10-701 Introduction to Machine Learning

Machine Learning Department
School of Computer Science
Carnegie Mellon University

K-Means

Matt Gormley
Lecture 17
Mar. 20, 2019

K-MEANS

K-Means Outline

- **Clustering: Motivation / Applications**
- **Optimization Background**
  - Coordinate Descent
  - Block Coordinate Descent
- **Clustering**
  - Inputs and Outputs
  - Objective-based Clustering
- **K-Means**
  - K-Means Objective
  - Computational Complexity
  - K-Means Algorithm / Lloyd's Method
- **K-Means Initialization**
  - Random
  - Farthest Point
  - K-Means++

Clustering, Informal Goals

Goal: Automatically partition unlabeled data into groups of similar datapoints.

Question: When and why would we want to do this?

Useful for:

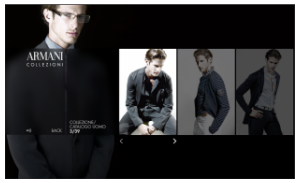
- Automatically organizing data.
- Understanding hidden structure in data.
- Preprocessing for further analysis.
    - Representing high-dimensional data in a low-dimensional space (e.g., for visualization purposes).

Slide courtesy of Nina Balcan

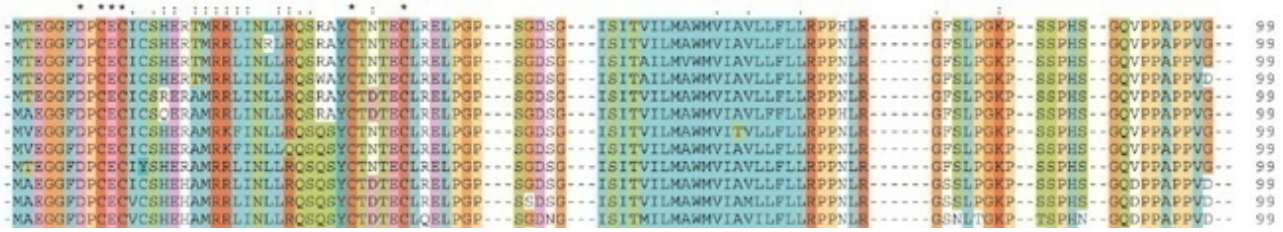
Applications (Clustering comes up everywhere…)

- Cluster news articles or web pages or search results by topic.

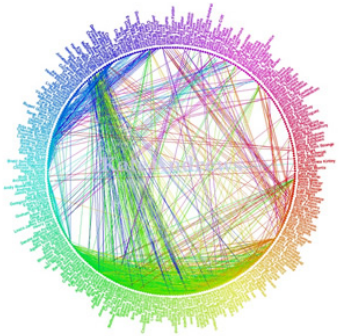




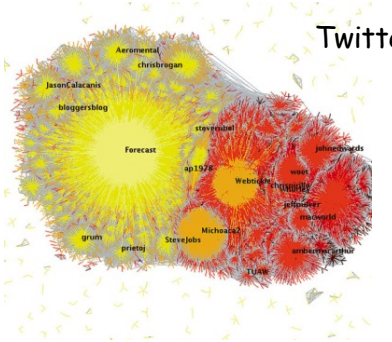
- Cluster protein sequences by function or genes according to expression profile.

- Cluster users of social networks by interest (community detection).



Facebook network



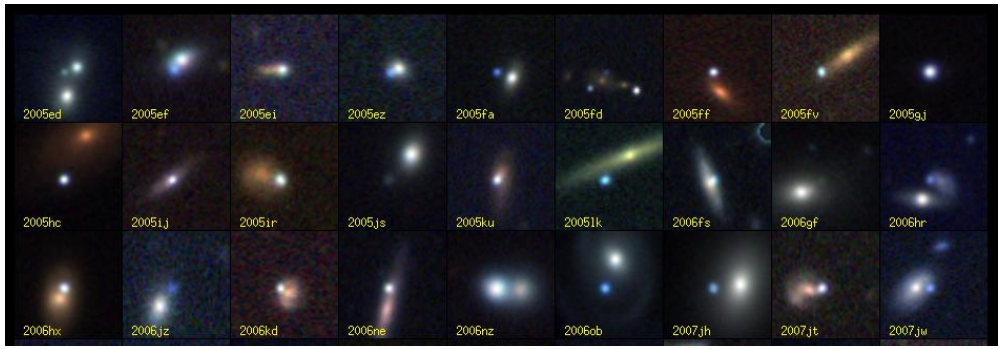
Twitter Network

Slide courtesy of Nina Balcan

Applications (Clustering comes up everywhere…)

- Cluster customers according to purchase history.

- Cluster galaxies or nearby stars (e.g. Sloan Digital Sky Survey)



- And many many more applications….

