K-means

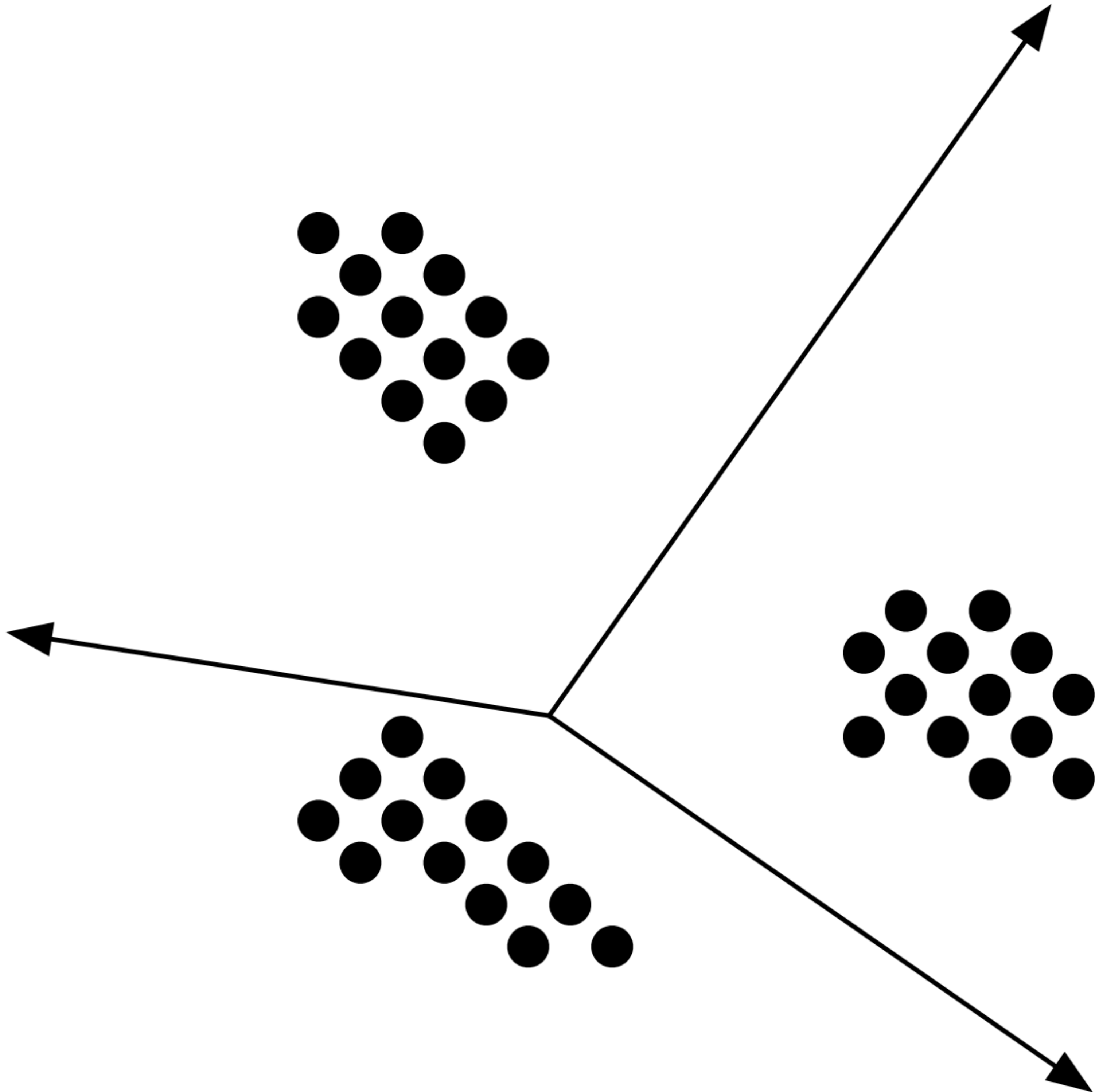
Sewoong Oh

CSE446
University of Washington

Unsupervised learning and K-means

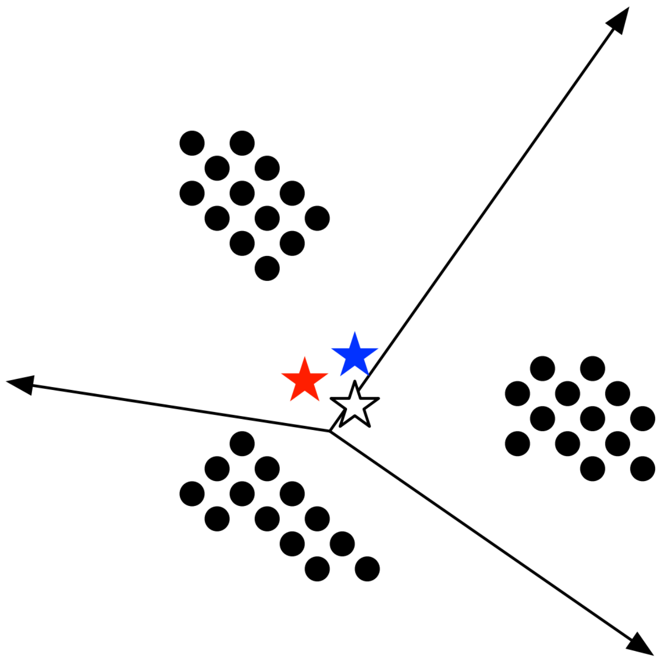
- in unsupervised learning, the dataset consists of only inputs $\{x_i\}_{i=1}^{n}$
- and we **do not have any labels**
