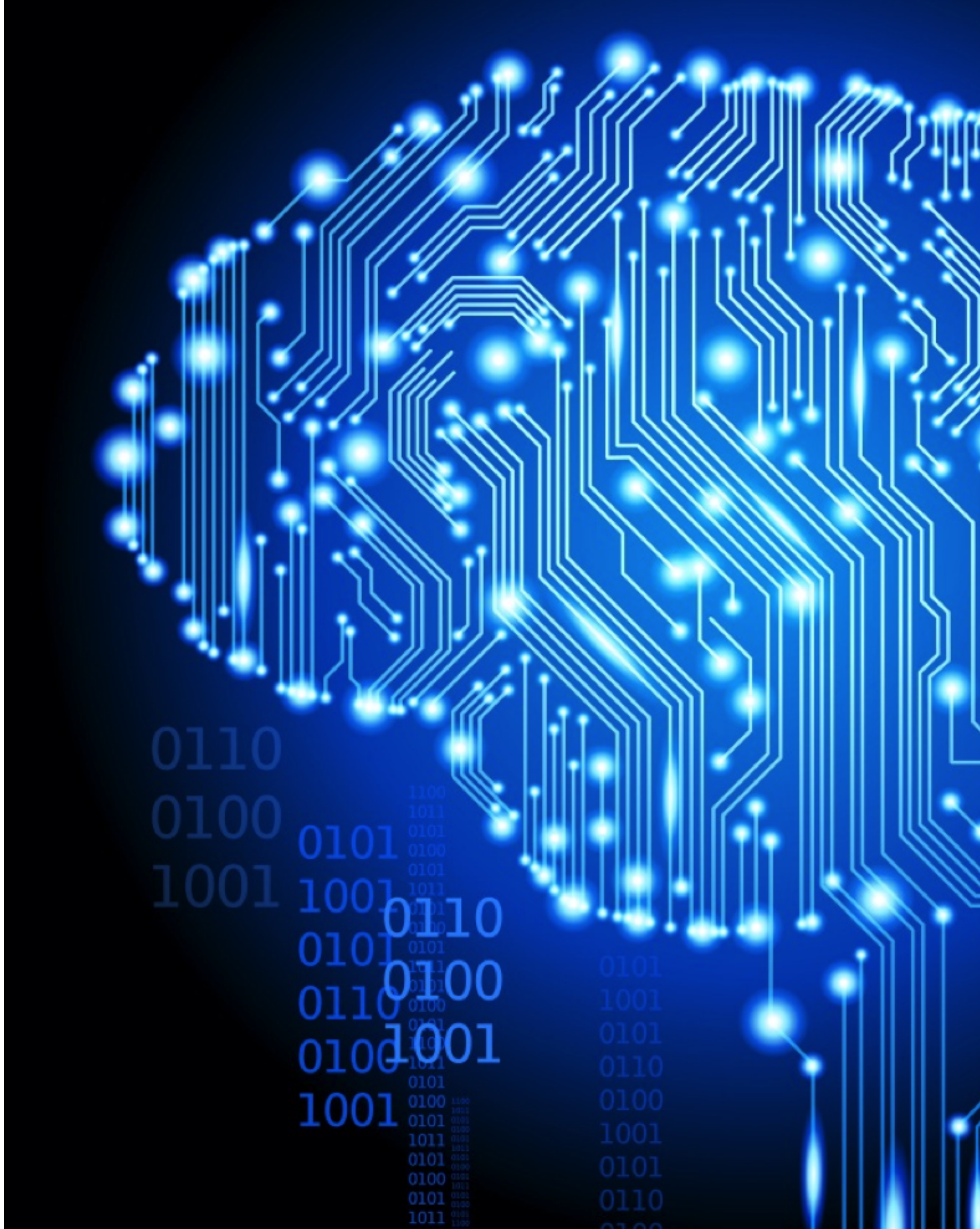
Introduction to Machine Learning

Instructor: Warasinee Chaisangmongkon, PhD



Day 3-4

- Day 3 Morning : Nearest-Neighbor Methods, Feature Selection
- Day 3 Afternoon : Recommender System, Unsupervised Learning
- Day 4 Morning : Neural Network
- Day 4 Afternoon : Advanced Concepts in Machine Learning

- Recommender in Business
- User-based CF
- Item-based CF
- Performance

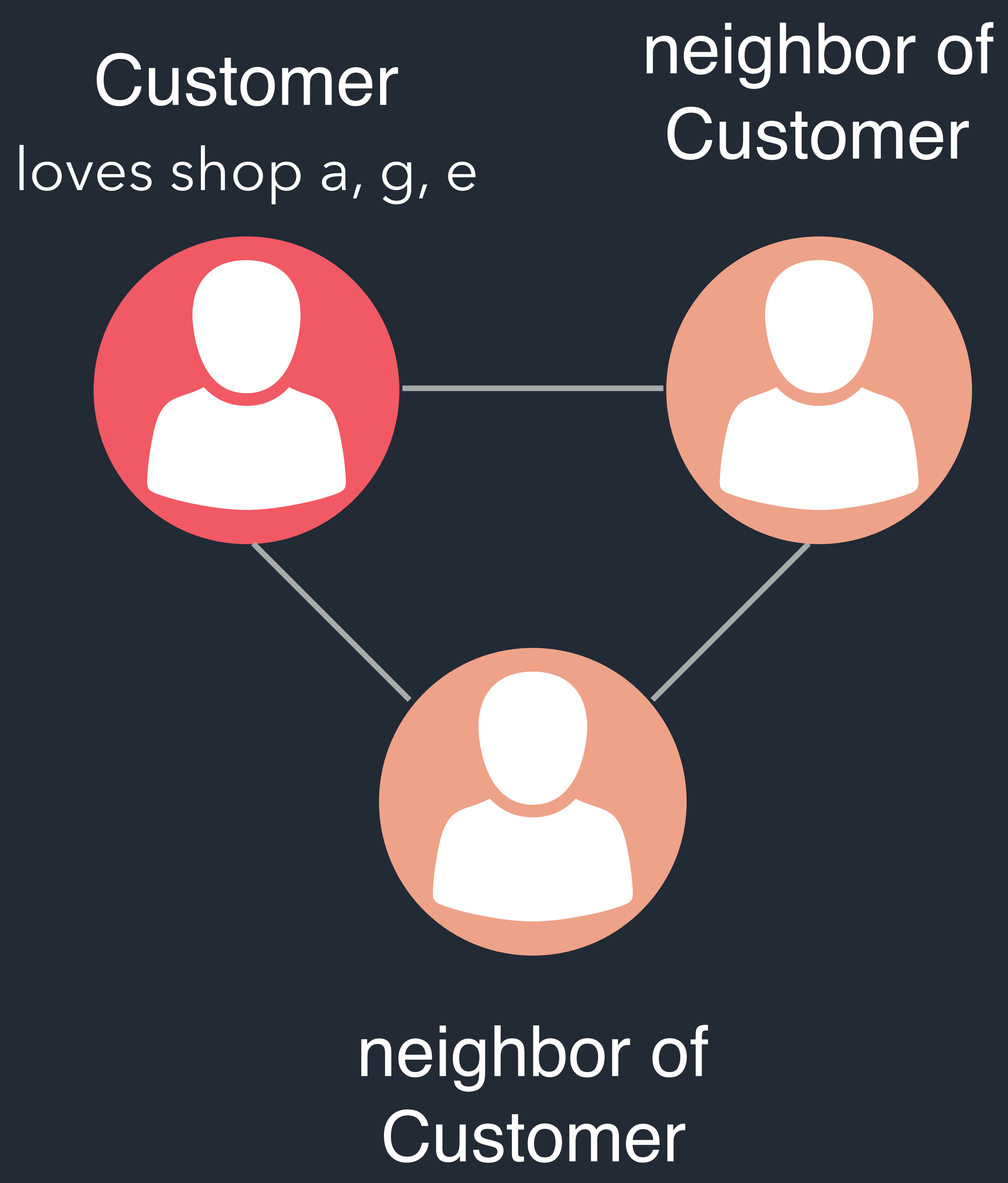
RECOMMENDER SYSTEM

RECOMMENDATIONS

RECOMMENDATIONS

Will customer A like shop B and C?

RECOMMENDATIONS



Find customers
with similar behavior as
customer A.

If neighbors love shop B&C,
customer A should love
shop B&C too.

Problem Definition

	u1	u2	u3	u4
m1	5	5	0	0
m2	4	?	?	0
m3	?	4	0	?
m4	0	0	5	4
m5	0	0	5	?

- Given the past ratings predict the ratings of movies that users have not watched.
- Once you predict their preference, recommend the movie you think they like.

Content-Based Recommendation

- Suppose we can define ‘features’ of each movie based on its genre:
  - x1: the amount of romantic content in the movie
  - x2: the amount of action in the movie

Content-Based Recommendation

	u1	u2	u3	u4
m1	5	5	0	0
m2	4	?	?	0
m3	?	4	0	?
m4	0	0	5	4
m5	0	0	5	?

	x1	x2
m1	0.9	0
m2	0.99	0.01
m3	0.8	0.05
m4	0.05	0.9
m5	0.1	0.9

Content-Based Recommendation

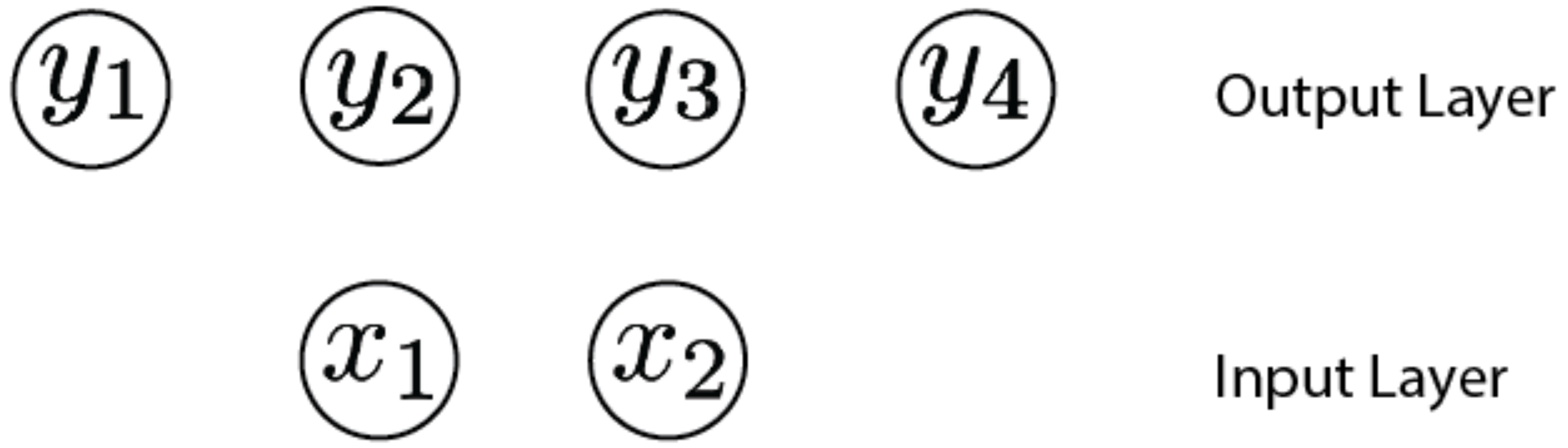
- This problem would be like linear regression problem with multiple inputs:

	u1 (y1)	u2 (y2)	u3 (y3)	u4 (y4)
m1	5	5	0	0
m2	4	?	?	0
m3	?	4	0	?
m4	0	0	5	4
m5	0	0	5	?

	x1	x2
m1	0.9	0
m2	0.99	0.01
m3	0.8	0.05
m4	0.05	0.9
m5	0.1	0.9

Content-Based Recommendation

- Or you might view it as a neural network model

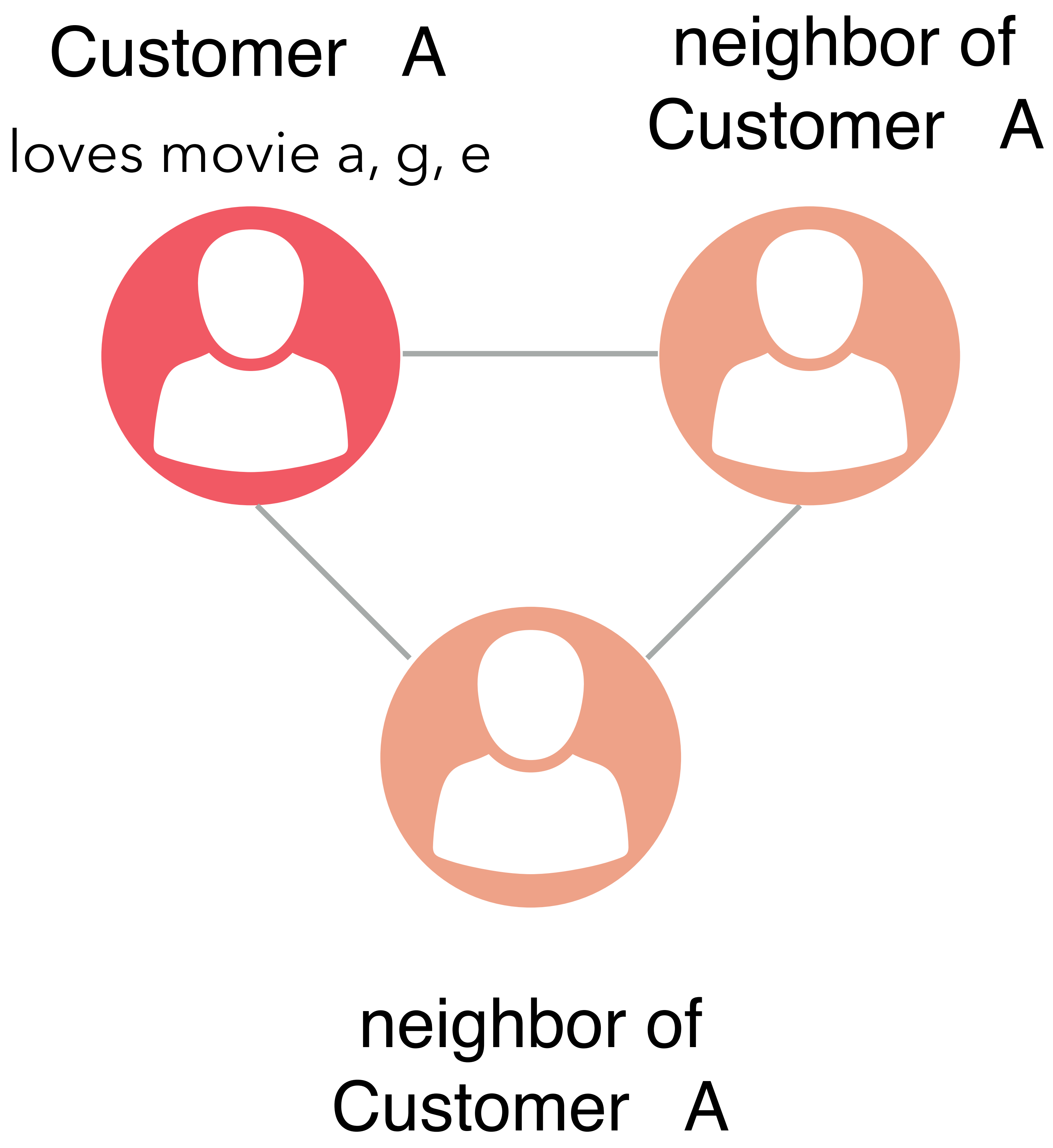


- How many weight parameters are there in this neural network?

Collaborative Filtering

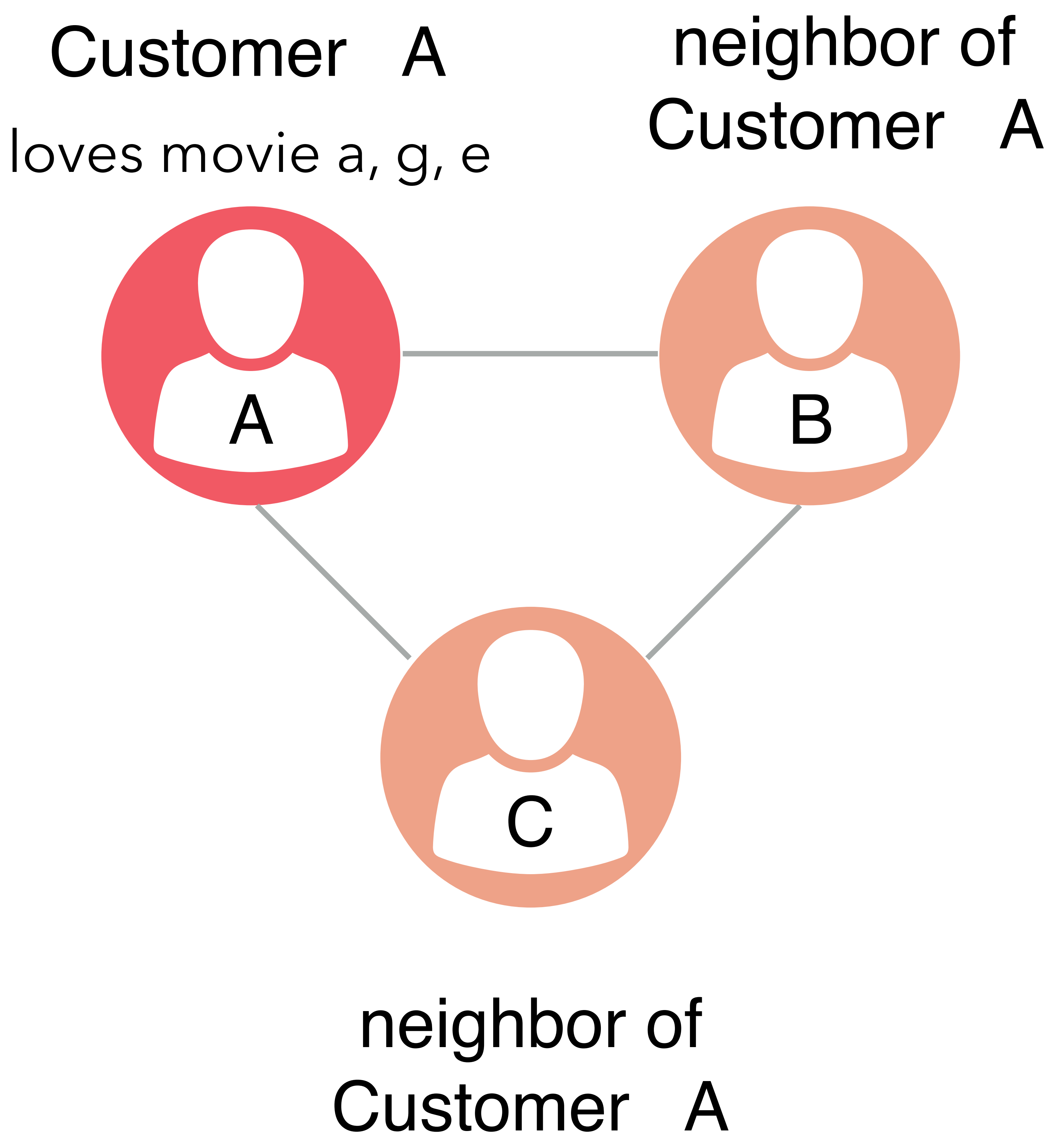
- **Collaborative Filtering:** filtering a bunch of items down to only a few items based on collaborative knowledge (knowledge about other people in the population)
- There are several ways to achieve collaborative filtering objectives. Some of the most popular ways are:
    - User-based Collaborative Filtering
    - Item-based Collaborative Filtering
    - Matrix Factorization

User-Based Collaborative Filtering



- Search for nearest neighbors of A, we could use n=1, 5, 50. (n depends on the number of total populations).
- Use the nearest neighbor formula to predict customer A's preference from neighbors' preference.

