Special Topics in Machine Learning

Shawn Schwartz
UCLA EEB C177/C234
Week 9
03/03/2020

What is Machine Learning?

- Subset of artificial intelligence (AI)

- Uses *patterns* and *inference* to complete a task (w/o explicit instructions from humans)

- Machine learning (ML) algorithms can essentially <u>learn</u> and <u>improve</u> from experience

Unsupervised vs. Supervised ML

- 2 main types of ML algorithms

- **<u>Unsupervised</u>**:

  - Learning the inherent structure of data w/o pre-assigned labels

  - Most common task is <u>clustering</u>

  - Very good way to conduct exploratory analyses on your data set

  - Difficult to compare model performance b/w different unsupervised ML algorithms since no labels are provided *a priori*

Unsupervised vs. Supervised ML

- **<u>Supervised</u>**:

  - Classification, mapping input to output labels

  - Models can be trained

  - Some common examples:
    - Logistic regression
    - Random forests
    - Naive bayes
    - Neural networks
    - Support vector machines

Clustering (K-means)

- Attempts to classify data w/o first having been trained with labeled data

- Once the algorithms has been run and groups are defined, newly presented data can be assigned to the corresponding, relevant groups
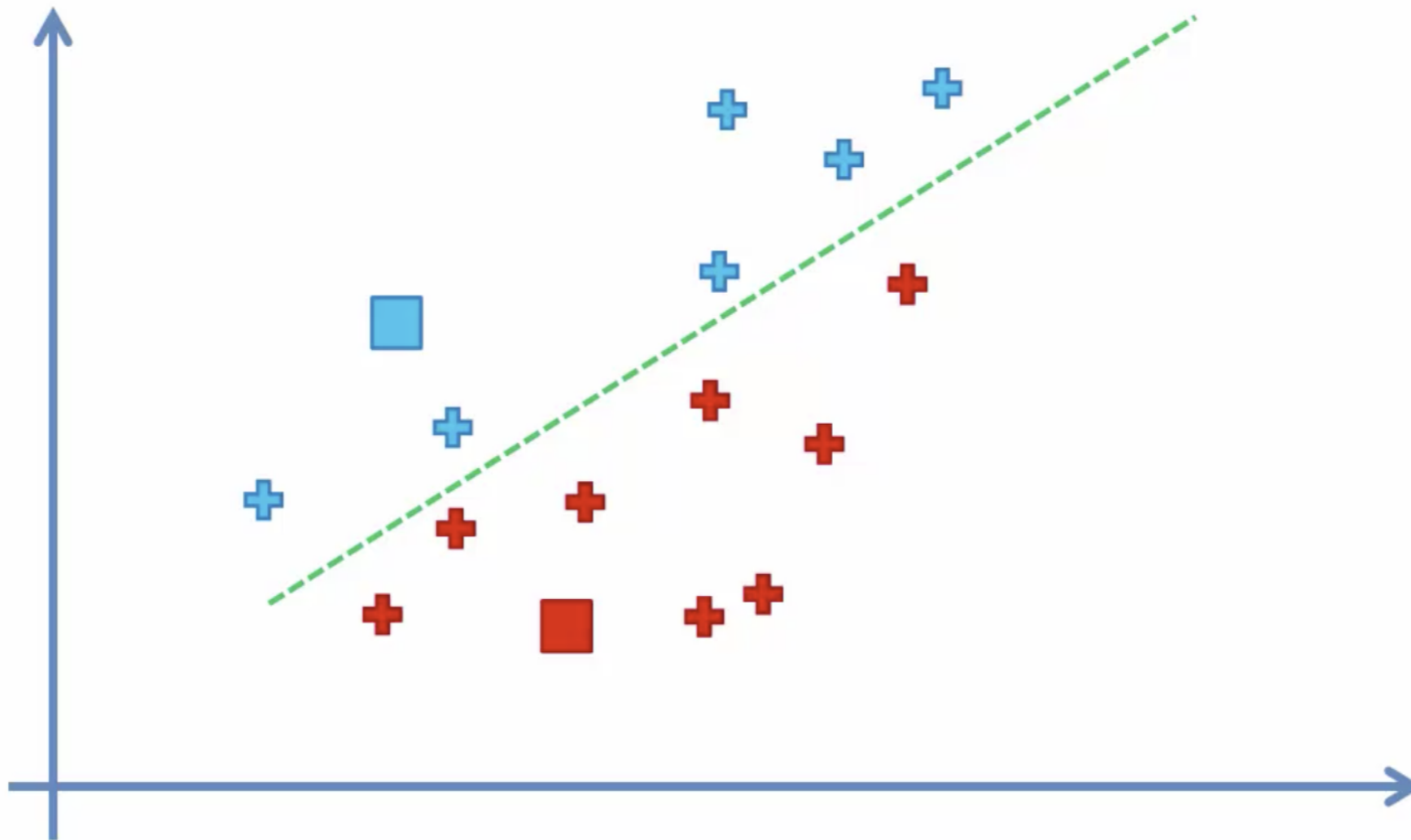
K-means Algorithm