- objective: find some pattern
- in particular, most popular unsupervised learning task is clustering: cluster the data into K groups

K-means: an iterative algorithm for clustering



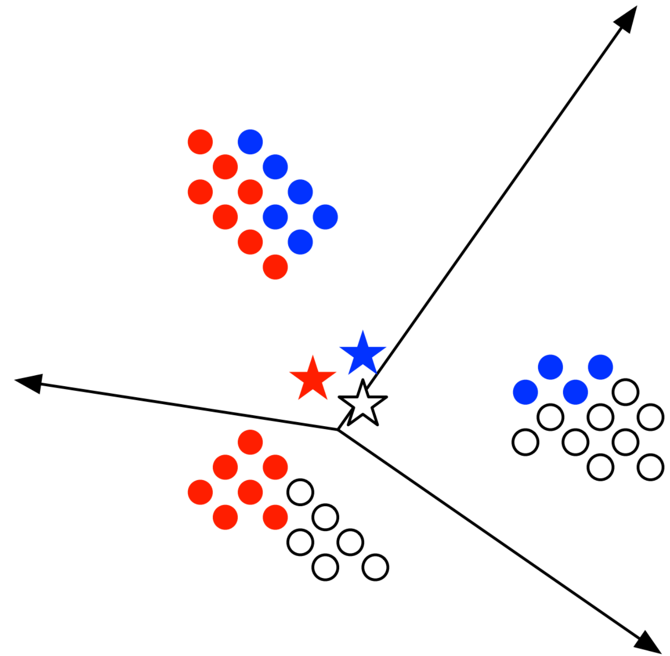
demo: https://www.naftaliharris.com/blog/visualizing-k-means-clustering/

