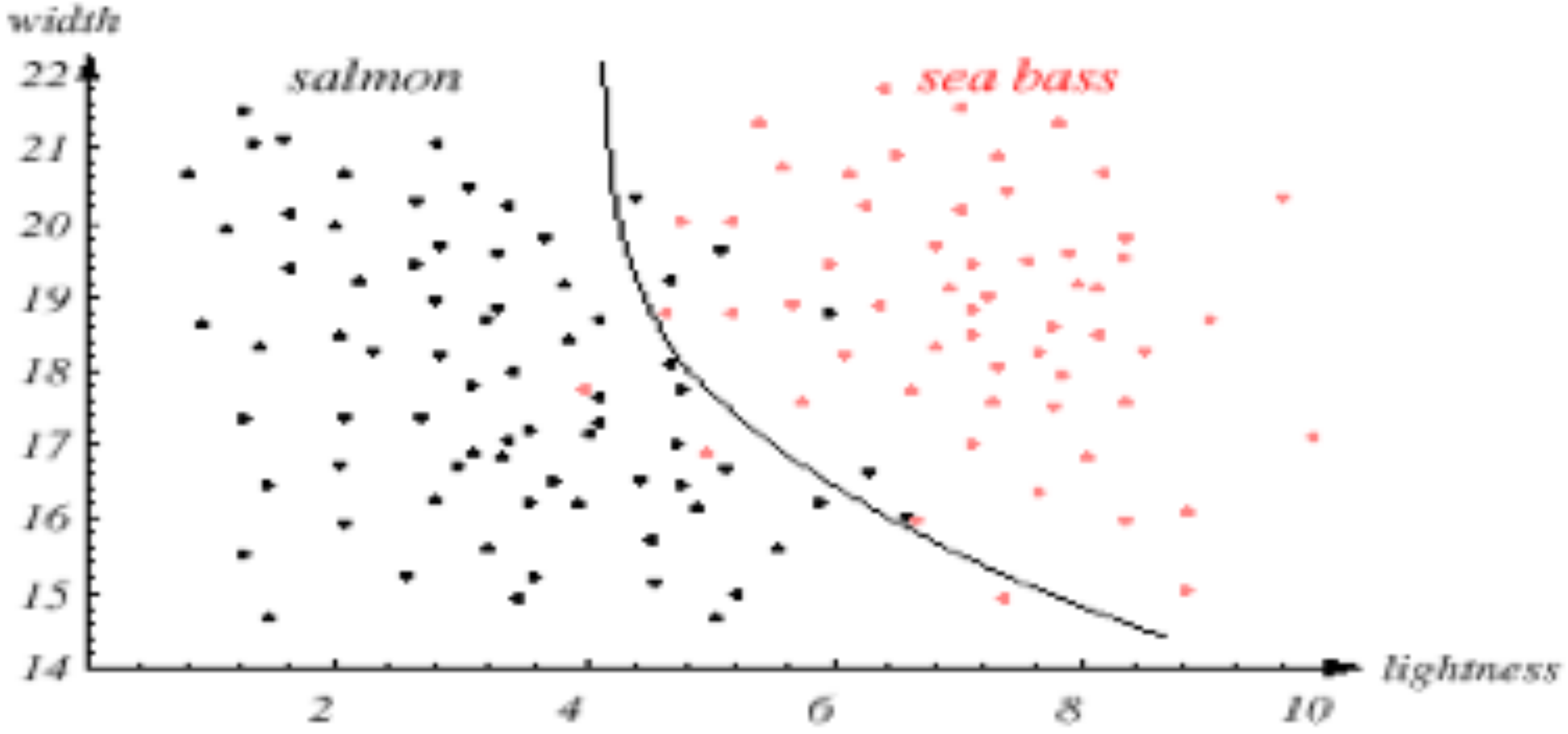
width
salmon
sea bass
lightness
22
21
20
19
18
17
16
15
14
2
4
6
8
10

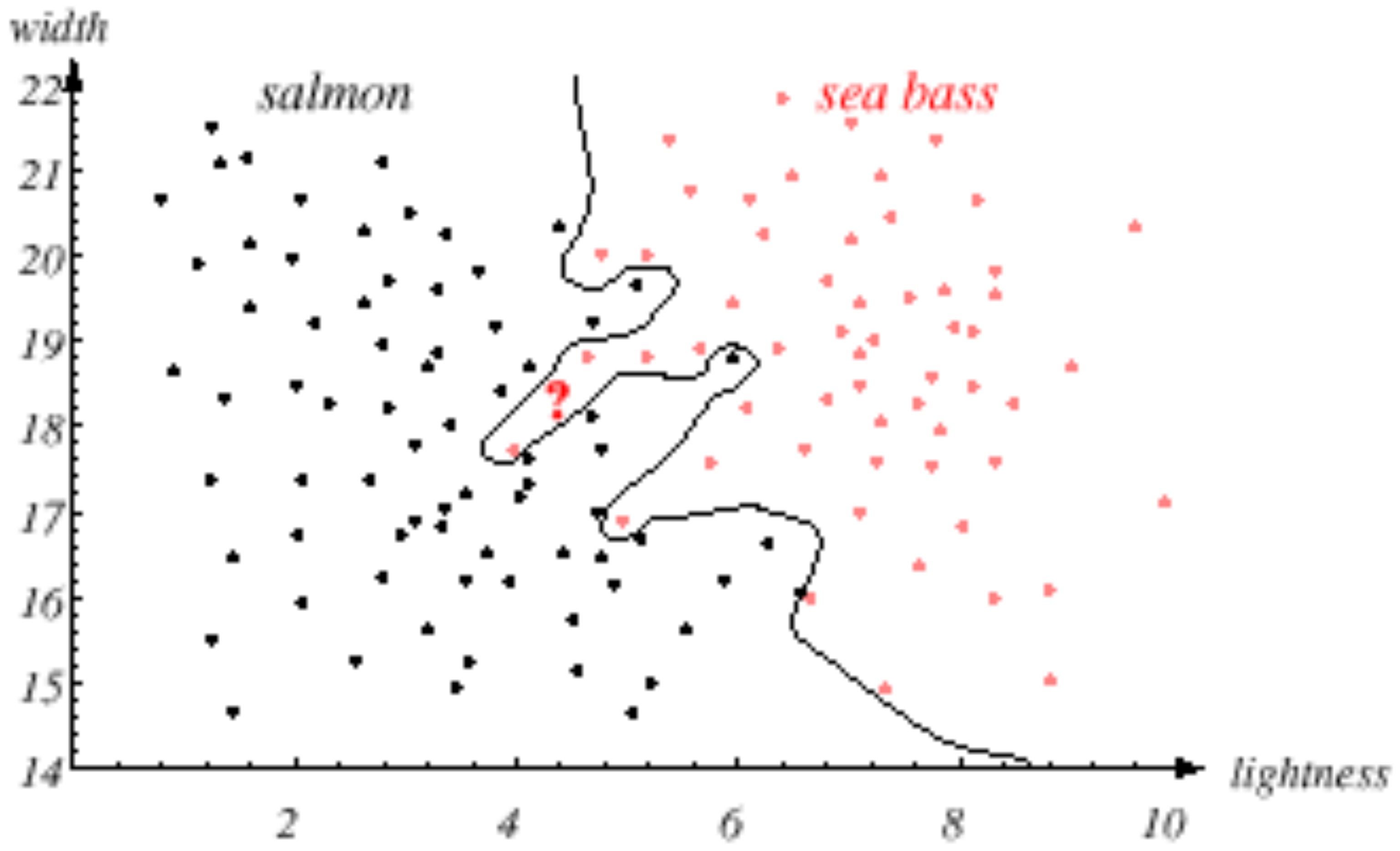
Optimal Performance?



Comparison with “New” Fish







Part of the Art of Machine Learning

- To what extent can we generalize examples to **nearby** examples or novel instances?
- What does **nearby** mean in the context of the current goals?

FEATURES

Features

- a representation of data
- for images it could be; raw pixel values, gradients, color histograms, edges, etc.
- for email/spam classification it could be; letters, words, histograms, n-grams, etc.
- for a person it could be; age (insurance risk), address (languages spoken), etc. Task dependent!
- usually a vector, e.g. age (34,54,87…)

$$\mathbf{x} = [x_1, x_2, x_3, \ldots, x_m] \qquad \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$$

Features & Labels

- Labels can be associated with images OR features
- e.g. this image is a person (or not-person [negative class])
- e.g. this feature was computed over a region of the image containing a cat, dog, carpet etc.
- Generally represented numerically e.g. +1 = positive example, -1 = negative example

