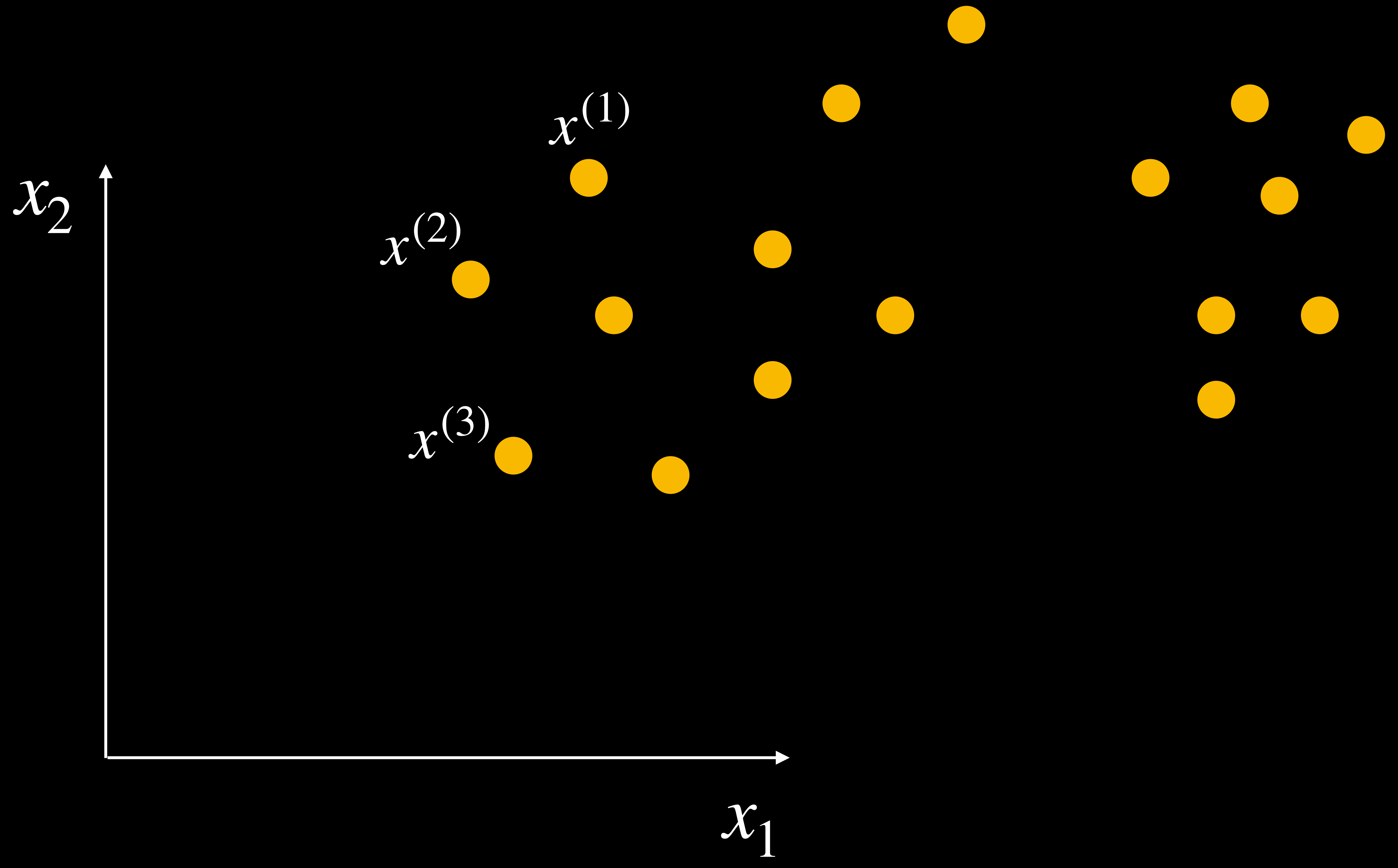
K-Means Clustering

Dataset

x_1	x_2
1.2	1.2
3.2	5.4
4.3	6.4
3.2	5.4
…	…



K-Means Clustering - Algorithm

Dataset

x_1	x_2
1.2	1.2
3.2	5.4
4.3	6.4
3.2	5.4
…	…

Initialize cluster centroids: $\mu_1, \mu_2, \ldots, \mu_k \in \mathbb{R}^d$

Repeat until convergence:

