UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

COMS 4030A/7047A Adaptive Computation and Machine Learning

Hima Vadapalli

Semester I, 2022

So far:
Supervised Learning

Today:
Unsupervised Learning

Slides based heavily on course material by Eric Eton Andrew Moore and Andrew Ng

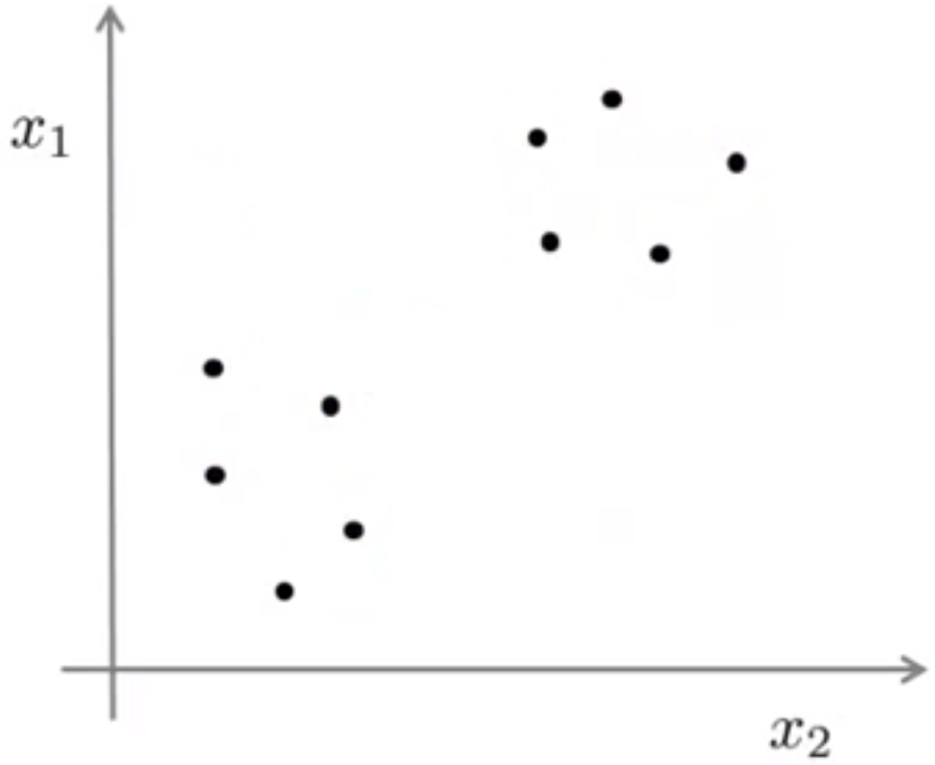
Unsupervised Learning

- Supervised learning used labeled data pairs $(\mathbf{x}, \mathrm{y})$ to learn a function $f : \mathrm{X} \rightarrow \mathrm{Y}$
  - But, what if we don't have labels?

- No labels = **unsupervised learning**
- Only some points are labeled = **semi-supervised learning**
  - Labels may be expensive to obtain, so we only get a few

- **Clustering** is the unsupervised grouping of data points.It can be used for **knowledge discovery**.

