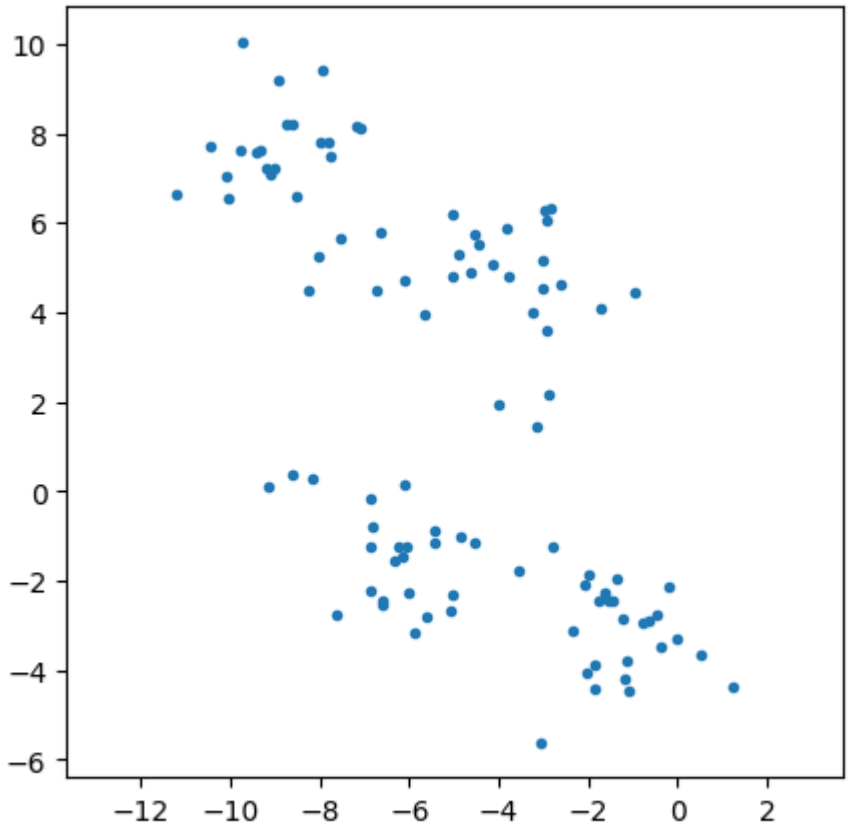
SDSE Lab 3

K Means Clustering

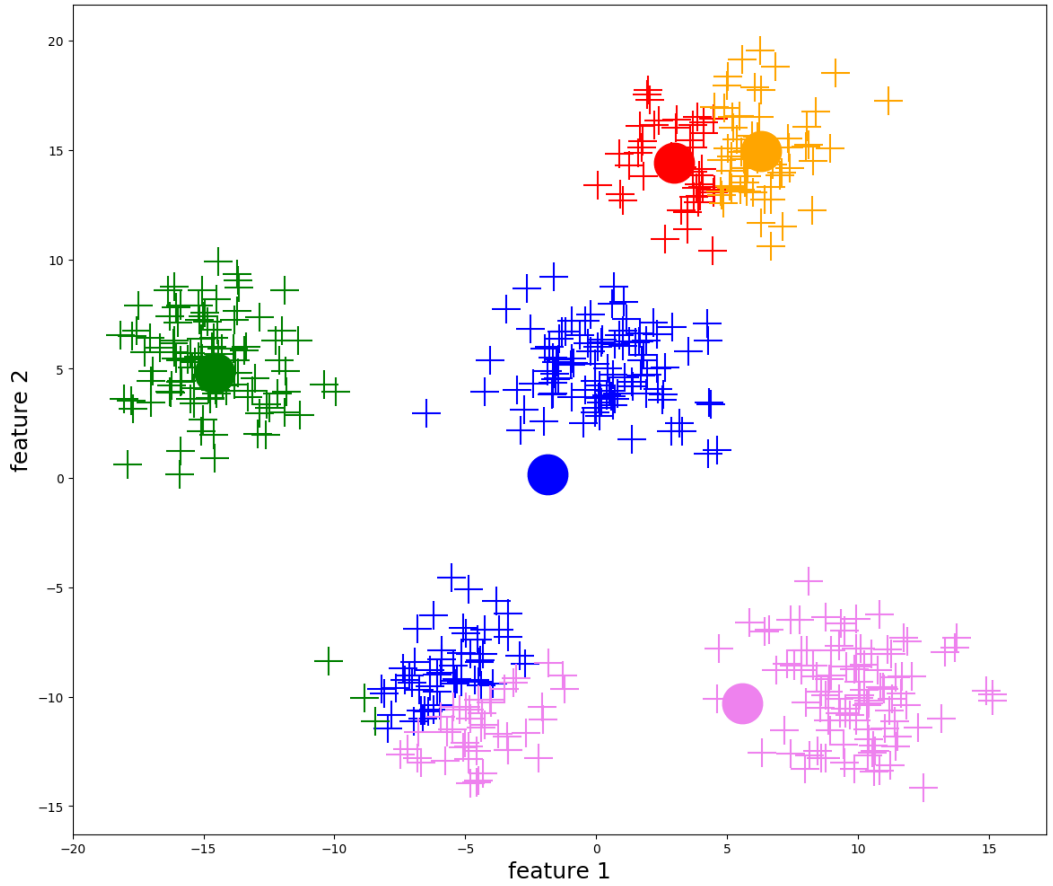
K-Means Clustering Algorithm

- *Unsupervised Learning*
  - No data labels
  - Goal to cluster data into meaningful groups