For every i, set:

$$c^{(i)} := \arg\min_j \|x^{(i)} - \mu_j\|^2.$$

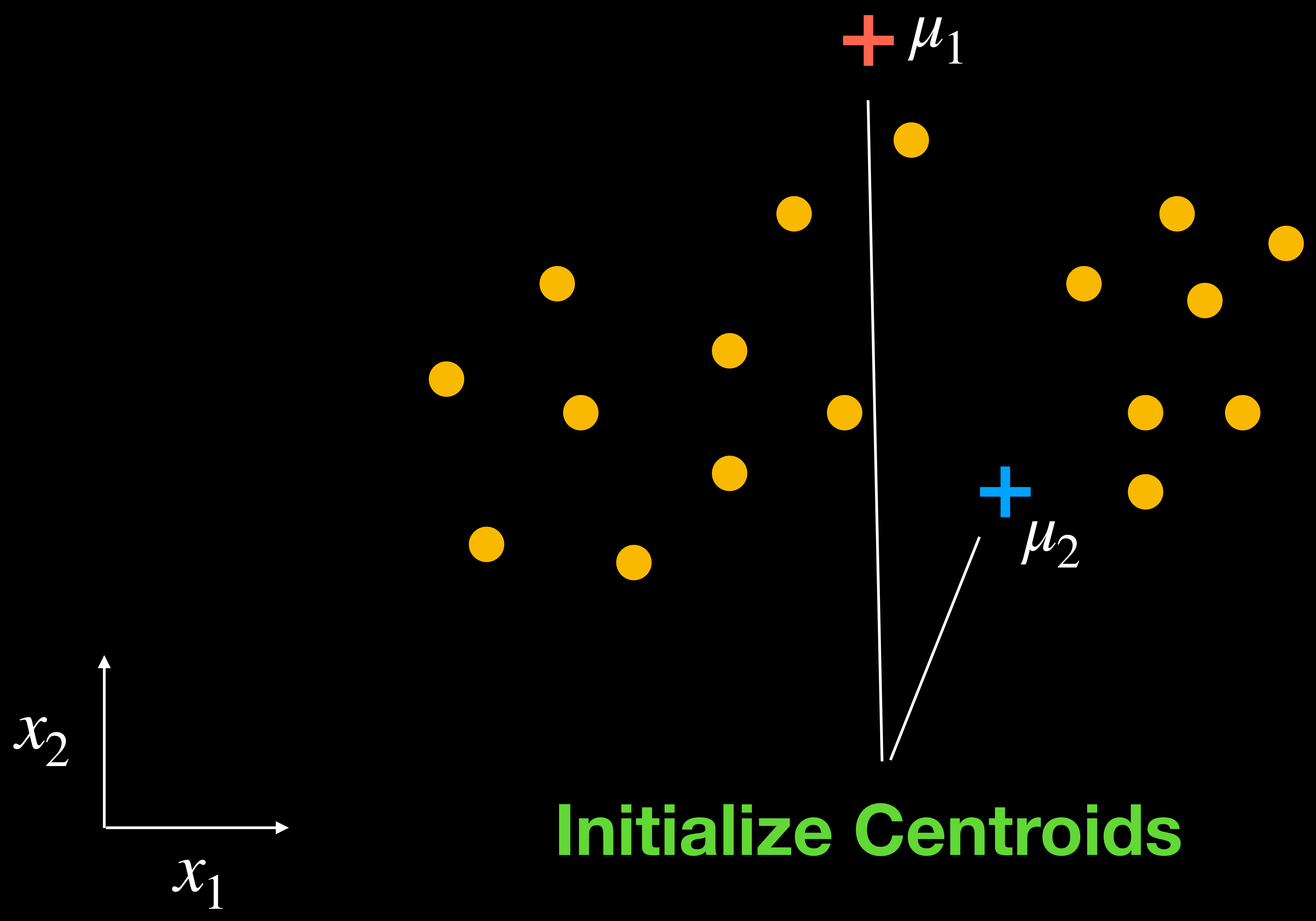
For every j, set:

$$\mu_j := \frac{\sum_{i=1}^n \mathbf{1}\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^n \mathbf{1}\{c^{(i)} = j\}}$$

K-Means Clustering

Dataset

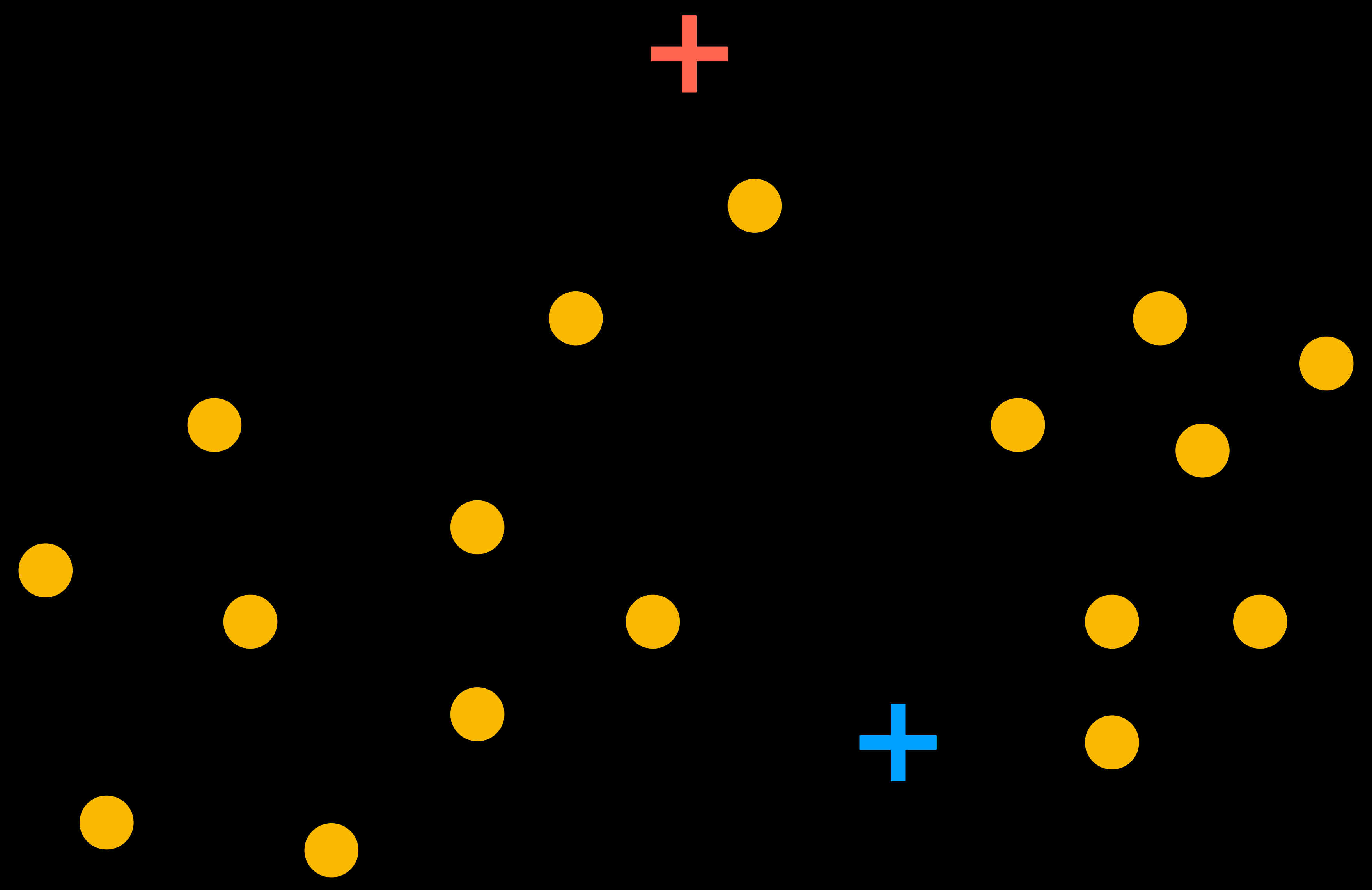
x_1	x_2
1.2	1.2
3.2	5.4
4.3	6.4
3.2	5.4
...	...