User-Based Collaborative Filtering



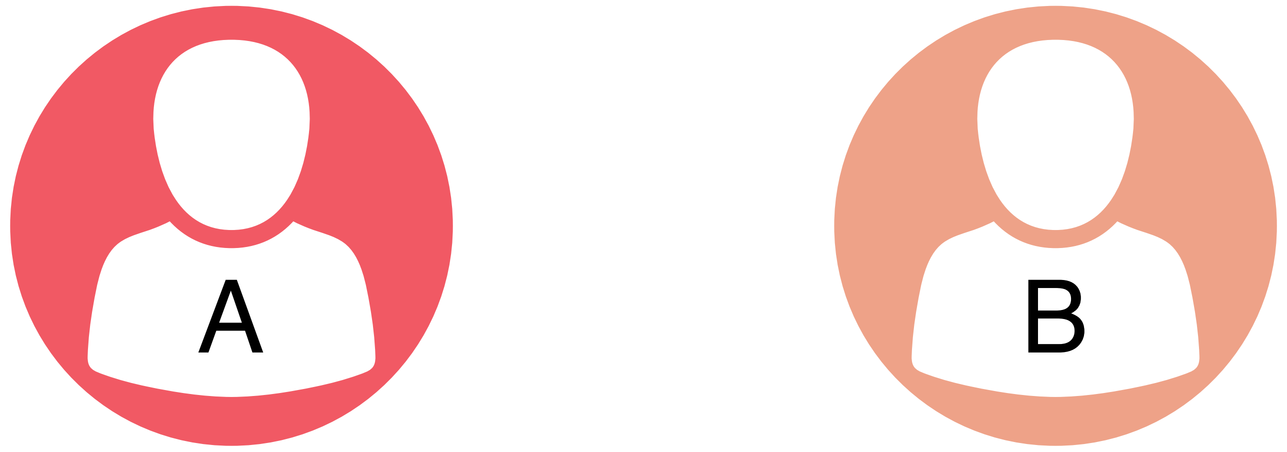
How much user A would rate movie "The Matrix"?

$$r(A, \text{Matrix}) = \frac{w(A, B) * r(B, \text{Matrix}) + w(A, C) * r(C, \text{Matrix})}{w(A, B) + w(A, C)}$$

$w(A, B)$ = how similar A and B are in their taste

This is basically equivalent to K-Nearest Neighbor Regression model

Similarity Between A and B



- Demographic similarity
- Past preference similarity
- Browsing history similarity
- Facebook likes similarity

- Similarity can be thought of as the inverse of distance.

$$w(A, B) = \frac{1}{d(A, B)}$$

- So any distance metrics we have learned about can be used here.
- Or you can look for other similarity metrics.

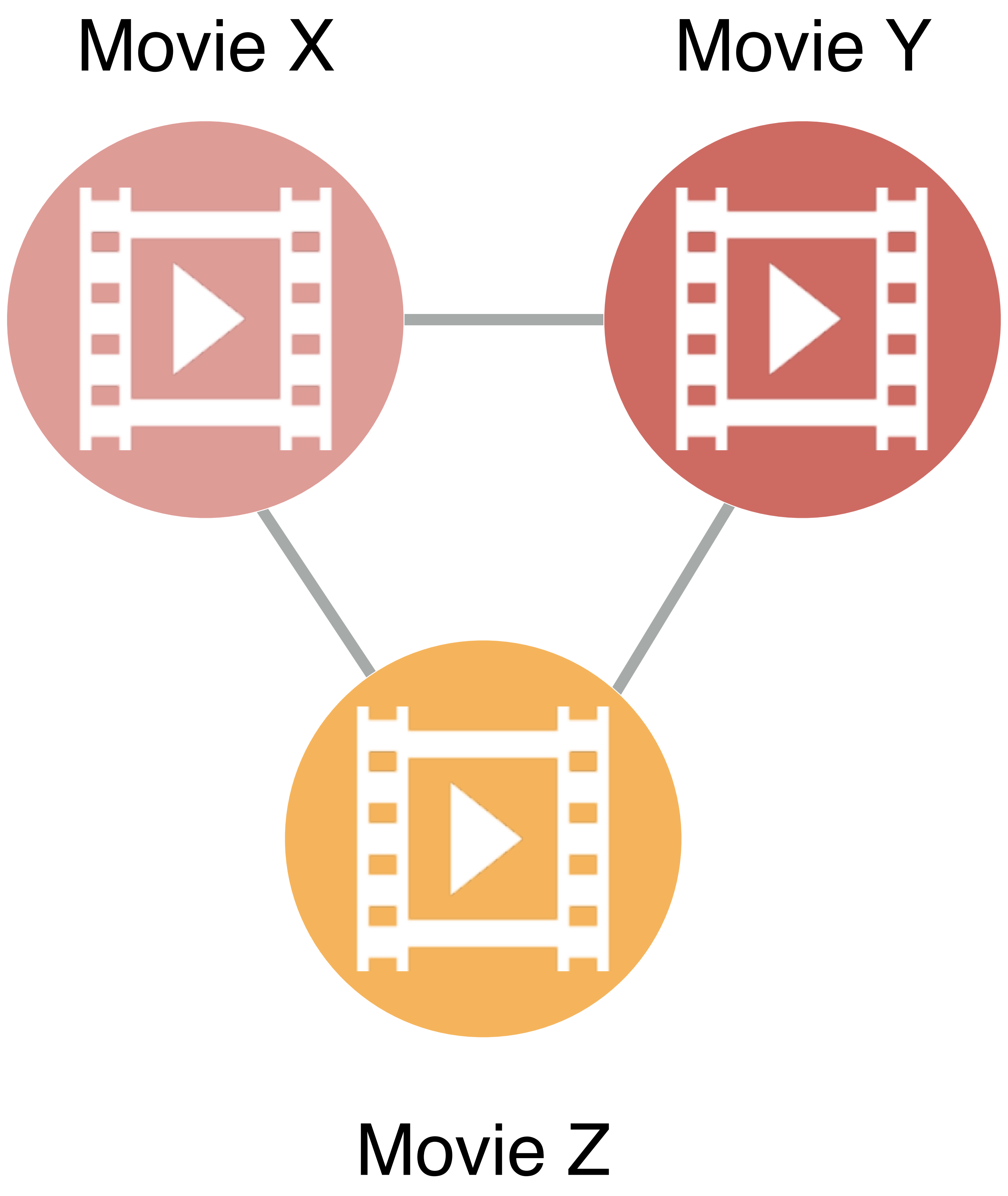
http://scikit-learn.org/stable/modules/metrics.html#metrics

User-Based CF Quiz

	u1	u2	u3	u4
m1	5	5	0	0
m2	4	?	?	0
m3	?	4	0	?
m4	0	0	5	4
m5	0	0	5	?

- Which user is most similar to user 1?
- Do you think user 2 would like movie 2?
- Which user is most similar to user 3?
- Do you think user 4 would like movie 3?

Item-Based Collaborative Filtering

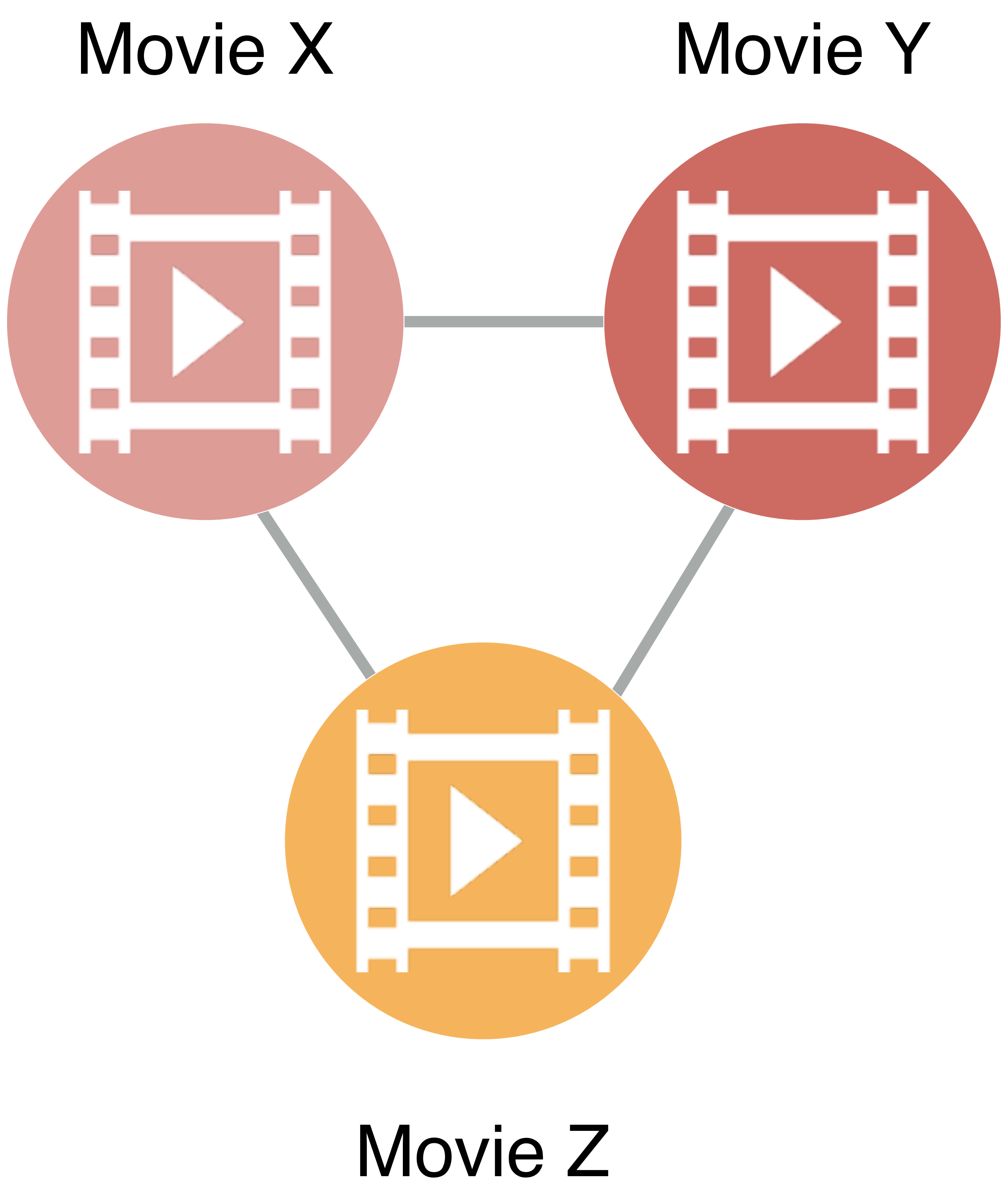


How much user A would rate movie X?

- If user A loves the movie "Beauty and The Beast", will user A like "Frozen"?
- If user A loves the movie "Fight Club", will user A like "Frozen"?

$$r(A, X) = \frac{w(X, Y) * r(A, Y) + w(X, Z) * r(A, Z)}{w(X, Y) + w(X, Z)}$$

Item-Based Collaborative Filtering



- Content / genre similarity
- Fan base similarity
- Common favorite lists
- Products:
  - Buyers
  - Browsers
  - Product spec (brand, qualification)

Item-Based vs. User-Based

- Users change all the time.
- Similarity matrix computation is super expensive
  (1M customer would result in 1M*1M computation)
- Expensive computation happening too often -> not good!
- Items rarely change, so item-based is much less expensive than user-based.
- Precompute to be used over time.

Collaborative Filtering Performance

- Recommender system solves a regression problem, so we use regression to measure performance.

$$E(W) = \frac{1}{2m} \sum_{c=1}^{k} \sum_{i=1}^{m} [y_c^i - h(x^i)_c]^2$$

- In business setting you can also measure 'click-through rate' or rank items from most favorite to least favorite and measure how often people click the top-ranked items versus low-ranked items.

Variations of Collaborative Filtering

- **Normalization:** each person as different 'toughness' in the rating so typically you won't use the raw rating but use normalized rating instead.

- **Hybrid collaborative filtering:** if you cannot decide on which aspect of similarity you should focus on (demographic or preference?), combine them.

- **Alternative Lease Square Model (ALS):** solve multi-output regression problem in a smarter way, using matrix factorization technique instead of KNN. Much faster to train with similar performance.

Software Library for Recommender

- There's no simple Python library for recommender. Although *Surprise* is under active development.

http://surpriselib.com/

- Corporations use library like Apache Spark ML to build big-data recommender system (beyond scope of this course).

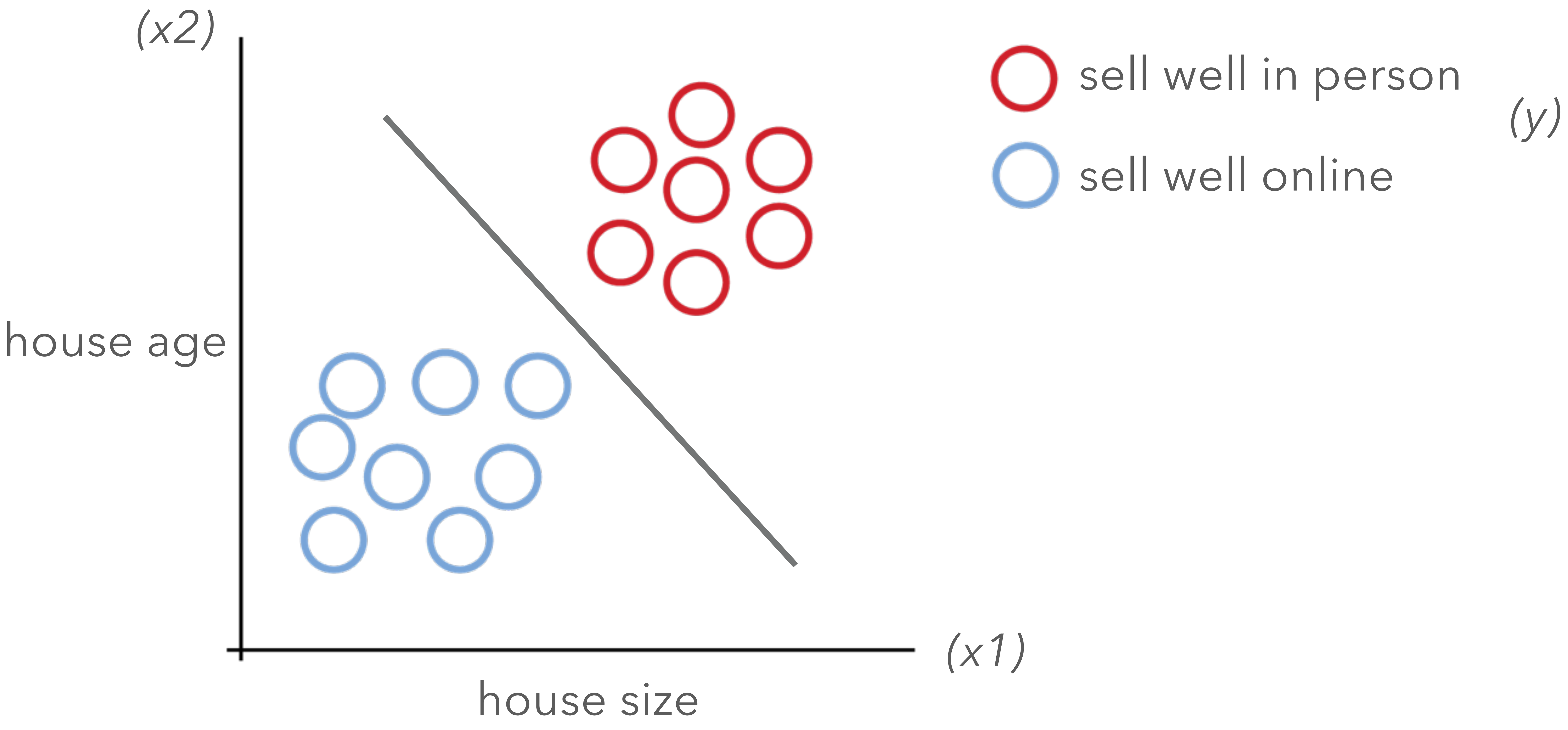
https://spark.apache.org/docs/latest/mllib-collaborative-filtering.html

- Unsupervised Learning
- K-Means Algorithm
- Clustering Performance

K-MEANS CLUSTERING

Supervised vs. Unsupervised Learning

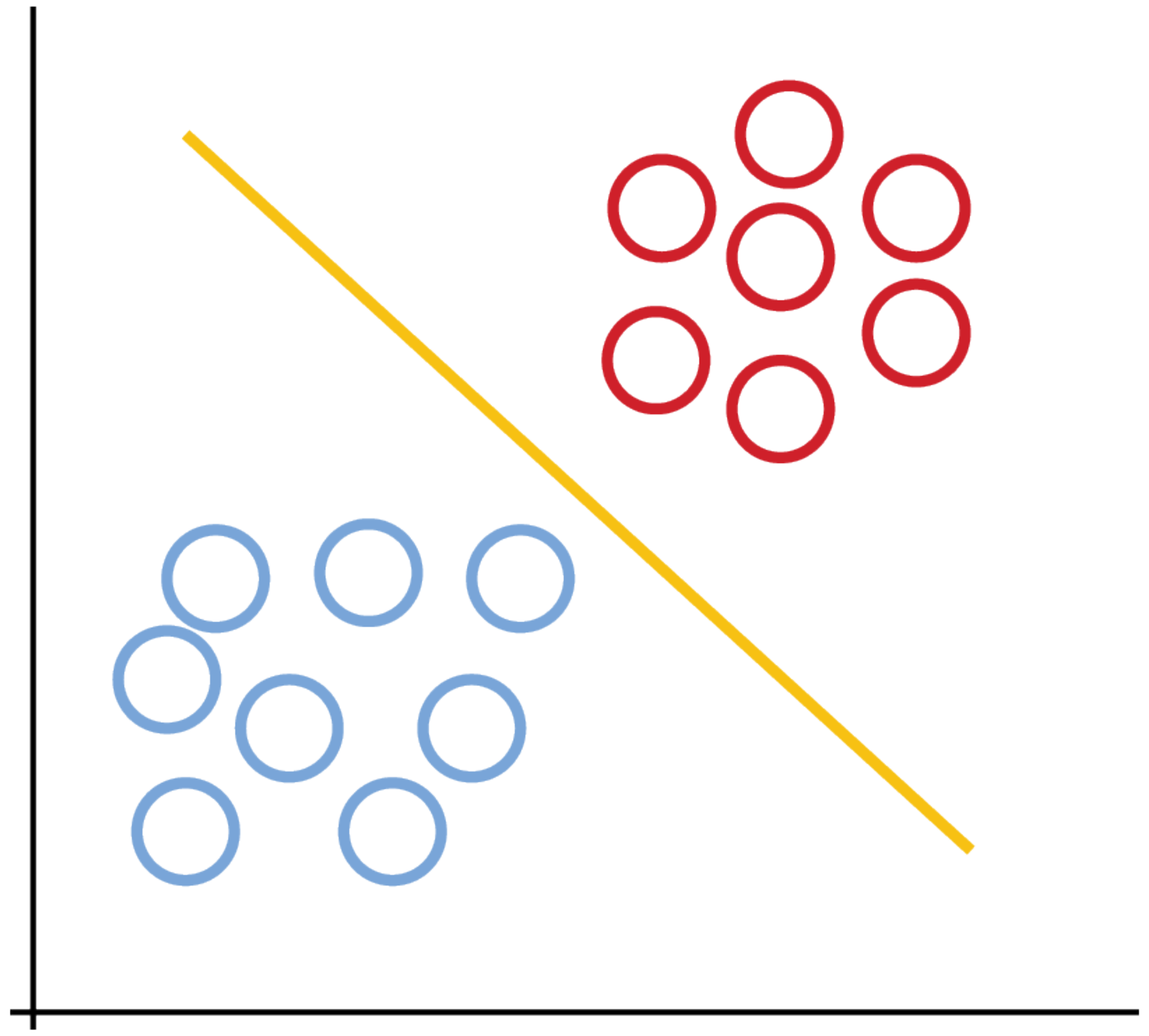
- **Classification:** predicting **discrete** output from input
    - Example: predicting whether a house would sell well in person or online