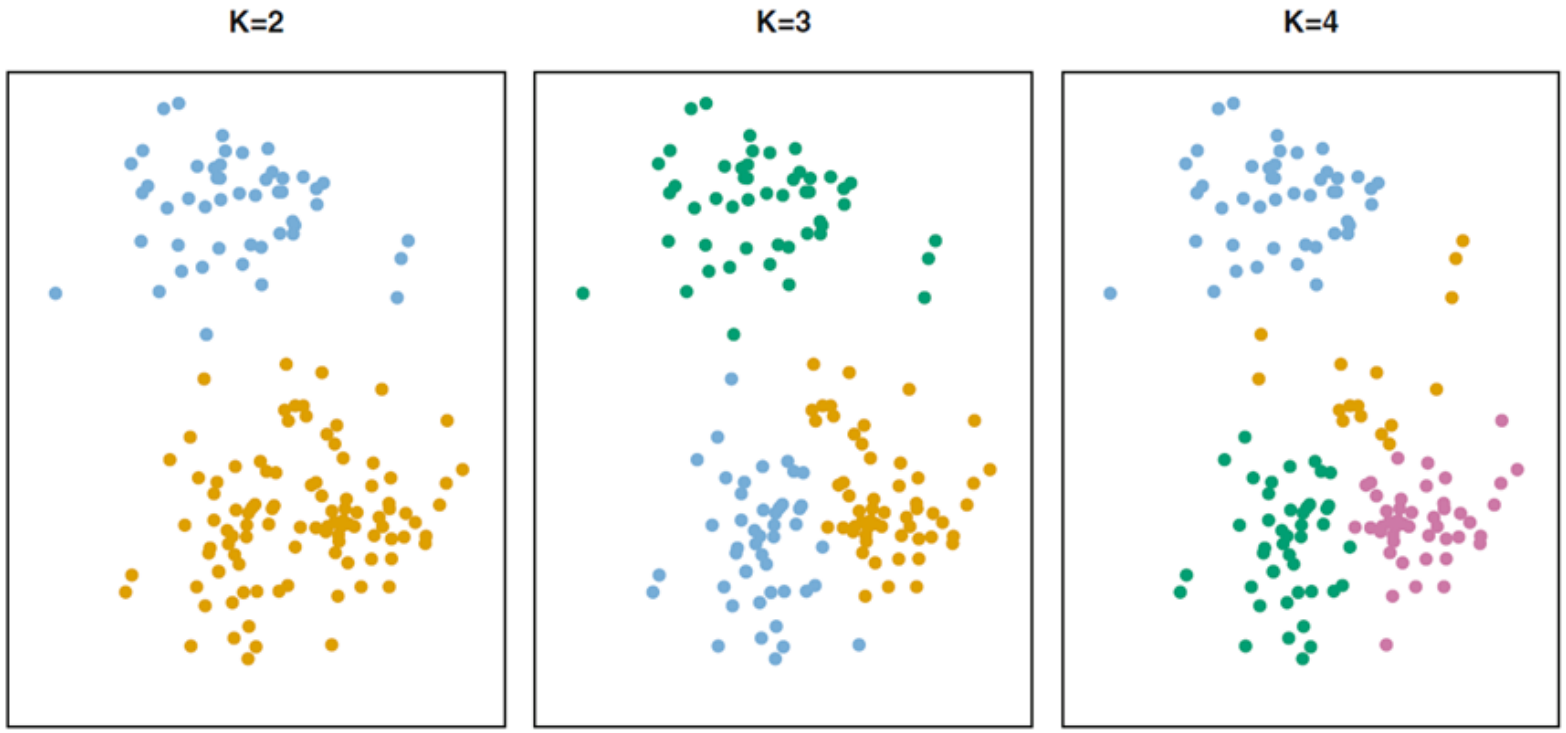
K-Means Clustering Algorithm

- Step 0: randomly initialize centroids
- Alternate between
    1. Assigning datapoints to a cluster based on closest centroid
    2. Re-calculating the centroids as the mean of each cluster



Extensions

- *How to select number of clusters, K?*
- *How does initial position of centroids affect the solution?*



Logistics

- Lab 2 is due **Sunday Oct 1 (10/1) at 11:59PM**
- Lab 3 is due Sunday Oct 15 at 11:59PM
- Export:
  - **IMPORTANT:** Kernel>Restart&Run All
  - Make sure you pass everything
  - Save using File>Save
  - Run the last cell, which generates a .zip file in the same folder as your lab
  - Upload this .zip to Gradescope
    - One assignment per group
    - After submitting, select group members in gradescope
  - Eventually, the auto-grader will tell you in gradescope your score
    - Submit early so you can troubleshoot