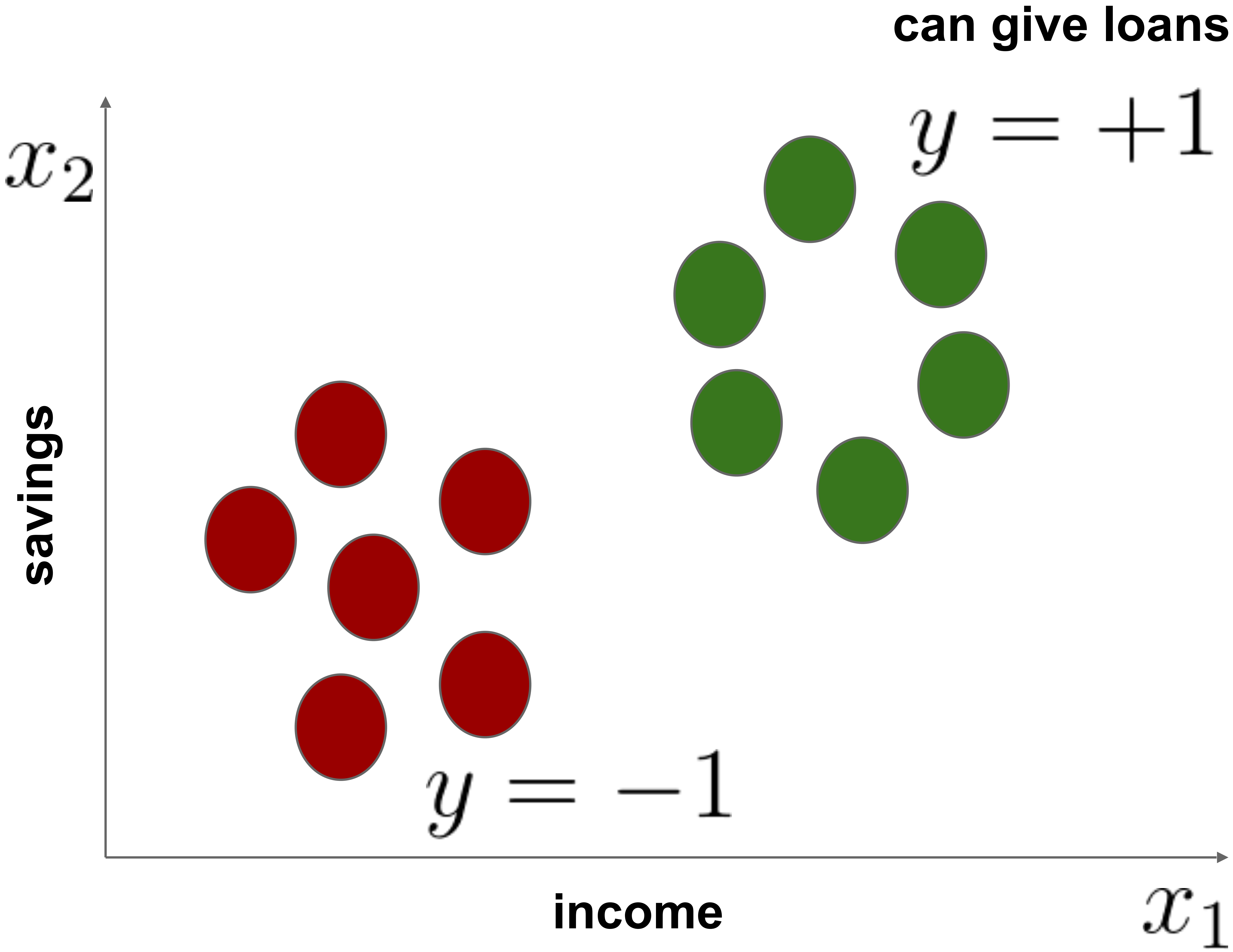
$$y_i = \{+1, -1\}$$

Features & Labels

- Often we are considering a dataset of “tuples”:
  N datapoints, **N** labels

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)$$

Features & Labels

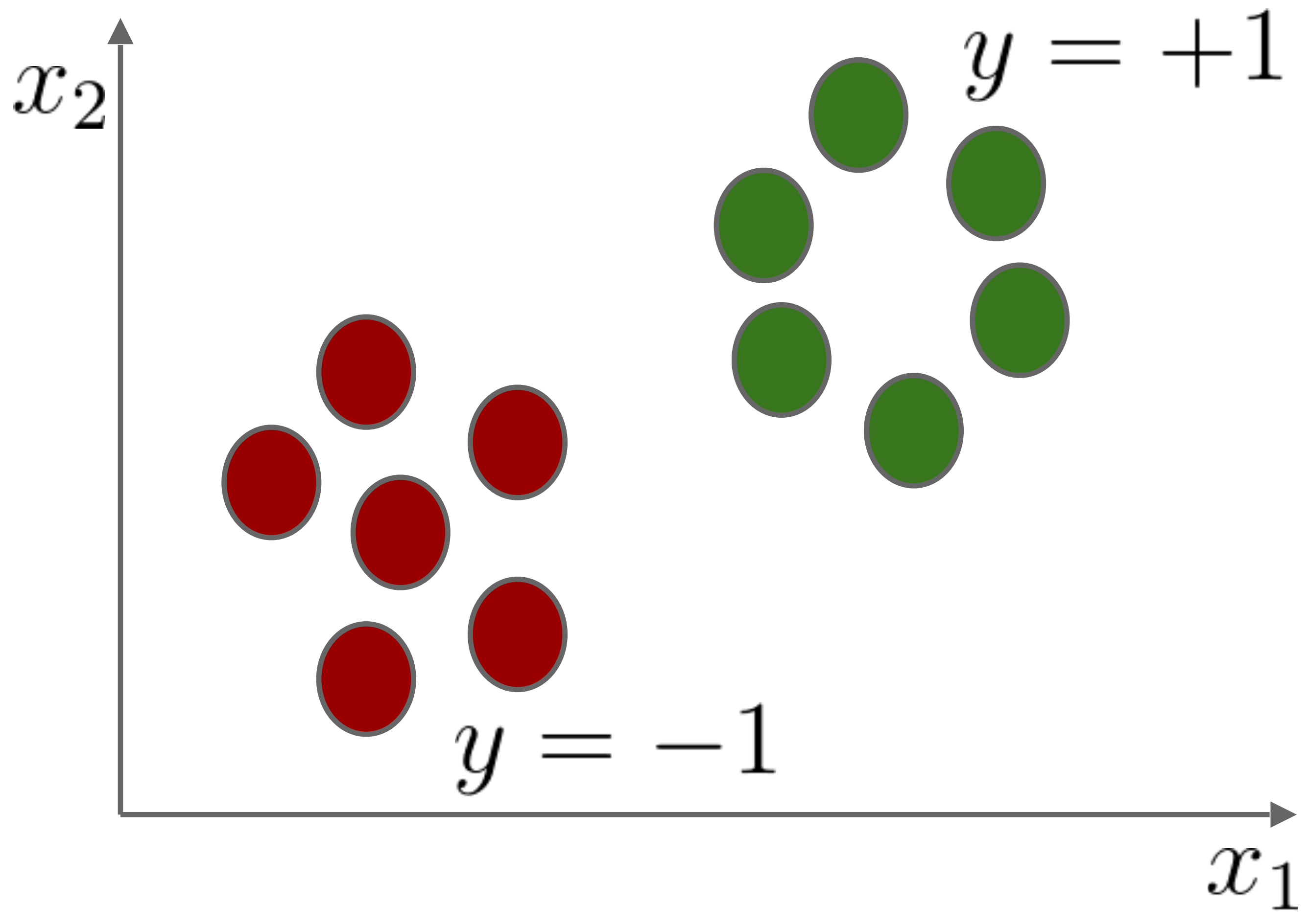


Supervised vs Unsupervised

Supervised vs Unsupervised

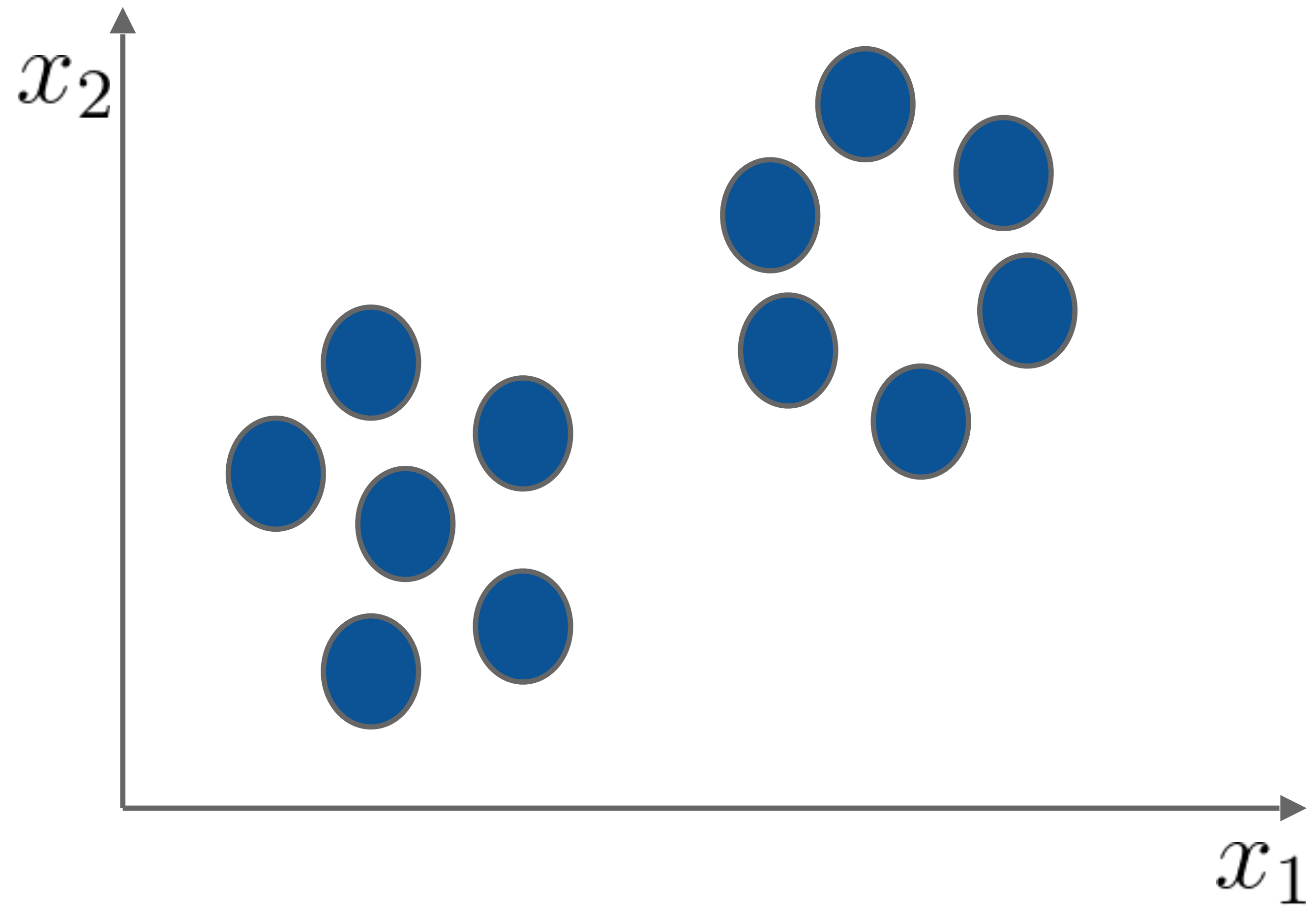
Supervised:

- for every **x**, we have a **y**

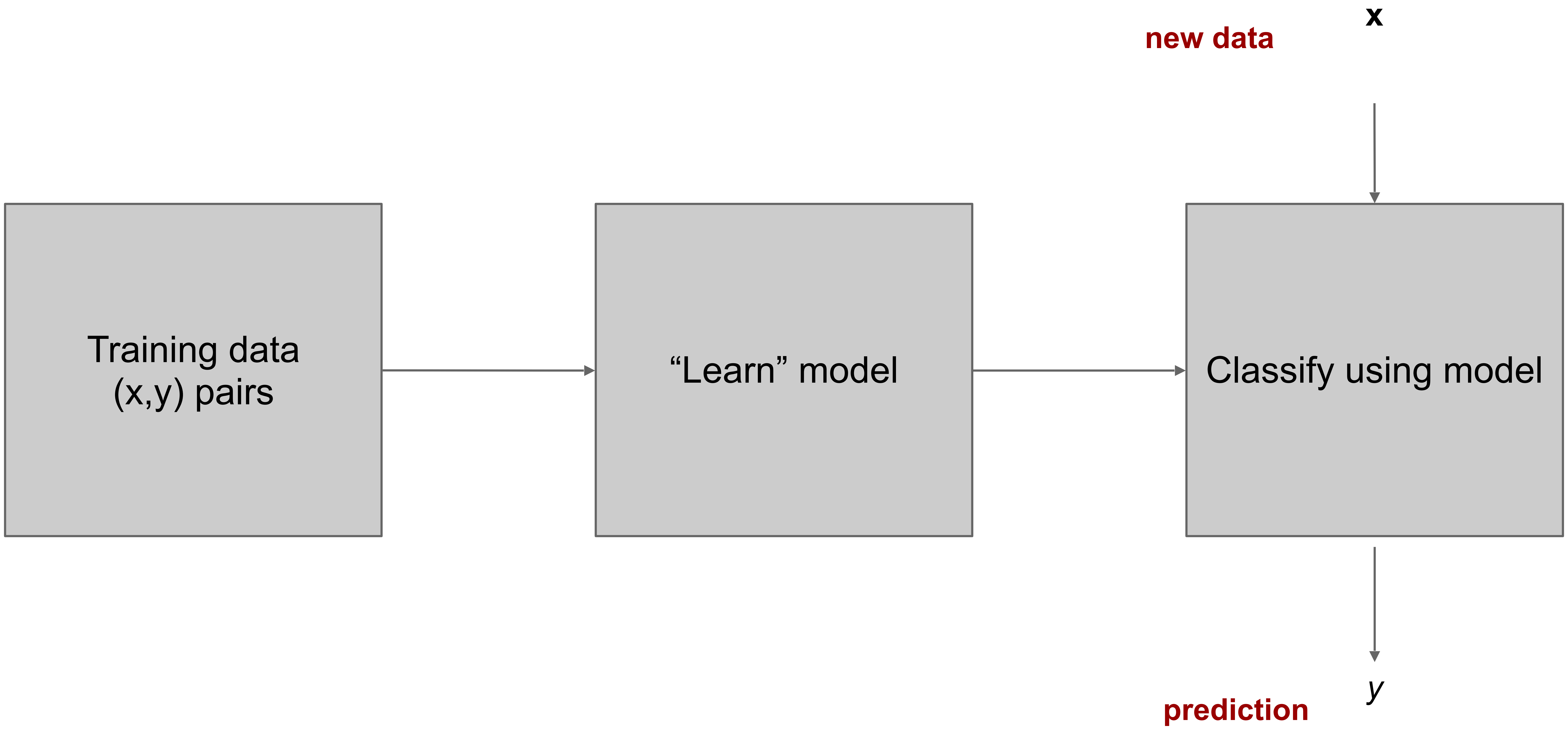


Unsupervised:

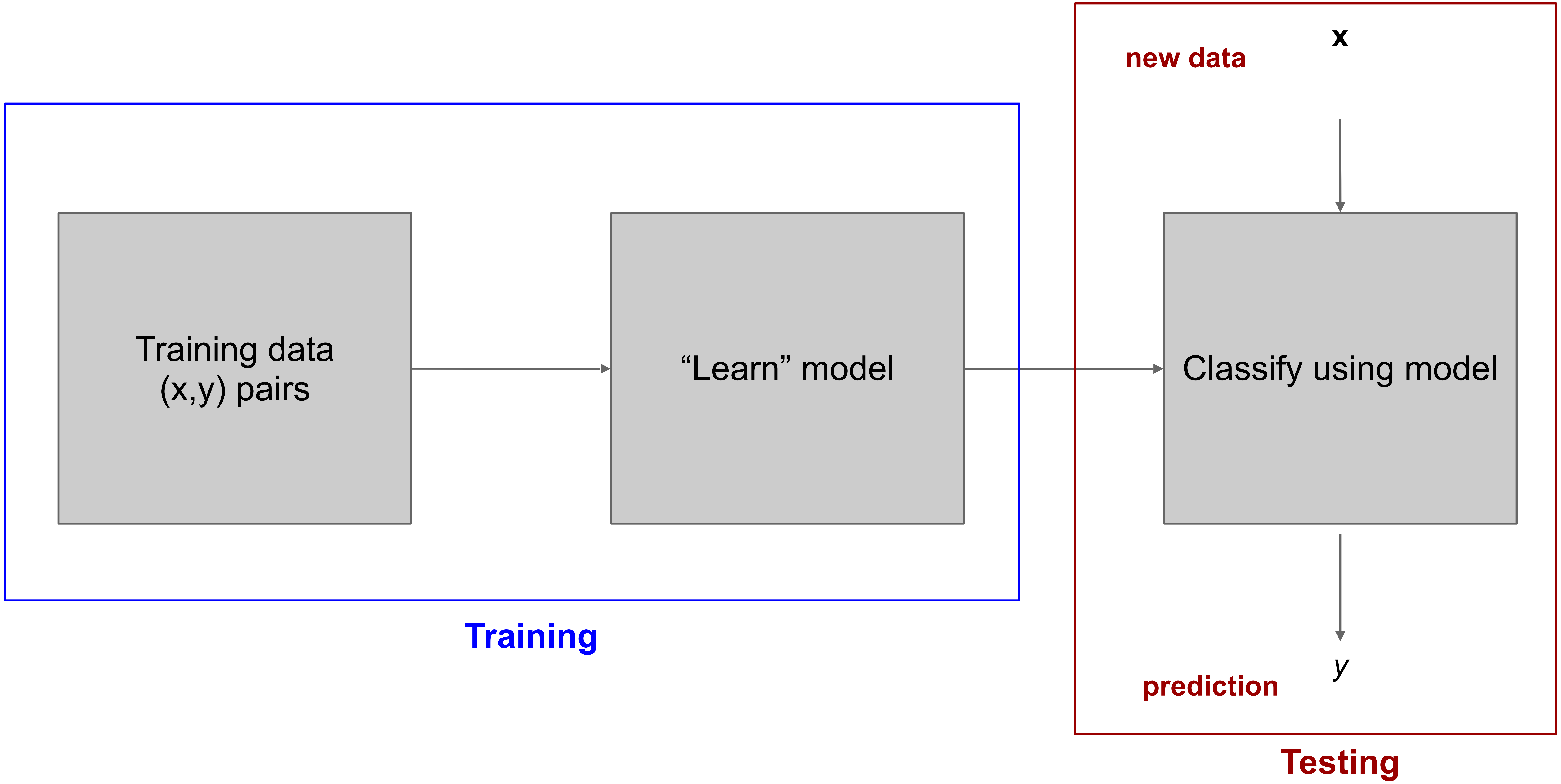
- only **x**, no labels



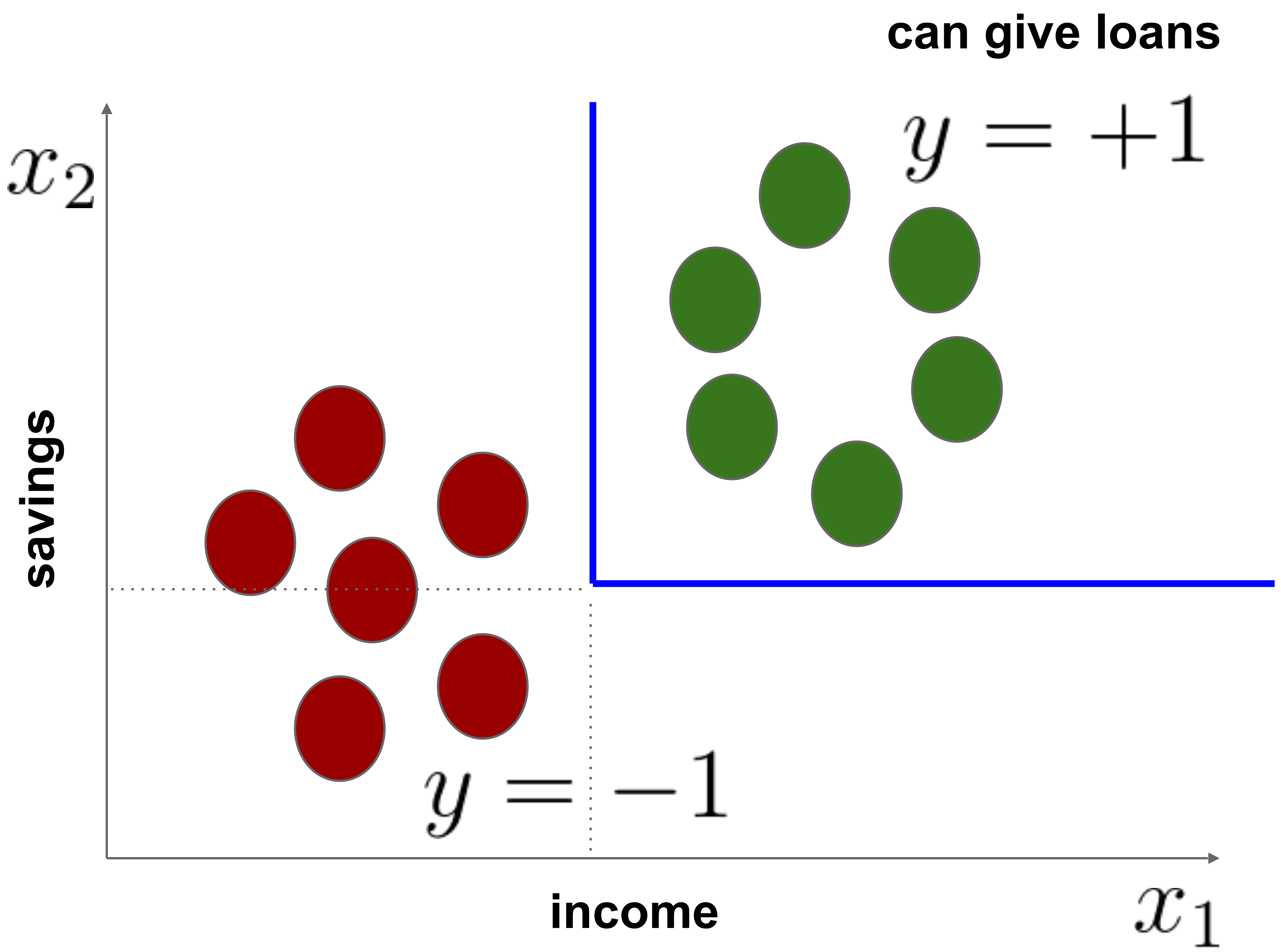
Classification



Classification



Classification



if **(income > T &**

savings > S) then

giveLoan()