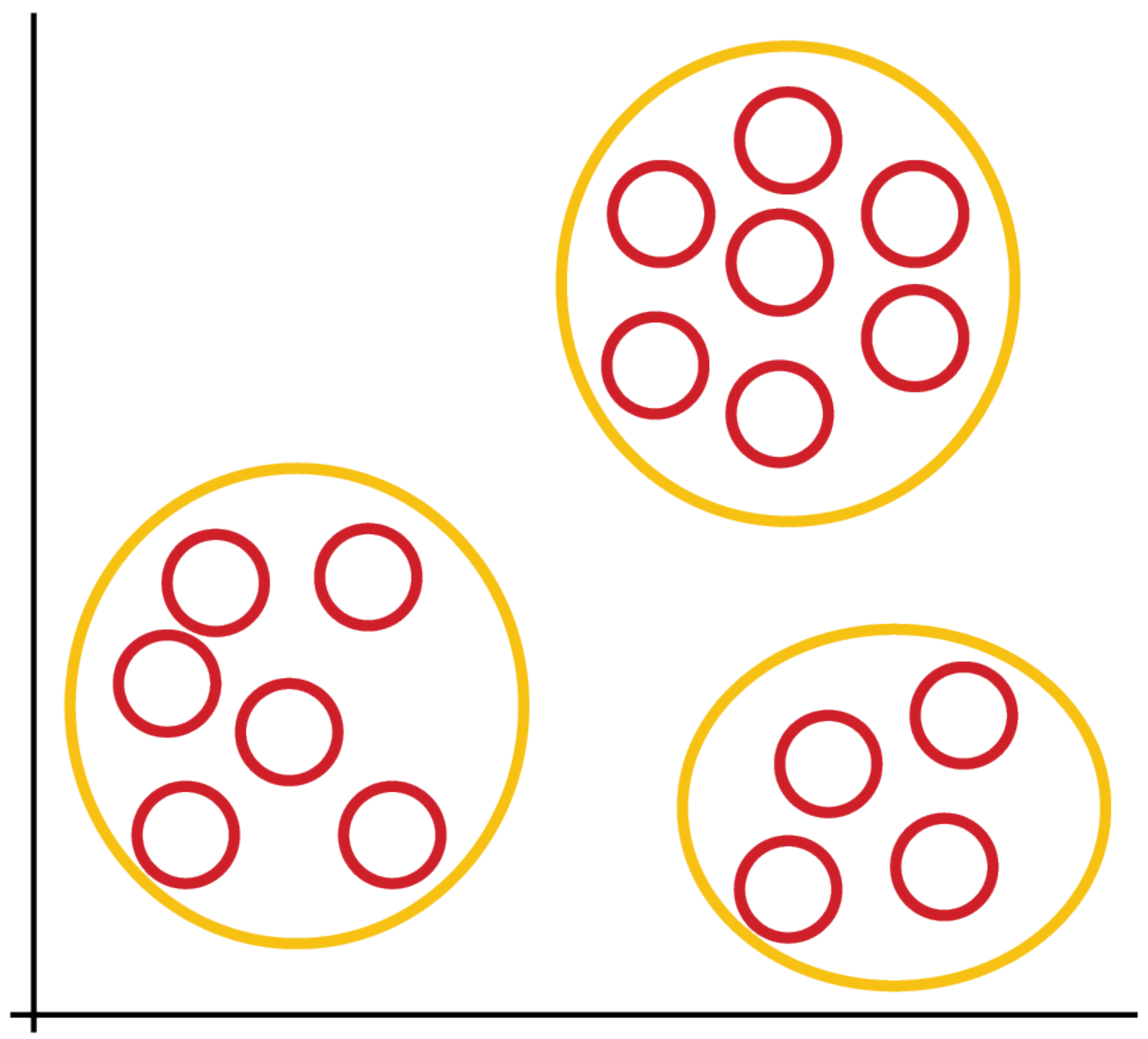


Unsupervised Learning

In **supervised learning,** your data come with labels indicating what class corresponds to each sample.

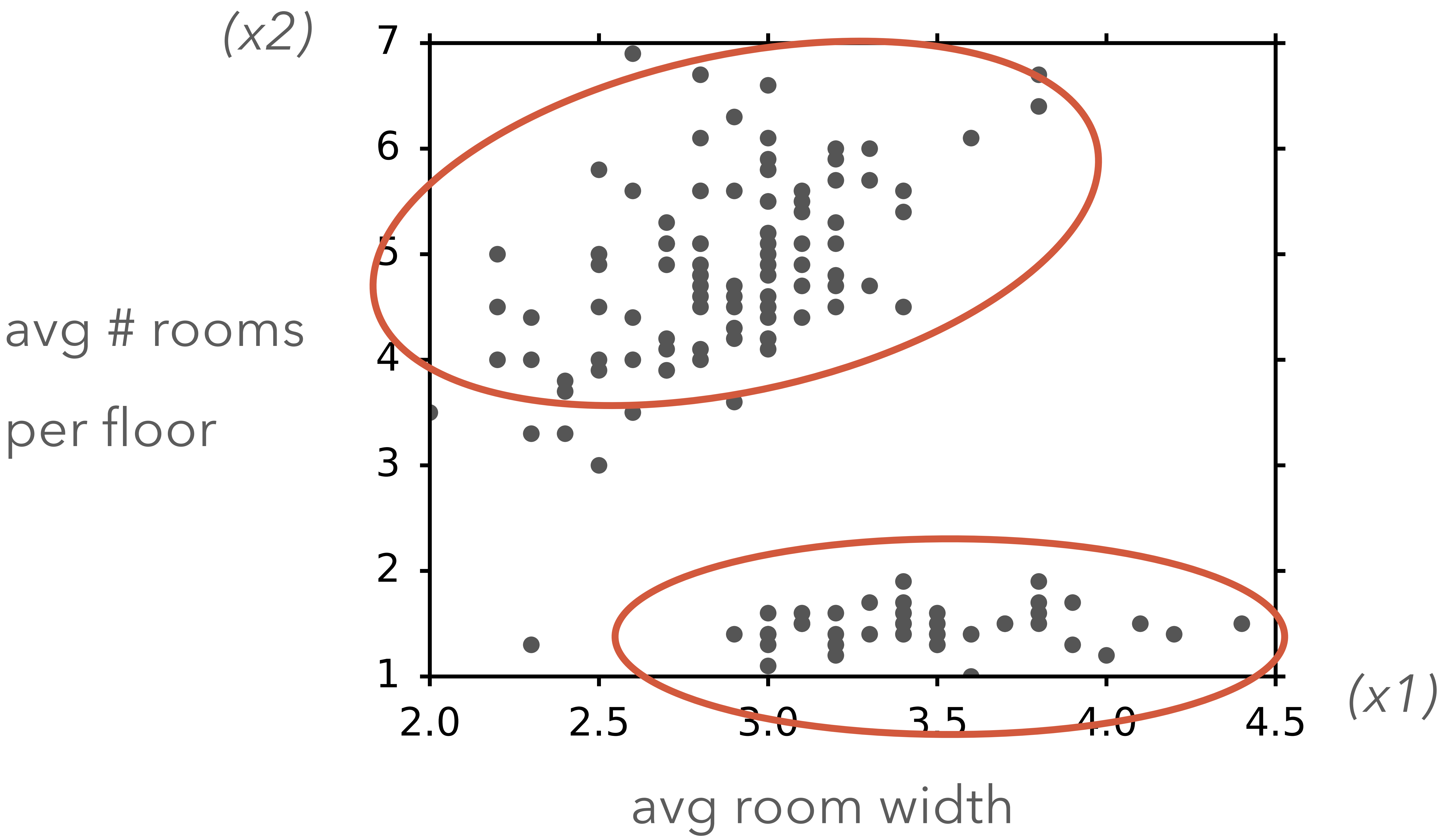
Sometimes, data do not come with categorical labels, but you can tell that there is a grouping structure.

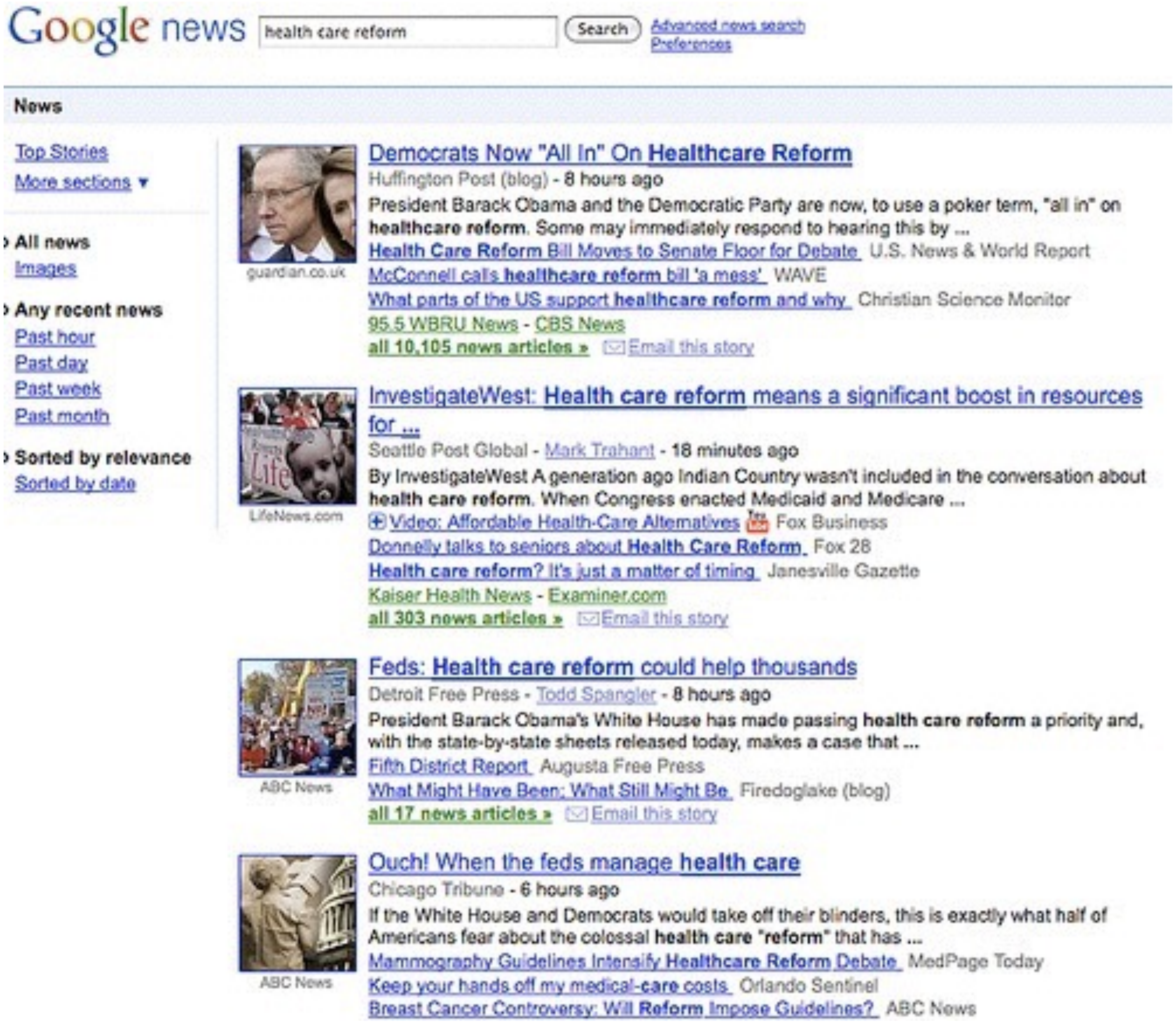




Unsupervised Learning

- **Clustering:** grouping unlabeled input by similarities
    - Example: grouping houses in the database




Google news [health care reform] Search Advanced news search Preferences

News

Top Stories
More sections ▾

All news
Images

Any recent news
Past hour
Past day
Past week
Past month

Sorted by relevance
Sorted by date

guardian.co.uk

Democrats Now "All In" On Healthcare Reform
Huffington Post (blog) - 8 hours ago
President Barack Obama and the Democratic Party are now, to use a poker term, "all in" on **healthcare reform**. Some may immediately respond to hearing this by ...
Health Care Reform Bill Moves to Senate Floor for Debate U.S. News & World Report
McConnell calls healthcare reform bill 'a mess' WAVE
What parts of the US support healthcare reform and why Christian Science Monitor
95.5 WBRU News - CBS News
all 10,105 news articles » Email this story

LifeNews.com

InvestigateWest: Health care reform means a significant boost in resources for ...
Seattle Post Global - Mark Trahant - 18 minutes ago
By InvestigateWest A generation ago Indian Country wasn't included in the conversation about **health care reform**. When Congress enacted Medicaid and Medicare ...
Video: Affordable Health-Care Alternatives Fox Business
Donnelly talks to seniors about Health Care Reform Fox 28
Health care reform? It's just a matter of timing Janesville Gazette
Kaiser Health News - Examiner.com
all 303 news articles » Email this story

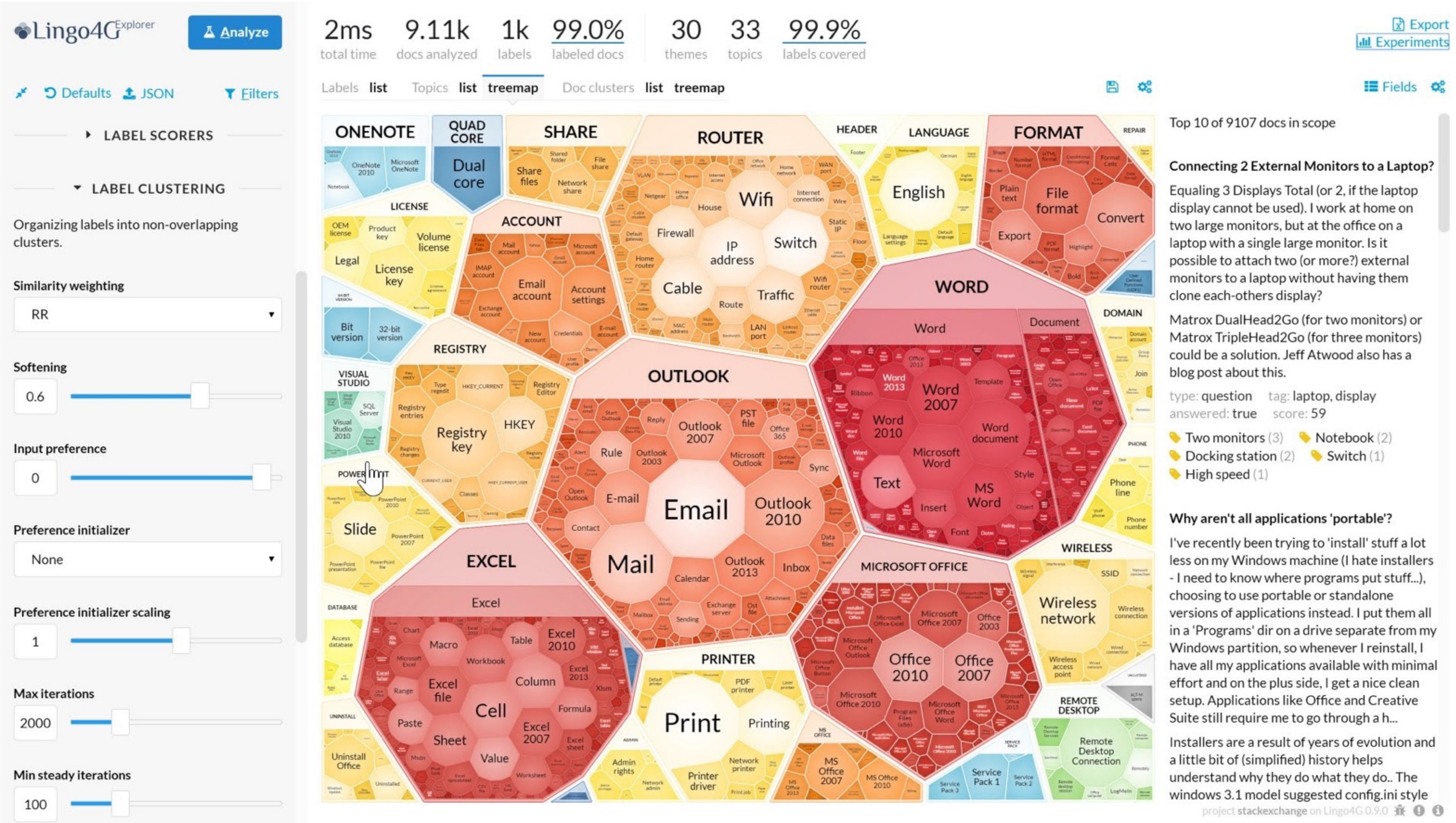
ABC News

Feds: Health care reform could help thousands
Detroit Free Press - Todd Spangler - 8 hours ago
President Barack Obama's White House has made passing **health care reform** a priority and, with the state-by-state sheets released today, makes a case that ...
Fifth District Report Augusta Free Press
What Might Have Been; What Still Might Be Firedoglake (blog)
all 17 news articles » Email this story