1. Select k (i.e., 2; number of clusters) random points as cluster centers called "centroids"
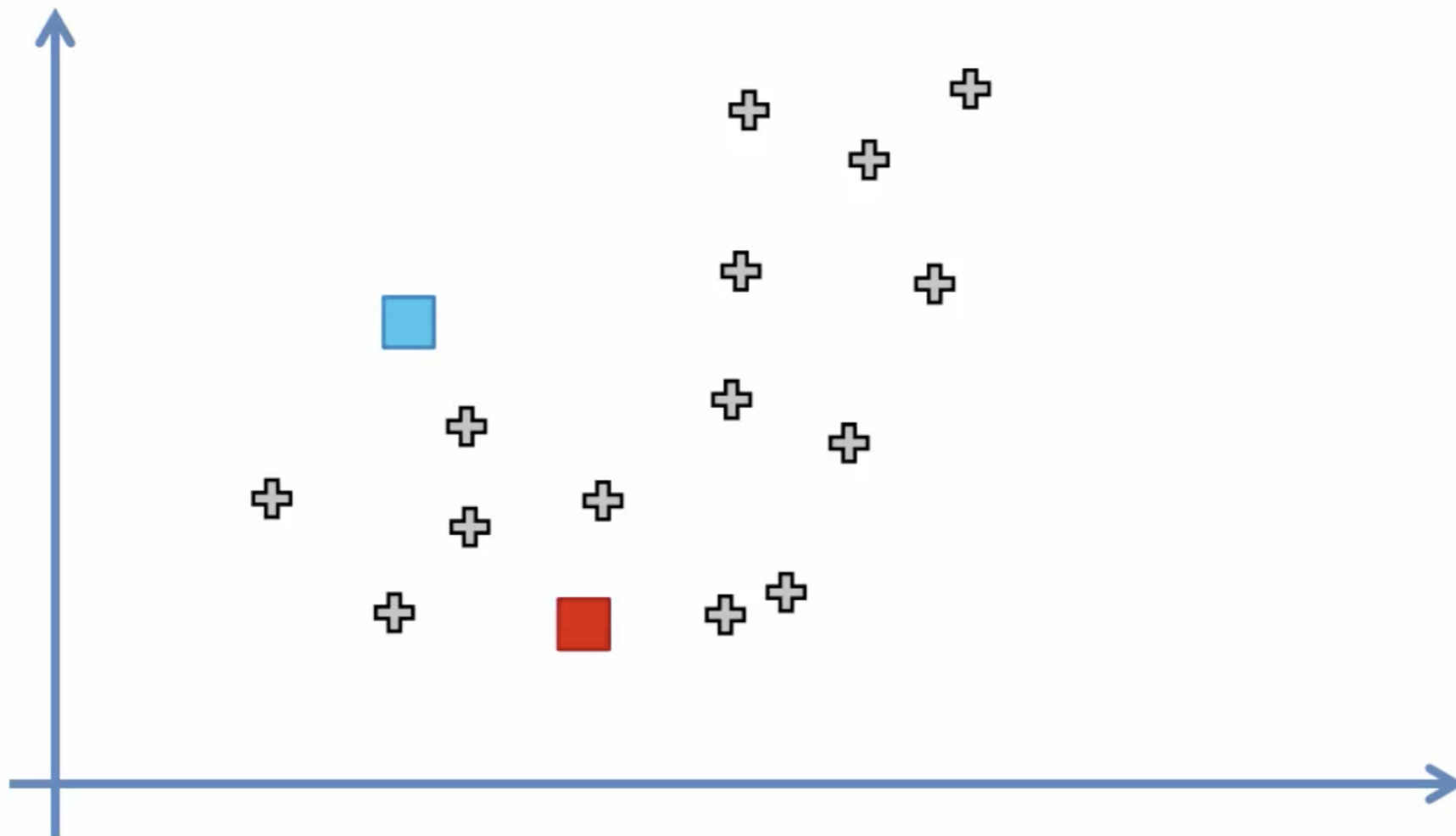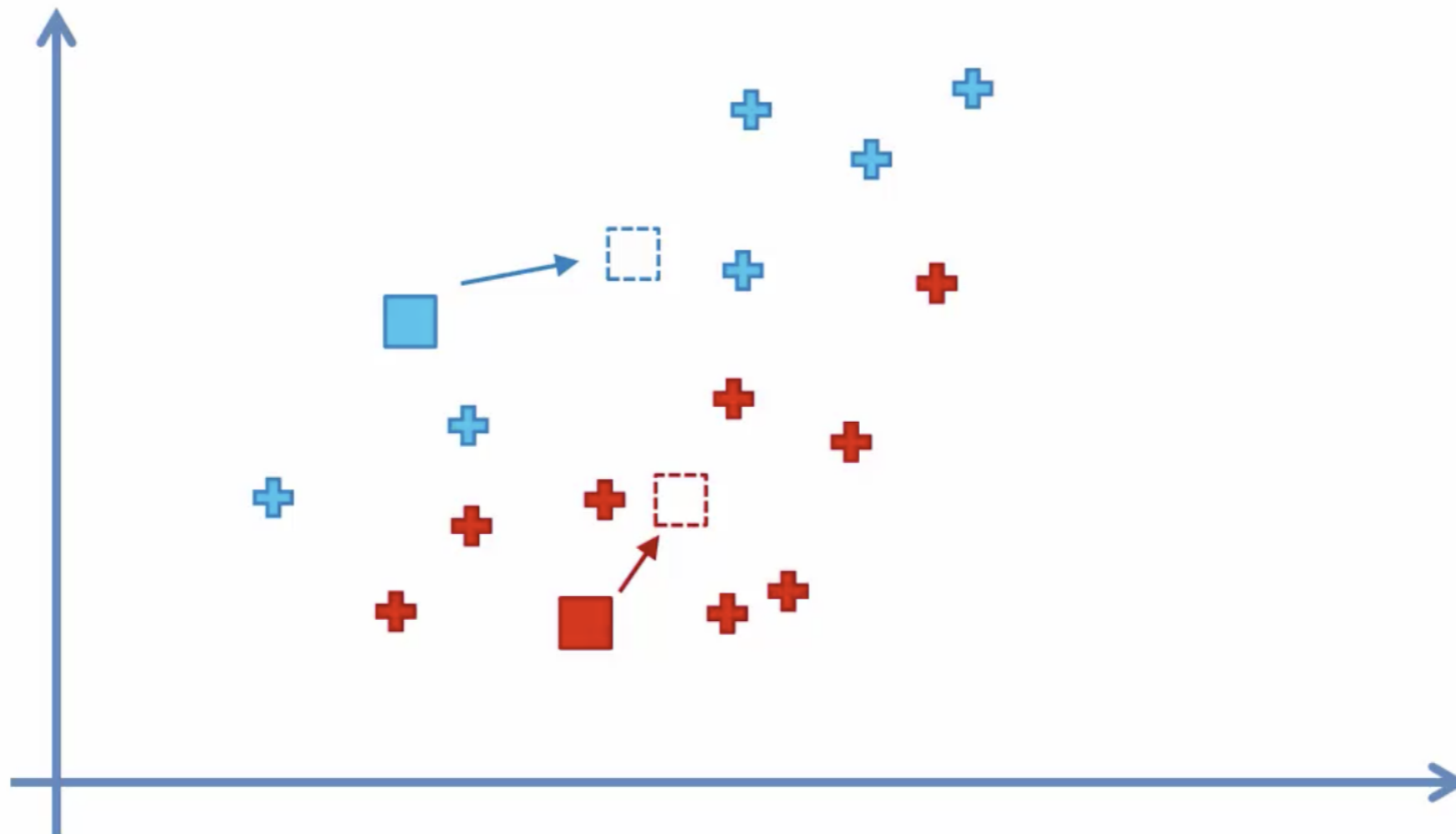
K-means Algorithm