Unsupervised Learning

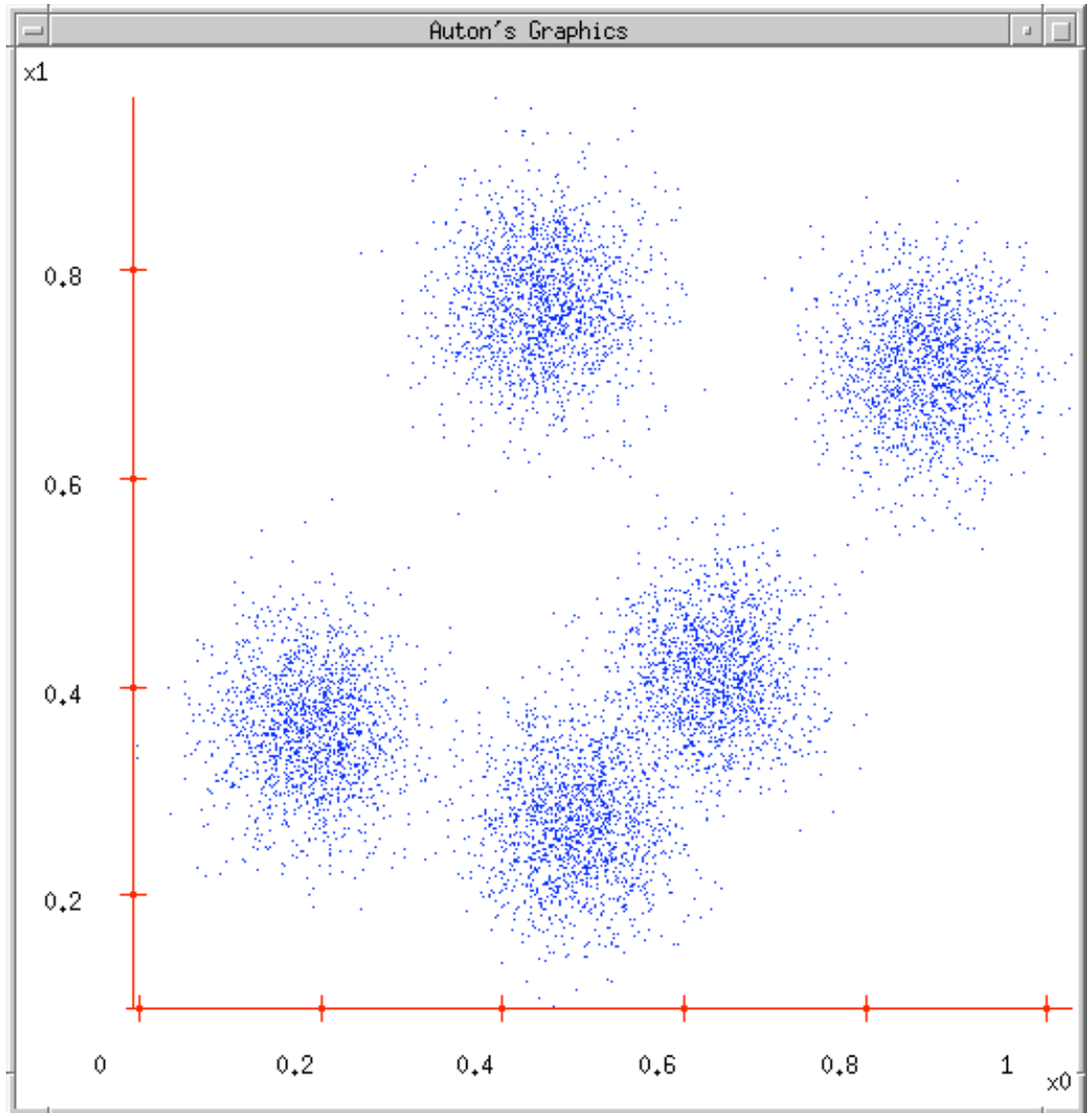


K-Means Clustering

Some material adapted from slides by Andrew Moore, CMU.

Visit http://www.autonlab.org/tutorials/ for
Andrew's repository of Data Mining tutorials.

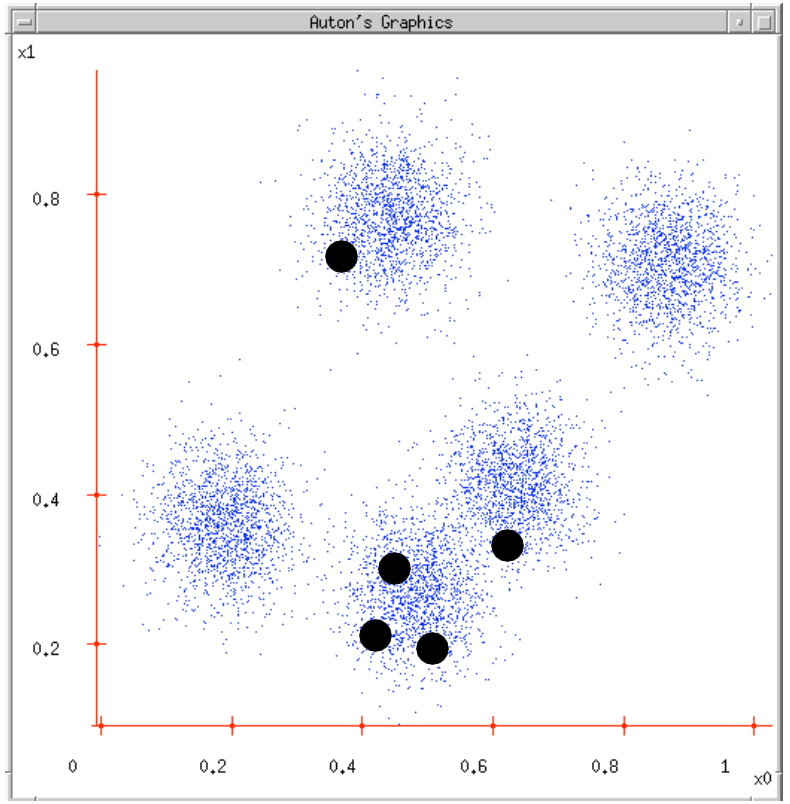
Clustering Data



K-Means Clustering

K-Means (k , X)