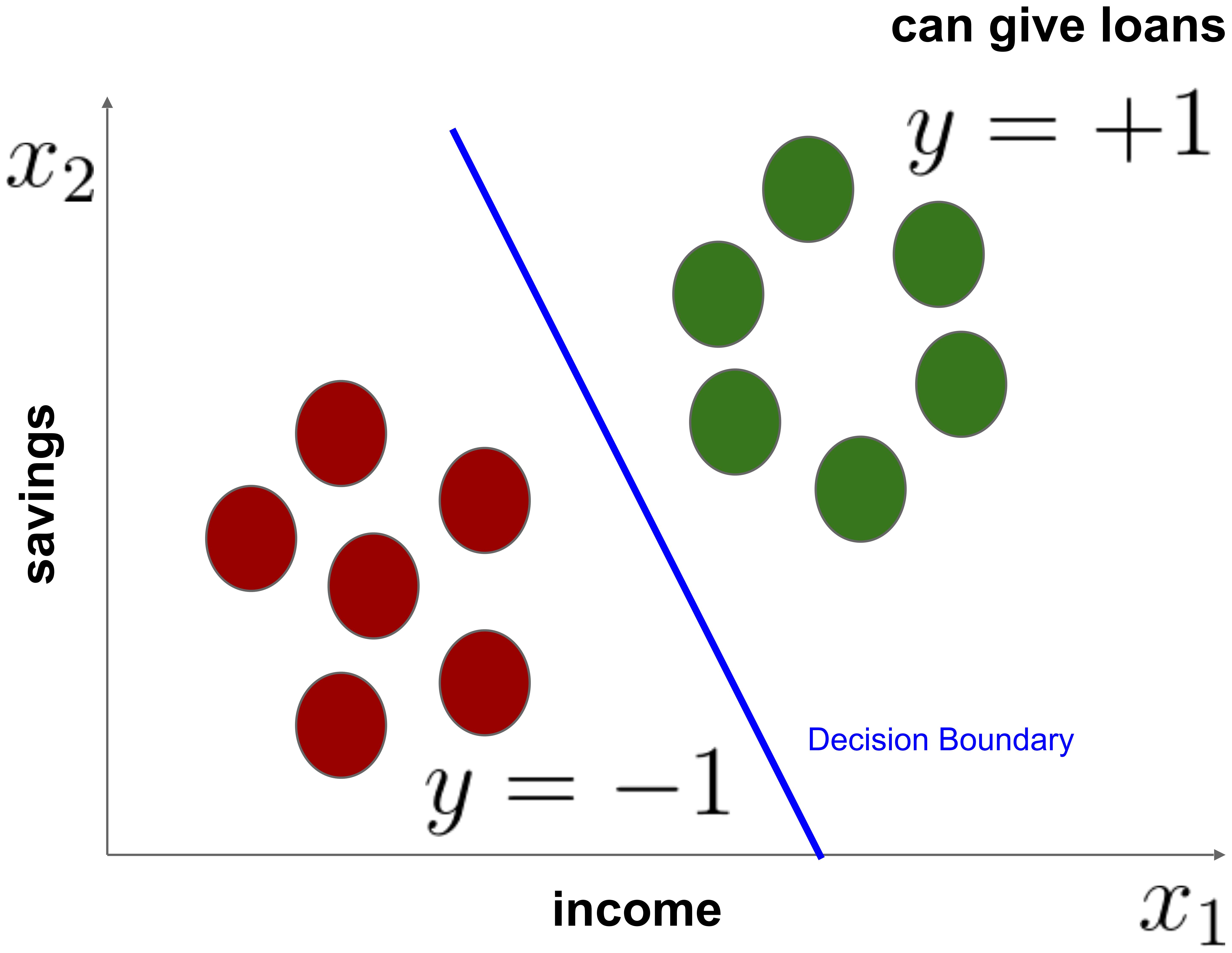
else

openTrapdoor()