K-means: an iterative algorithm for clustering



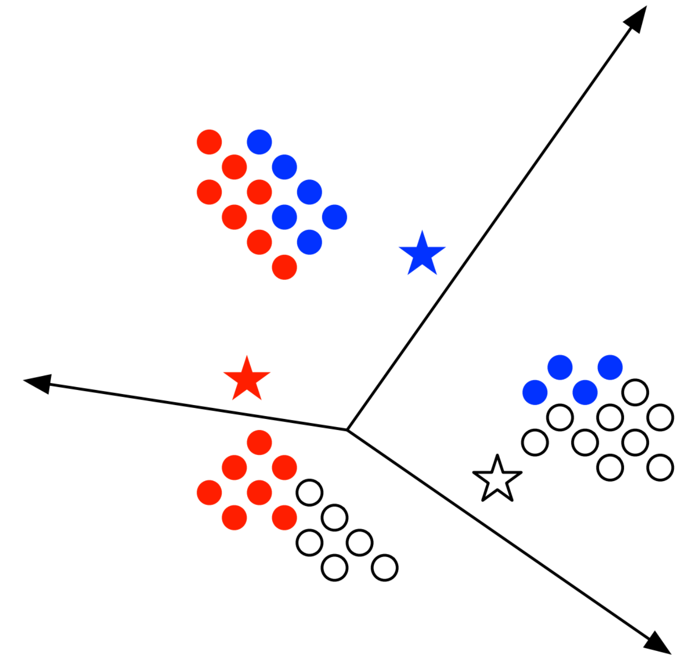
step 1

start with K centers at arbitrary positions



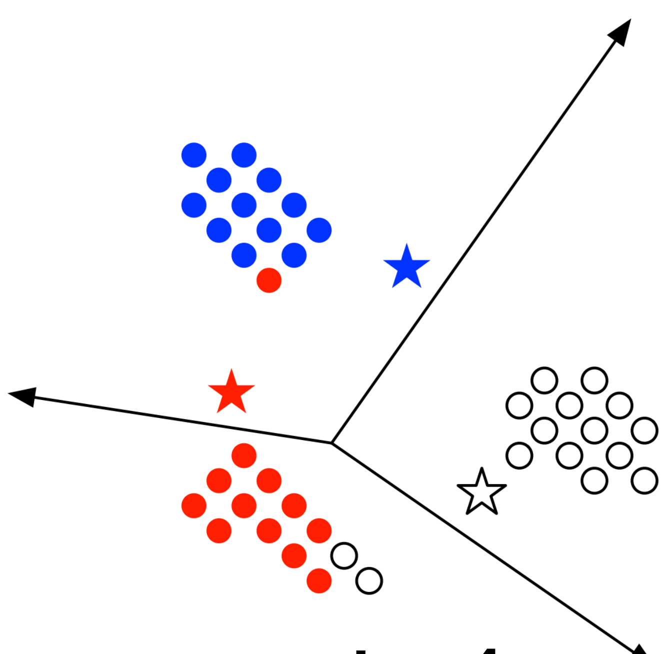
step 2

assign each point to its closest center

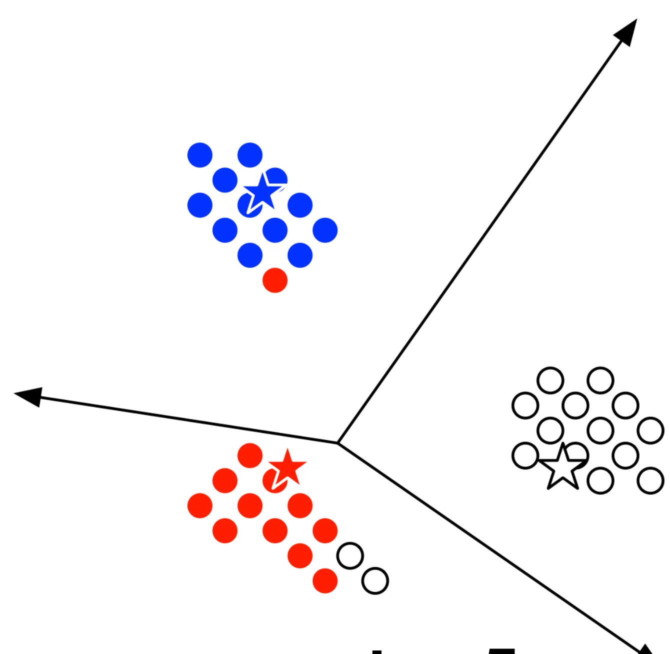


step 3

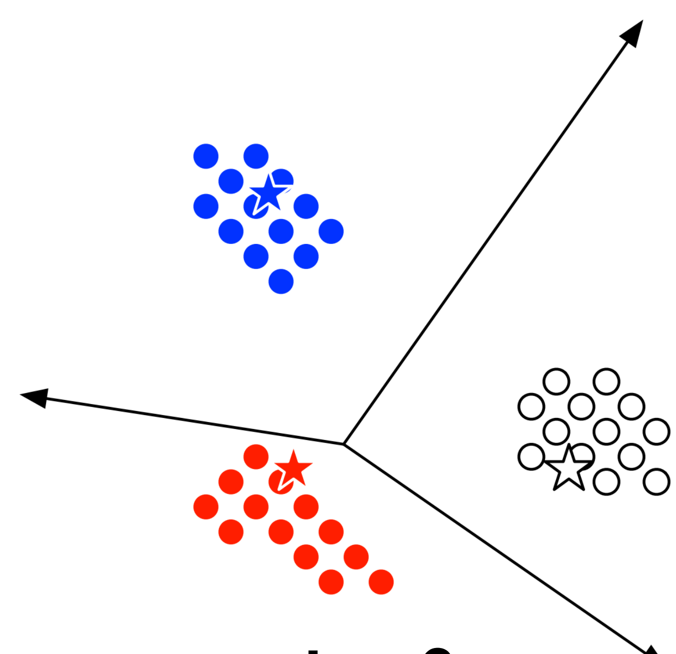
recalculate the centers each as the center of its cluster



step 4



step 5



step 6

K-means: an iterative algorithm for clustering

- in this example K-means converged, i.e. it does not change after this point
- will it always converge? Yes
- does it converge to the right answer (whatever that means)? no