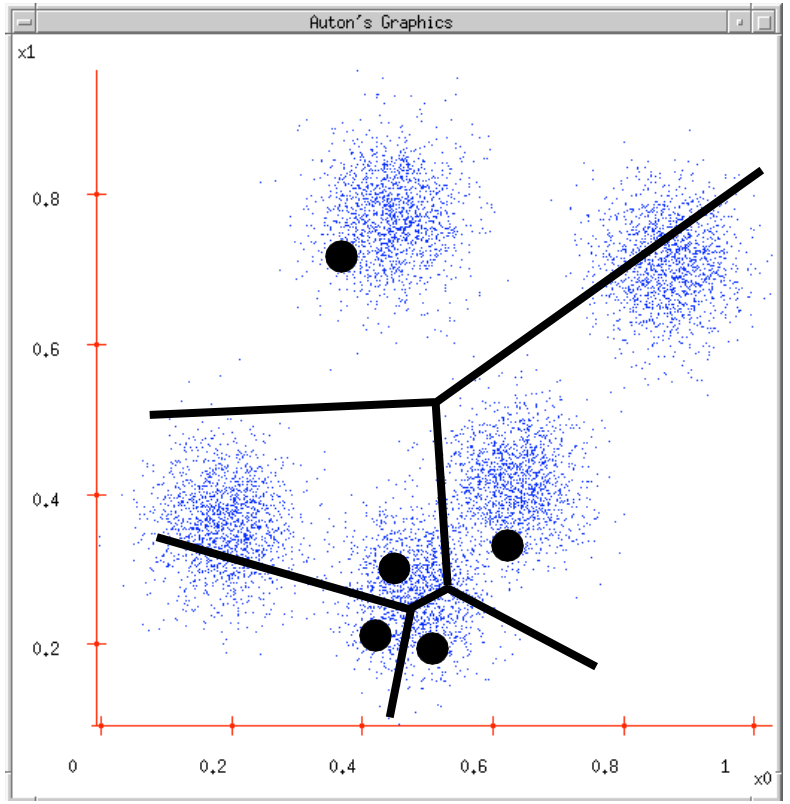
- Randomly choose k cluster center locations (centroids)
- Loop until convergence
  - Assign each point to the cluster of the closest centroid
  - Re-estimate the cluster centroids based on the data assigned to each cluster



K-Means Clustering

K-Means (k , X)

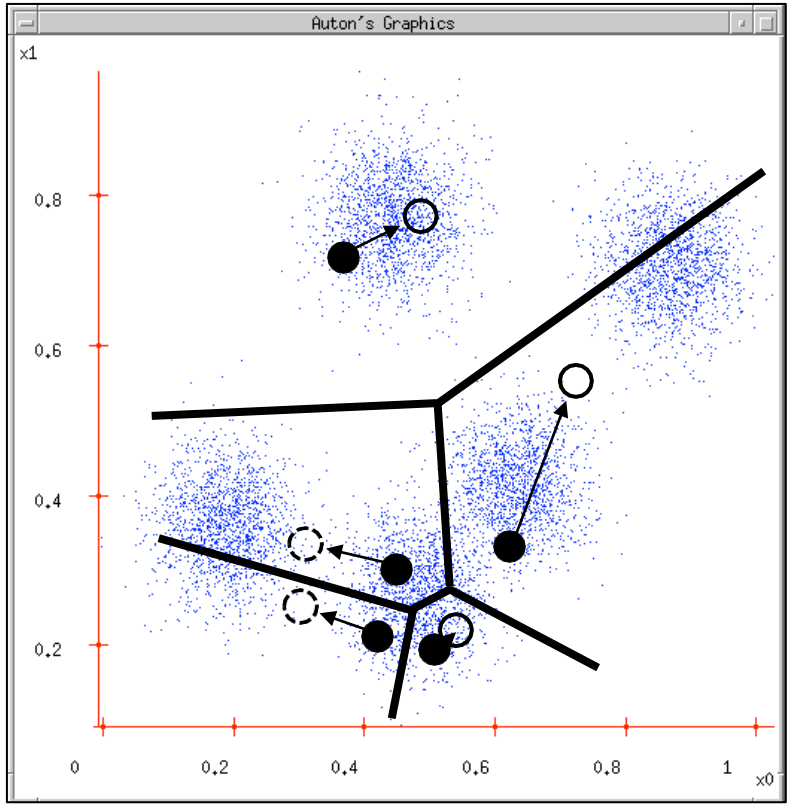
- Randomly choose k cluster center locations (centroids)
- Loop until convergence
  - Assign each point to the cluster of the closest centroid
  - Re-estimate the cluster centroids based on the data assigned to each cluster



K-Means Clustering

K-Means (k , X)