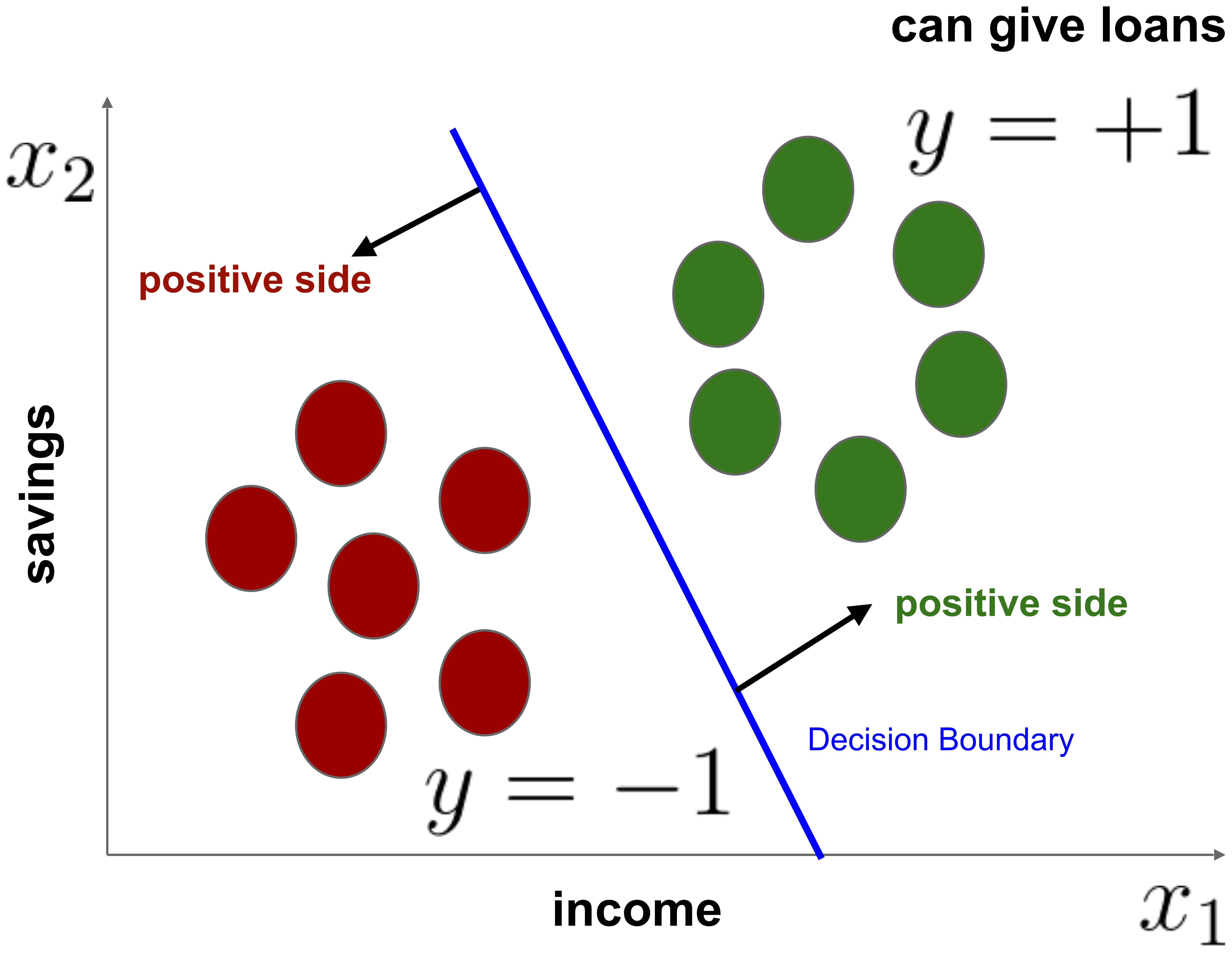
end

Linear Classifiers and Support Vector Machines

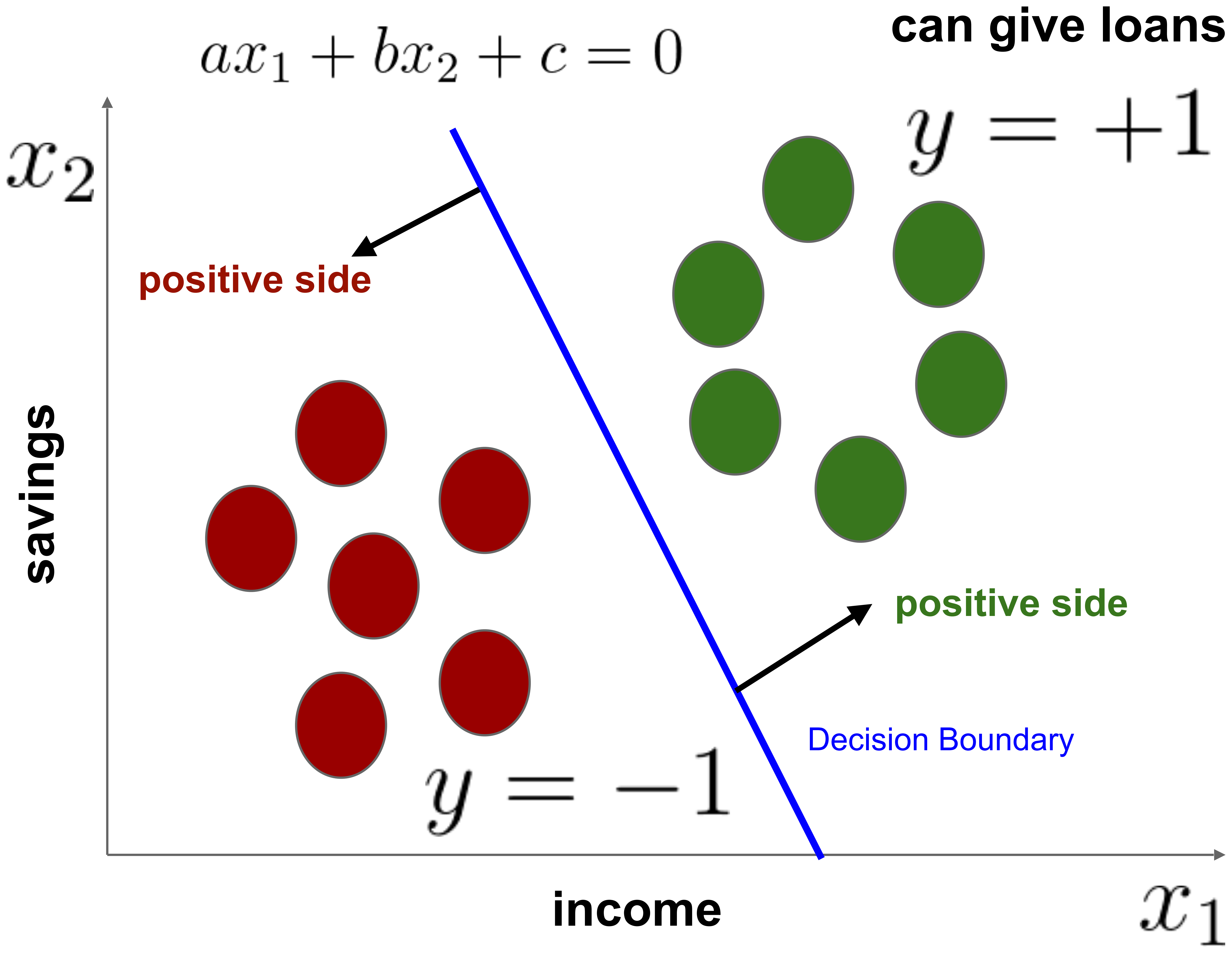
Linear Classifier



Linear Classifier



Linear Classifier



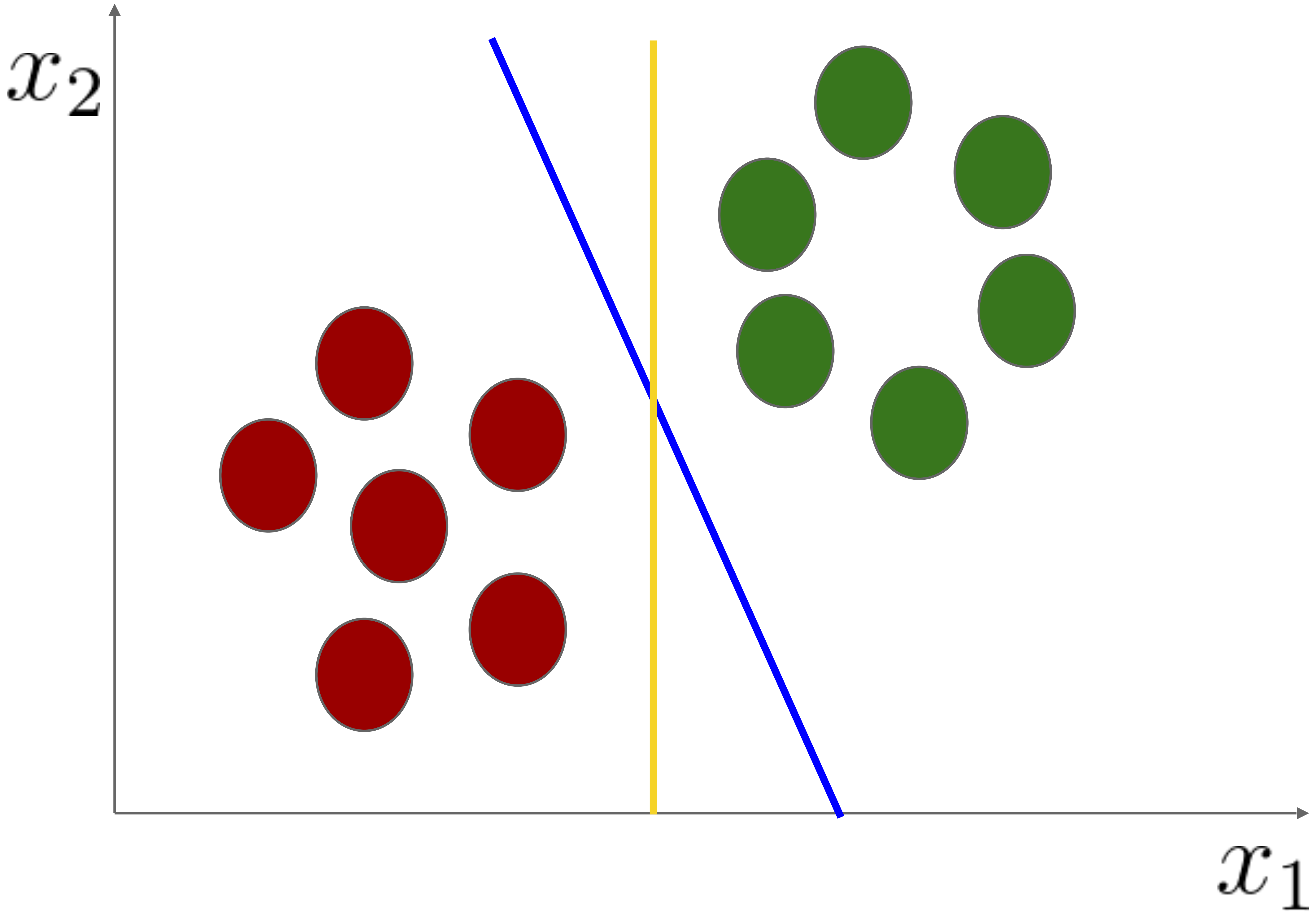
Linear Classifier

$$ax_1 + bx_2 + c = 0$$

$$\mathbf{w} = [a\ b]$$
$$\mathbf{x} = [x_1\ x_2]$$

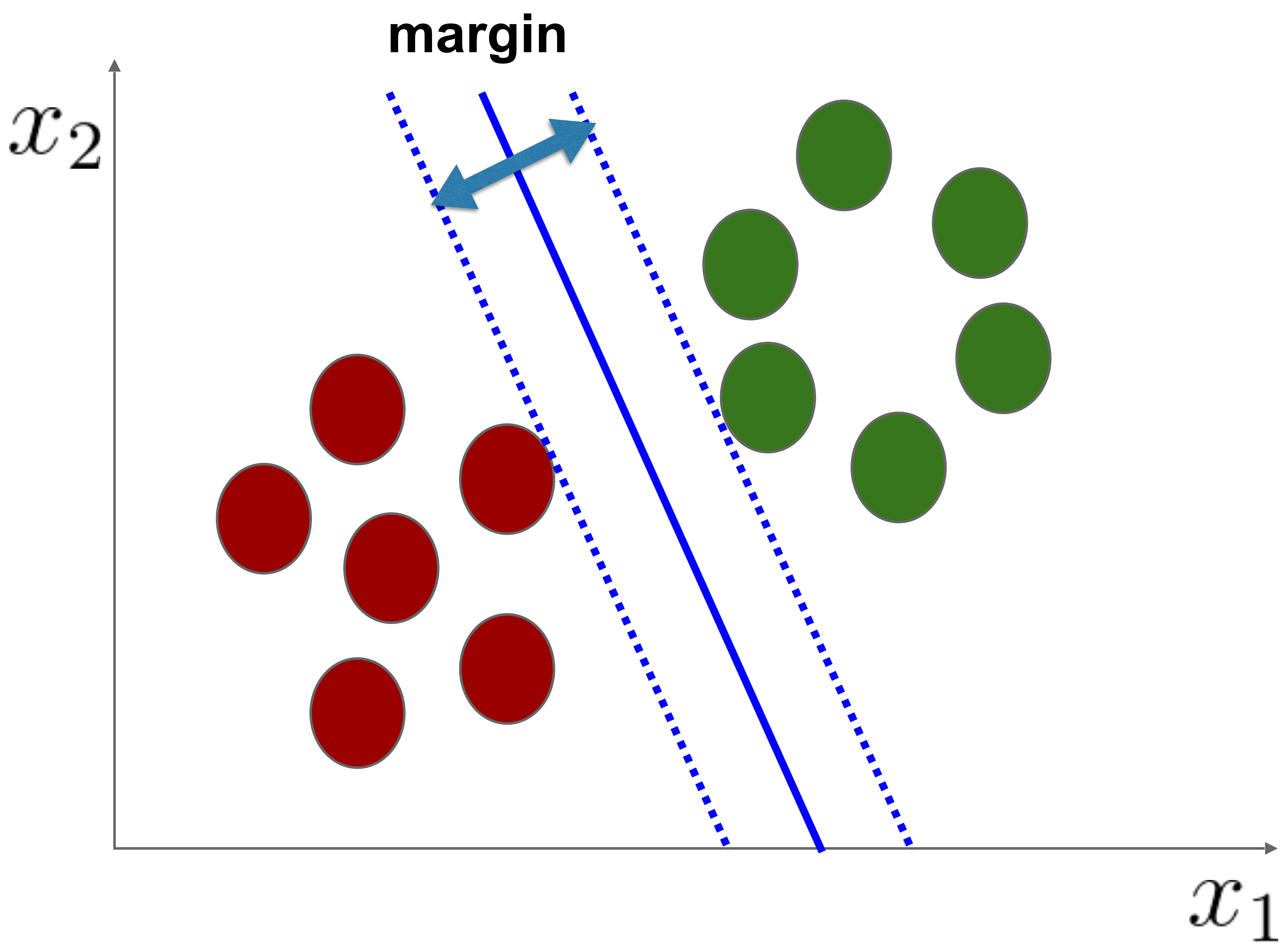
$$\mathbf{w}^T\mathbf{x} + c > 0\ \text{(positive)}$$
$$\mathbf{w}^T\mathbf{x} + c < 0\ \text{(negative)}$$

Linear Classifiers - which is better?



What's the Difference

Support Vector Machine (SVM)

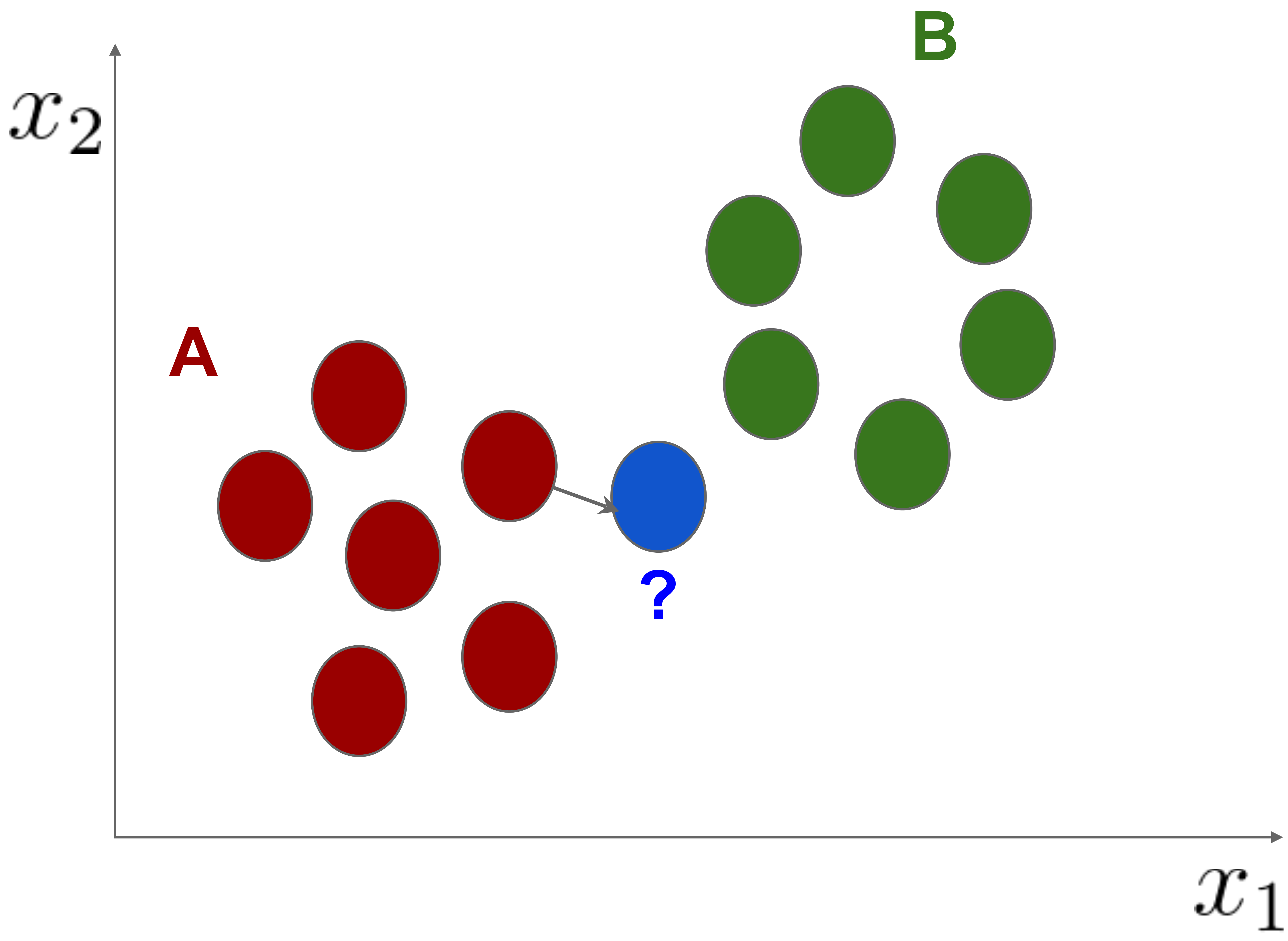


SVM Issues

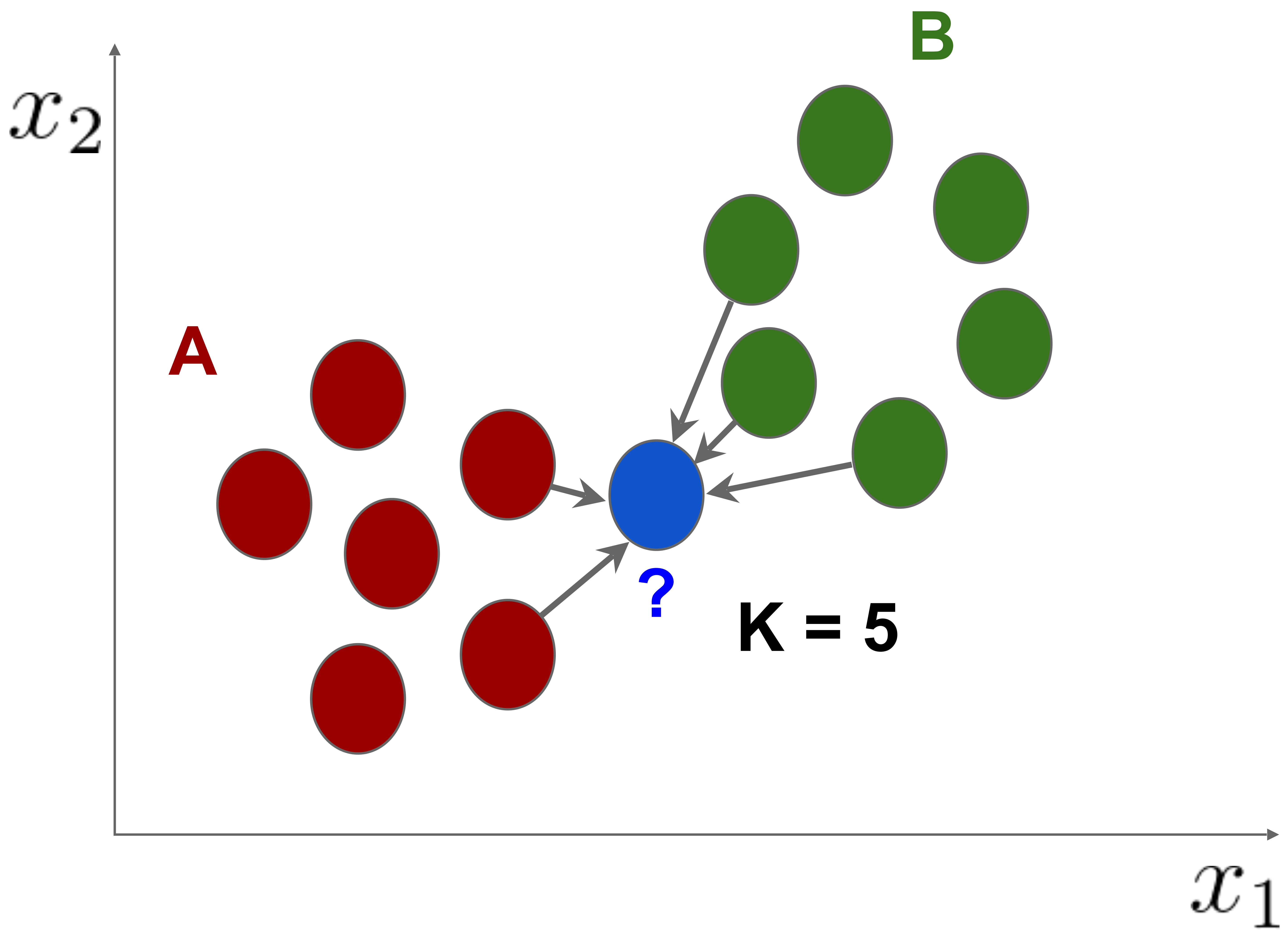
- What about multiple classes?
- How to get non-linear decision boundaries?
- Computation and training complexity
- Good-to-know : libSVM (GPU libraries also available)

Nearest Neighbor Classifiers

Nearest Neighbor



K - Nearest Neighbor



Unsupervised Learning

Unsupervised Learning

- Find hidden structure in unlabeled data
- Given data X, ask a good question:
  - Clustering
  - Component analysis
  - Commonality discovery
  - Novelty/anomaly detection
  - Sequence analysis
  - more...