Slide courtesy of Nina Balcan

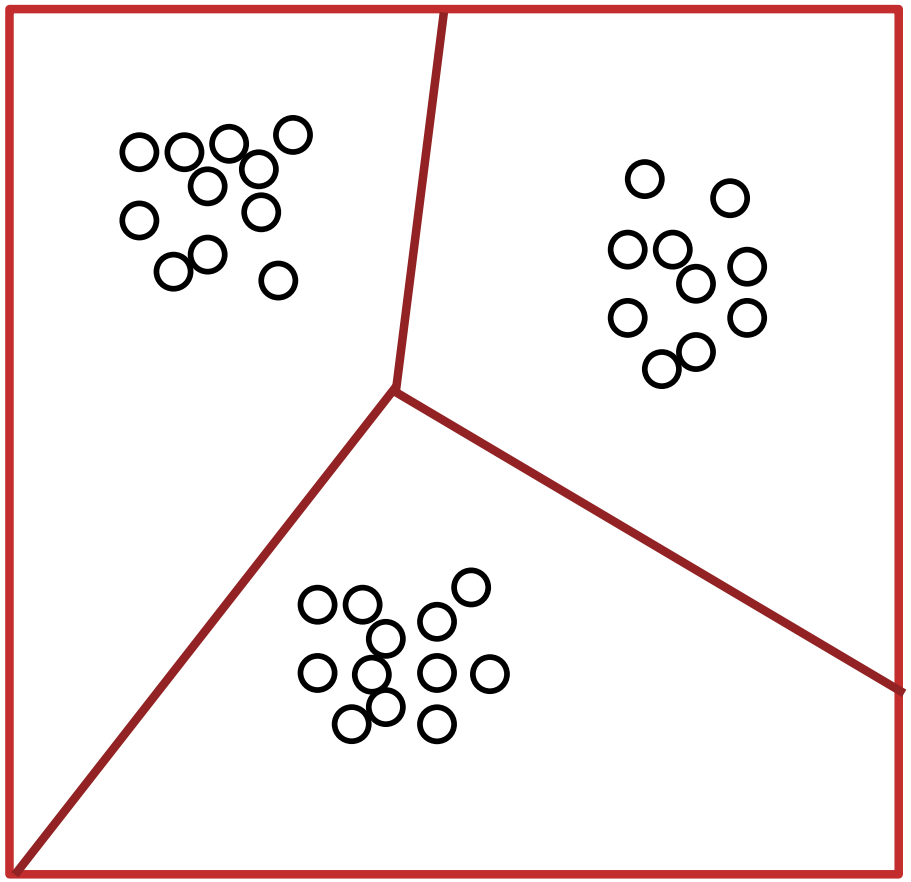
Optimization Background

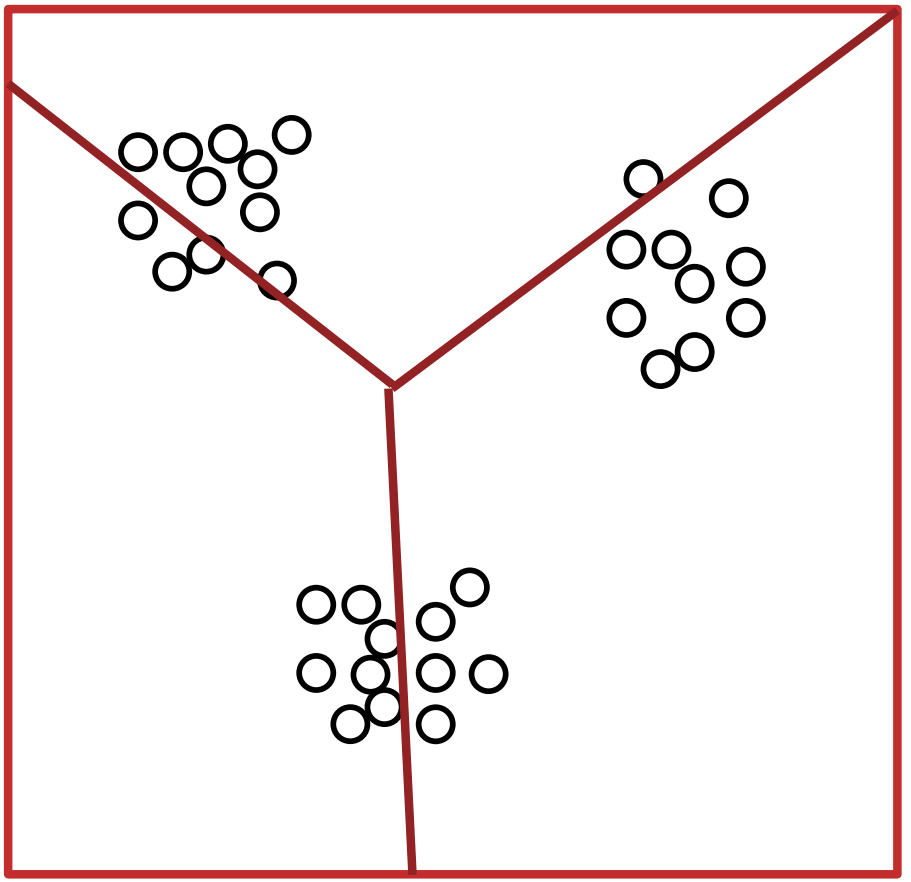
Whiteboard:

- Coordinate Descent
- Block Coordinate Descent

Clustering

Question: Which of these partitions is “better”?