ABC News

Ouch! When the feds manage health care
Chicago Tribune - 6 hours ago
If the White House and Democrats would take off their blinders, this is exactly what half of Americans fear about the colossal **health care "reform"** that has ...
Mammography Guidelines Intensify Healthcare Reform Debate MedPage Today
Keep your hands off my medical-care costs Orlando Sentinel
Breast Cancer Controversy: Will Reform Impose Guidelines? ABC News

credit: searchengineland.com


Lingo4G Explorer
Analyze
Defaults
JSON
Filters
LABEL SCORERS
LABEL CLUSTERING
Organizing labels into non-overlapping clusters.
Similarity weighting
RR
Softening
0.6
Input preference
0
Preference initializer
None
Preference initializer scaling
1
Max iterations
2000
Min steady iterations
100
2ms total time
9.11k docs analyzed
1k labels
99.0% labeled docs
30 themes
33 topics
99.9% labels covered
Export
Experiments
Labels list
Topics list treemap
Doc clusters list treemap
Fields
ONENOTE
QUAD CORE
Dual core
SHARE
Share files
Network share
ROUTER
Wifi
Firewall
IP address
Switch
Cable
Traffic
Route
House
HEADER
LANGUAGE
English
FORMAT
File format
Convert
Export
Plain text
LICENSE
Volume license
License key
Legal
ACCOUNT
Email account
Account settings
REGISTRY
Registry key
HKEY
VISUAL STUDIO
OUTLOOK
Email
Mail
Outlook 2010
Outlook 2007
Outlook 2013
Outlook 2003
E-mail
Rule
Reply
Contact
Calendar
Inbox
Sync
WORD
Word
Document
Word 2007
Word 2010
Word 2013
Microsoft Word
Word document
Text
MS Word
Insert
Template
Style
Font
DOMAIN
Join
Phone line
Phone number
WIRELESS
Wireless network
SSID
POWERPOINT
Slide
EXCEL
Excel
Cell
Excel file
Macro
Workbook
Table
Column
Excel 2010
Excel 2007
Excel 2013
Formula
Paste
Sheet
Value
Range
Chart
DATABASE
Access database
UNINSTALL
Uninstall Office
PRINTER
Print
Printing
PDF printer
Printer driver
Network printer
Admin rights
MICROSOFT OFFICE
Office 2010
Office 2007
Office 2003
Microsoft Office 2007
Microsoft Office 2010
MS Office 2007
REMOTE DESKTOP
Remote Desktop Connection
Service Pack 1
Top 10 of 9107 docs in scope
Connecting 2 External Monitors to a Laptop?
Equaling 3 Displays Total (or 2, if the laptop display cannot be used). I work at home on two large monitors, but at the office on a laptop with a single large monitor. Is it possible to attach two (or more?) external monitors to a laptop without having them clone each-others display?
Matrox DualHead2Go (for two monitors) or Matrox TripleHead2Go (for three monitors) could be a solution. Jeff Atwood also has a blog post about this.
type: question tag: laptop, display
answered: true score: 59
Two monitors (3) Notebook (2)
Docking station (2) Switch (1)
High speed (1)
Why aren't all applications 'portable'?
I've recently been trying to 'install' stuff a lot less on my Windows machine (I hate installers - I need to know where programs put stuff...), choosing to use portable or standalone versions of applications instead. I put them all in a 'Programs' dir on a drive separate from my Windows partition, so whenever I reinstall, I have all my applications available with minimal effort and on the plus side, I get a nice clean setup. Applications like Office and Creative Suite still require me to go through a h...
Installers are a result of years of evolution and a little bit of (simplified) history helps understand why they do what they do.. The windows 3.1 model suggested config.ini style
project stackexchange on Lingo4G 0.9.0

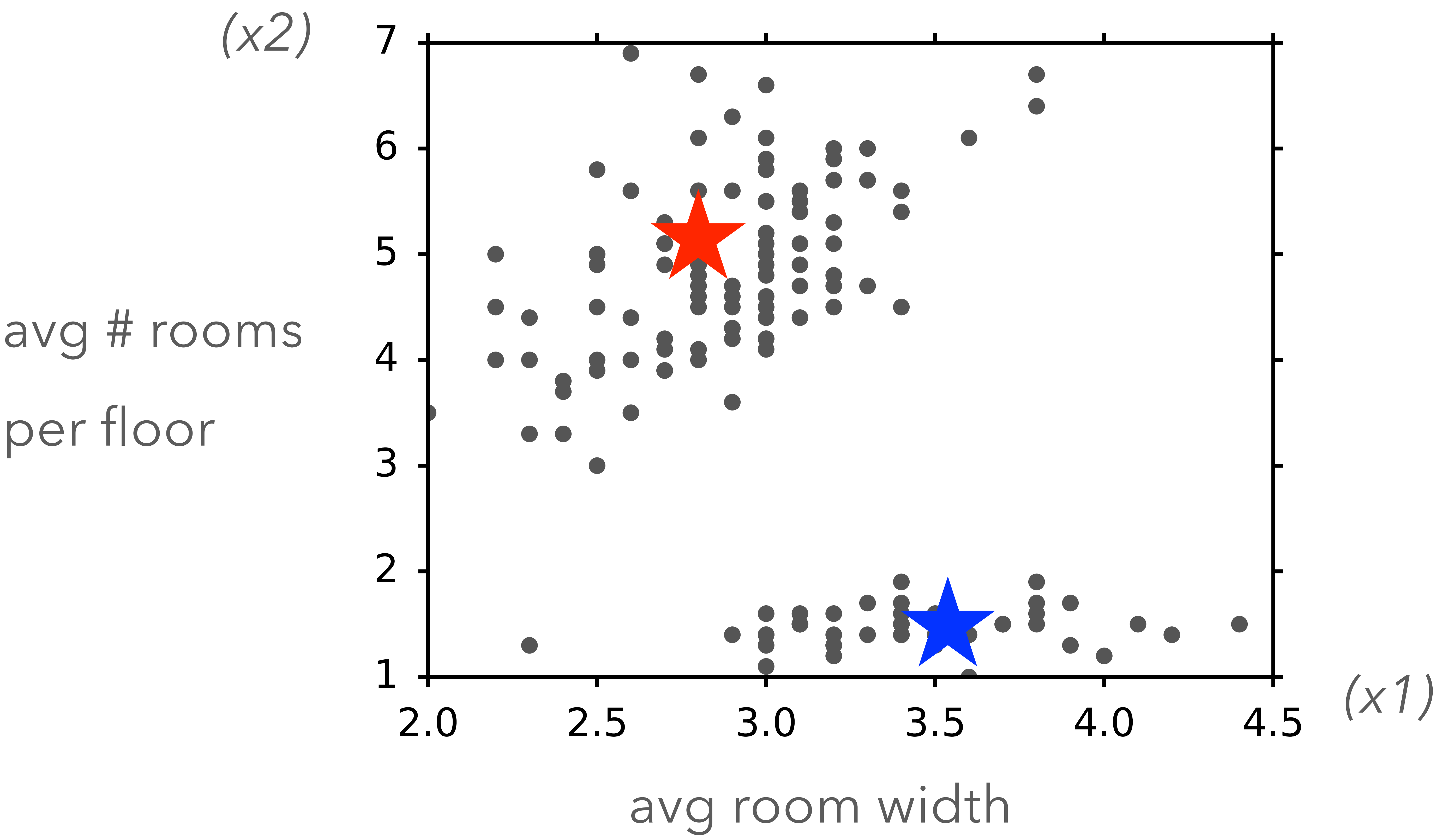
Supervised vs. Unsupervised Learning

- Supervised learning problem: "given all history of feature data points (X) and their labels (Y), find a model that computes label (y′) for a given sample data point (x′). The prediction must minimize a cost function (minimize errors)."
- Clustering problem: "given all feature data points (X), find their labels (Y). The label Y must minimize a given objective function."

	X	Y	x′	y′	Predict	Cost Function
Supervised	Yes	Yes	Yes	No	y′	Error
Unsupervised	Yes	No	Yes	No	Y, y′	Objective F.

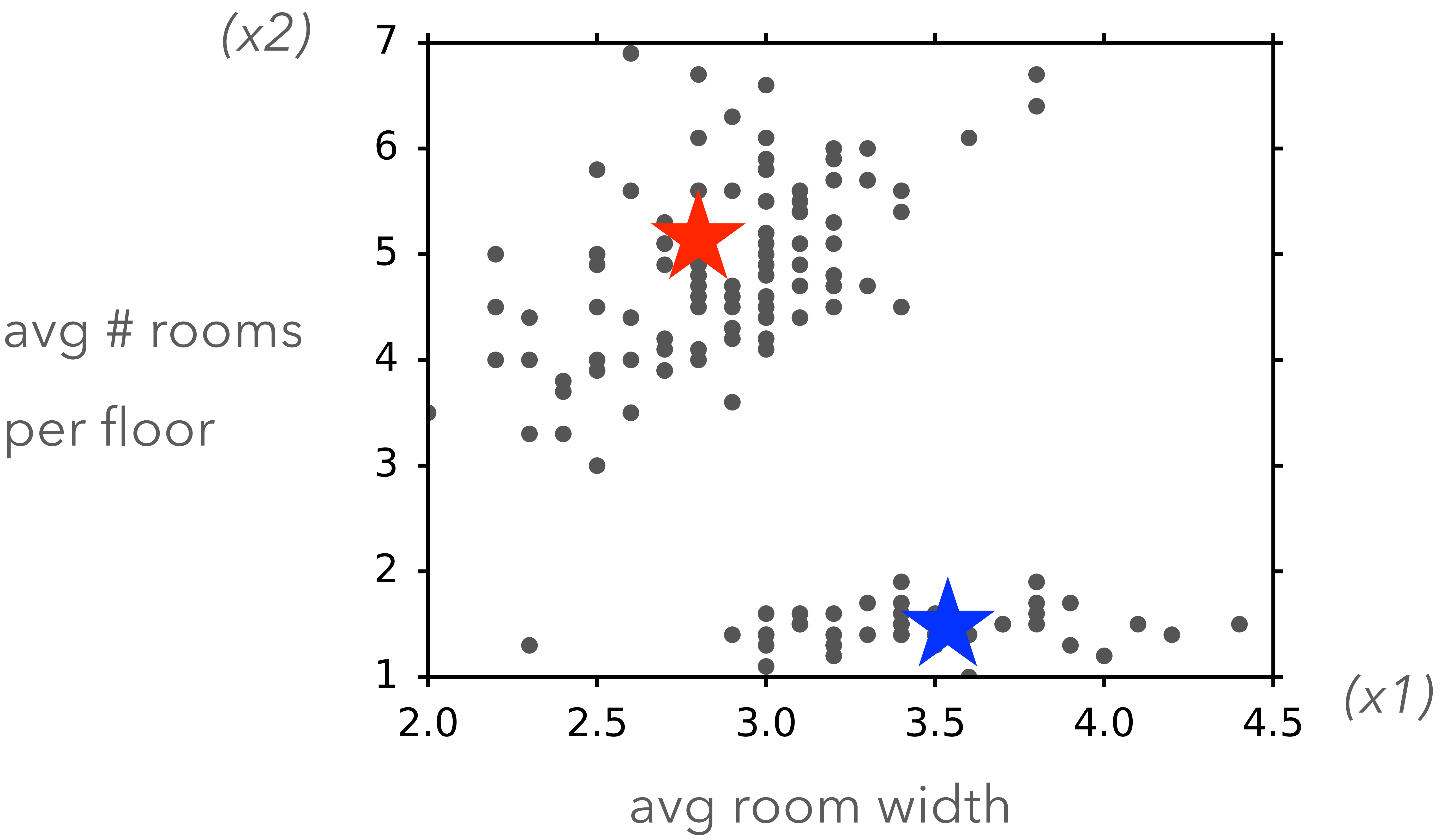
K-Means Clustering

- K-means clustering is the most popular clustering algorithm.
- K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.



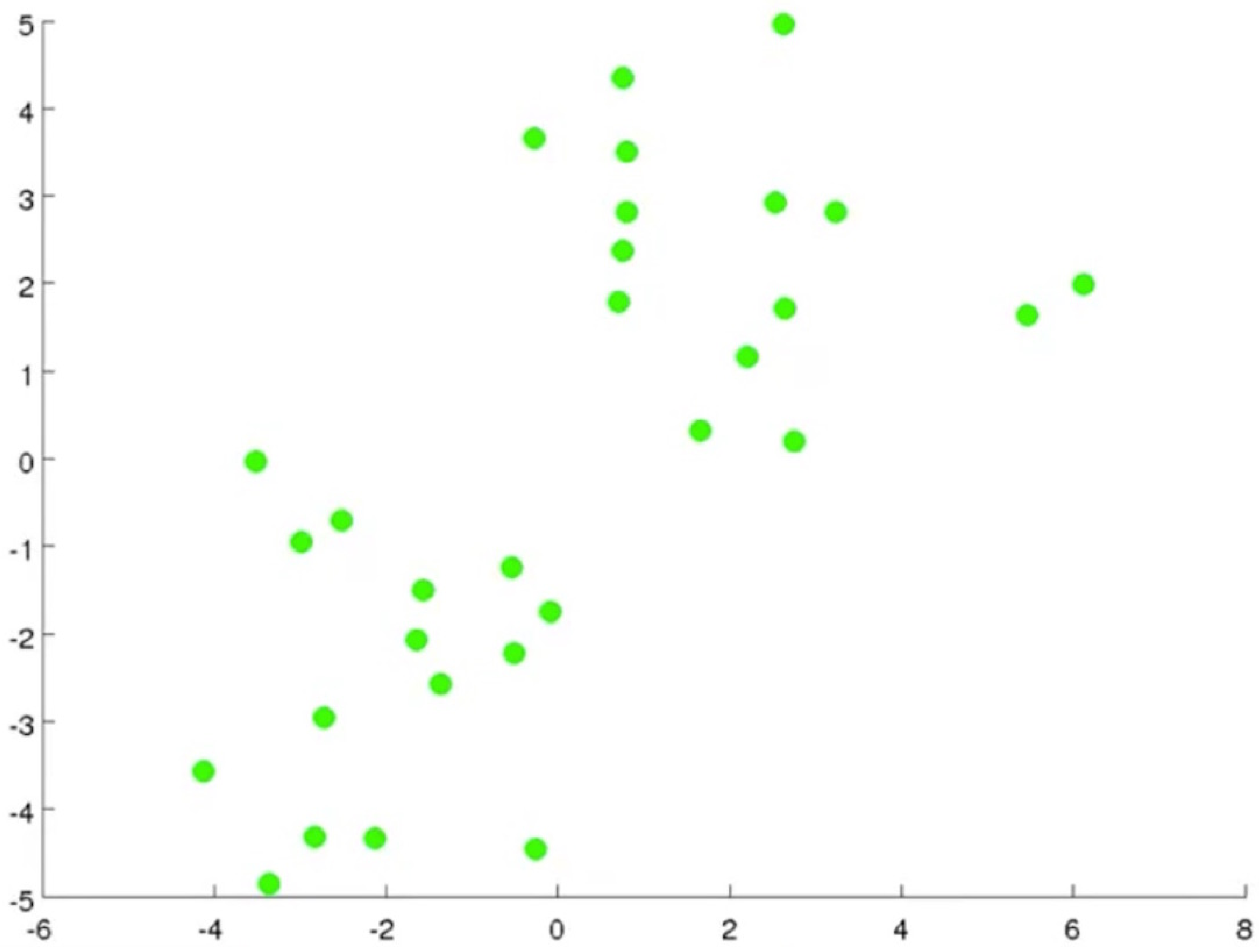
K-Means Clustering

- The red and blue stars are called “centroids”.
- The red star is the centroid of top cluster.
- The blue star is the centroid of the bottom cluster.
- The location of the centroid is at the mean of all points in the cluster.



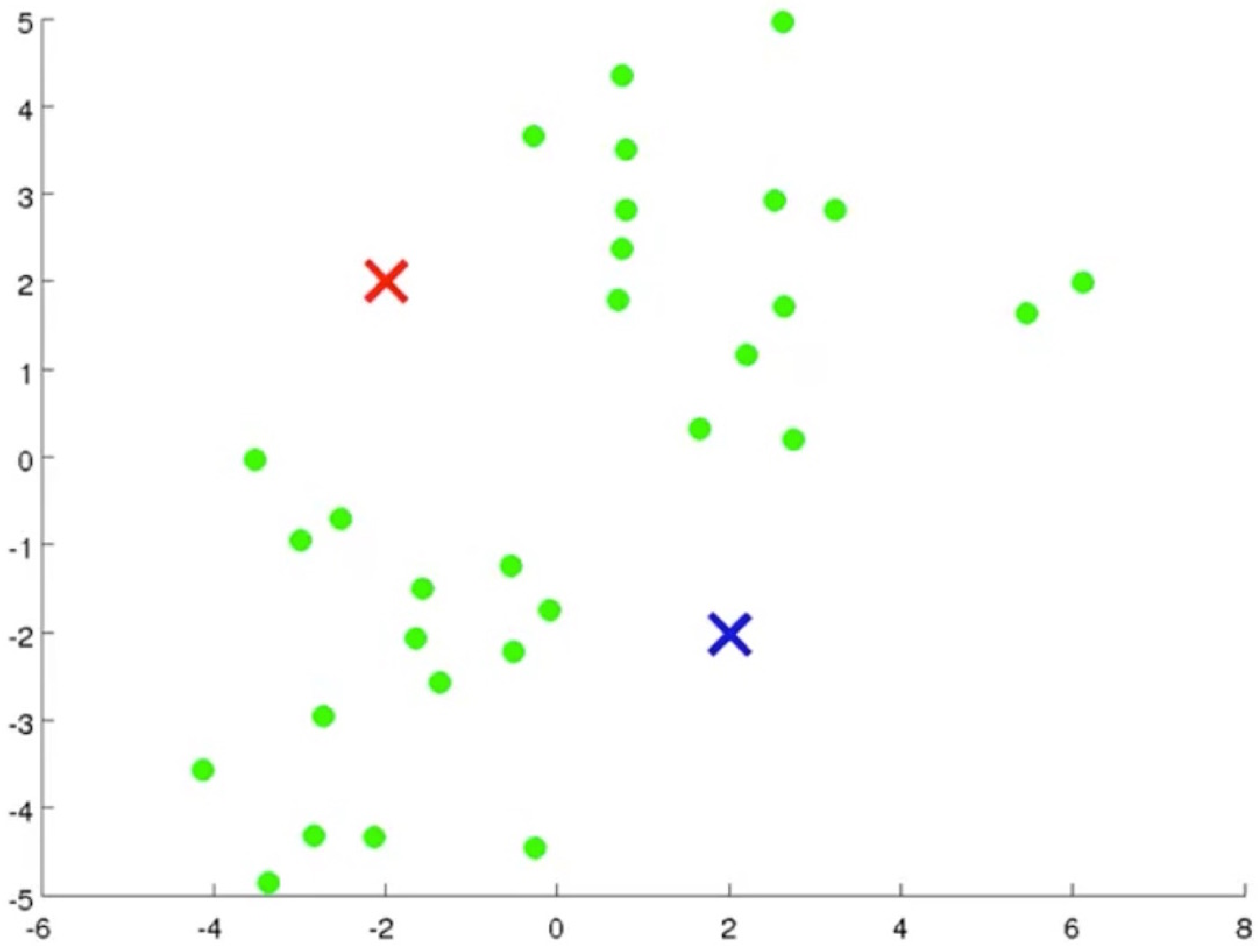
K-Means Algorithm

First, we have a bunch of unlabeled data points. We decide that we are going to find two clusters in this data.



