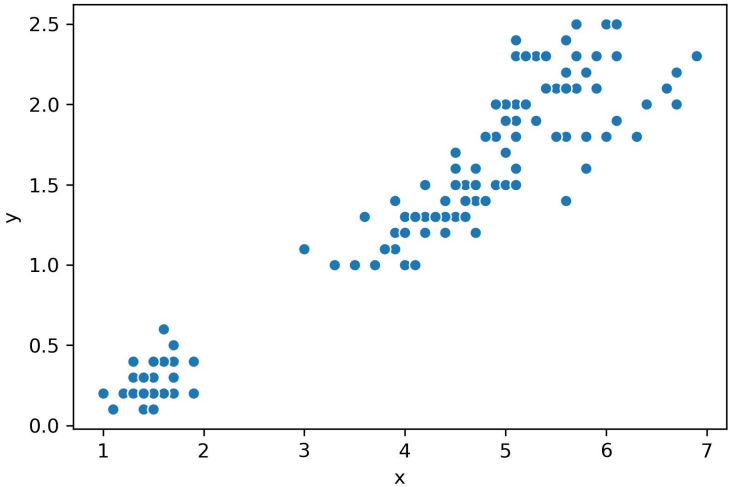
K-Means Clustering Algorithm

K-Means Clustering

Most popular clustering approach: K-Means

- Pick an arbitrary k, and randomly place k “centers”, each a different color
- Repeat until convergence:
  - Color points according to the closest center
  - Move center for each color to center of points with that color

Raw data →

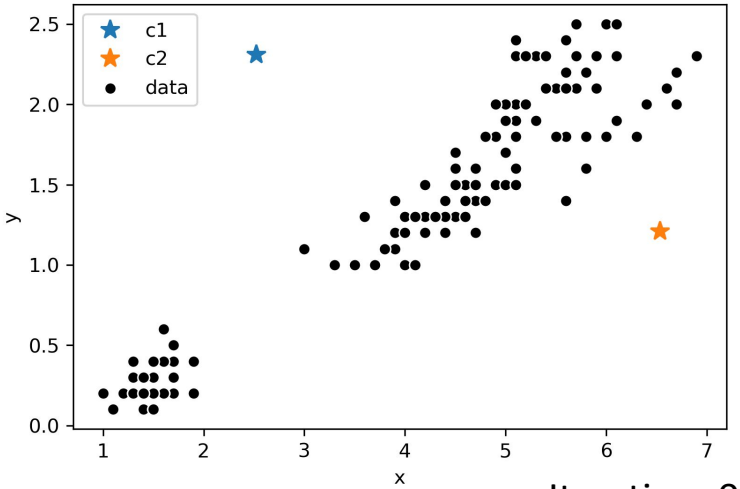


K-Means Clustering

Most popular clustering approach: K-Means.

- Pick an arbitrary k, and **randomly place k "centers",** each a different color
- Repeat until convergence:
  - Color points according to the closest center
  - Move center for each color to center of points with that color

Initial random placement of two centers



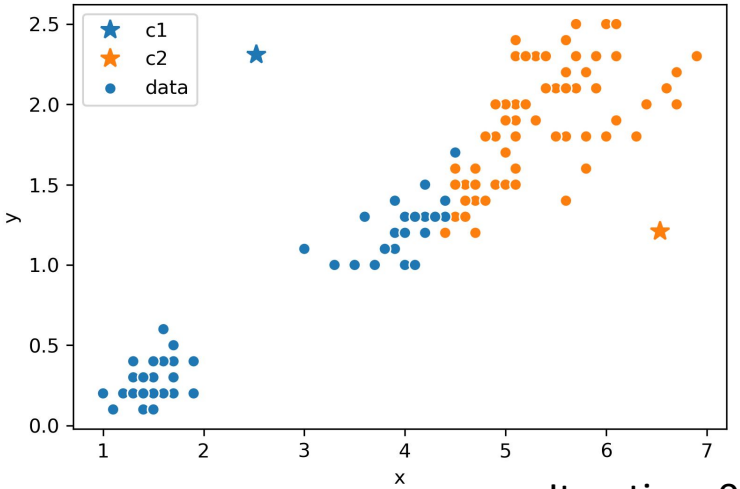
Iteration: 0

K-Means Clustering

Most popular clustering approach: K-Means

- Pick an arbitrary k, and randomly place k “centers”, each a different color
- Repeat until convergence:
  - **Color points according to the closest center**
  - Move center for each color to center of points with that color