Clustering

Whiteboard:

- Inputs and Outputs
- Objective-based Clustering

K-Means

Whiteboard:

- K-Means Objective
- Computational Complexity
- K-Means Algorithm / Lloyd's Method

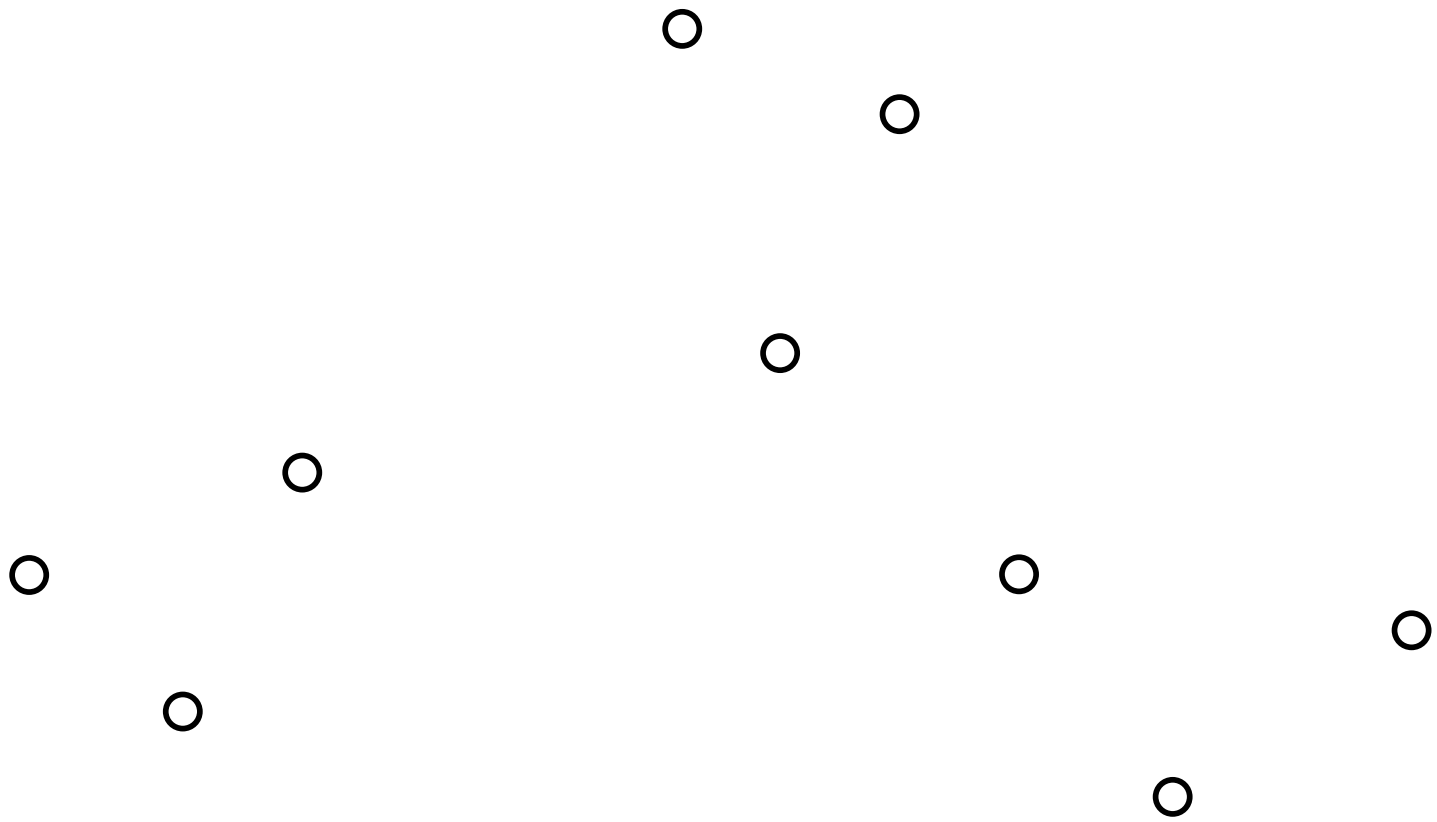
K-Means Initialization

Whiteboard:

- Random
- Furthest Traversal
- K-Means++

Lloyd's method: Random Initialization

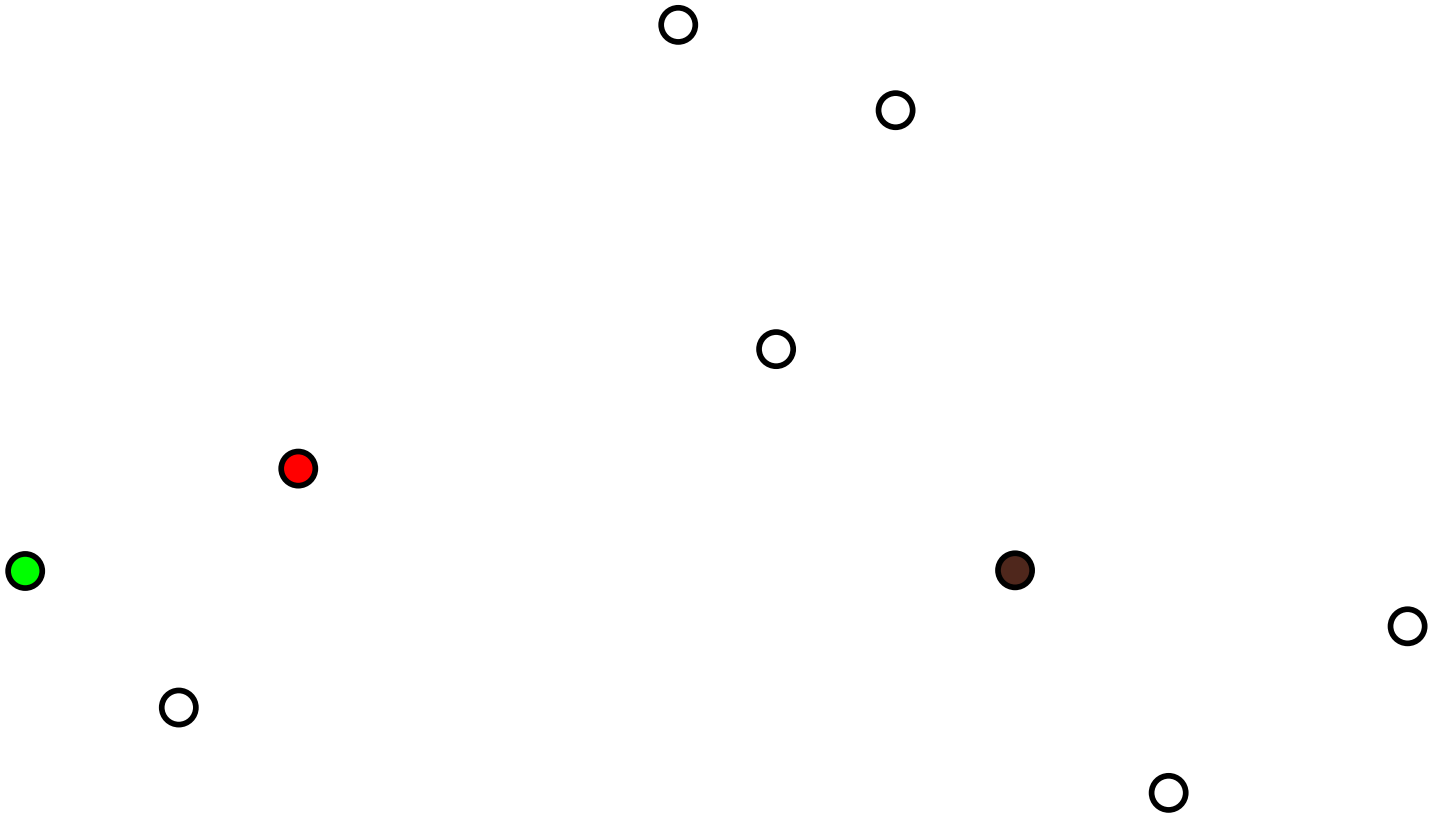
Example: Given a set of datapoints



Slide courtesy of Nina Balcan

Lloyd's method: Random Initialization

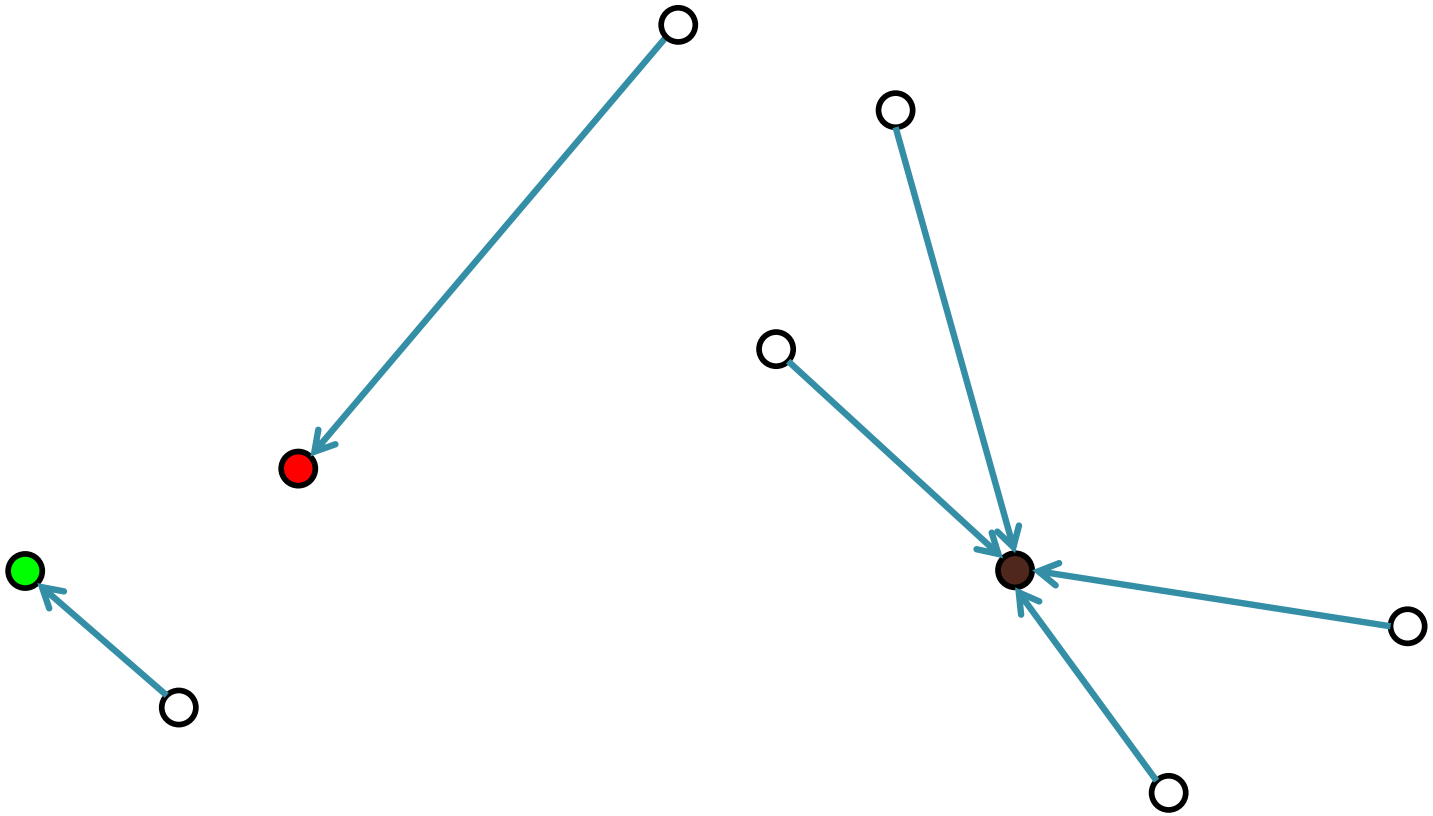
Select initial centers at random