- K-means algorithm
  - input: dataset $\{x_i\}_{i=1}^{n}$, number of clusters K
  - output: cluster assignment z_i for each data point x_i
  - Initialize each center μ_k to a random location for $k \in \{1,\ldots,K\}$
  - repeat
    - (assign each point to its nearest cluster-center)
      $z_i = \arg\min_k \|x_i - \mu_k\|_2$ for all $i \in \{1,\ldots,n\}$
    - (recenter each cluster-center)
      - for $k \in \{1,\ldots,K\}$
        $$X_k \leftarrow \{x_i \,|\, z_i = k\}$$
        $$\mu_k \leftarrow \text{mean}(X_k)$$
  - while any μ_k changed from previous value

What would we like to do?

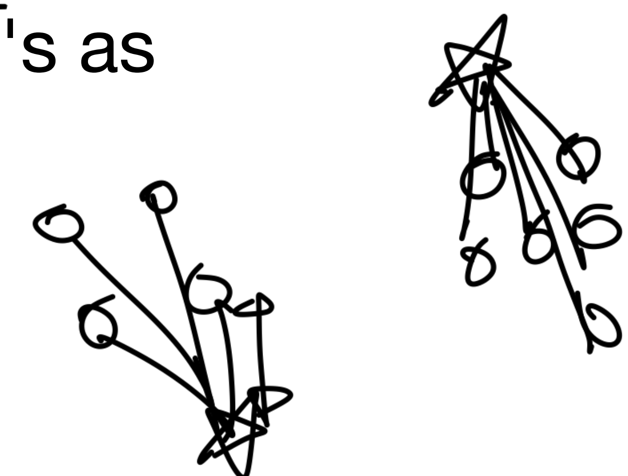
- K-means algorithm is trying to minimize the following objective function

$$\text{minimize}_{\mu_1,\ldots,\mu_K} \sum_{i=1}^{n} \left\{ \min_{k' \in \{1,\ldots,K\}} \|x_i - \mu_{k'}\|_2^2 \right\}$$

- this can be written in terms of the assignments z_i's as

$$\text{minimize}_{\mu_1,\ldots,\mu_K, z_1,\ldots,z_n} \boxed{\sum_{i=1}^{n} \|x_i - \mu_{z_i}\|_2^2}$$

objective function



- as K-means is alternatively minimizing
  - (the assignment step)
    fix μ_k's and minimize the objective over z_i's
  - (the re-centering step)
    fix z_i's and minimize the objective over μ_k's

- in particular, the objective can only decrease at each step of K-means

Proof of convergence

$z_i \in \{1, \dots, K\}$

- there is only a finite set of values that $\boxed{\{z_i\}_{i=1}^{n}}$ can take (K^n is large but finite)

K^n choices

- so there is only finite, K^n at most, values for cluster-centers also
- each time we update them, we will never increase the objective function $\mathscr{L}(z_1, \dots, z_n, \mu_i, \dots, \mu_K) = \sum_{i=1}^{n} \|x_i - \mu_{z_i}\|_2^2$
- the objective is lower bounded by zero
- after at most K^n steps, the algorithm must converge (as the assignments z_i's cannot return to previous assignments in the course of K-means iterations)