Clustering

Find a set of prototypes representing the data

- Clustering More than Two Million Biomedical Publications

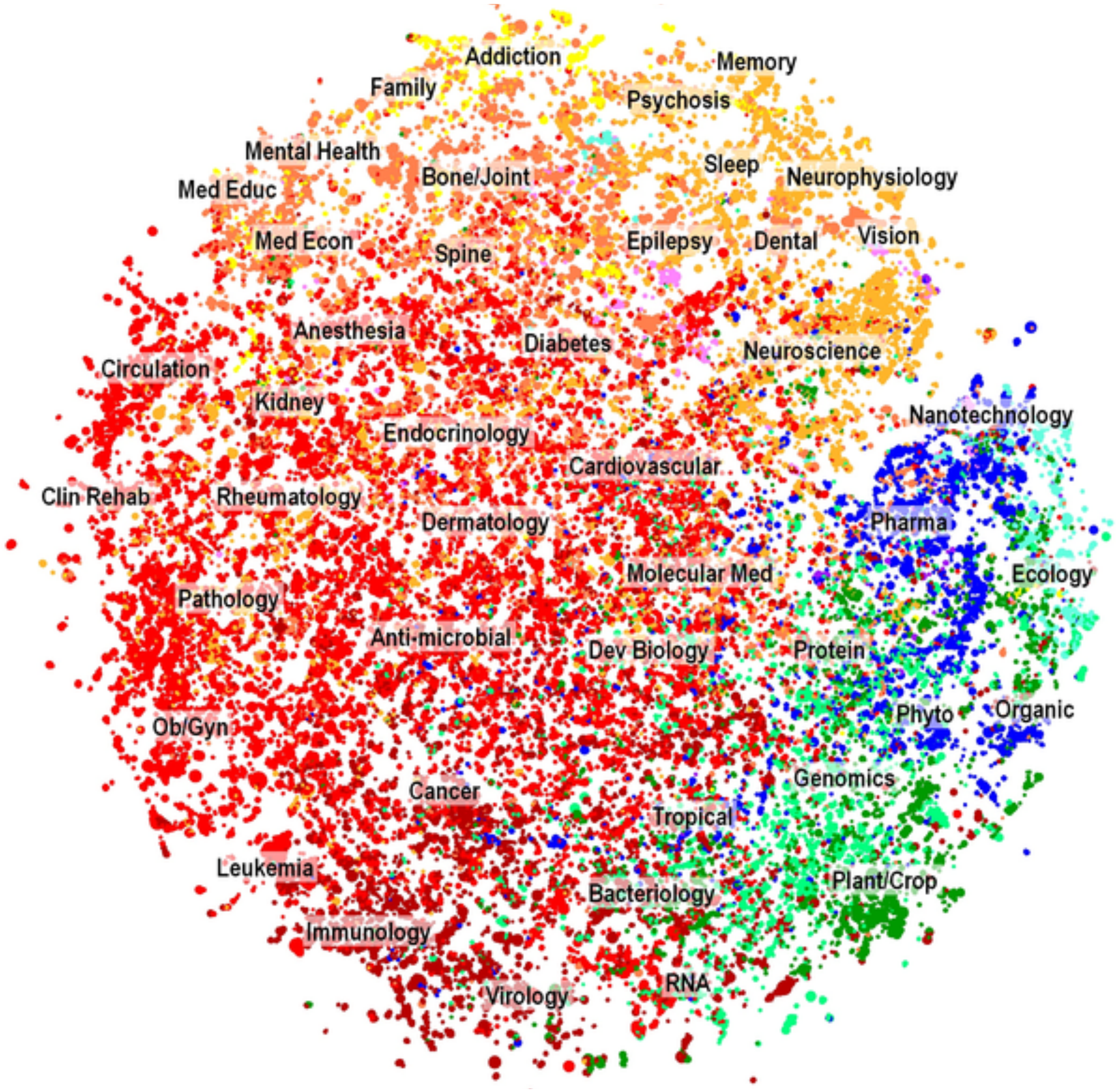
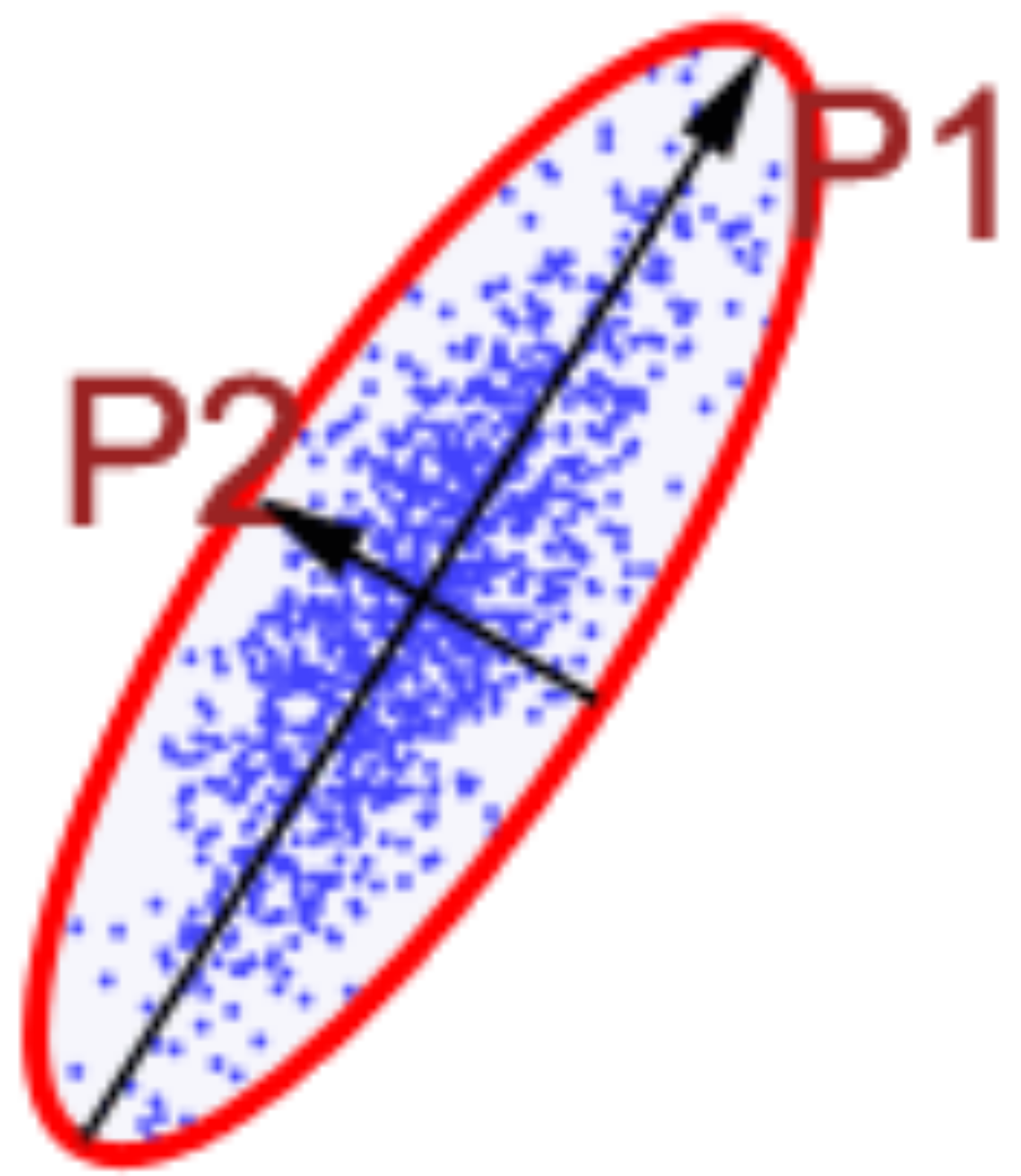