2. Assign each data point to the closest cluster by calculating its distance with respect to each centroid
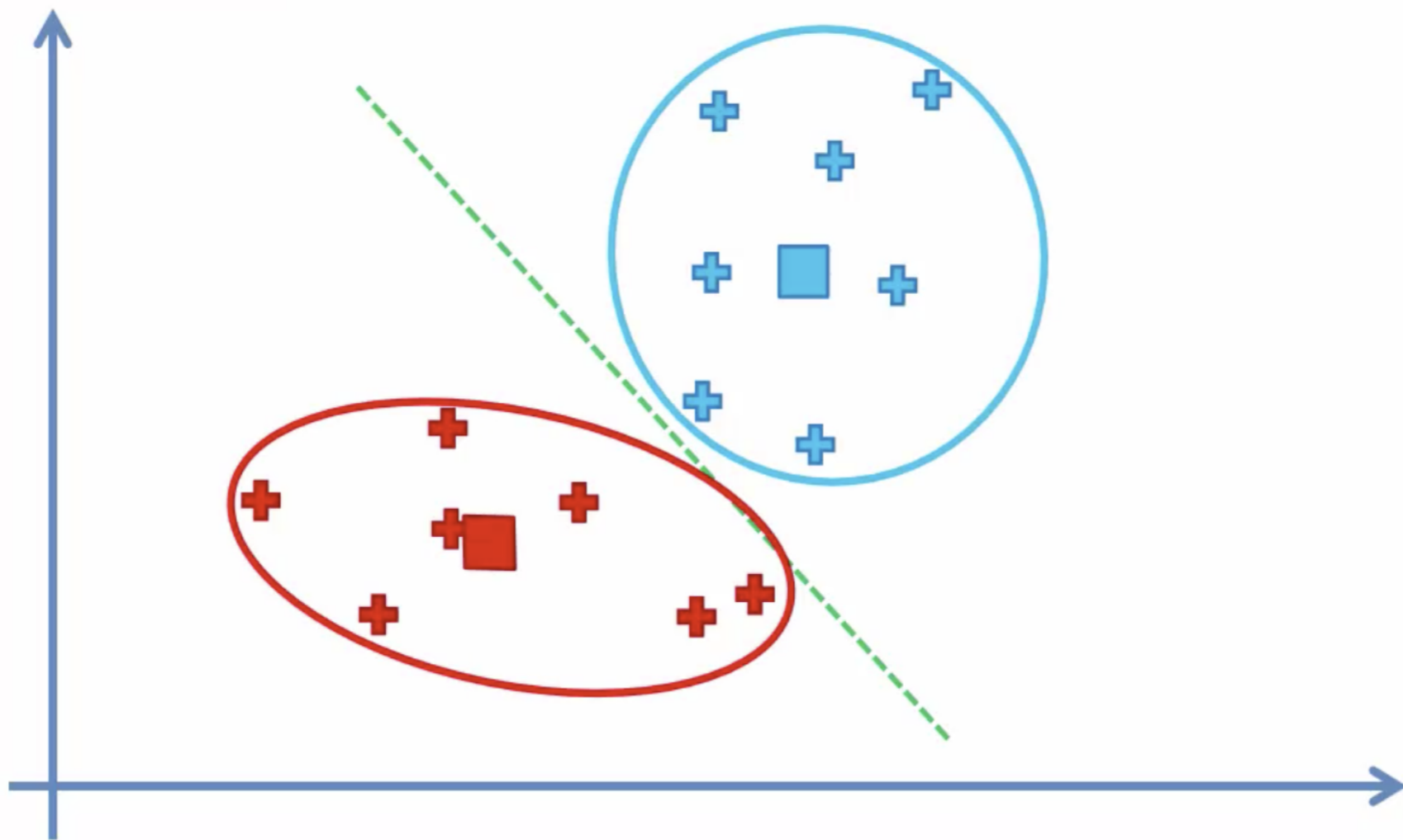
K-means Algorithm