Data colored by closest center



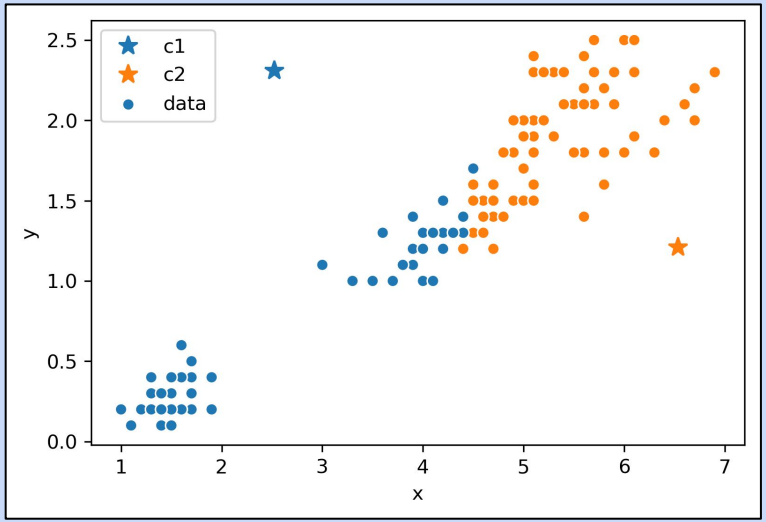
Iteration: 0

K-Means Clustering

Most popular clustering approach: K-Means

- Pick an arbitrary k, and randomly place k “centers”, each a different color
- Repeat until convergence:
  - Color points according to the closest center
  - **Move center for each color to center of points with that color**

Where should the centers go next?

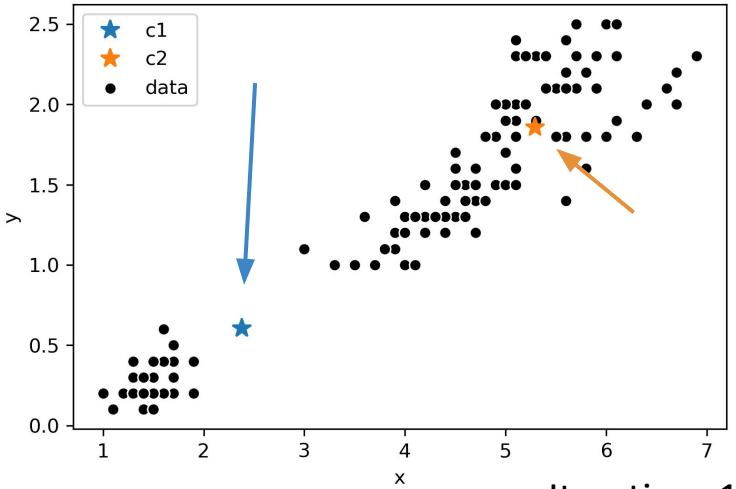


K-Means Clustering

Most popular clustering approach: K-Means

- Pick an arbitrary k, and randomly place k “centers”, each a different color
- Repeat until convergence:
  - Color points according to the closest center
  - **Move center for each color to center of points with that color**