Slide courtesy of Nina Balcan

Lloyd's method: Random Initialization

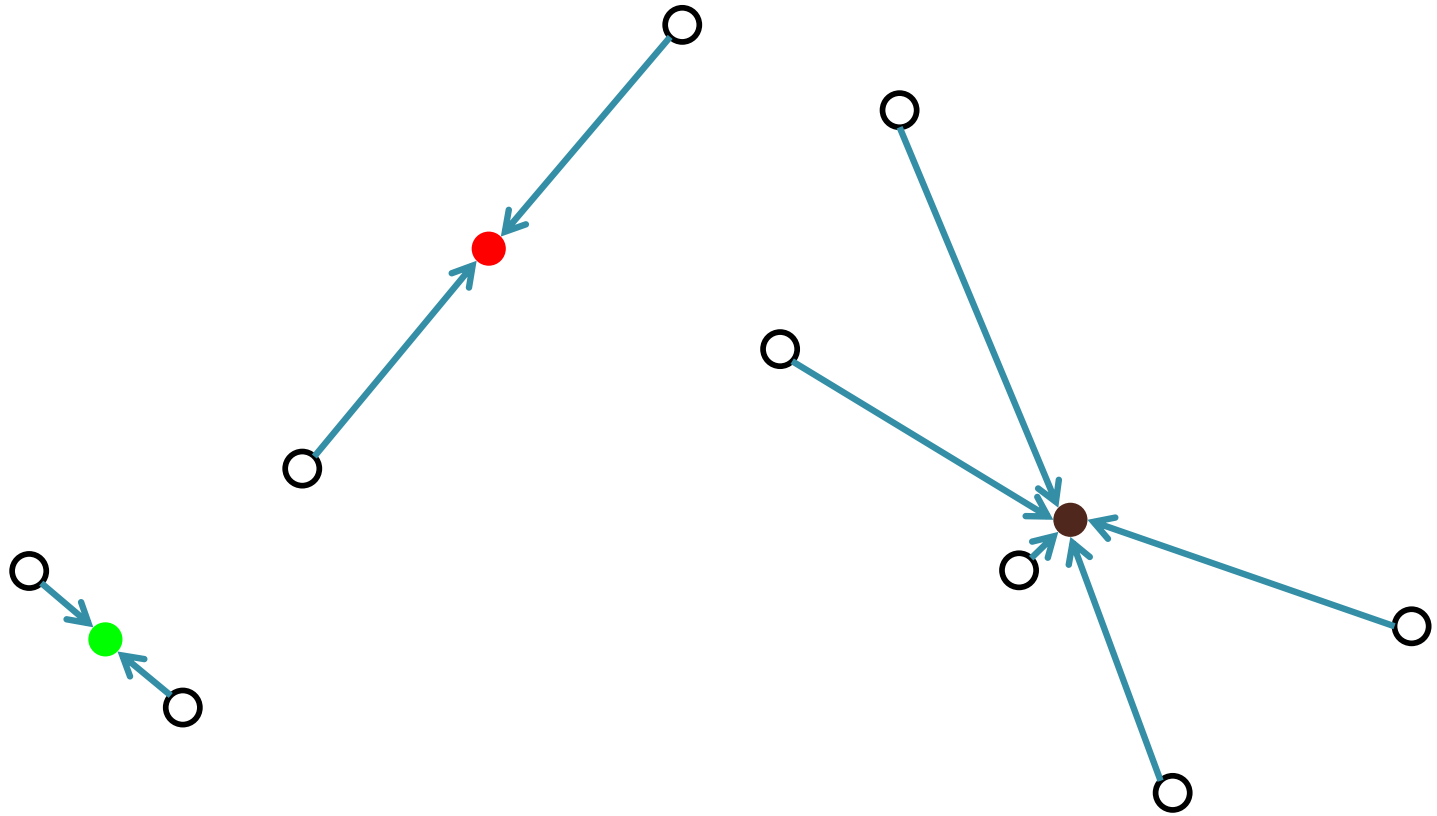
Assign each point to its nearest center



Slide courtesy of Nina Balcan

Lloyd's method: Random Initialization

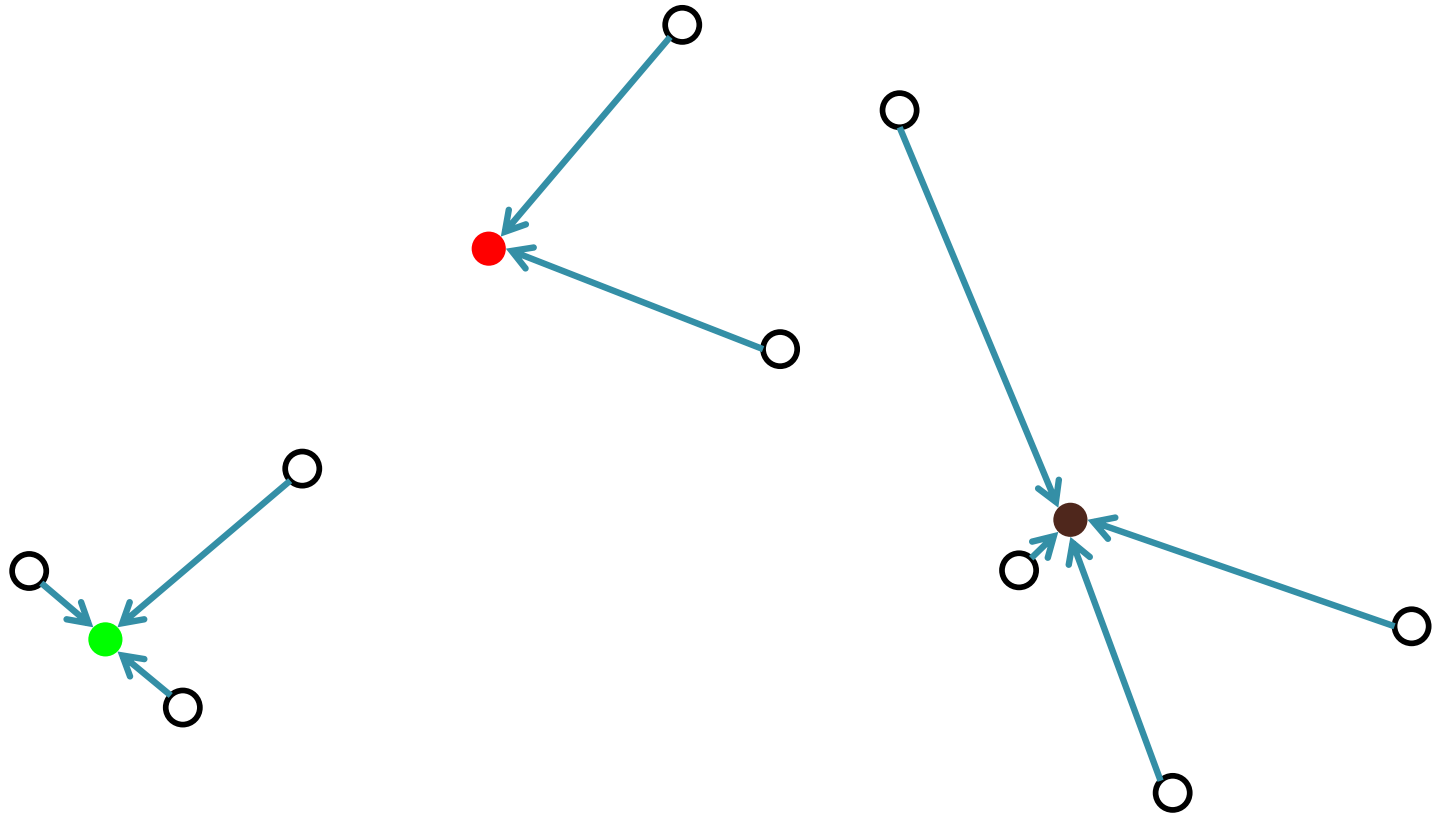
Recompute optimal centers given a fixed clustering