Image from http://www.plosone.org/article/info:doi/10.1371/journal.pone.0018029

Component Analysis

Find a subspace representing the data



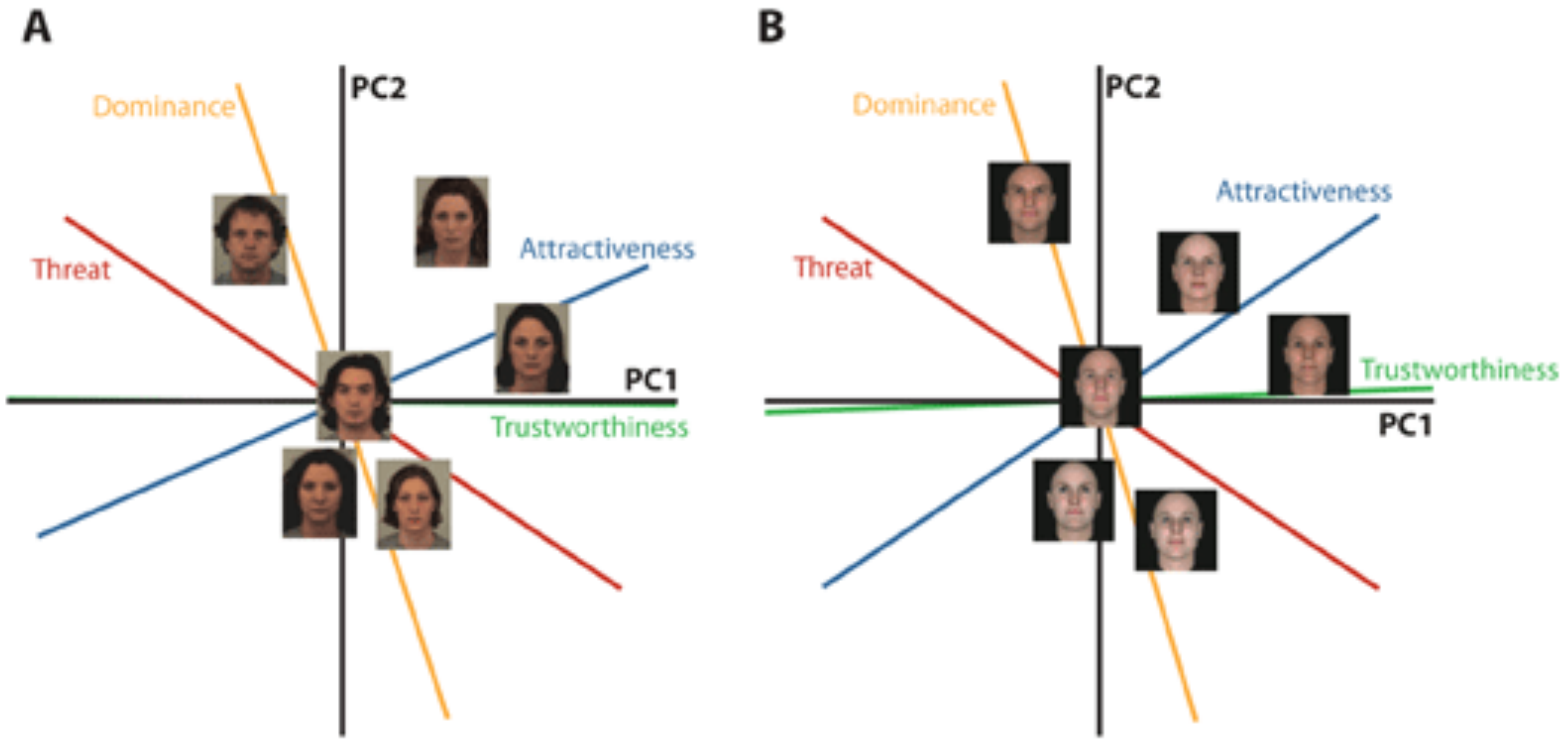


Image from http://www.apa.org/science/about/psa/2010/03/sci-brief.aspx

Dictionary Learning

Find a small set of factors from observation

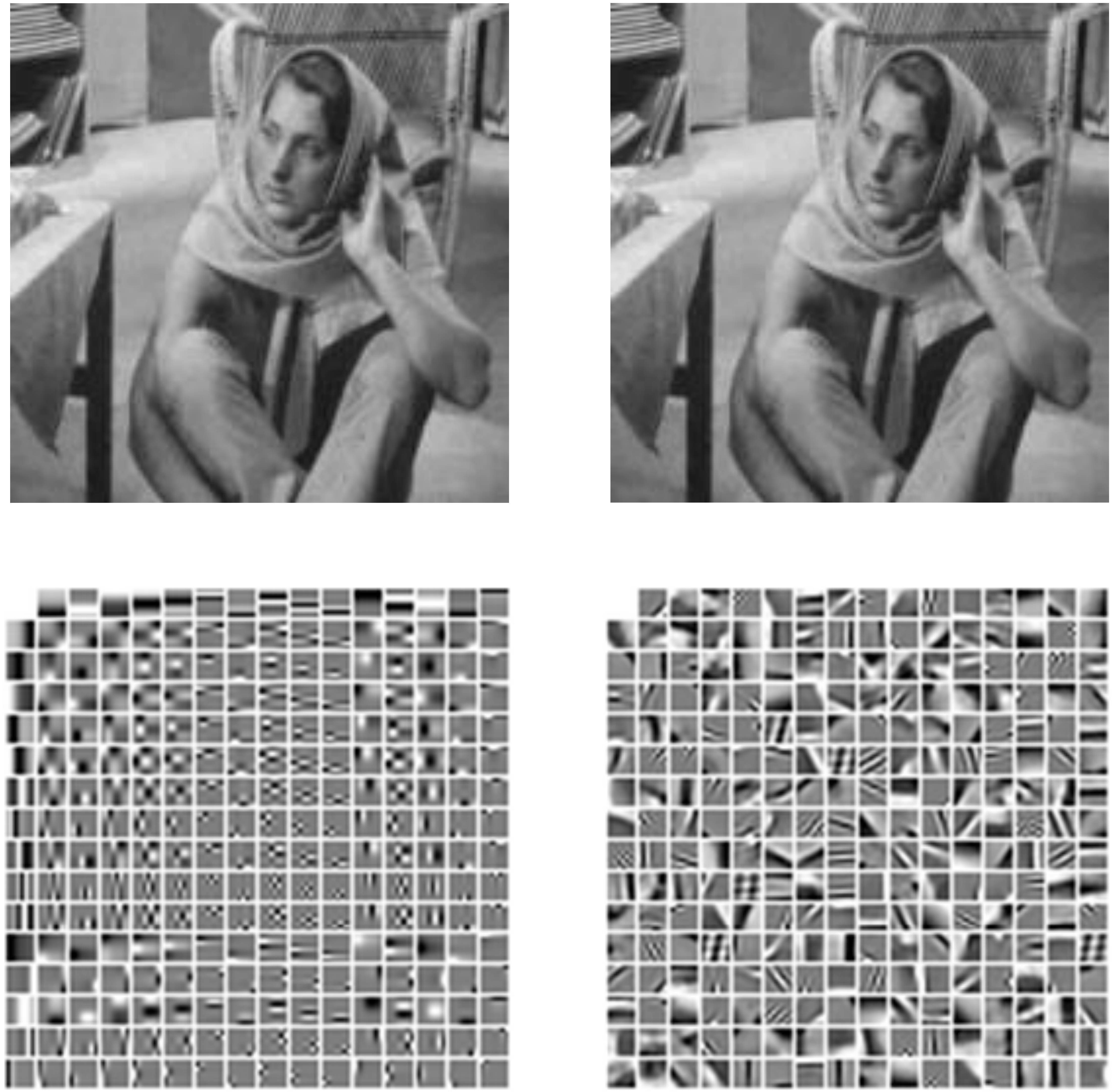


Image from http://www.gol.ei.tum.de/index.php?id=25&type=98

Commonality Discovery

Find commonalities in images

Image from http://humansensing.cs.cmu.edu/wschu/

Novelty/Anomaly Detection

Find the odd ones

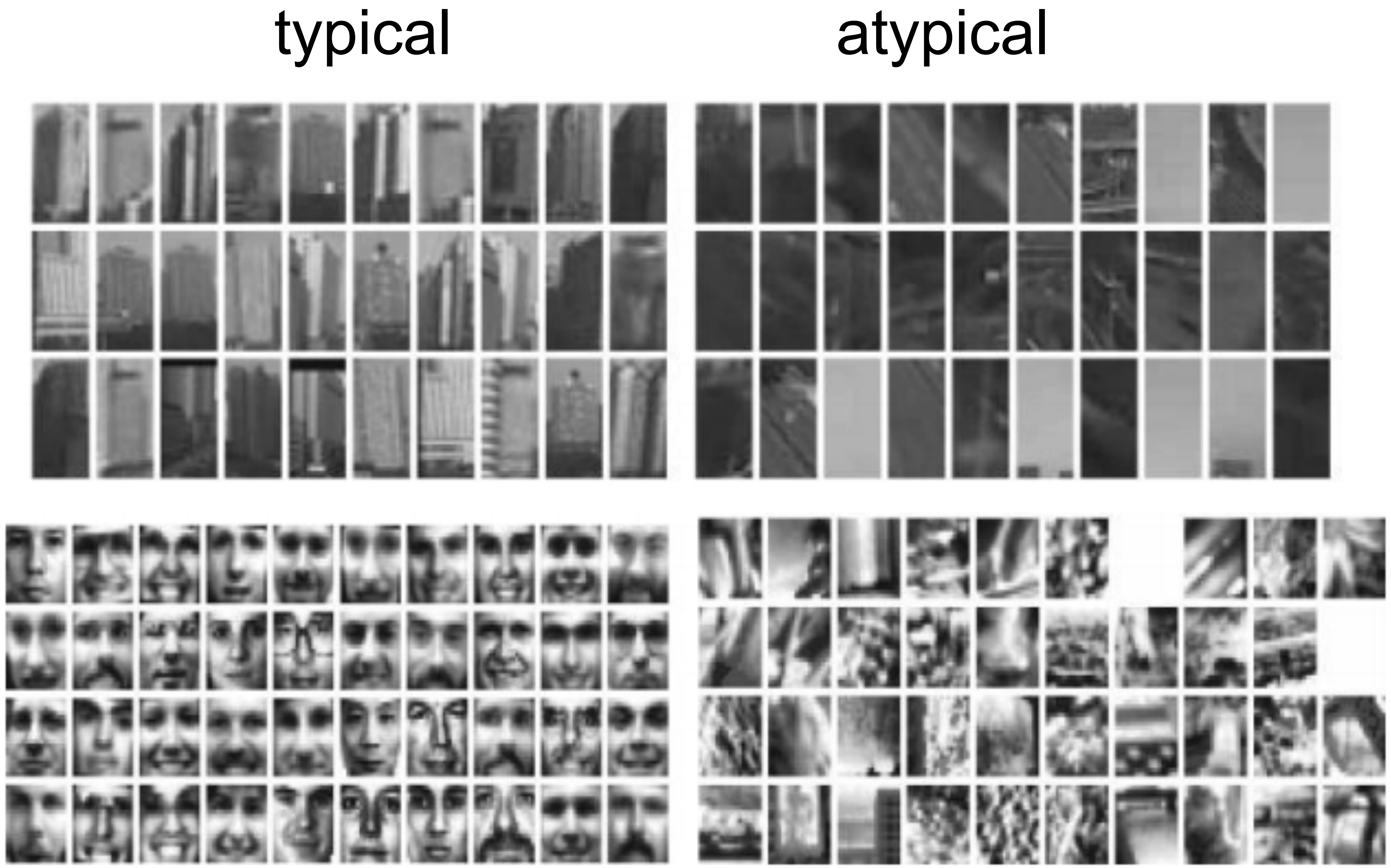


Image from Alex Smola's slides for 10-701 Machine Learning

Sequence Analysis

Find a latent sequence for observation

- Identification and analysis of functional elements in 1% of the human genome

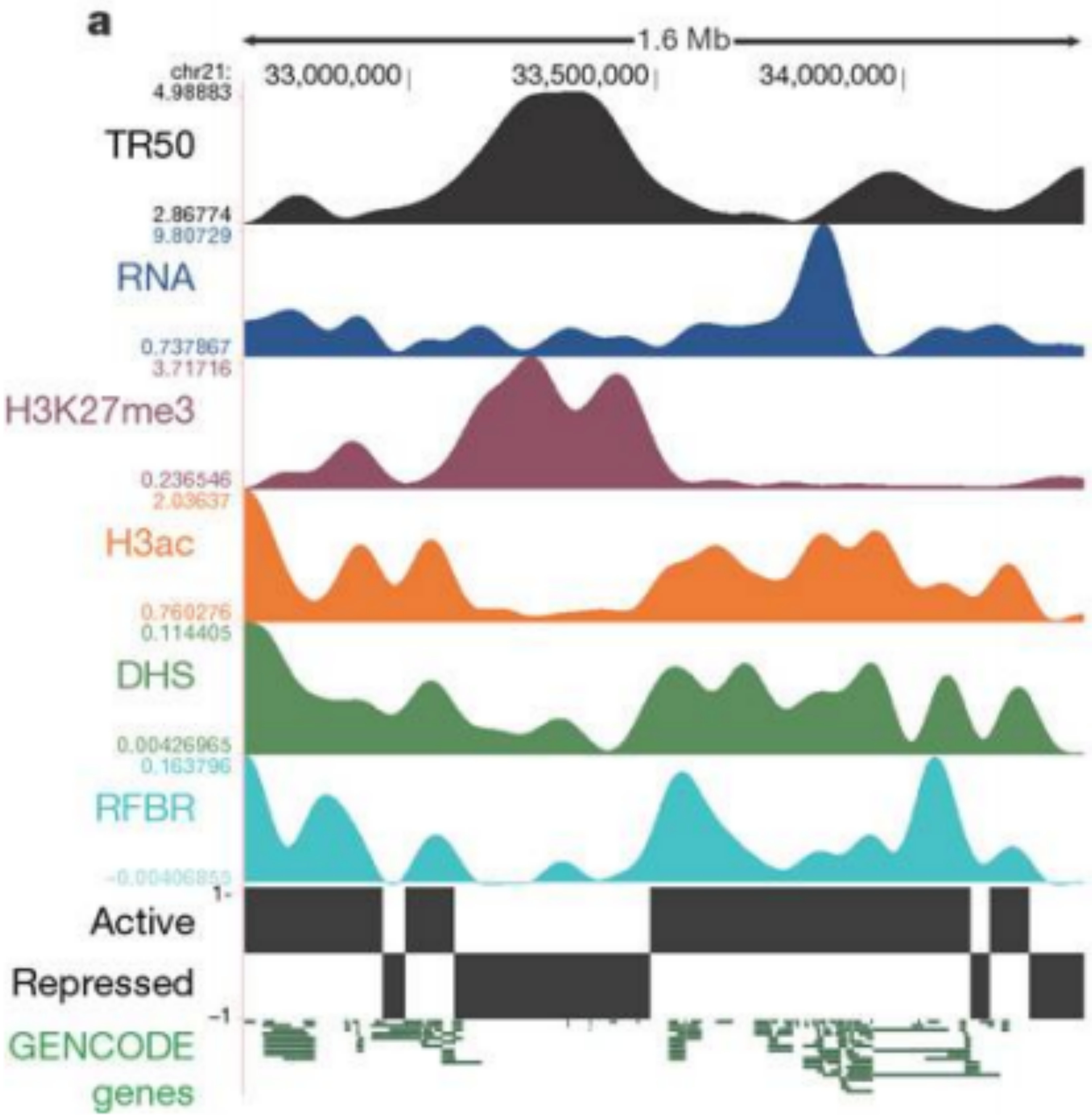


Image from http://www.nature.com/nature/journal/v447/n7146/full/nature05874.html

Clustering

- The process of grouping an object set into classes of similar objects
- Intra-class similarity $\Uparrow$, inter-class similarity $\Downarrow$

Similarity


- It is a philosophical question. Hard to define!
- Let's just take the distance in our feature space.

Clustering Definition

- **Input:** data set $\{x_1, ..., x_n\}$ where $x_i \in \mathbb{R}^d$

- **Goal:** partition data into K disjoint sets $\{S_1, ..., S_K\}$ that the intra-cluster variance is minimized.

- The problem is NP hard, but there are good heuristic algorithms.

Heuristic Algorithms

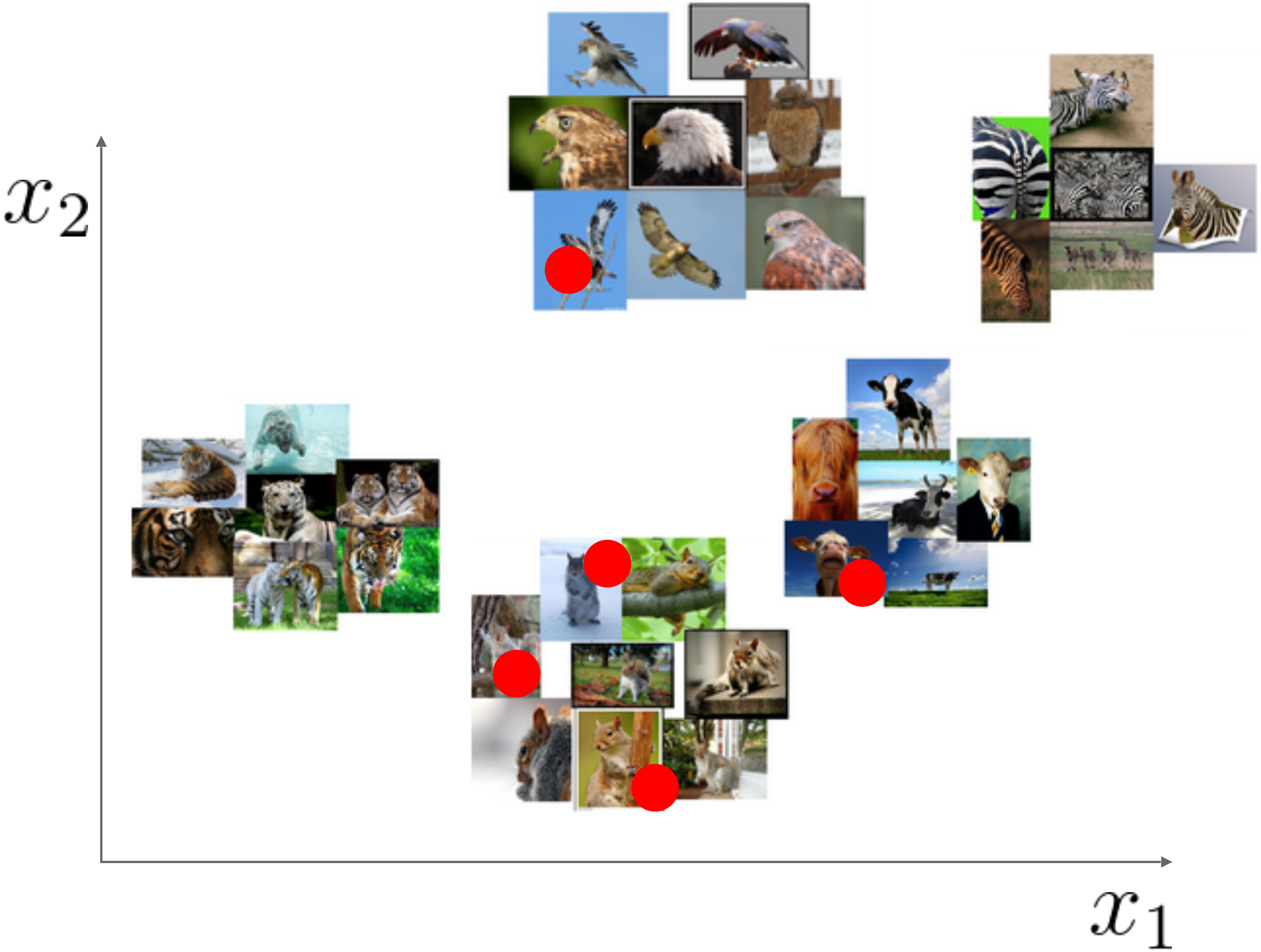
- K-means algorithm
- Hierarchical clustering
- Mean-shift (will be taught in class)
- Mixture of Gaussians (solved by EM algorithm)

K-Means (1/4)

Ask for k (e.g. k = 5)

K-Means (2/4)