Centers moved to their new homes



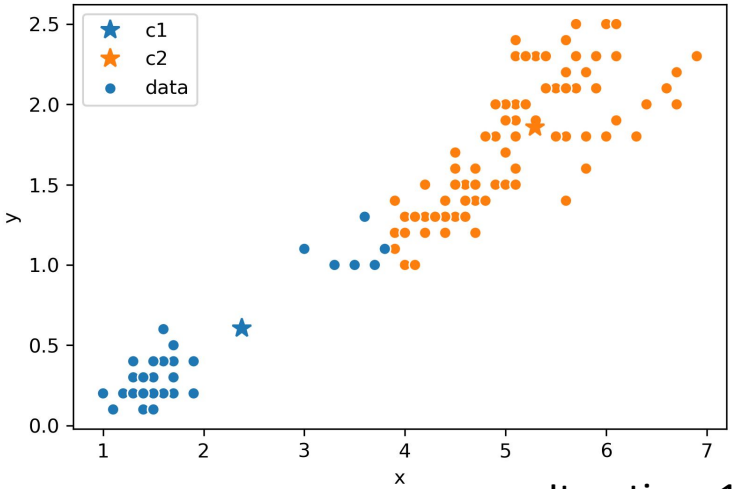
Iteration: 1

K-Means Clustering

Most popular clustering approach: K-Means

- Pick an arbitrary k, and randomly place k “centers”, each a different color
- Repeat until convergence:
  - **Color points according to the closest center**
  - Move center for each color to center of points with that color

Data colored by closest center (in new position)



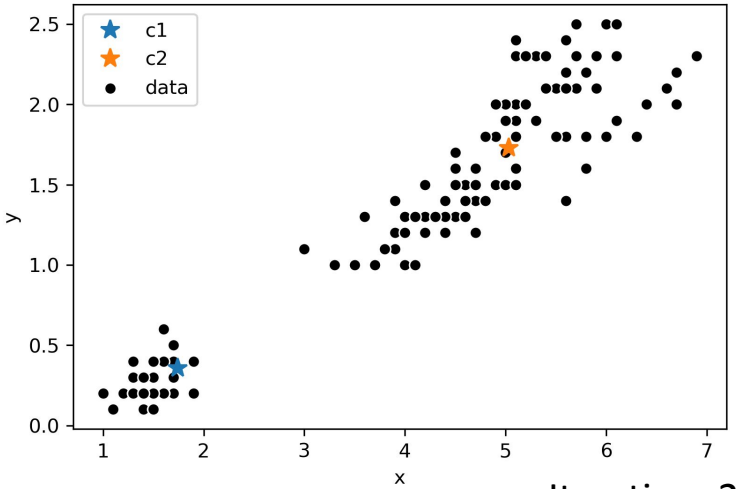
Iteration: 1

K-Means Clustering

Most popular clustering approach: K-Means

- Pick an arbitrary k, and randomly place k “centers”, each a different color
- Repeat until convergence:
  - Color points according to the closest center
  - **Move center for each color to center of points with that color**

Centers moved to new position



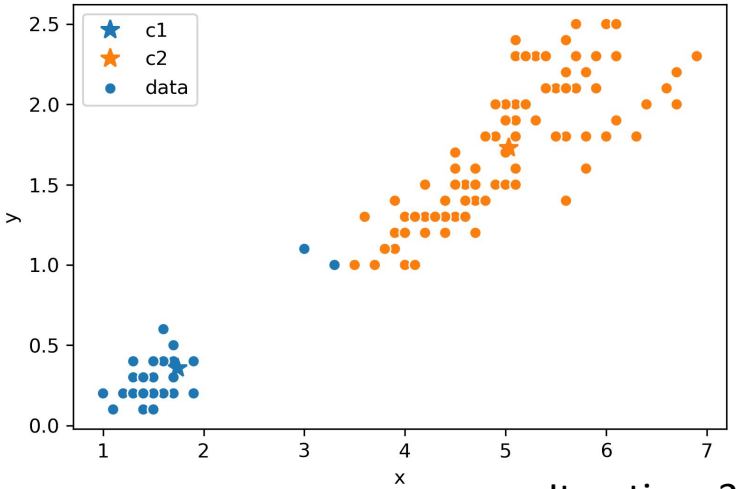
Iteration: 2

K-Means Clustering

Most popular clustering approach: K-Means.