Slide courtesy of Nina Balcan

Lloyd's method: Random Initialization

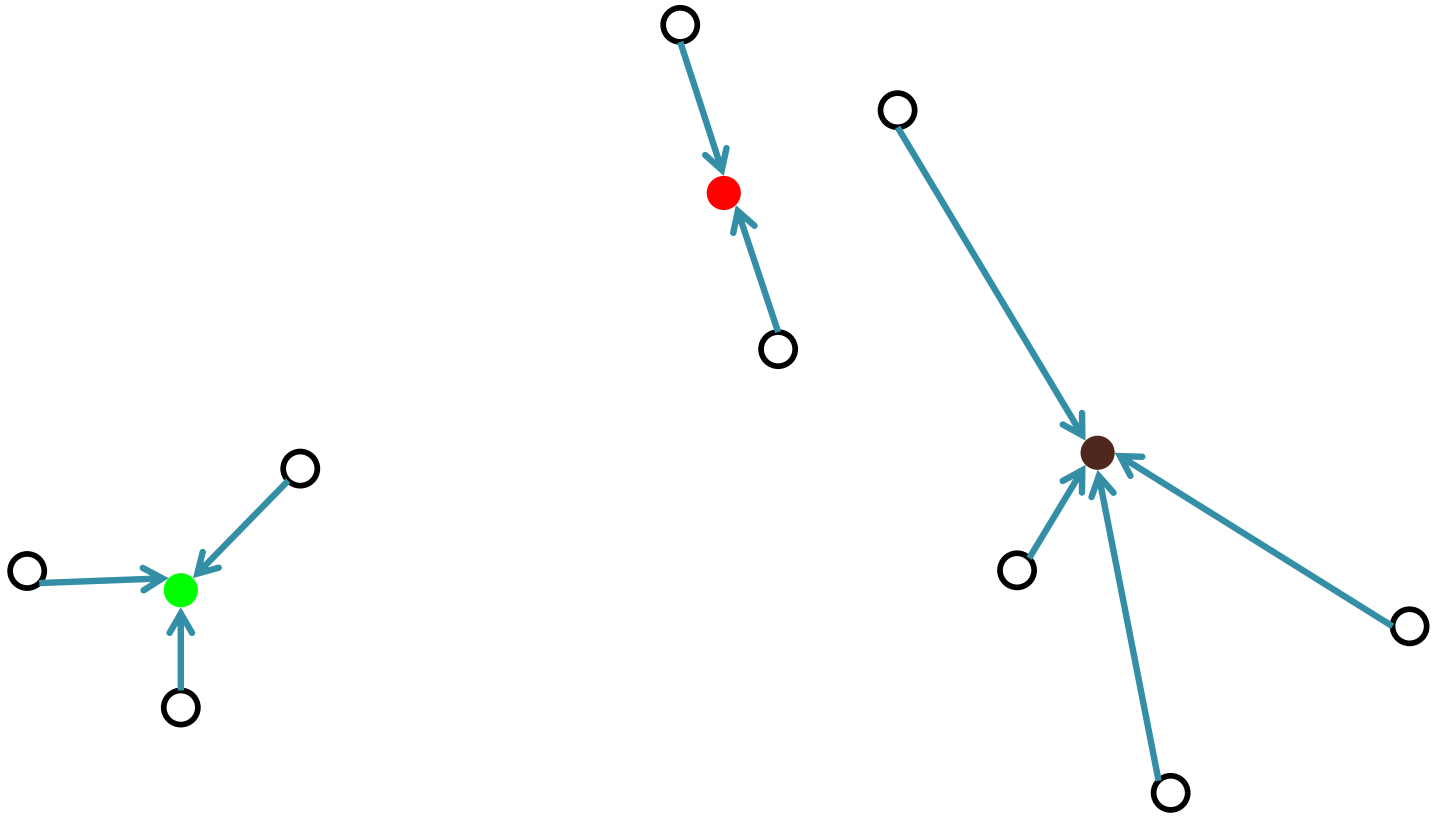
Assign each point to its nearest center



Slide courtesy of Nina Balcan

Lloyd's method: Random Initialization

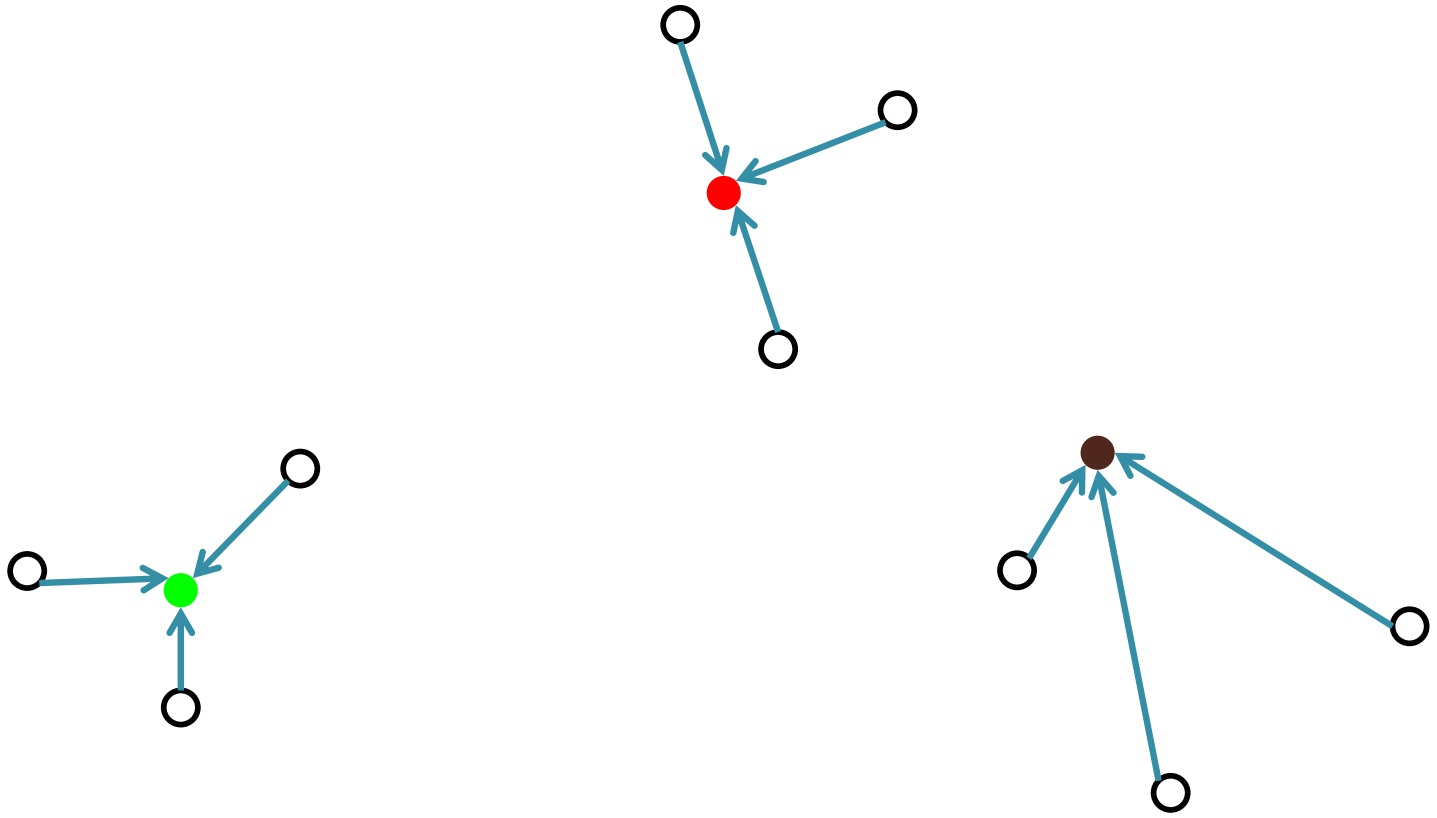
Recompute optimal centers given a fixed clustering