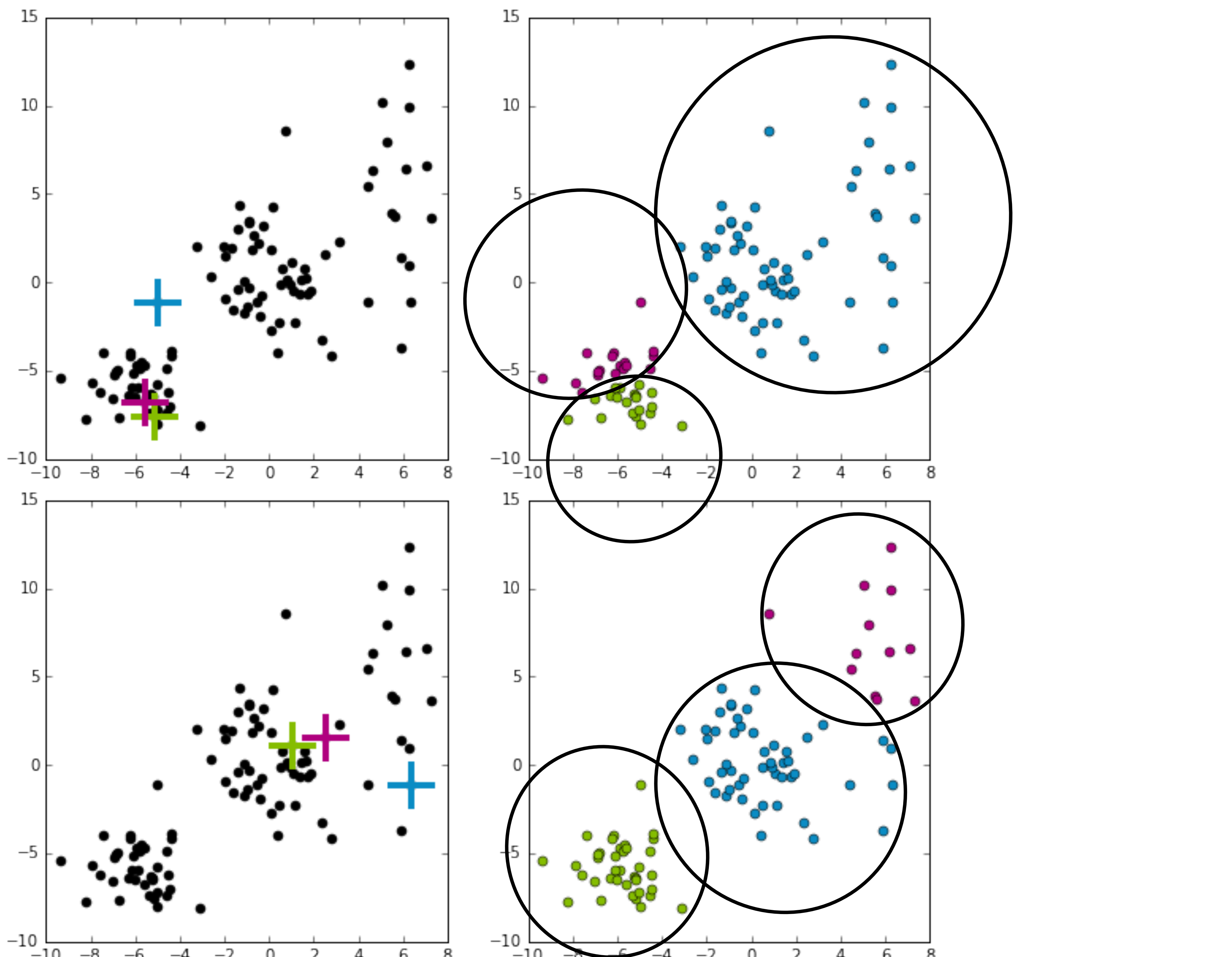
downsides of K-means

- it requires the number of clusters K to be specified by us
- the final solution depends on the initialization
  (does not find global minimum of the objective)

Initial position of centers　　final converged assignment

Trial 1

Trial 2



K-means++: a smart initialization

Smart initialization:

1. Choose first cluster center uniformly at random from data points
2. Repeat ***K-1*** times
    3. For each data point $\boldsymbol{x_i}$, compute distance $\boldsymbol{d_i}$ to nearest cluster center
    4. Choose new cluster center from amongst data points, with

probability of $\boldsymbol{x_i}$ being chosen proportional to $\boldsymbol{(d_i)^2}$

- apply standard K-means after the initialization

$$x_1 = \mu_1$$

$$\mu_2 = \begin{cases} x_2 \\ \vdots \\ x_n \end{cases} \quad \text{w-p} \quad \frac{\|\mu_1 - x_2\|^2}{\sum_{i=2}^{n} \|\mu_1 - x_i\|^2}$$