- Pick an arbitrary k, and randomly place k “centers”, each a different color.
- Repeat until convergence:
  - **Color points according to the closest center.**
  - Move center for each color to center of points with that color

Data colored by closest center (in new position)



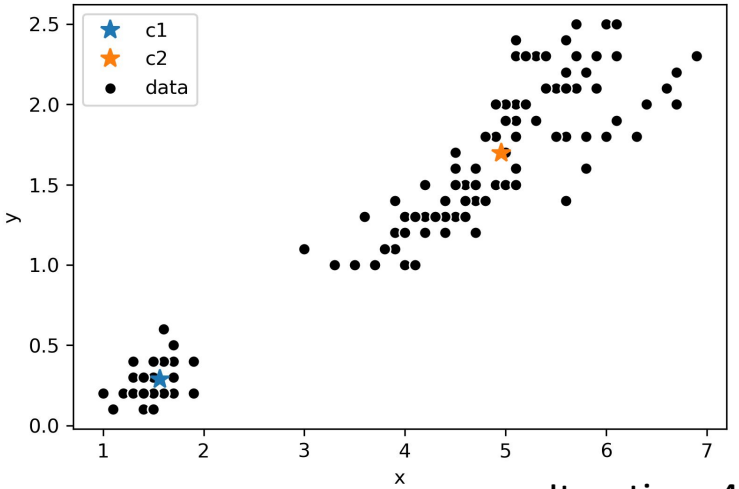
Iteration: 3

K-Means Clustering

Most popular clustering approach: K-Means

- Pick an arbitrary k, and randomly place k “centers”, each a different color
- Repeat until convergence:
  - Color points according to the closest center
  - **Move center for each color to center of points with that color**

Centers moved to new position



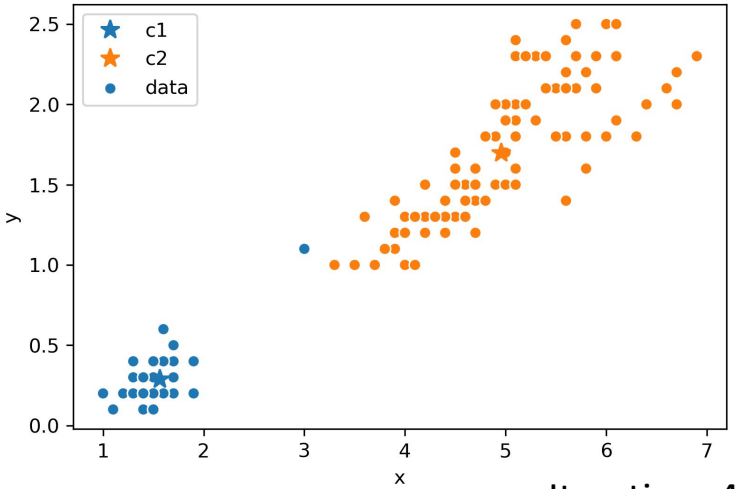
Iteration: 4

K-Means Clustering

Most popular clustering approach: K-Means

- Pick an arbitrary k, and randomly place k “centers”, each a different color
- Repeat until convergence:
  - **Color points according to the closest center**
  - Move center for each color to center of points with that color

Data colored by closest center (in new position)