Slide courtesy of Nina Balcan

Lloyd's method: Random Initialization

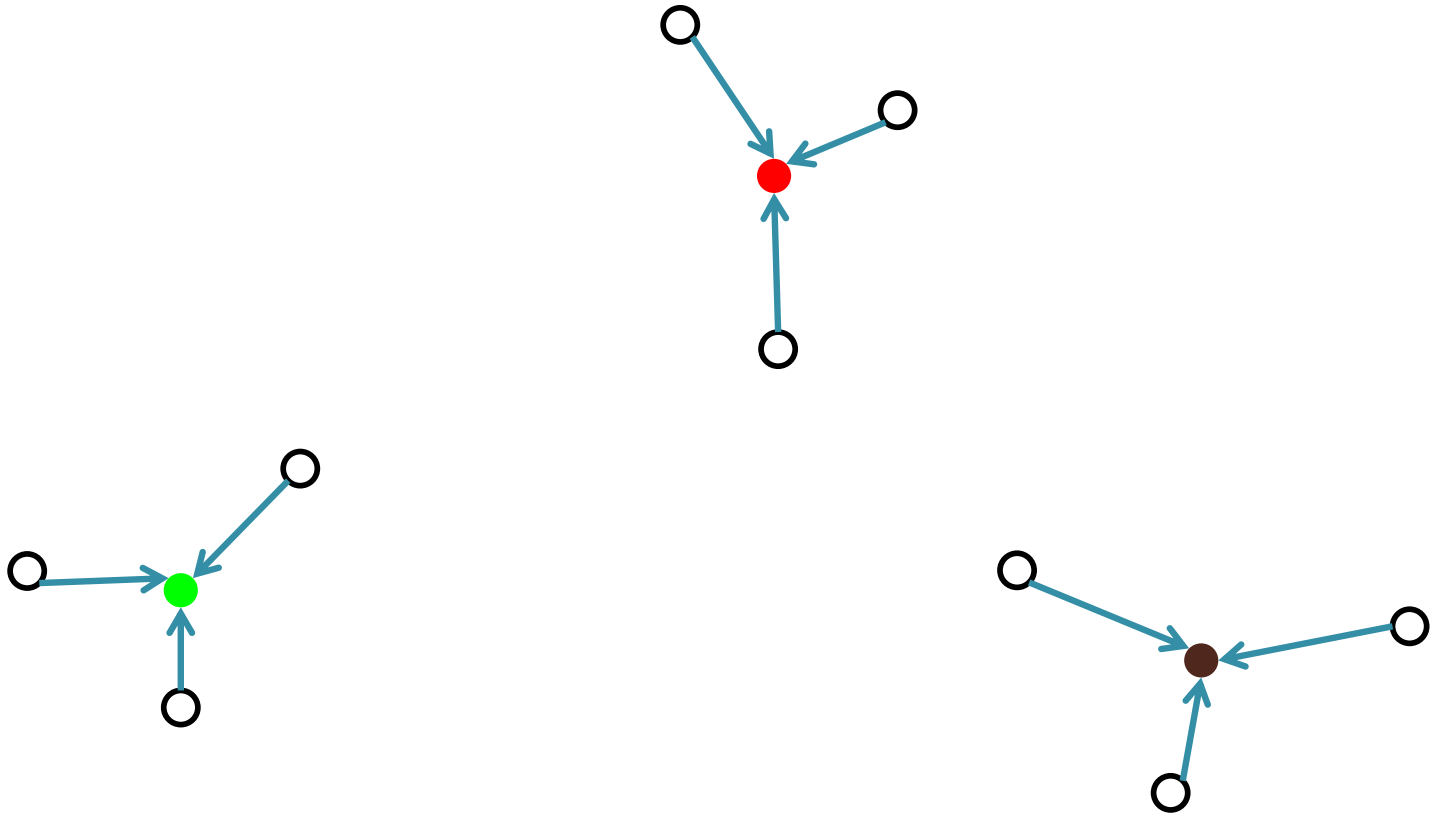
Assign each point to its nearest center



Slide courtesy of Nina Balcan

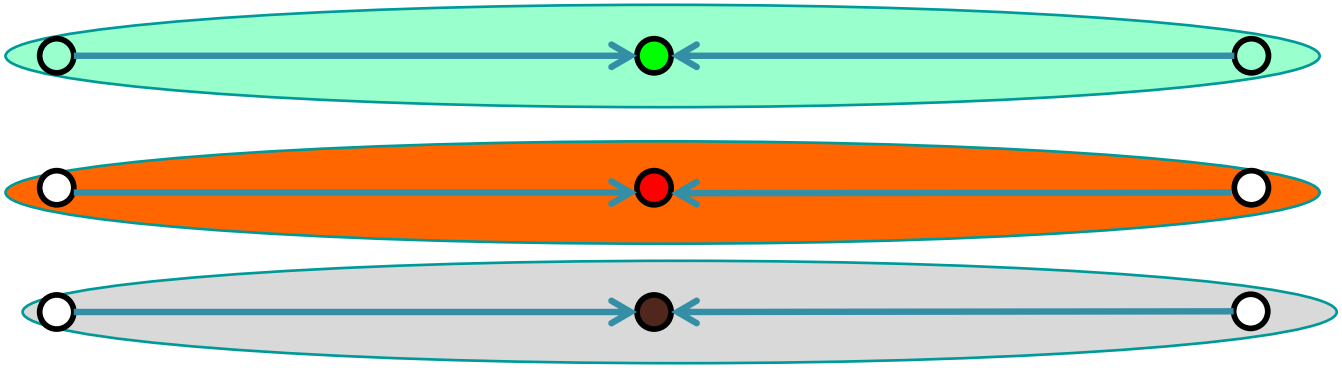
Lloyd's method: Random Initialization

Recompute optimal centers given a fixed clustering



Get a good quality solution in this example.

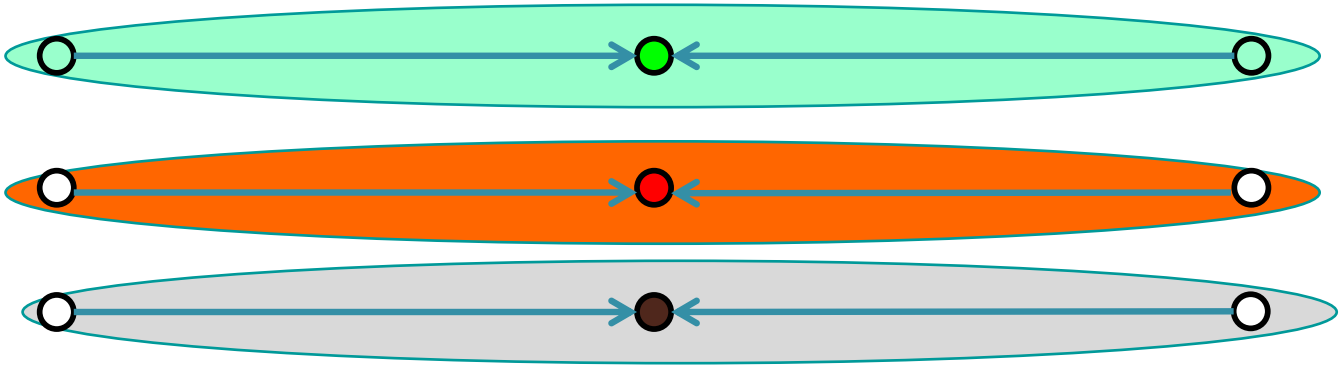
Lloyd's method: Performance



It always converges, but it may converge at a local optimum that is different from the global optimum, and in fact could be arbitrarily worse in terms of its score.