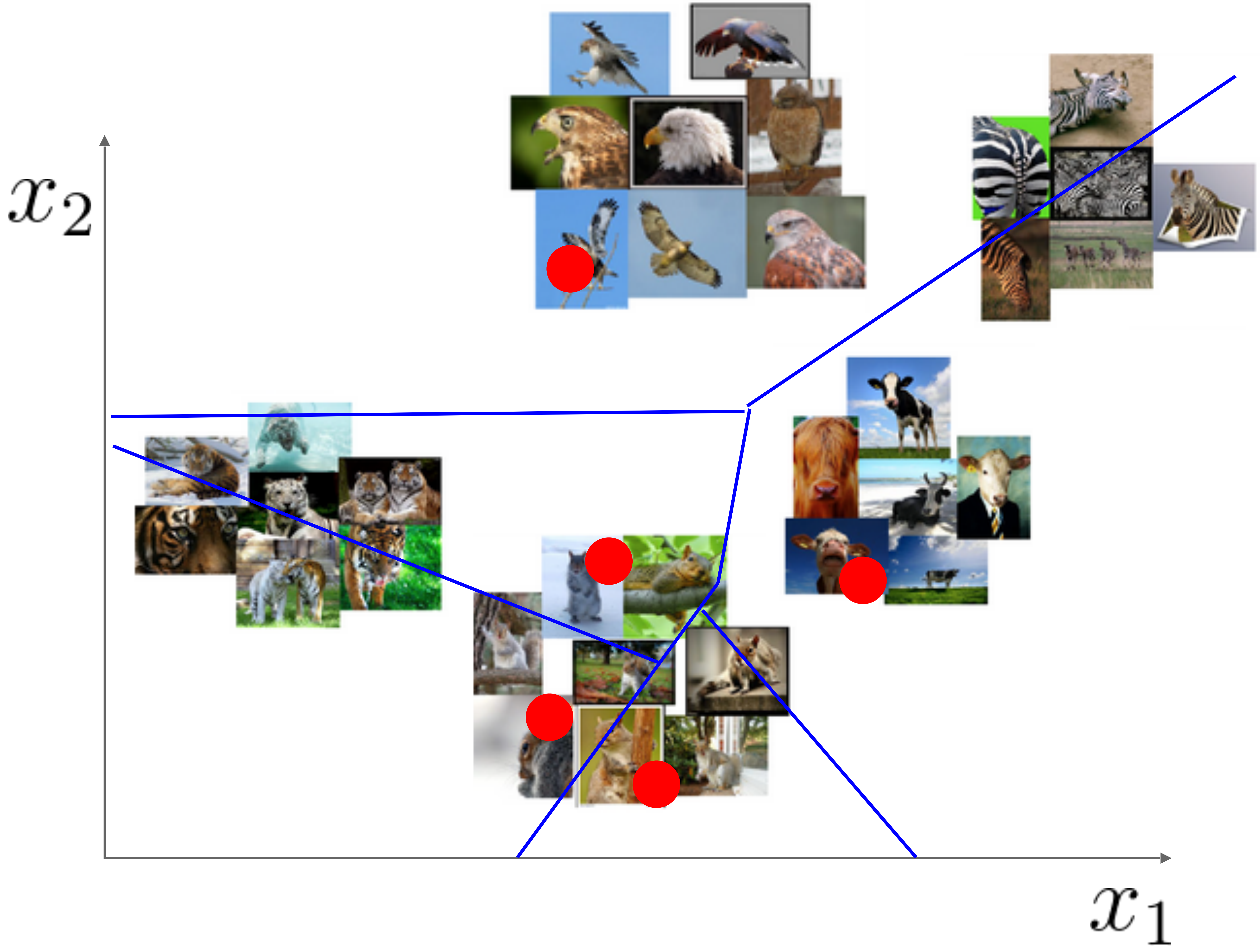
Randomly guess k locations



K-Means (3/4)

For each data

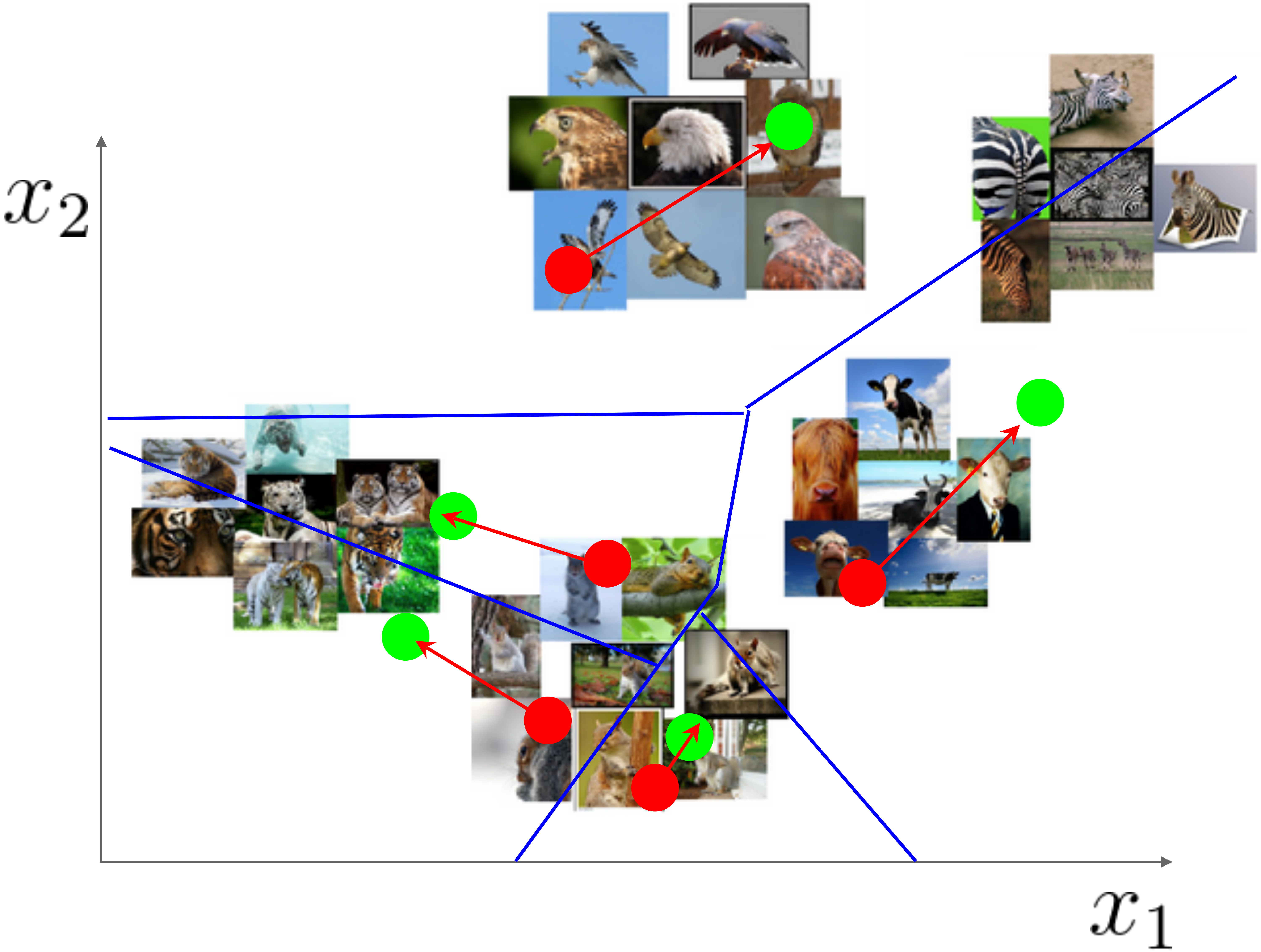
find its closest center



K-Means (4/4)

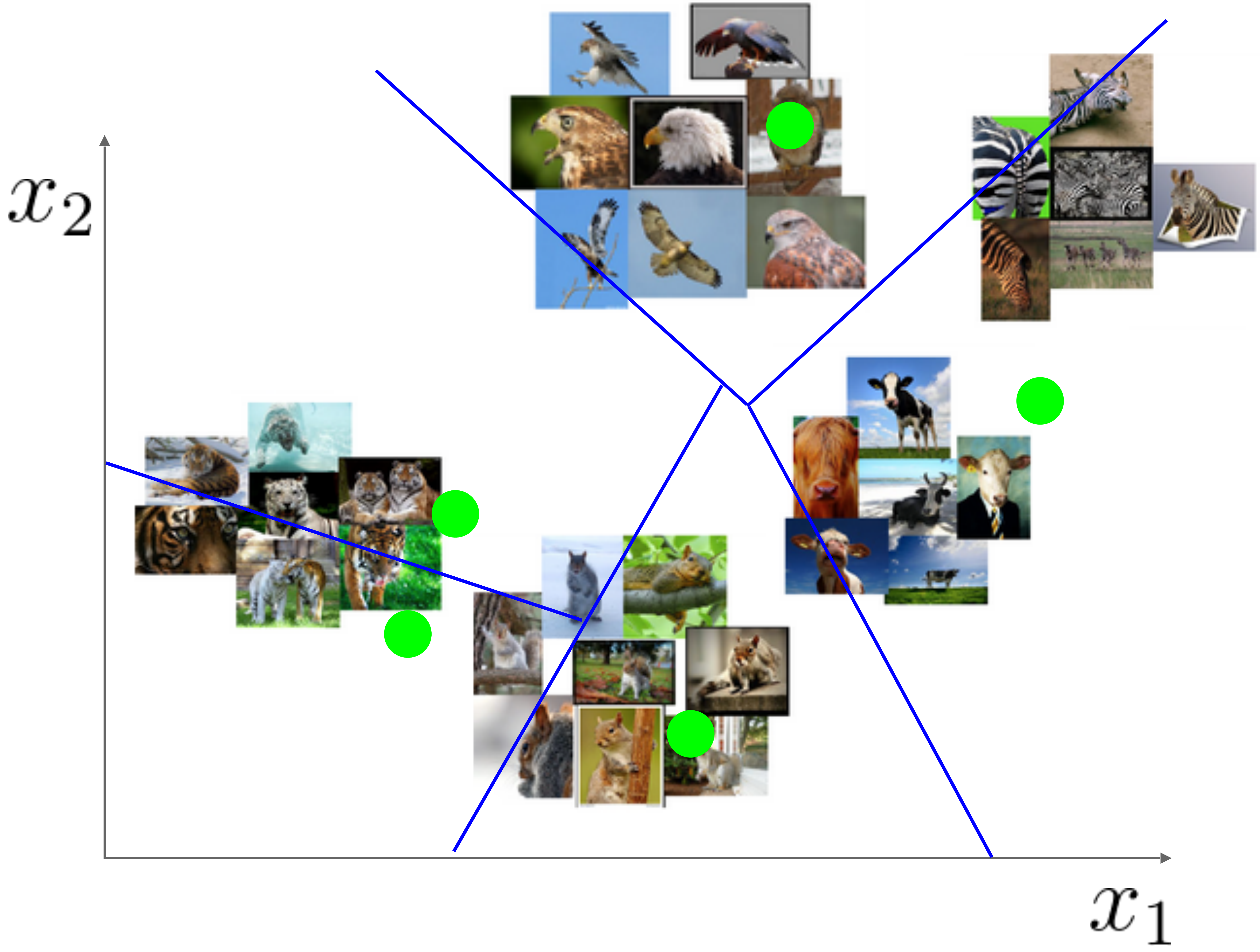
Update centers

for each cluster



K-Means

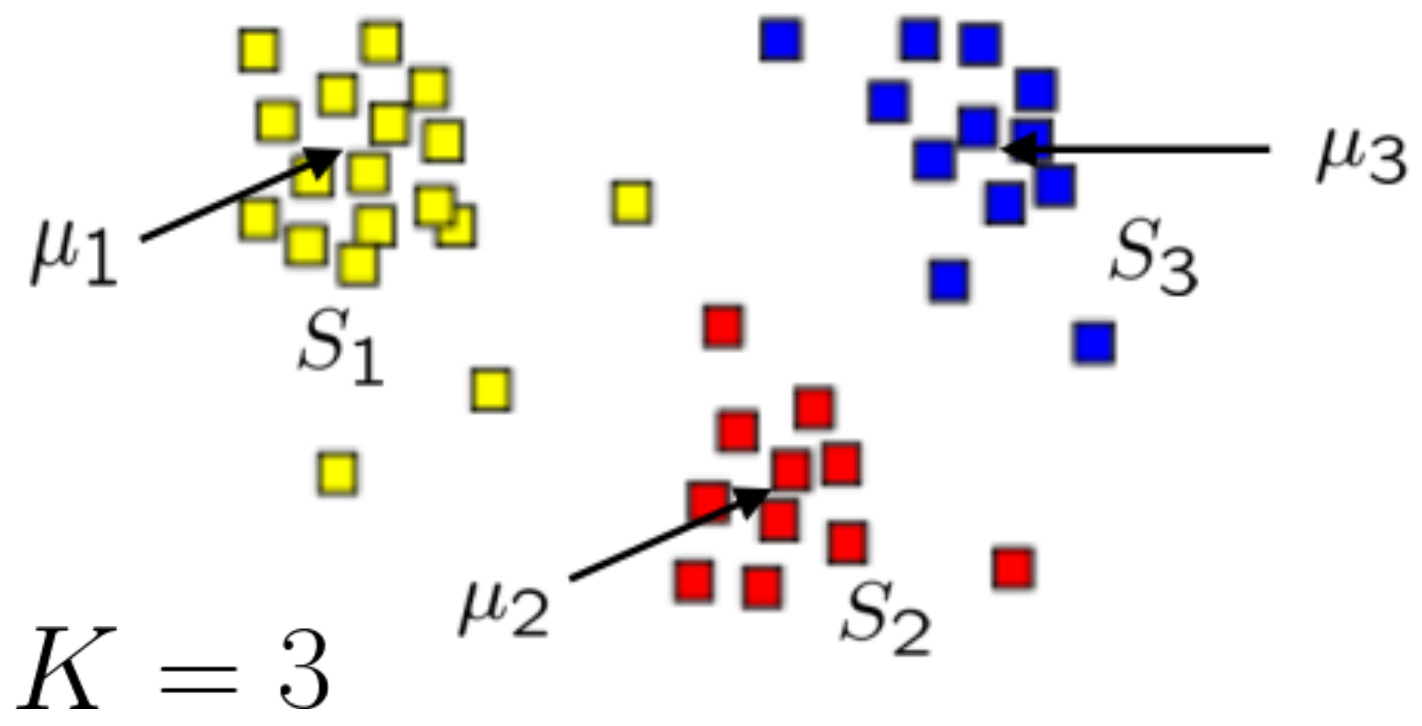
Repeat until convergence



What does K-means Optimize?