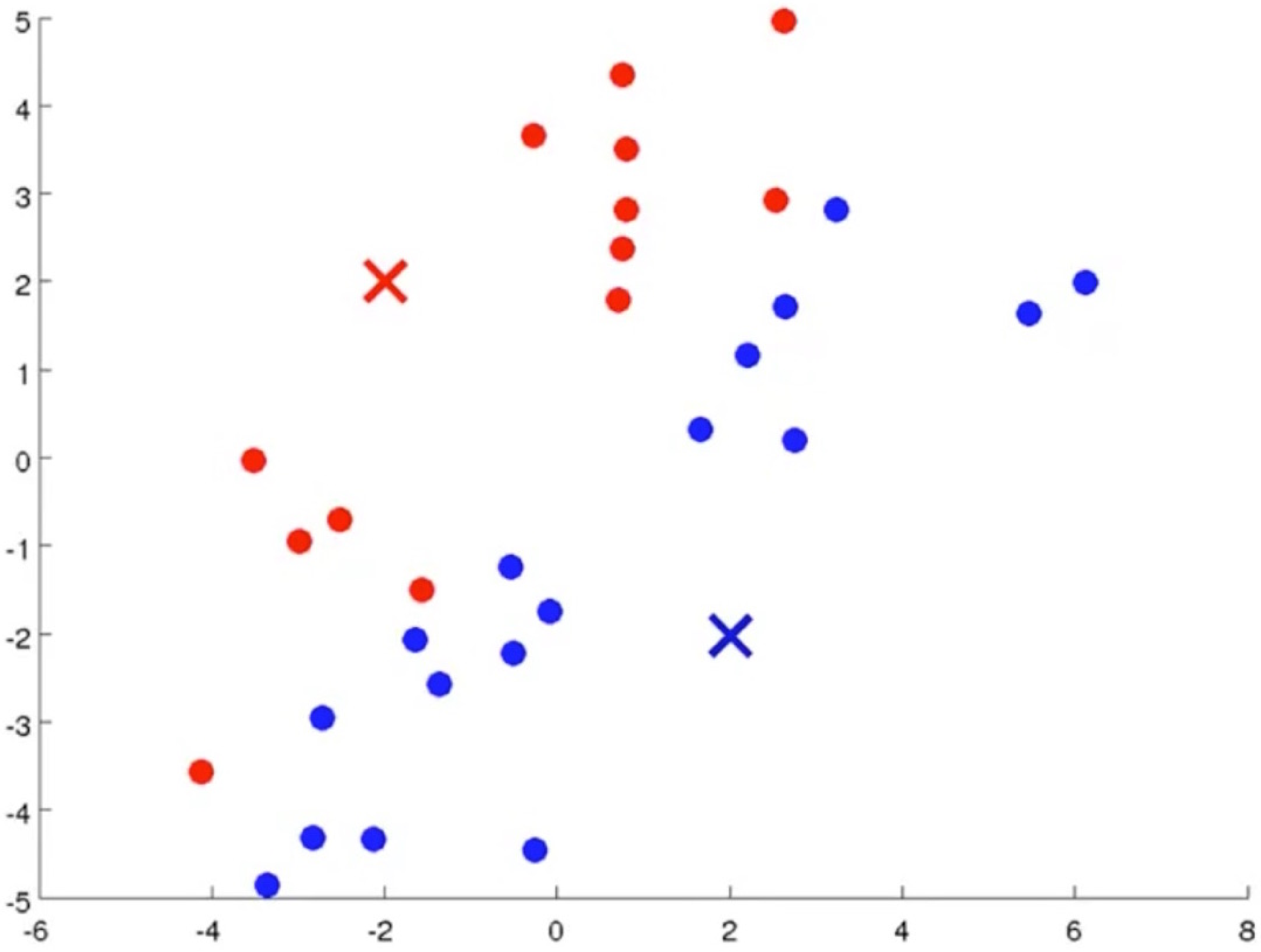
K-Means Algorithm

The first step is to pick two random locations to be our cluster centroids.



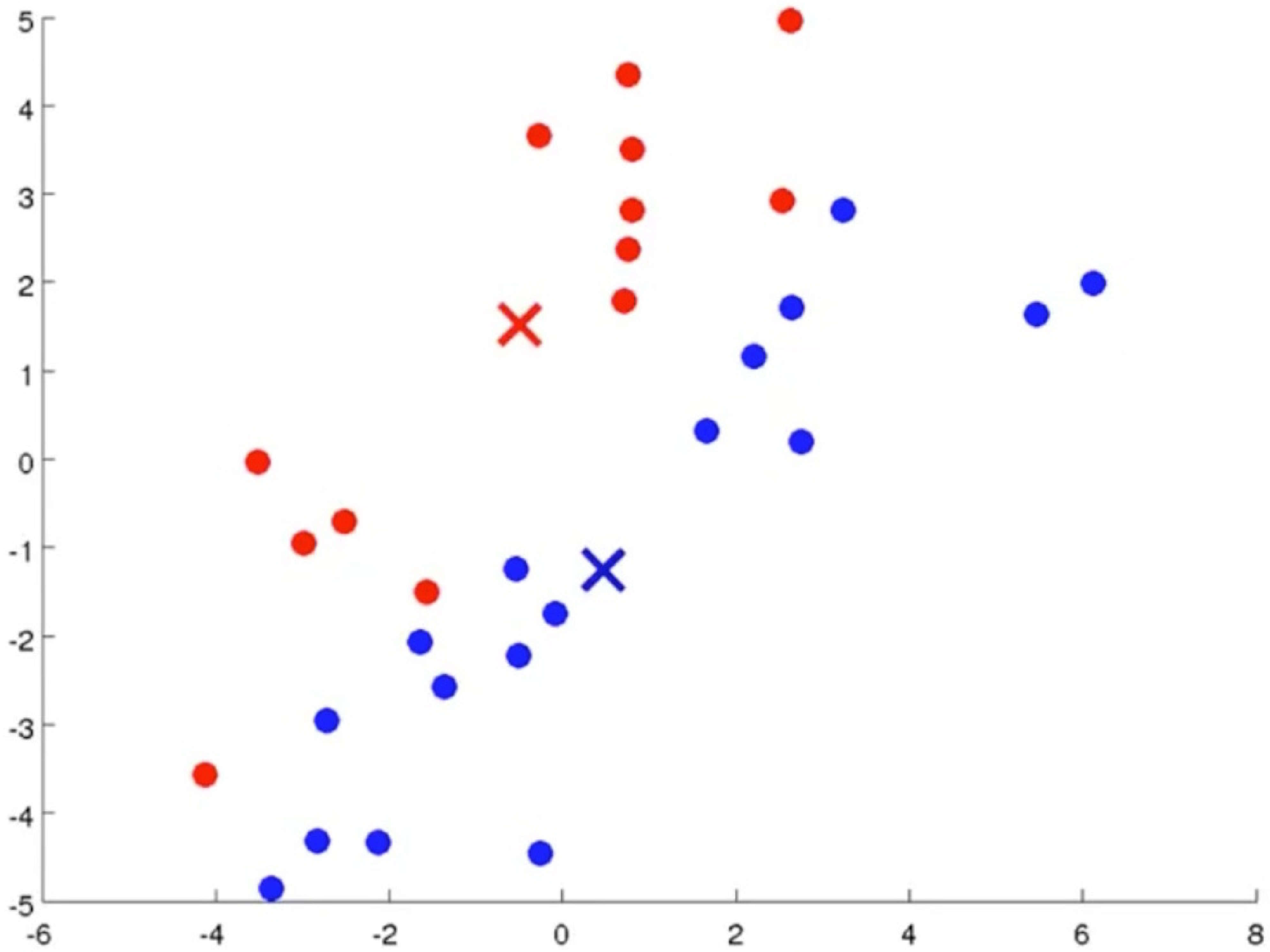
K-Means Algorithm

Then, we do cluster assignment step, where we go through every sample and determine whether each dot is closer to red or blue centroid. Label the sample to red/blue accordingly.



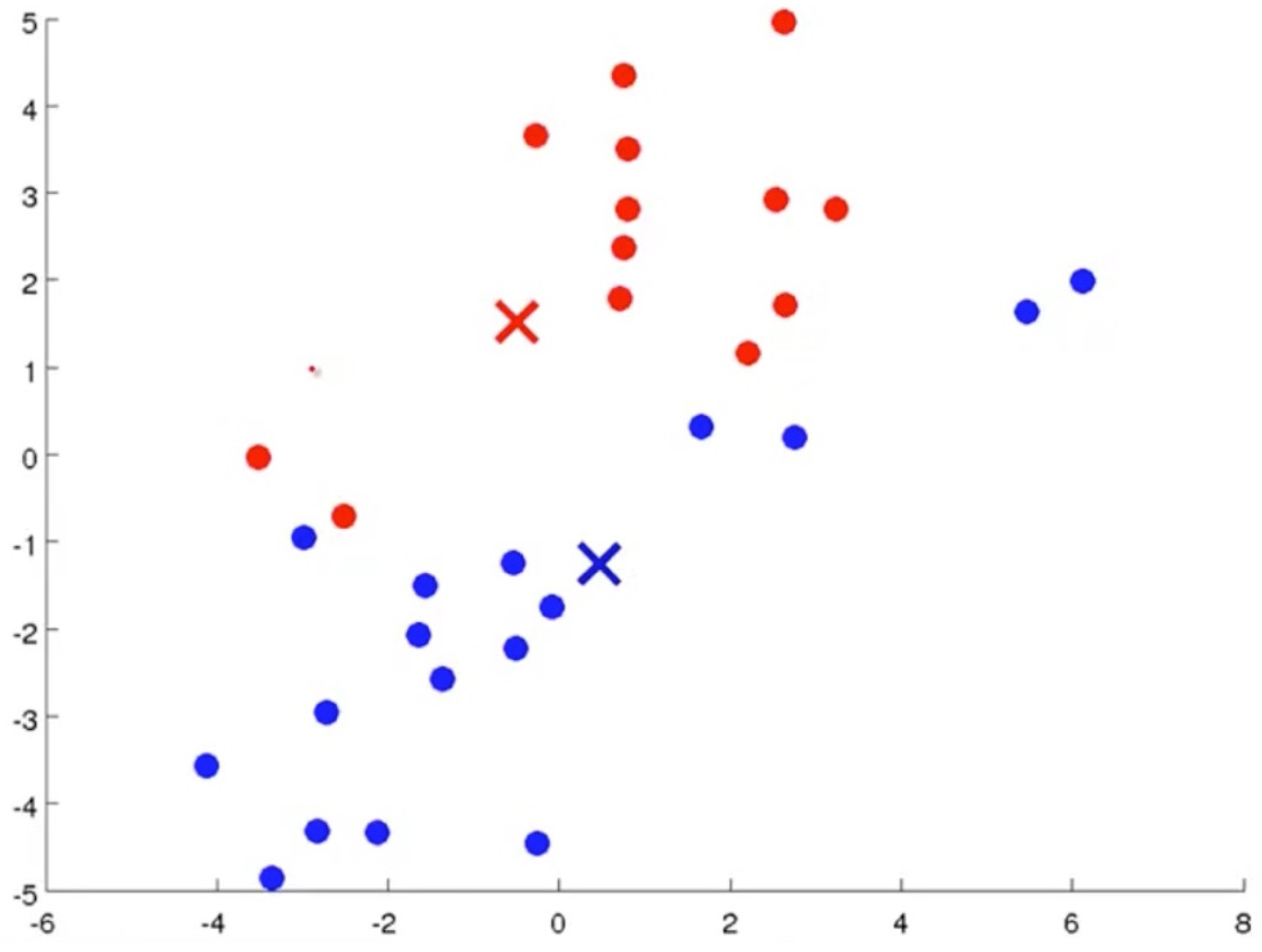
K-Means Algorithm

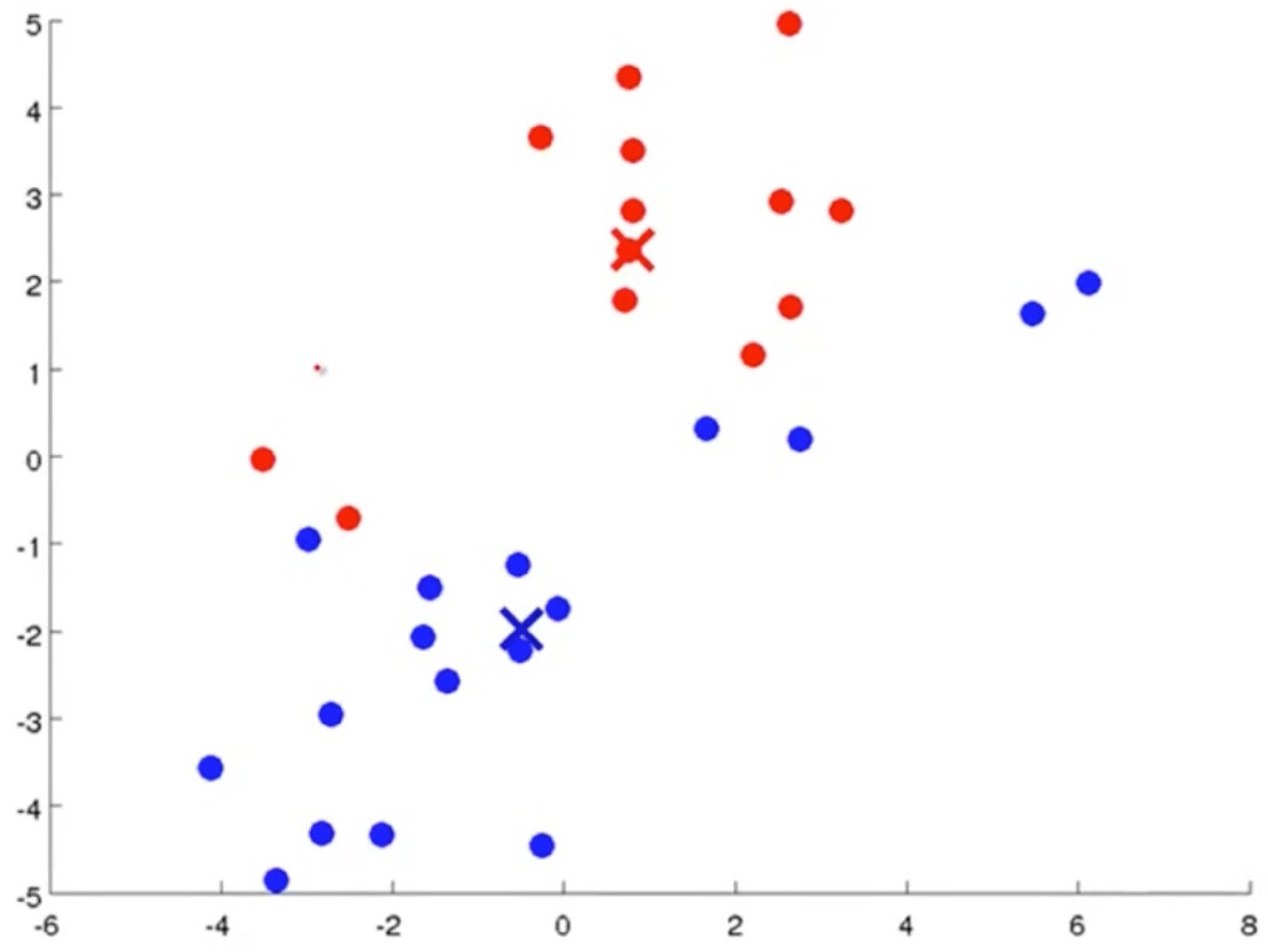
After the assignment step, we do centroid movement step where we move the red and blue centroids to the means of our clusters.



K-Means Algorithm

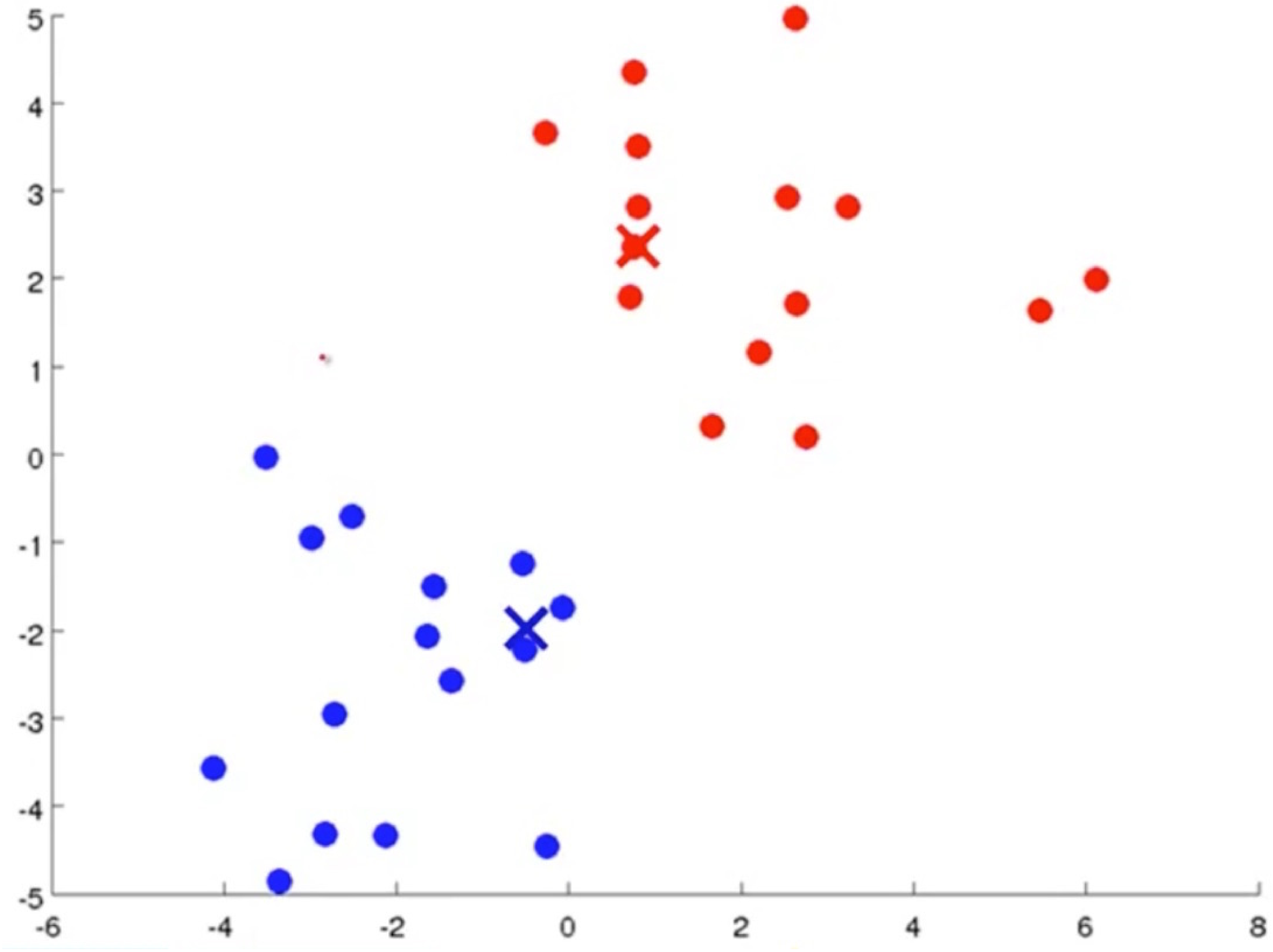
Repeat the cluster assignment step and centroid movement step, alternately.





K-Means Algorithm

The algorithm converges to the solution when cluster centroids are not changed anymore.