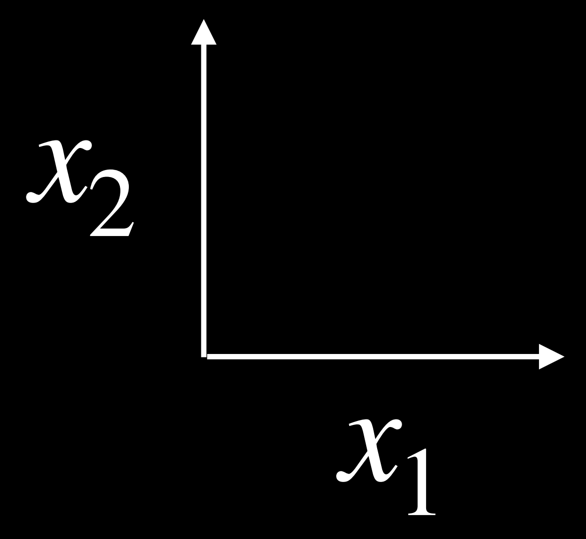


K-Means Clustering

Dataset

x_1	x_2
1.2	1.2
3.2	5.4
4.3	6.4
3.2	5.4
...	...





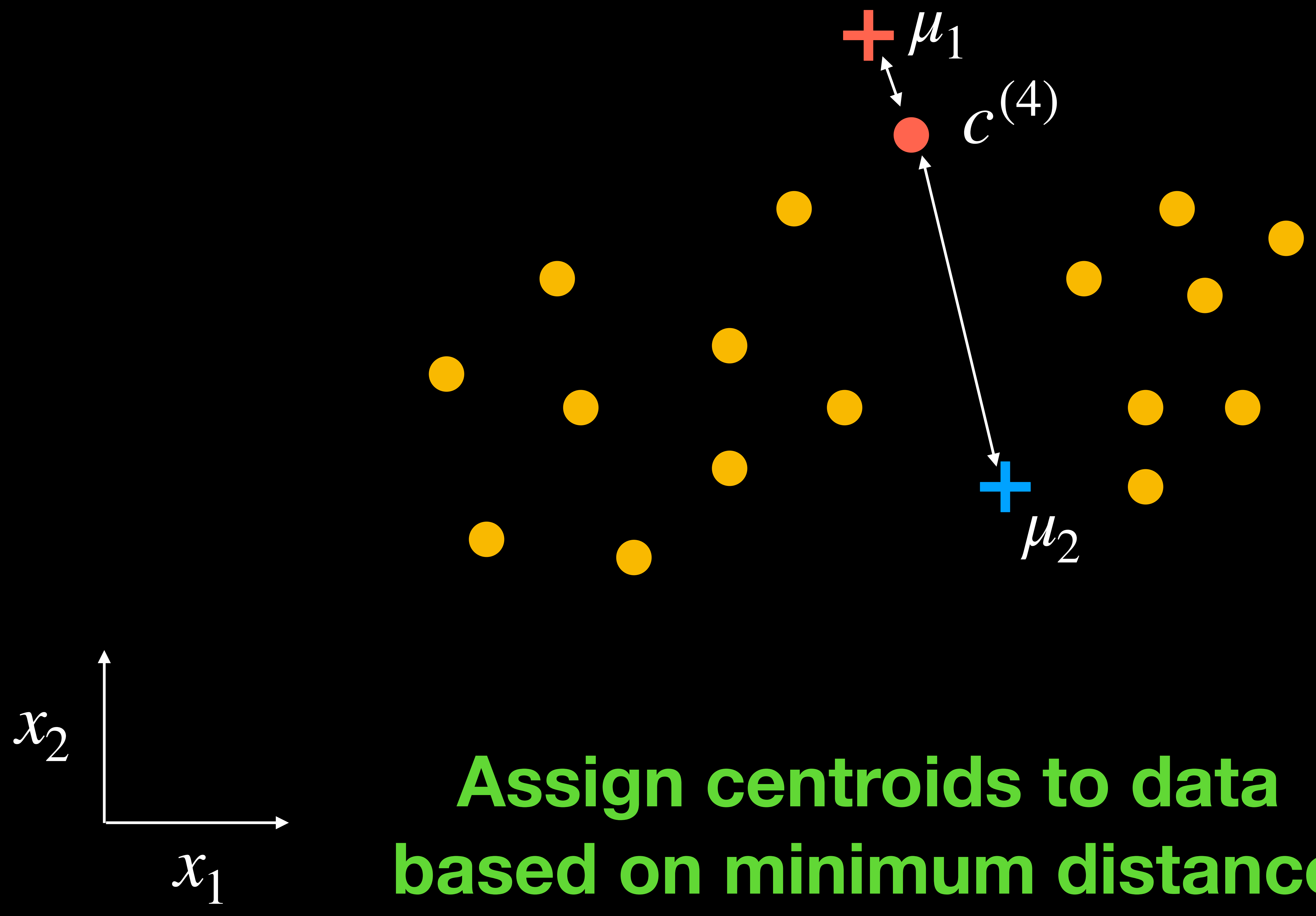
Assign centroids to data based on minimum distance

$$c^{(4)} = \arg\min_{1,2} \left(\|x^{(4)} - \mu_1\|^2, \|x^{(4)} - \mu_2\|^2 \right) = 1$$

K-Means Clustering

Dataset

x_1	x_2
1.2	1.2
3.2	5.4
4.3	6.4
3.2	5.4
...	...