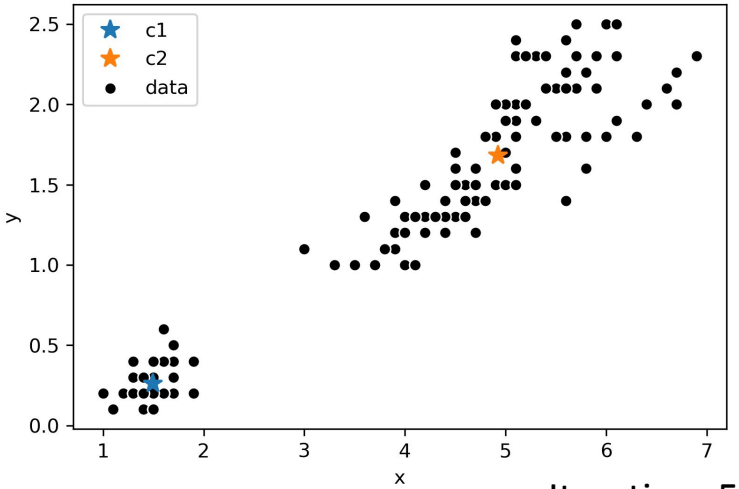
Iteration: 4

K-Means Clustering

Most popular clustering approach: K-Means

- Pick an arbitrary k, and randomly place k “centers”, each a different color
- Repeat until convergence:
  - Color points according to the closest center
  - **Move center for each color to center of points with that color**

Centers moved to new position



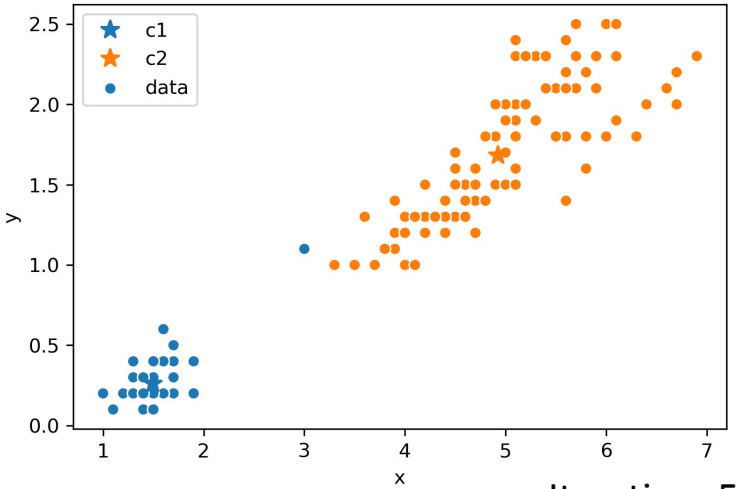
Iteration: 5

K-Means Clustering

Most popular clustering approach: K-Means

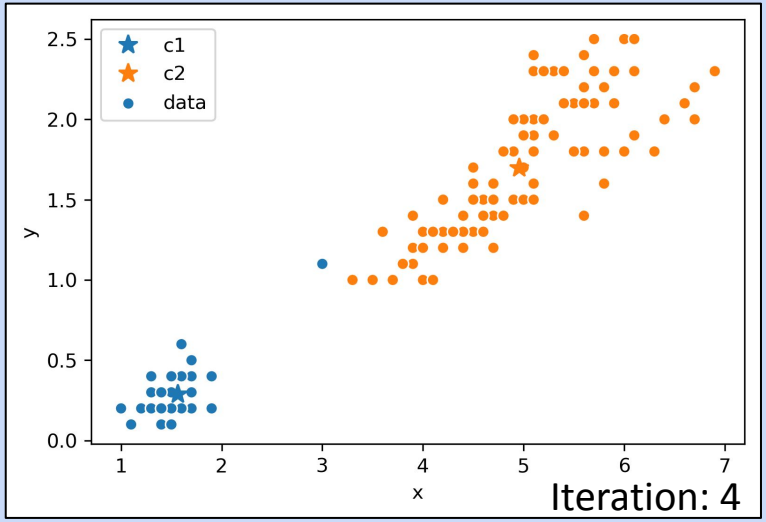
- Pick an arbitrary k, and randomly place k “centers”, each a different color
- Repeat until convergence:
  - **Color points according to the closest center**
  - Move center for each color to center of points with that color