Optimize the potential function:

$$\min_{\{S,\mu_i\}} \sum_{i=1}^{K} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$



1.Fix $\{\mu_i\}_{i=1}^{K}$, optimize S (Expectation step)

- Step 3/4: Find each data its closest center

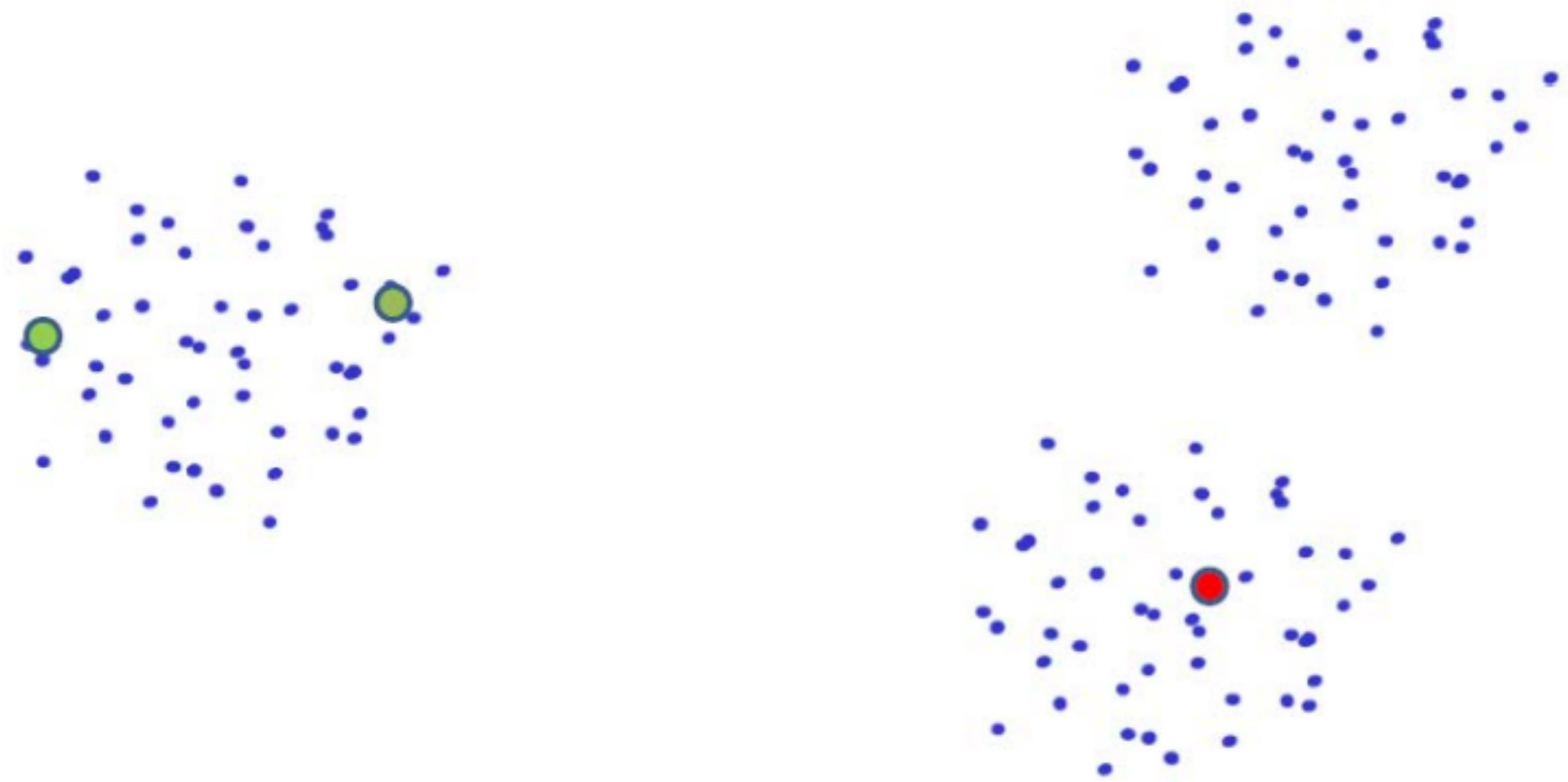
1.Fix S, optimize $\{\mu_i\}_{i=1}^{K}$ (Maximization step)

- Step 4/4: Update centers for each cluster

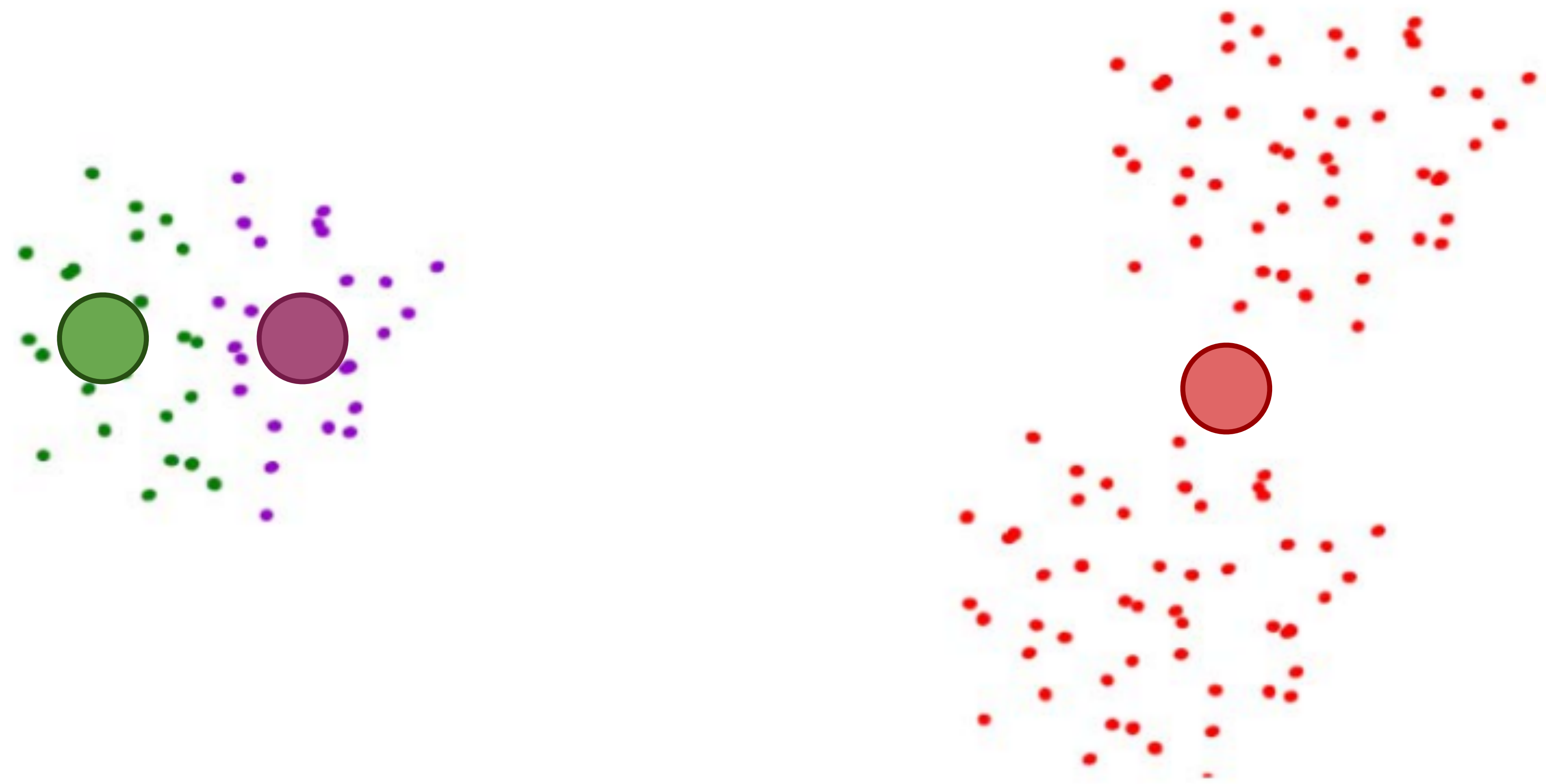
Optimal?

- The results vary based on seed selection.
- K-means can easily get stuck at local minima.
- Alternatives
  - try multiple seeds
  - initialize with the results of another method

Initialization



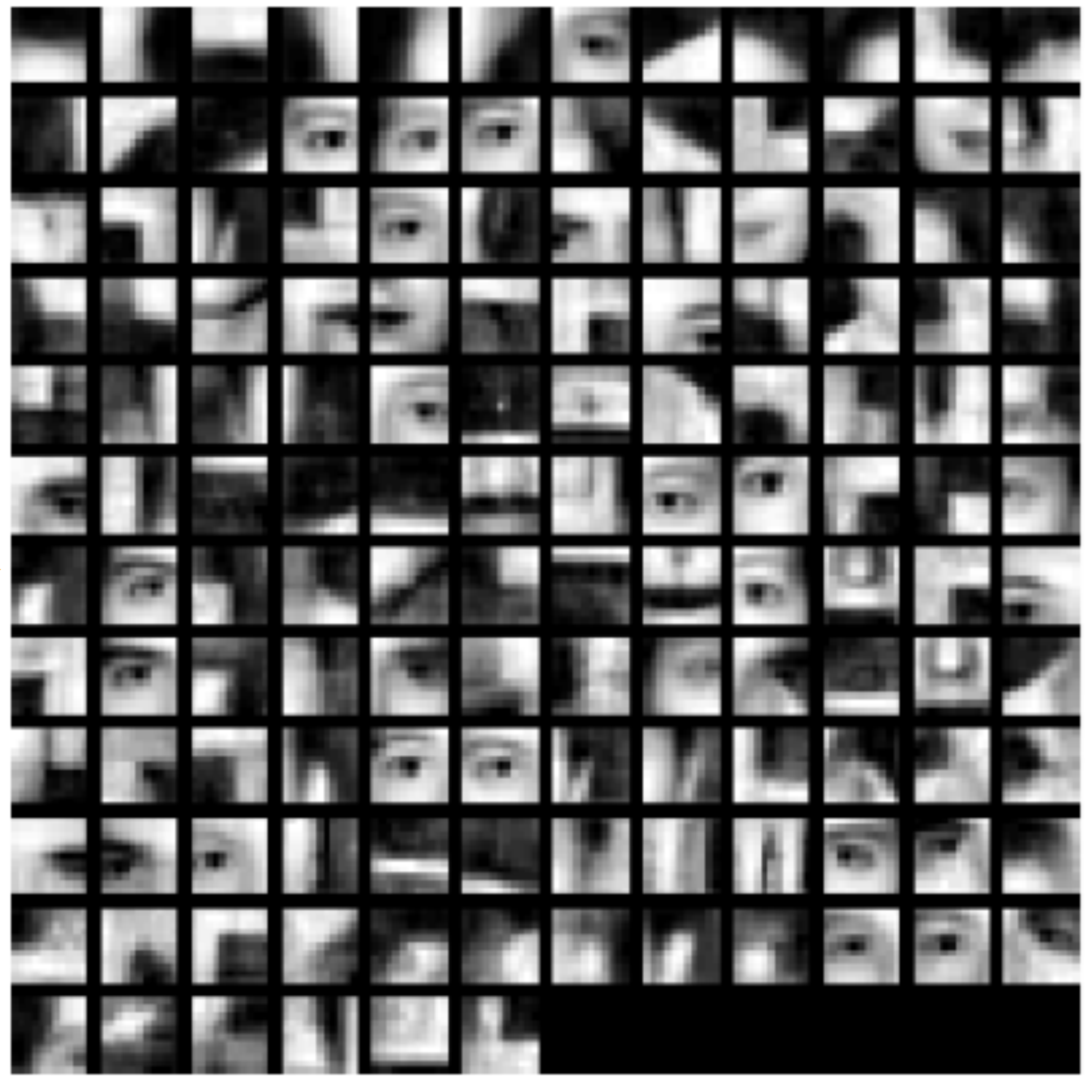
Convergence



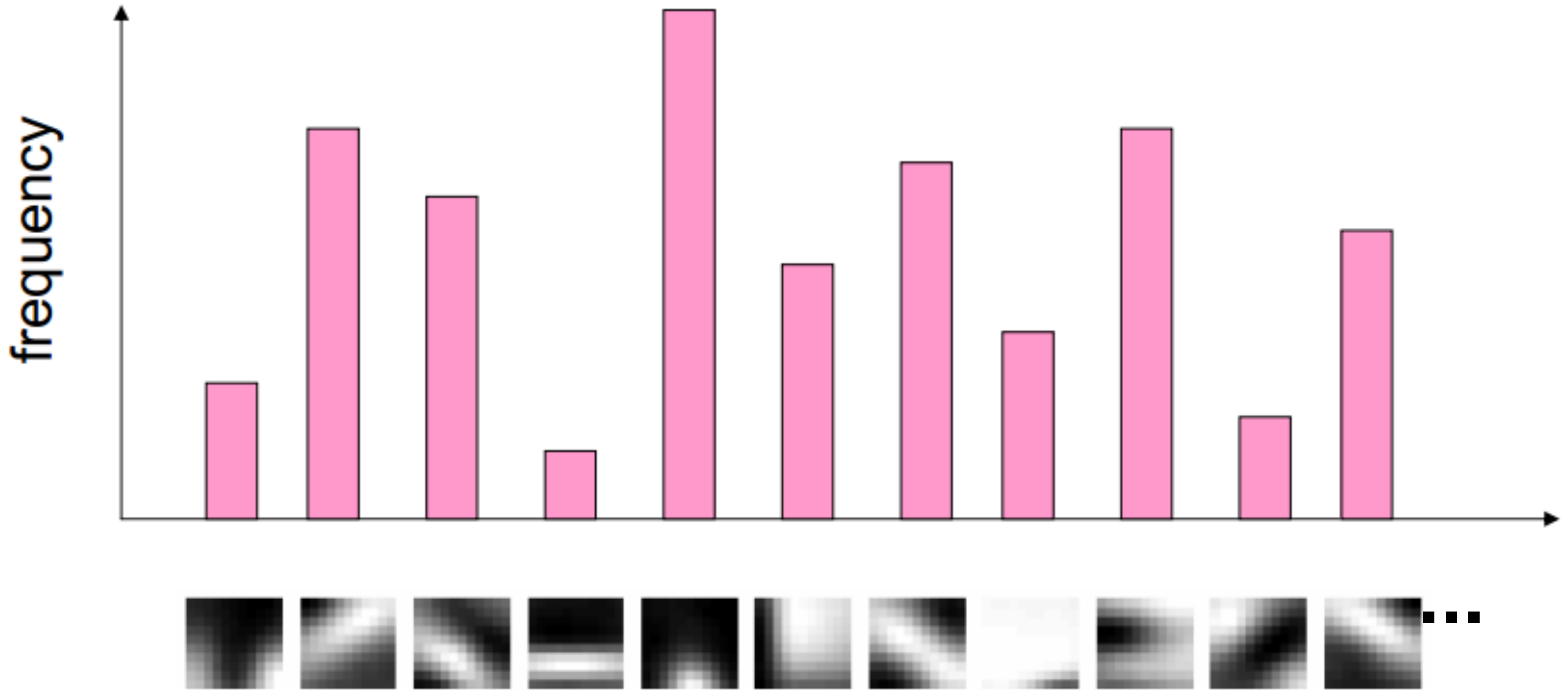
Application - Image Segmentation

Application - Image Representation (BoW)

100-1000 images



~100 visual words



codewords