Slide courtesy of Nina Balcan

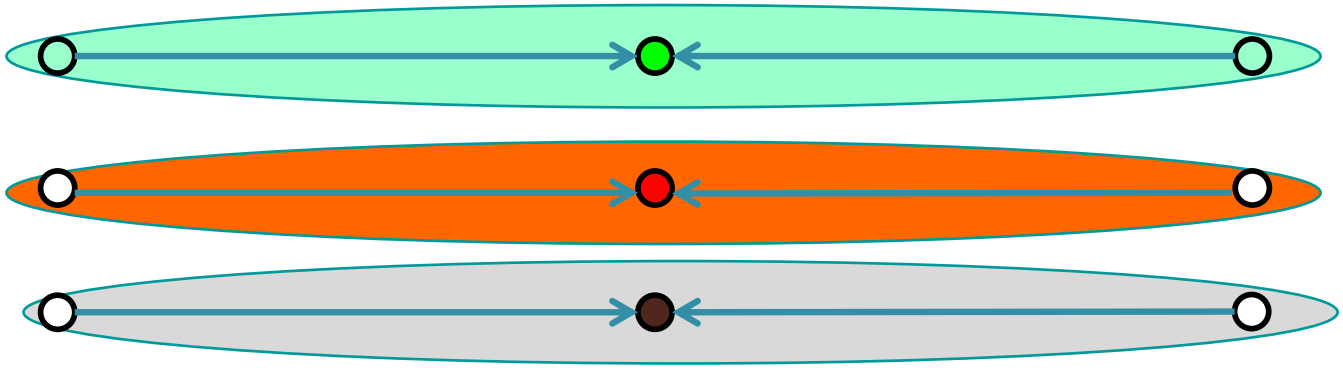
Lloyd's method: Performance



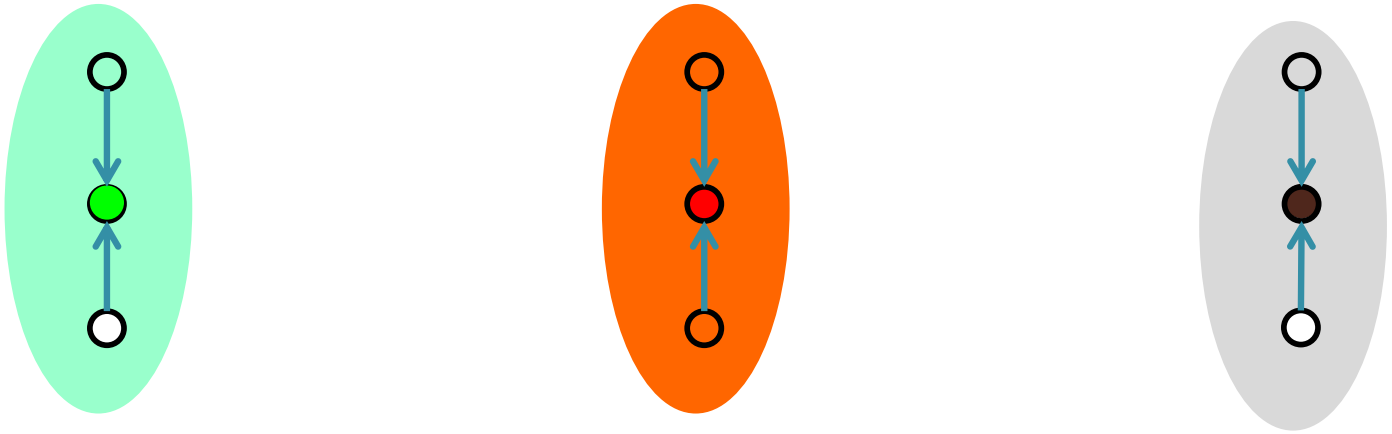
Local optimum: every point is assigned to its nearest center and every center is the mean value of its points.

Slide courtesy of Nina Balcan

Lloyd's method: Performance

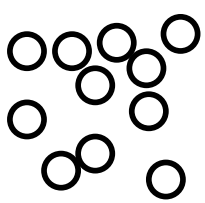


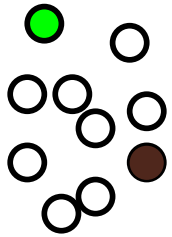
.It is arbitrarily worse than optimum solution….

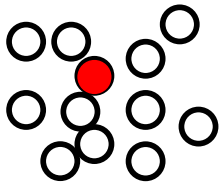


Slide courtesy of Nina Balcan

Lloyd's method: Performance



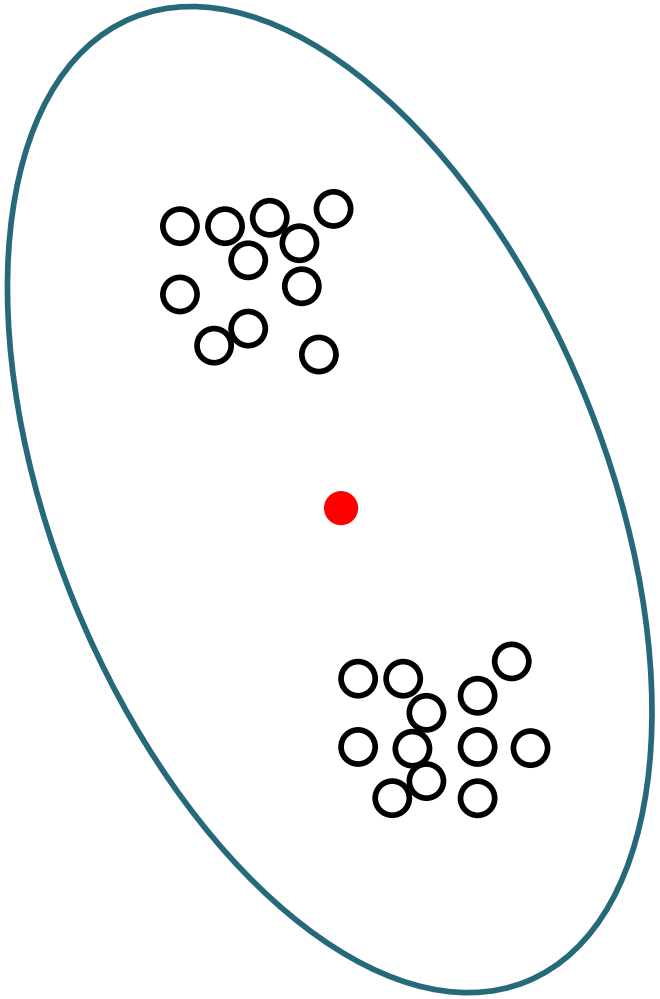


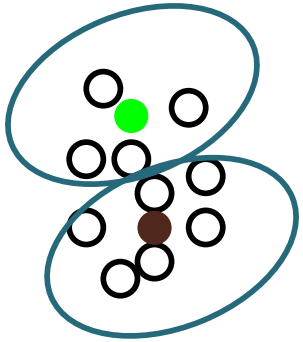


This bad performance, can happen even with well separated Gaussian clusters.

Slide courtesy of Nina Balcan

Lloyd's method: Performance





This bad performance, can happen even with well separated Gaussian clusters.

Some Gaussian are combined……

Slide courtesy of Nina Balcan

Learning Objectives

K-Means

You should be able to…

1. Distinguish between coordinate descent and block coordinate descent
2. Define an objective function that gives rise to a "good" clustering
3. Apply block coordinate descent to an objective function preferring each point to be close to its nearest objective function to obtain the K-Means algorithm
4. Implement the K-Means algorithm
5. Connect the nonconvexity of the K-Means objective function with the (possibly) poor performance of random initialization