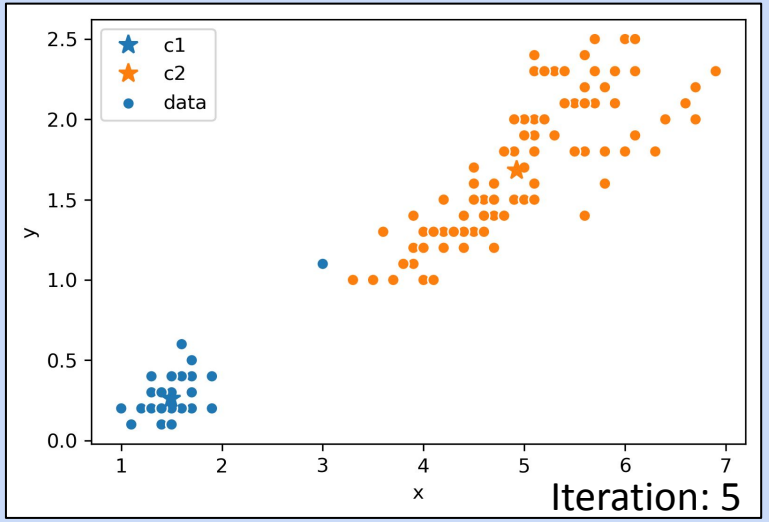
Data colored by closest center (in new position)



Iteration: 5

K-Means Clustering

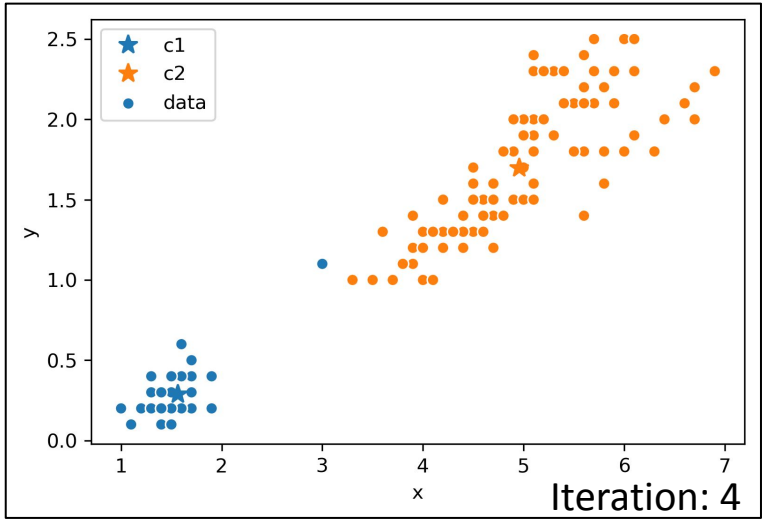


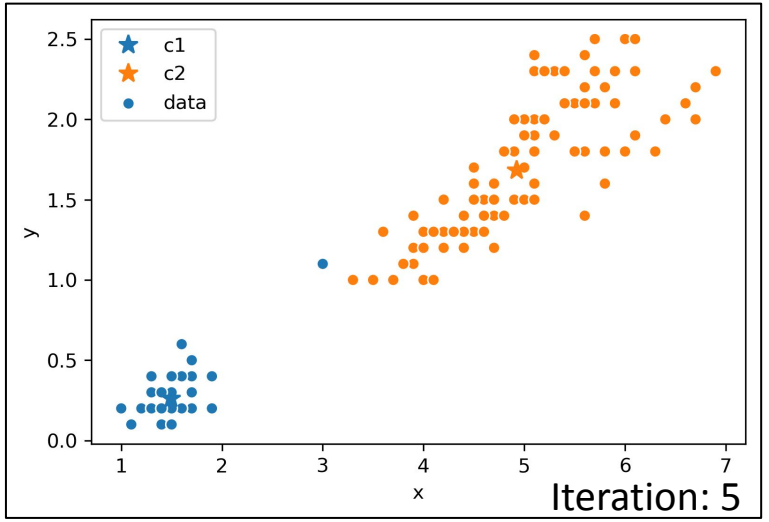


Above we see the results after iteration 4 and 5

- Centers moved slightly between iteration 4 and 5
- But no points changed color
- Are we done?

K-Means Clustering





Above we see the results after iteration 4 and 5

- Centers moved slightly between iteration 4 and 5
- But no points changed color
- Are we done?
  - Yes! If we tried iteration 6, we'd see that centers don't move at all