- Randomly choose k cluster center locations (centroids)
- Loop until convergence
  - Assign each point to the cluster of the closest centroid
    - Re-estimate the cluster centroids based on the data assigned to each cluster



K-Means Objective Function

- K-means finds a local optimum of the following objective function:

$$\arg\min_{\boldsymbol{\mathcal{S}}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in \mathcal{S}_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|_2^2$$

where $\boldsymbol{\mathcal{S}} = \{\mathcal{S}_1, \ldots, \mathcal{S}_k\}$ is a partitioning over $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ s.t. $X = \bigcup_{i=1}^{k} \mathcal{S}_i$ and $\boldsymbol{\mu}_i = \text{mean}(\mathcal{S}_i)$

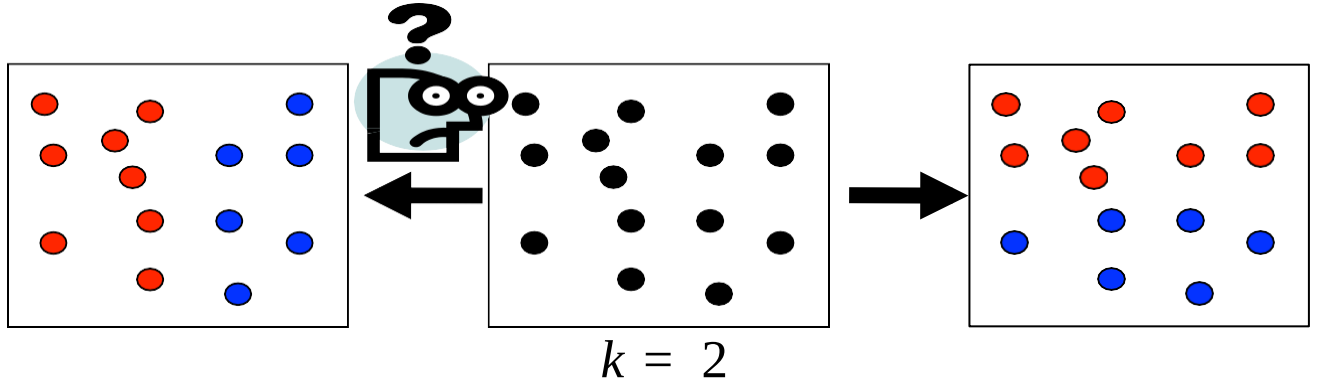
Problems with K–Means

- ***Very*** sensitive to the initial points
  - Do many runs of K–Means, each with different initial centroids
  - Seed the centroids using a better method than randomly choosing the centroids
    - e.g., Farthest--first sampling

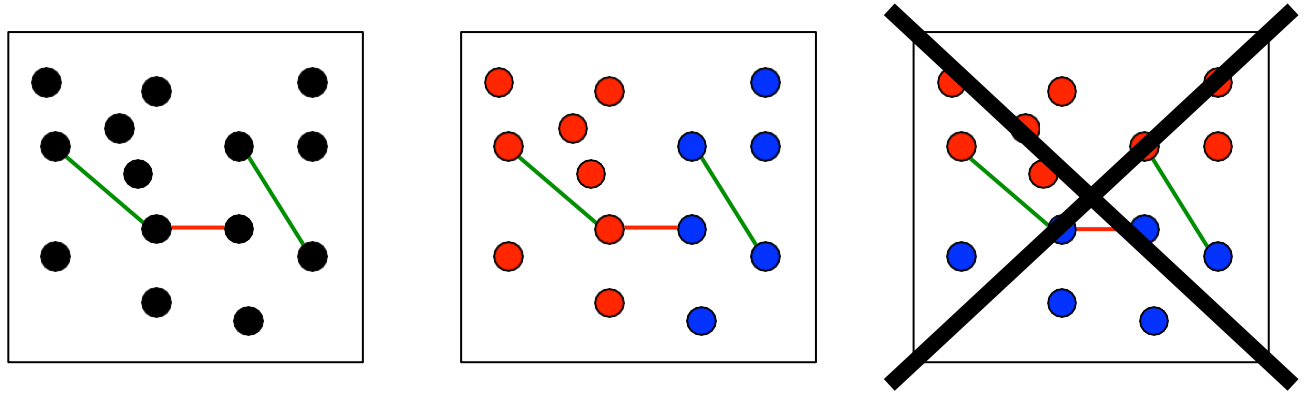
- Must manually choose k
  - Learn the optimal k for the clustering
    - Note that this requires a performance measure

Problems with K–Means

- How do you tell it which clustering you want?



Constrained clustering techniques (semi-supervised)



Same-cluster constraint (must–link)

Different--cluster constraint (cannot--link)