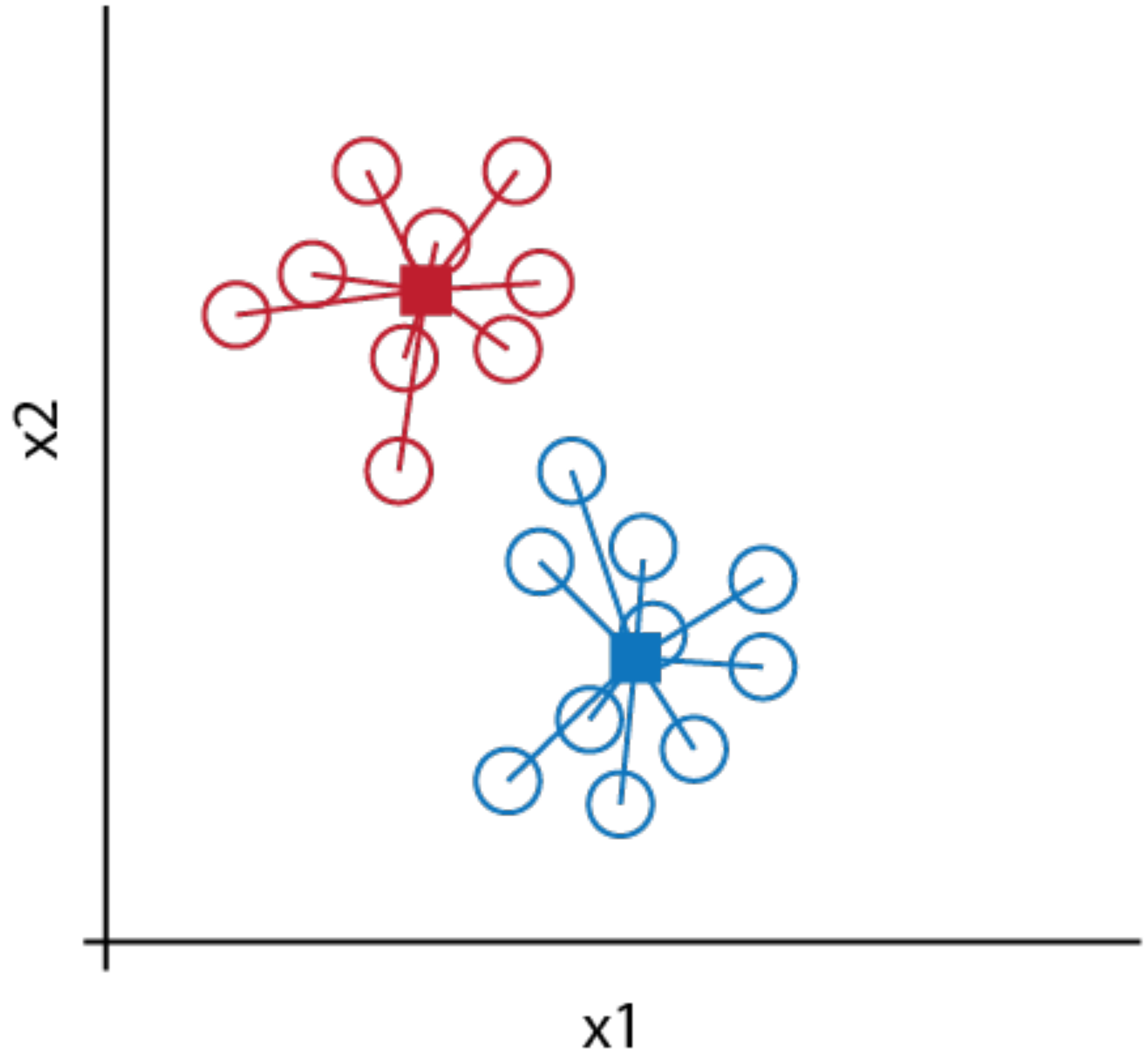


K-Means Algorithm

- To summarize, k-means algorithm has two steps:
  - In cluster assignment step, we fixed the centroids and label each data point to belong to the nearest cluster centroid.
  - In centroid movement step, we fixed data point labels and move each centroid to the mean of its data points.
- Iterating over these two steps and you will achieve the best clusters.

K-Means Algorithm

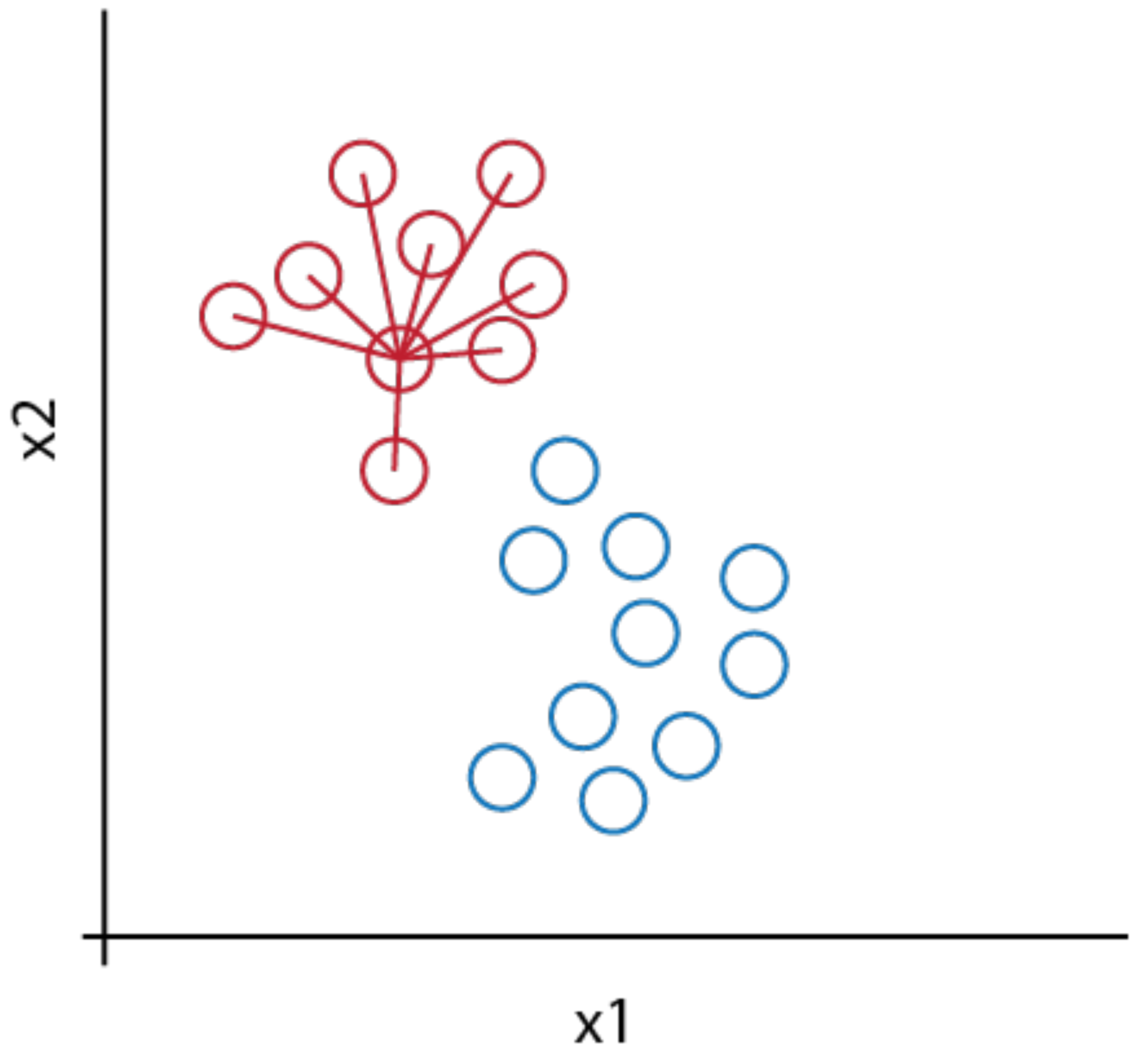
- What K-Means algorithm is doing?
- The algorithm minimize the total distances between all data points and the centroids of the clusters they belong to.



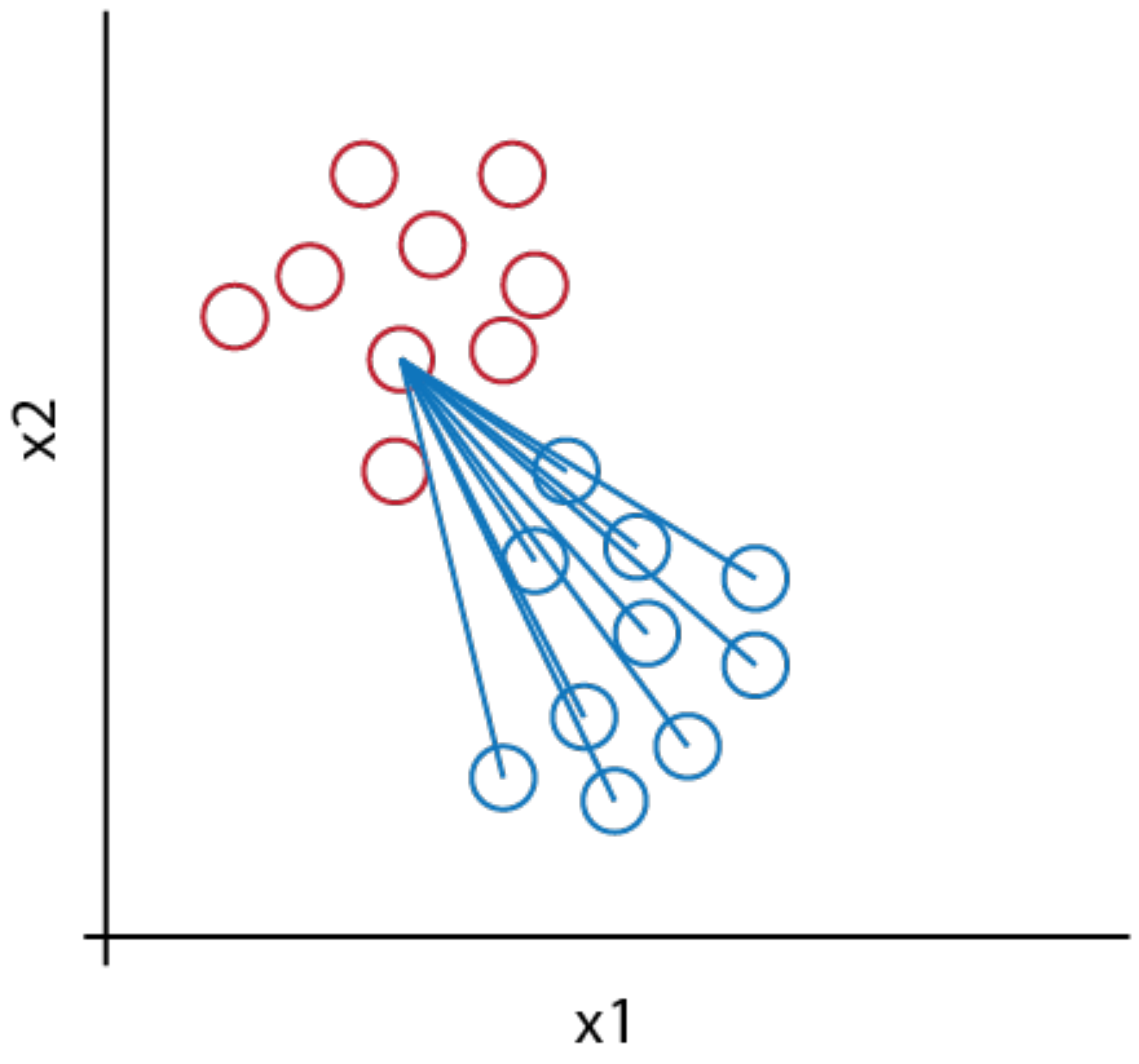
Clustering Performance

- How good is your clustering results? It is not as simple as counting right and wrong predictions. Performance of clustering algorithm is often judged based on how well you algorithm separates out the data into several clusters.

Cohesion

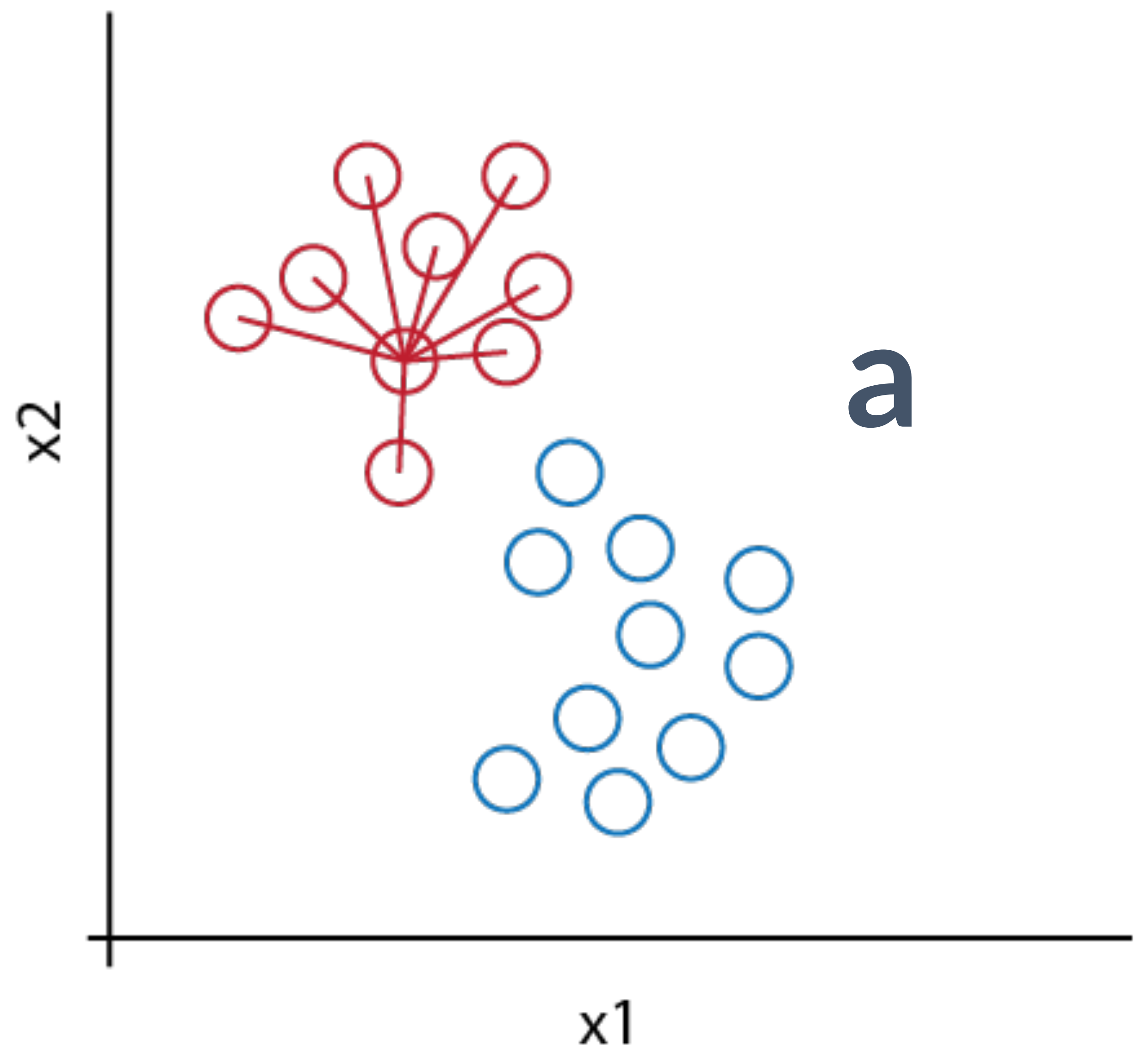


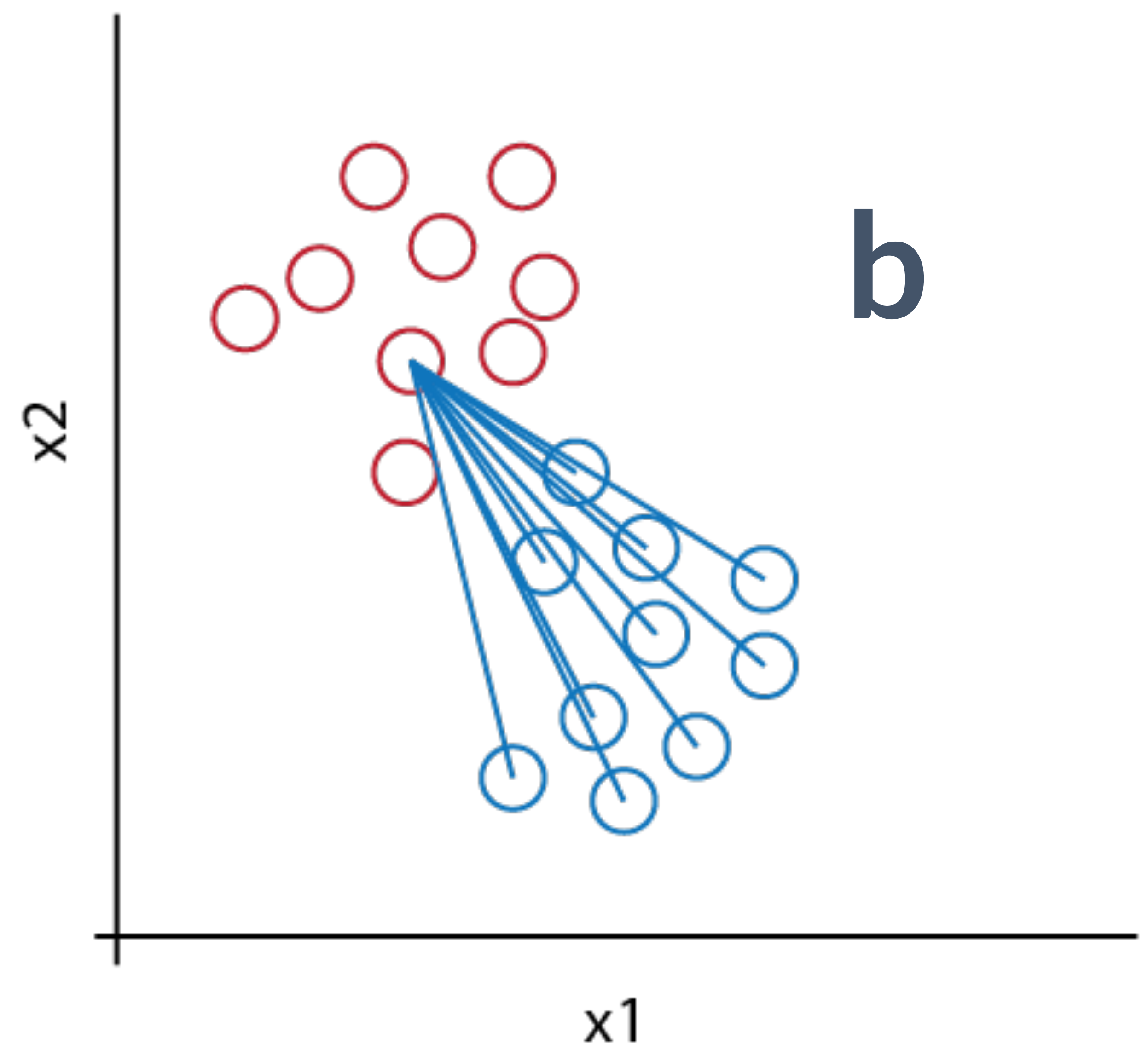
Separation



Clustering Performance

- Silhouette Coefficient is one such measure. It's defined by two separate scores:
  - a: mean distance between a sample and all other points in the same class.
  - b: mean distance between a sample and all other points in the next nearest cluster.
- Silhouette Coefficient
  $$= (b-a)/\max(b,a)$$
  $$= 1 - (a/b) \quad \text{- if } b>a$$