3. Determine the new cluster center by computing the average of the assigned points

K-means Algorithm

4. Repeat steps 2 and 3 until none of the cluster assignments change

Choosing the Right Number of Clusters

- Can compute the Within Cluster Sum of Squares (WCSS) for a range of clusters and determine where the change in WCSS begins to level off (known as the "elbow method")
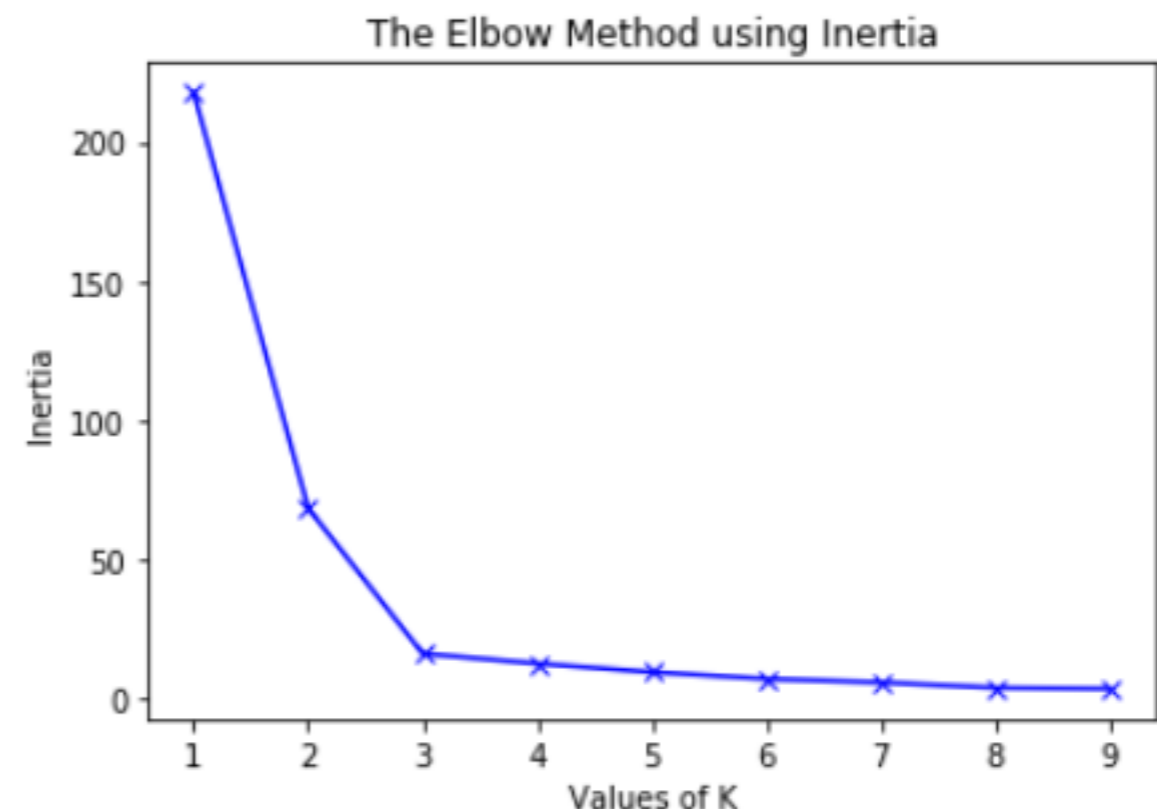
$$WCSS = \Sigma_{i=1}^{m}(x_i - c_i)^2$$

- The sum of the squared distance b/w each point of the cluster and its centroid

Choosing the Right Number of Clusters

- Can compute the Within Cluster Sum of Squares (WCSS) for a range of clusters and determine where the change in WCSS begins to level off (known as the "elbow method")

$$WCSS = \Sigma_{i=1}^{m}(x_i - c_i)^2$$

- The sum of the squared distance b/w each point of the cluster and its centroid

The Elbow Method using Inertia

Choosing the Right Number of Clusters

- Can compute the Within Cluster Sum of Squares (WCSS) for a range of clusters and determine where the change in WCSS begins to level off (known as the "elbow method")

$$WCSS = \Sigma_{i=1}^{m}(x_i - c_i)^2$$

- The sum of the squared distance b/w each point of the cluster and its centroid

The Elbow Method using Inertia