Clustering Performance

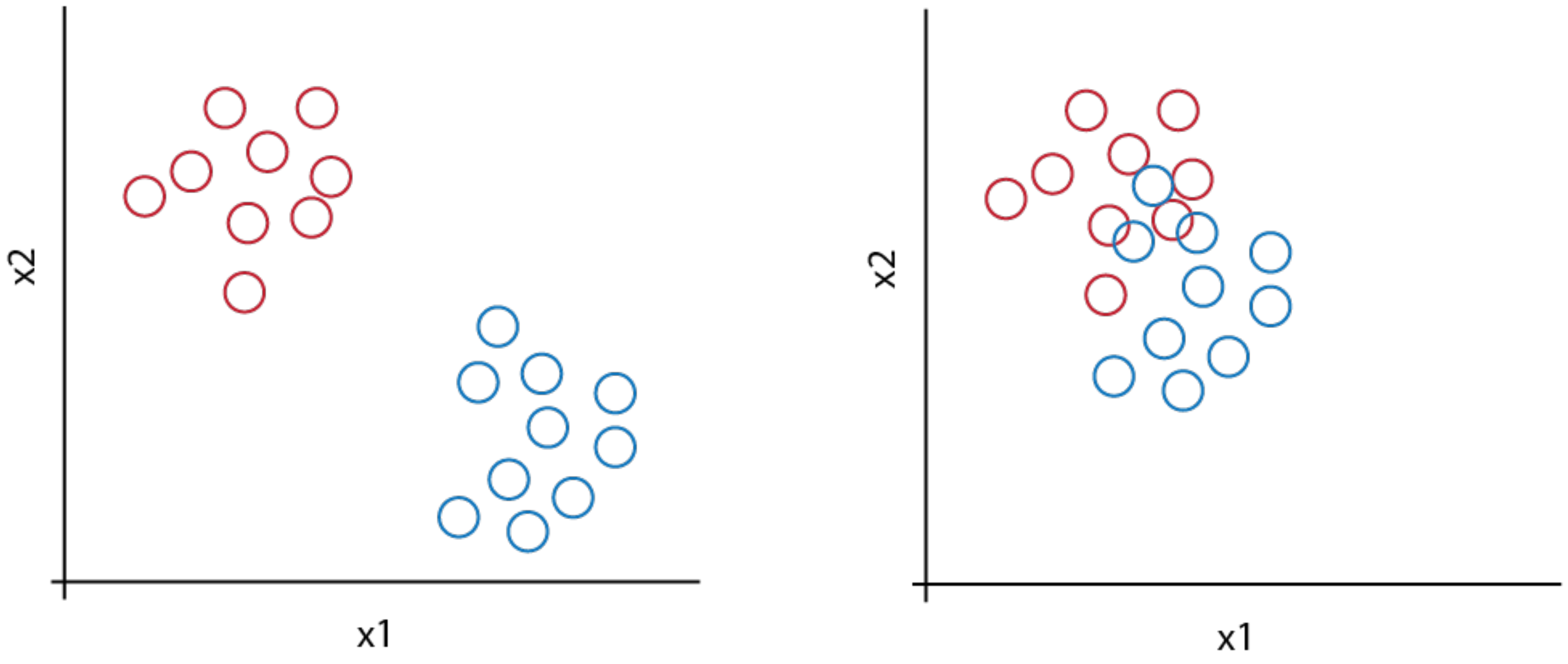
- Silhouette Coefficient = 1- (a/b)
- If b>>a, SC is close to 1.
- If a<<b, SC is close to -1.

Silhouette close to 1

Good Clusters

Silhouette close to -1

Bad Clusters



Clustering Performance

- Other methods of clustering performance include Calinski-Harabaz Index (dispersion from centroid), percentage of explained variance (between-group variance / total variance), and many others.
- We can calculate these performance metrics using any distance metrics we learned about (Euclidean, Manhattan, Jaccard).
- Typically, it would be best if we could find some ground truth to validate our clusters. For example, if we cluster 1M news articles, we could get people to label 5000 news articles to establish ground truth for validating our clusters.

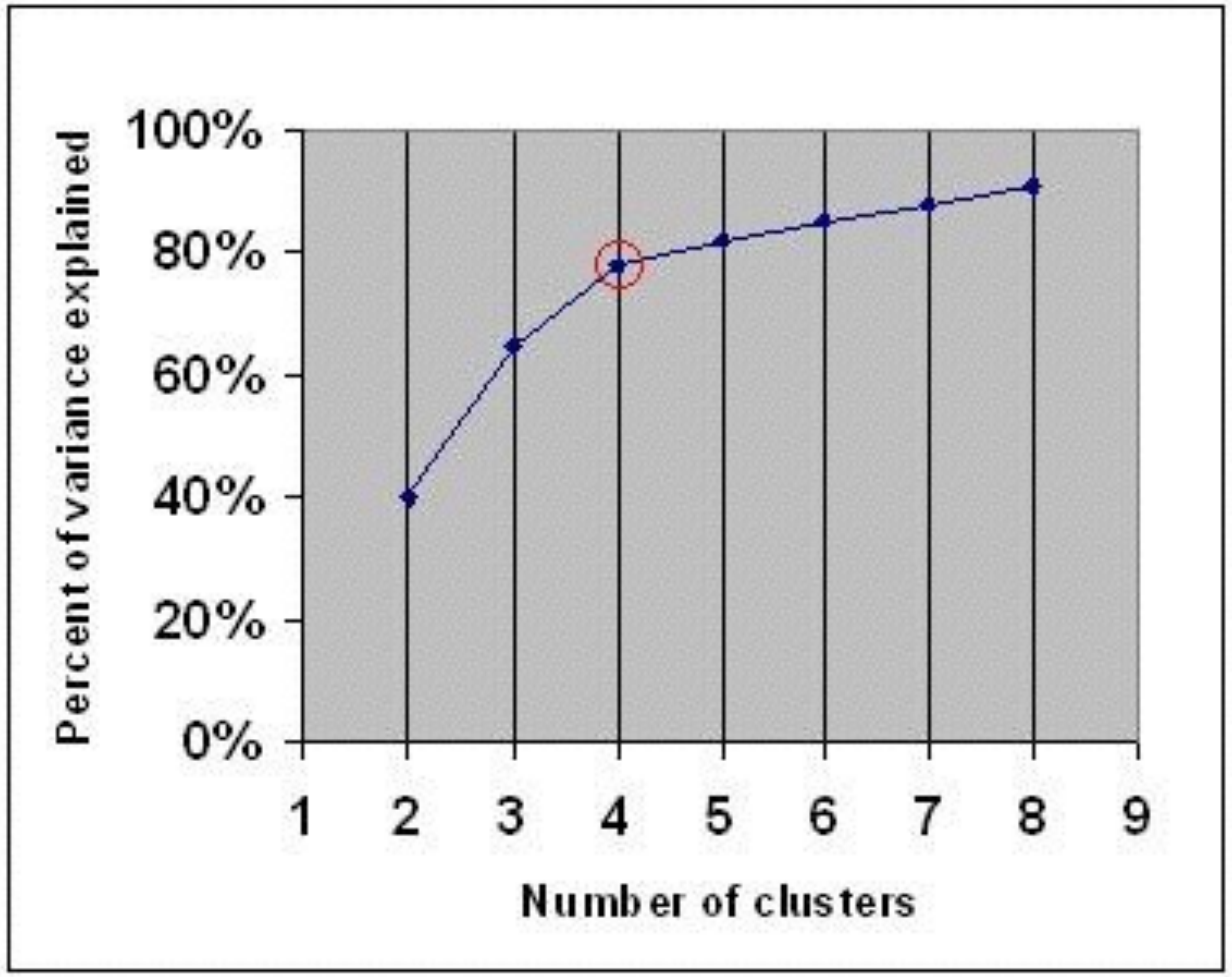
Clustering Performance

- **When ground-truth labels are provided, people talk about clustering performance in terms of homogeneity, completeness, and v-measure**
  - Homogeneity: each cluster contains only members of a single class.
  - Completeness: all members of a given class are assigned to the same cluster.
  - Both homogeneity and completeness are bounded below by 0.0 and above by 1.0 (higher is better)
  - **V-Measure Score:** the harmonic mean between homogeneity and completeness.

v = 2 * (homogeneity * completeness) / (homogeneity + completeness)

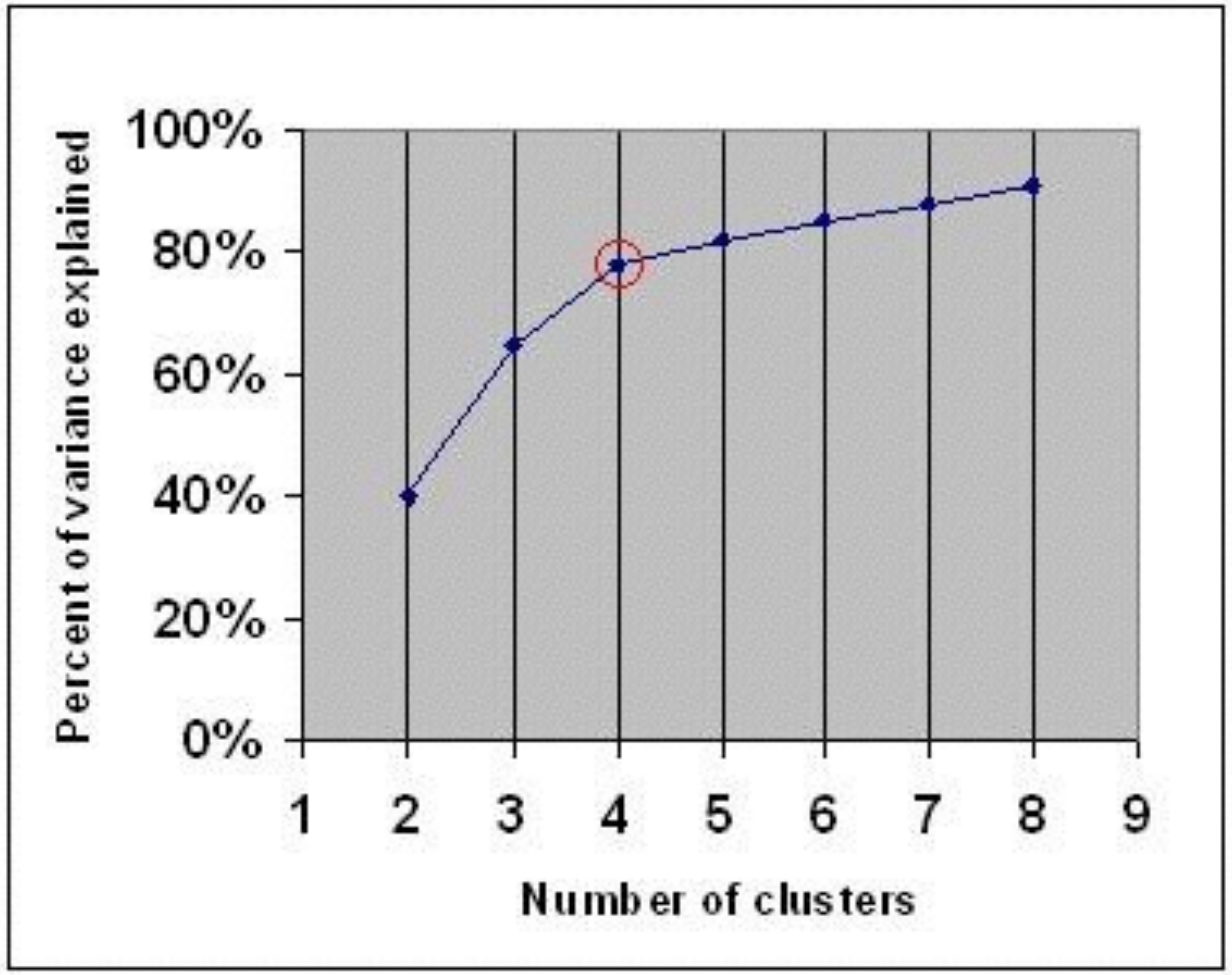
Choosing the Right K

- How to determine the number of clusters in the dataset?
- The most common way is called elbow methods.
- You compute performance metrics such as Silhouette Coefficient or Percentage of Variance (in the figure), while varying k.



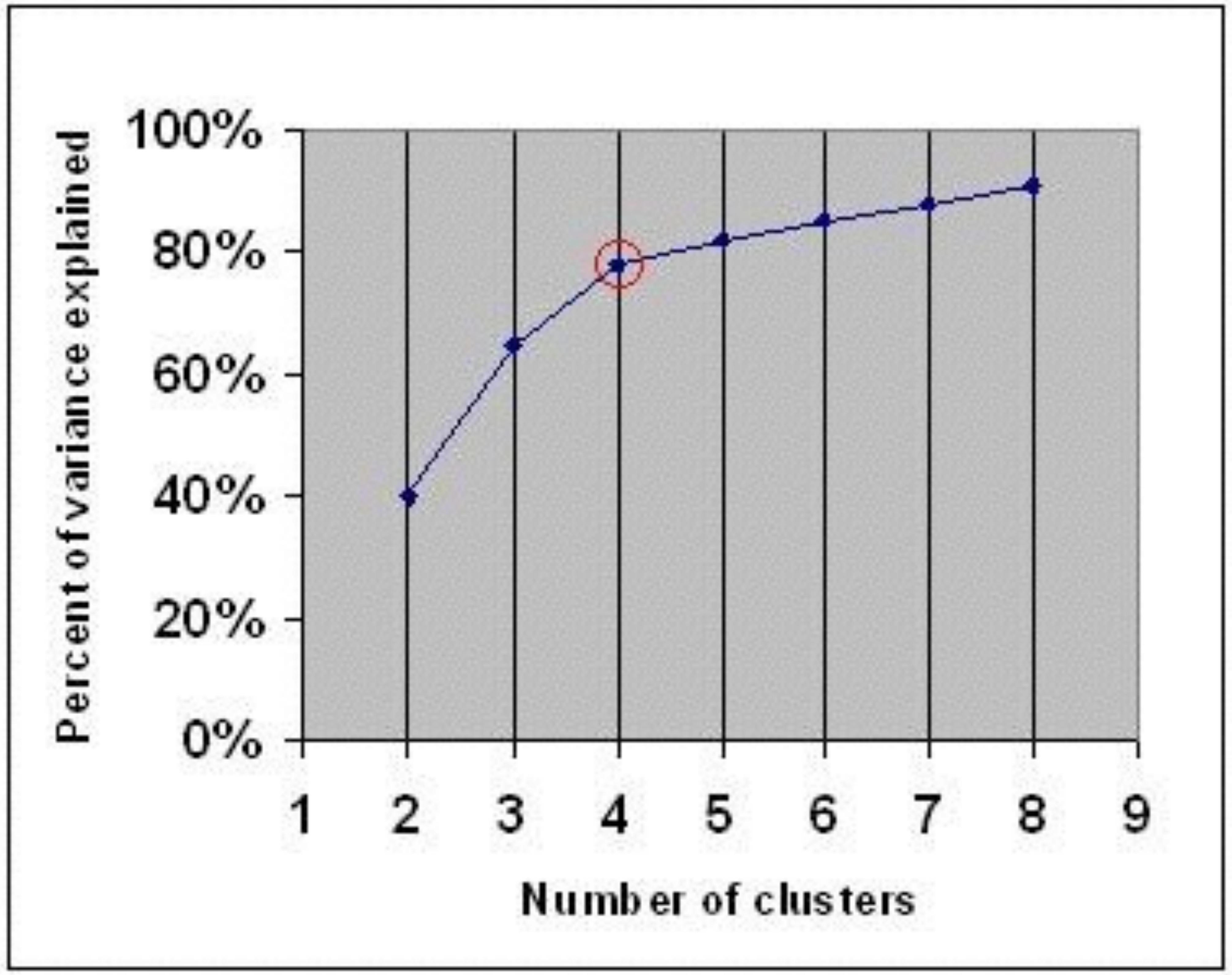
Choosing the Right K

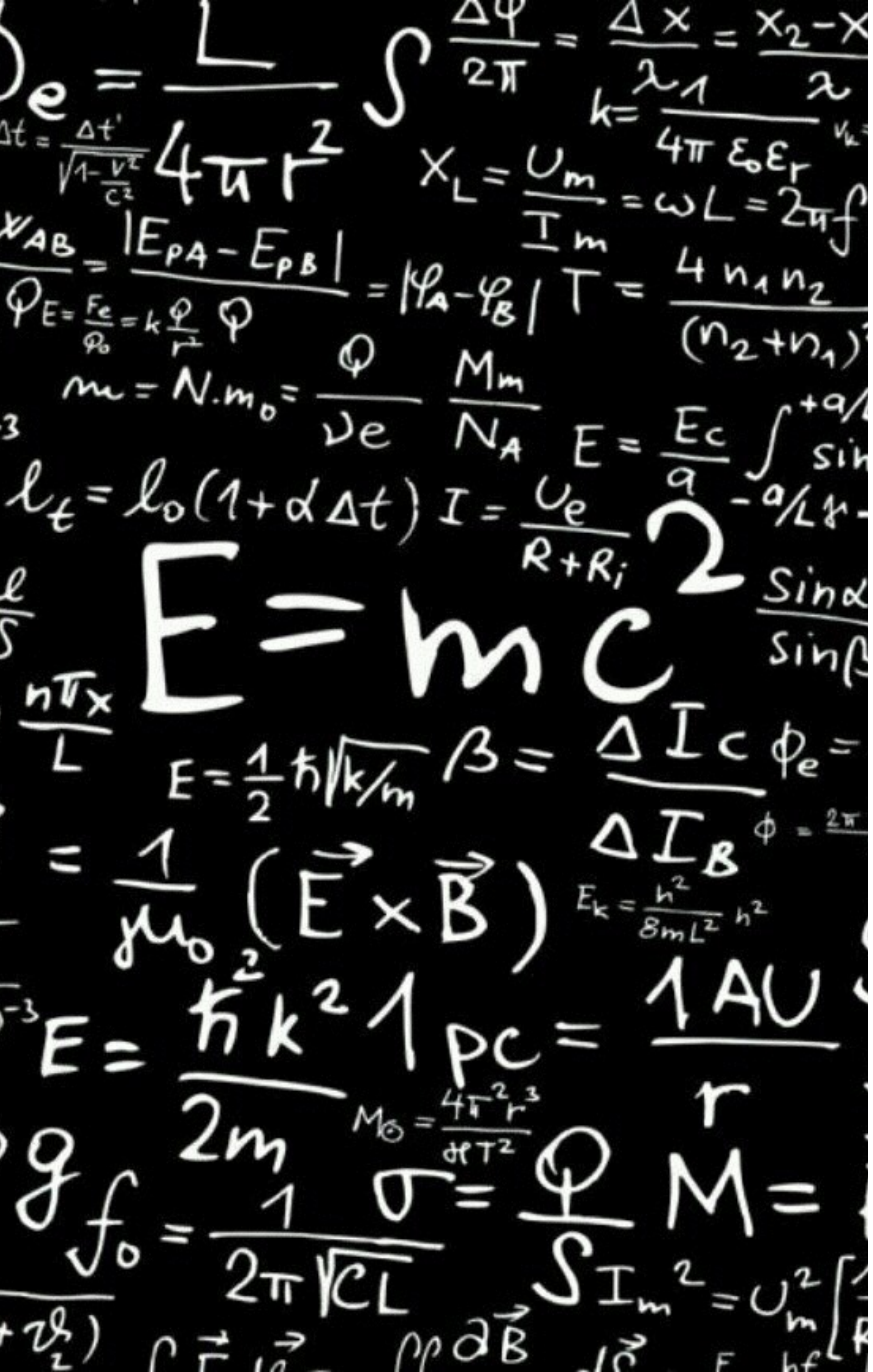
- Pick the k at the elbow point. At this point, more clusters do not necessarily mean higher performance.
- Similarly we use metrics like Information Criterion (the famous one being Bayesian - BIC, and Akaike - AIC). These are measures that find a good balance between the variance explained and the complexity of the model.



Choosing the Right K

- Pick the k at the elbow point. At this point, more clusters do not necessarily mean higher performance.
- Similarly we use metrics like Information Criterion (the famous one being Bayesian - BIC, and Akaike - AIC). These are measures that find a good balance between the variance explained and the complexity of the model.

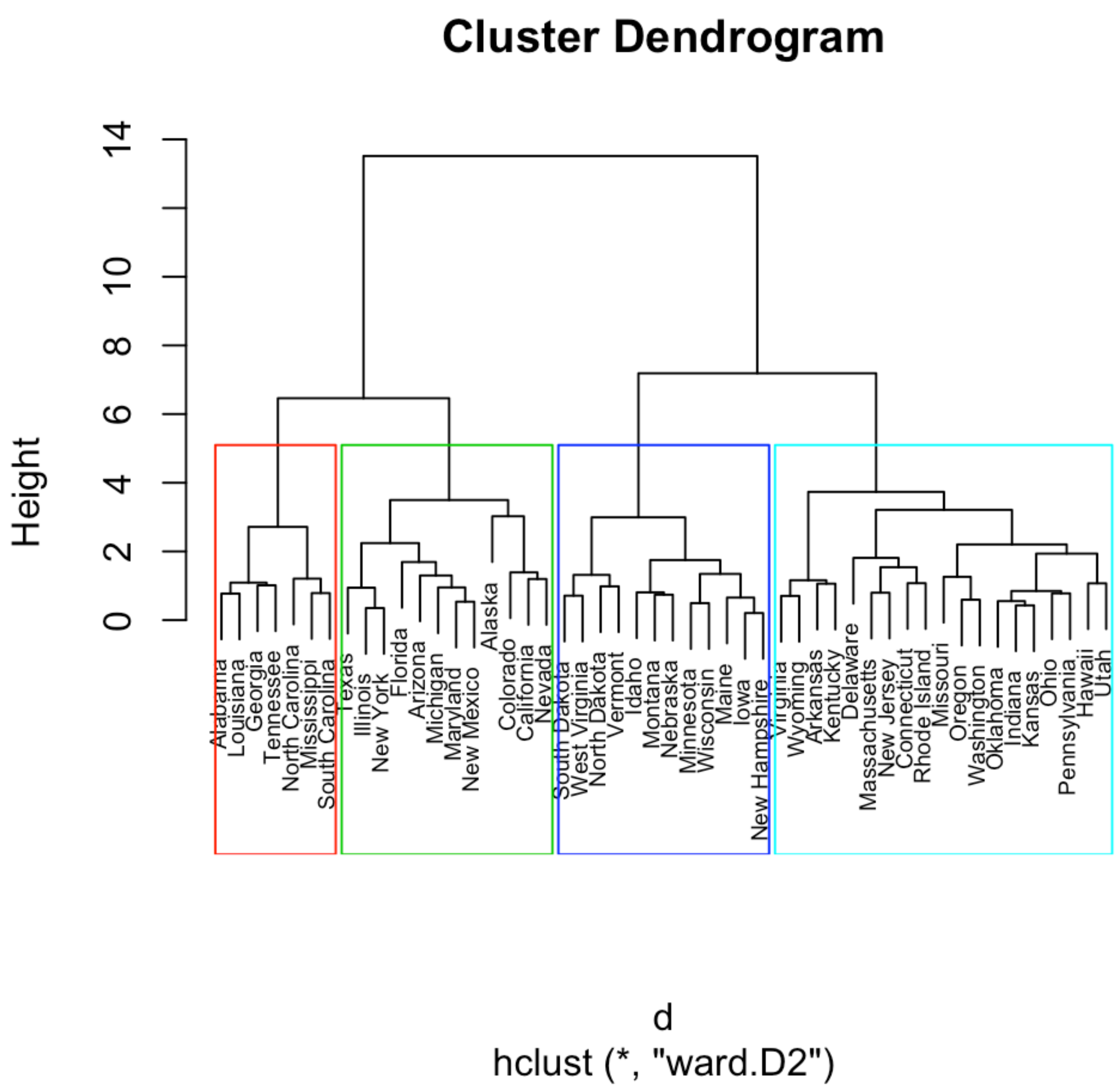




HIERARCHICAL CLUSTERING

Hierarchical Clustering

- Hierarchical clustering does not require to pre-specify the number of clusters to be generated.
- The result is a tree-based representation of the observations which is called a dendrogram.
- It uses pairwise distance matrix between observations as clustering criteria.



credit: sthda.com

Two Approaches

- Agglomerative: This is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
  - Each object is initially considered as a single-element cluster (leaf).
  - At each step of the algorithm, the two clusters that are the most similar are combined into a new bigger cluster (nodes).
  - Iterated until all points are member of just one single big cluster (root)
- Divisive: This is a "top down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.
  - It begins with the root, in which all objects are included in a single cluster. At each step of iteration, the most heterogeneous cluster is divided into two.

credit: sthda.com

Measuring Dissimilarity

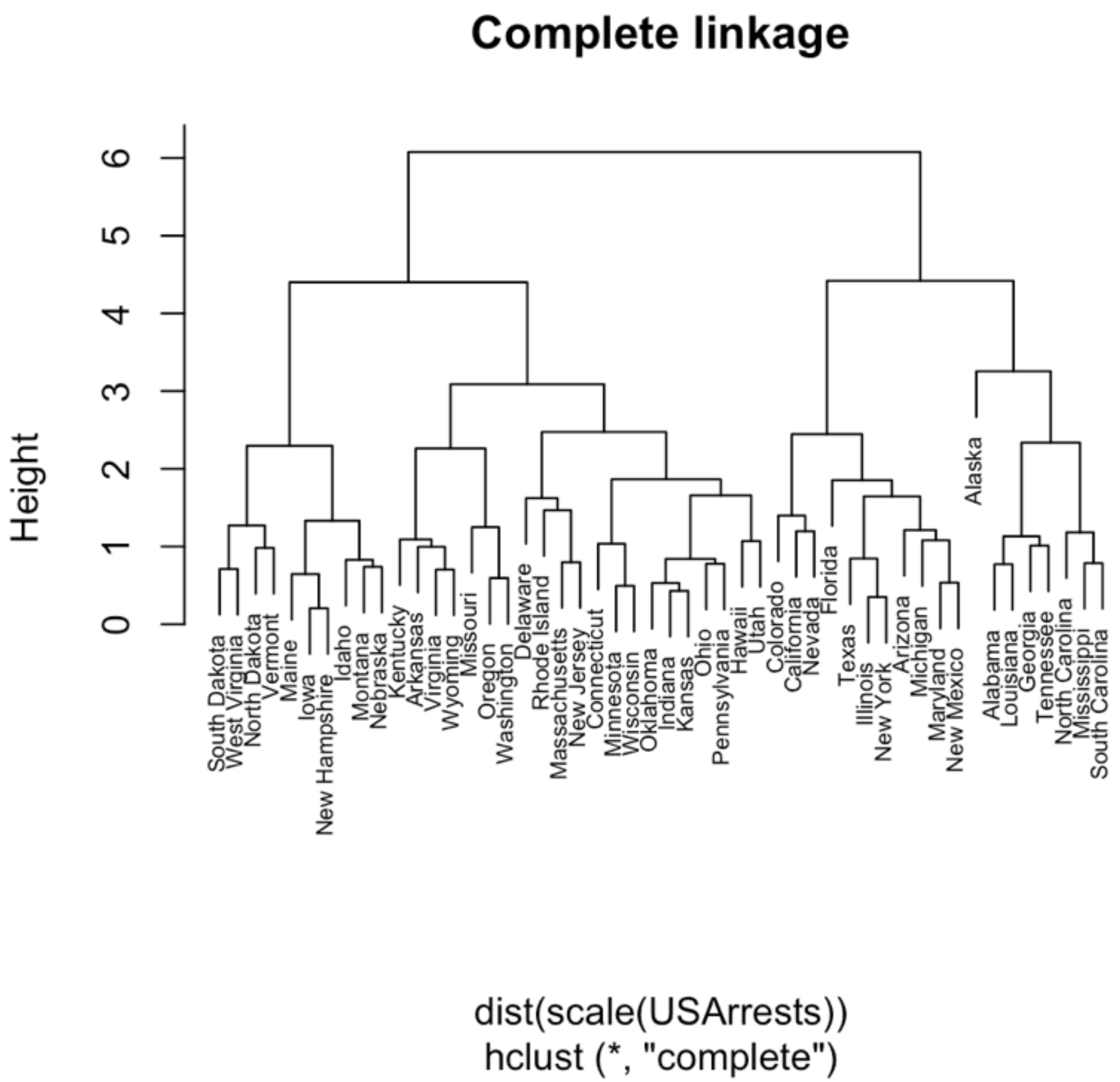
- **How to measure the dissimilarity between two clusters of observations?**
- **A number of different linkage methods has been developed to answer to this question.**
  - **Maximum or complete linkage clustering:** It computes all pairwise dissimilarities between cluster 1 and cluster 2, and considers the maximum dissimilarity as the distance between the two clusters. It tends to produce more compact clusters.
  - **Minimum or single linkage clustering:** Same as above, and considers the smallest of these dissimilarities as a linkage criterion. It tends to produce long, "loose" clusters.

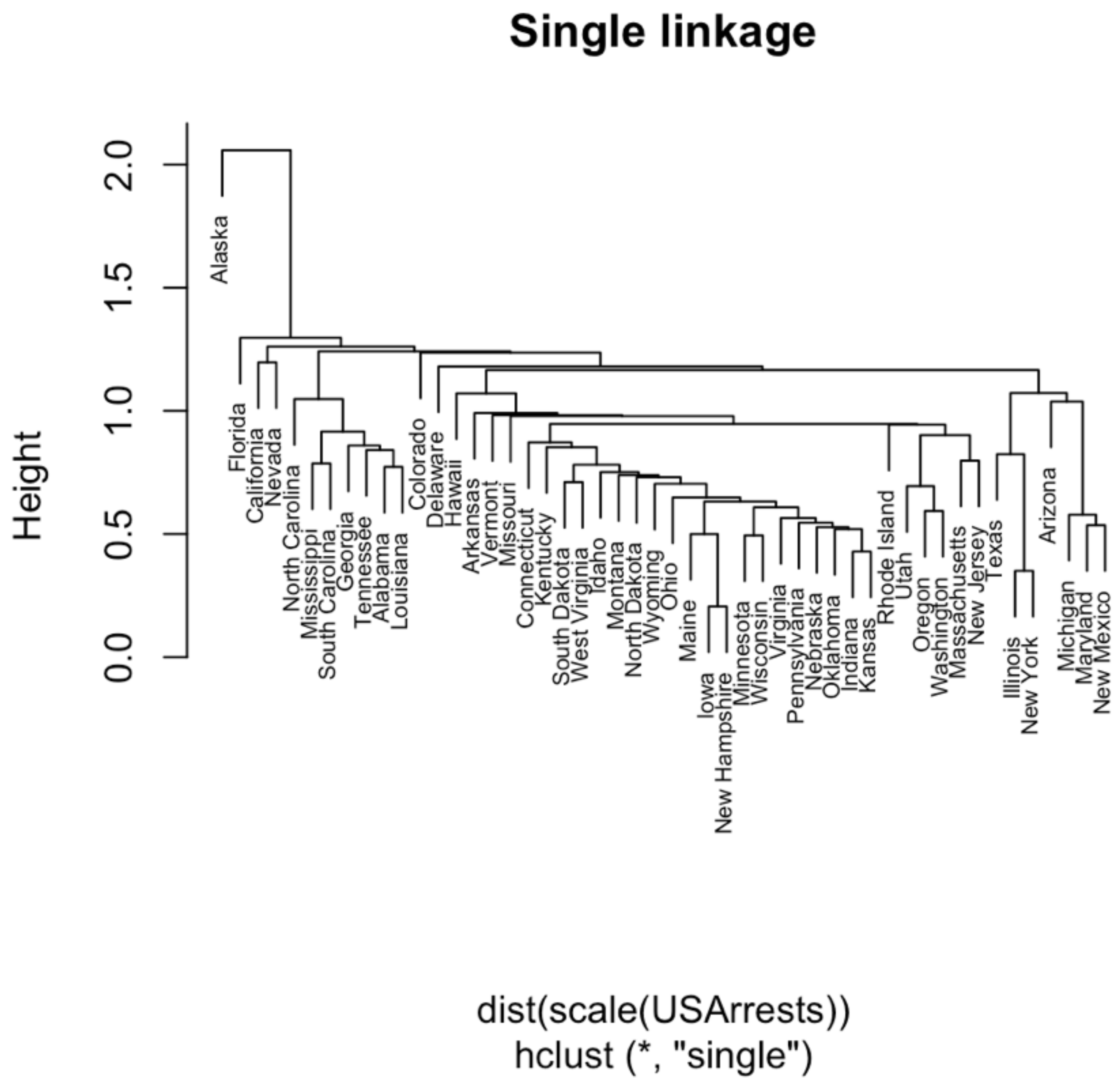
credit: sthda.com

Measuring Dissimilarity

- **How to measure the dissimilarity between two clusters of observations?**
- **A number of different linkage methods has been developed to answer to this question.**
  - **Mean or average linkage clustering:** Same as above, and considers the average of these dissimilarities as the distance between the two clusters.
  - **Centroid linkage clustering:** It computes the dissimilarity between the centroid for cluster 1 and the centroid for cluster 2.

Hierachical Clustering



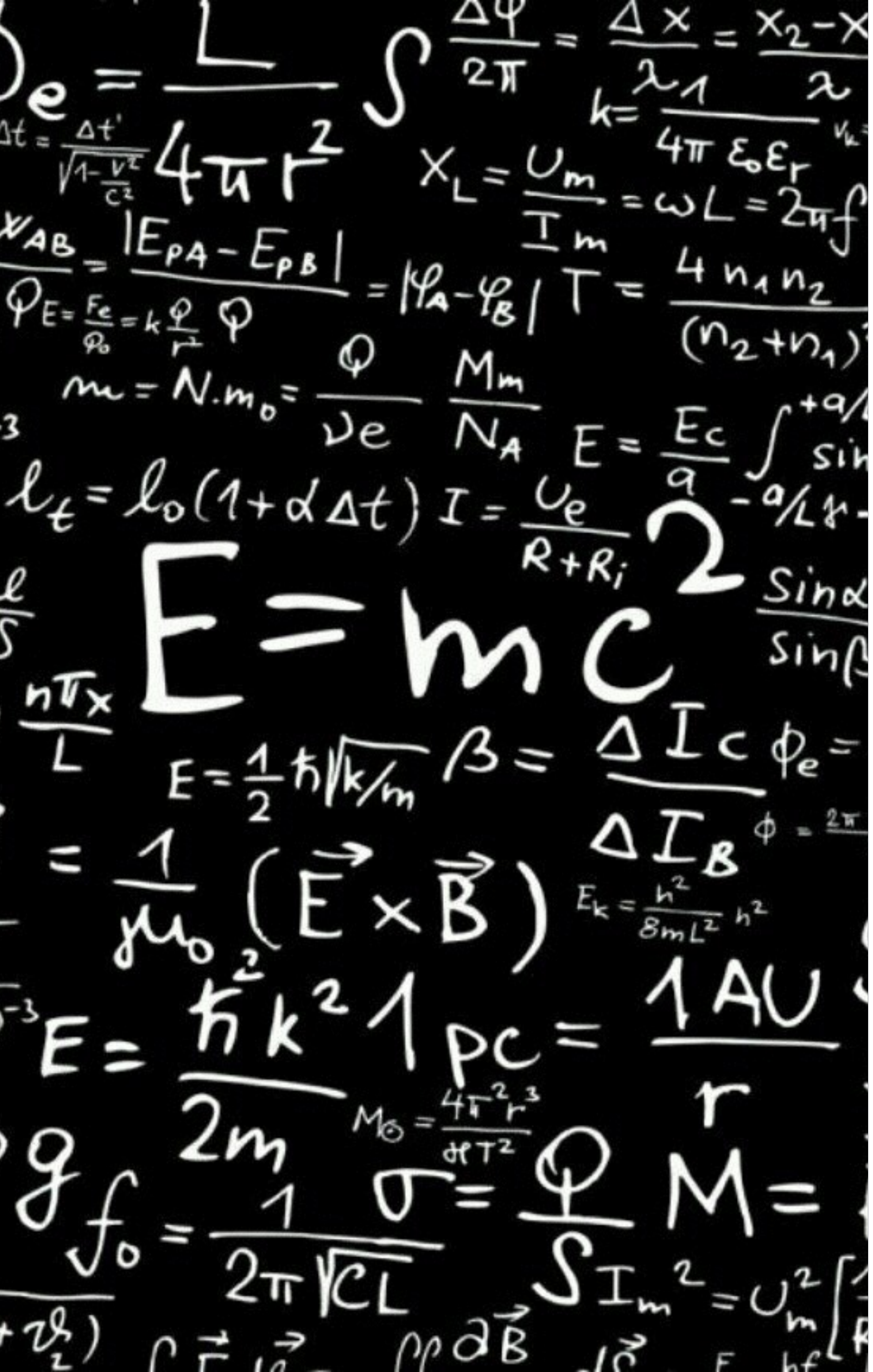


Other Unsupervised Learning Flavors

- **Random Optimization:**
  - Ways to find the right solution (x) from a large space of solutions
  - Example: factories need to find the right setting for their old refinement plants. There are 100 variables, which one would yield the best old quality?
  - Hill Climbing, GA, Simulated Annealing, etc.
- **Dimensionality Reduction**
  - Transform a set of huge data (1000 dimension) to a small data (3 dimensions) without information loss
  - Find a set of weights that optimize an objective function
  - Principle component analysis, Independent component analysis

Other Unsupervised Learning Flavors

- **Neural Network Unsupervised Learning**
  - Self-organizing map: a grid of neuron nodes being continuously updated until there are nodes that represent significant subclasses of the inputs.
  - Embedding techniques: creating neural-network-like representation of the inputs that might be abstract such as modeling relationships between words from reading articles.
  - Auto encoder: training neural network to photocopy the images in order to create a feature space of images.



K-MEANS LAB