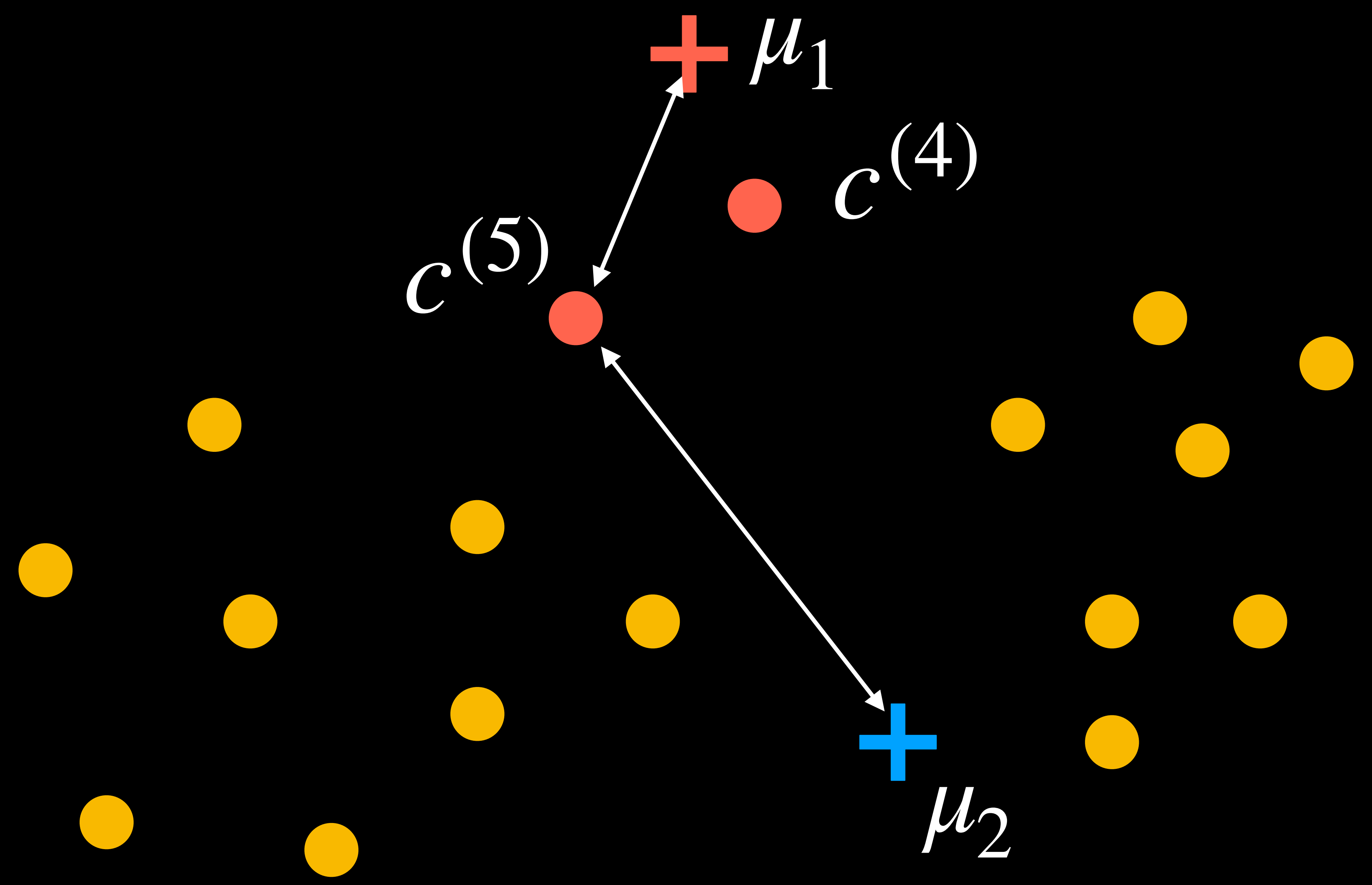


Assign centroids to data based on minimum distance

$$c^{(i)} := \arg\min_{j} \|x^{(i)} - \mu_j\|^2$$

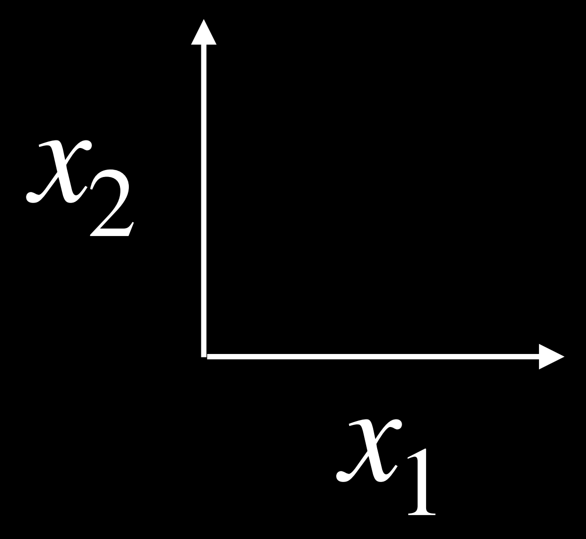
$$c^{(5)} = \arg\min_{1,2} \left(\|x^{(5)} - \mu_1\|^2, \|x^{(5)} - \mu_2\|^2 \right) = 1$$

K-Means Clustering



Dataset

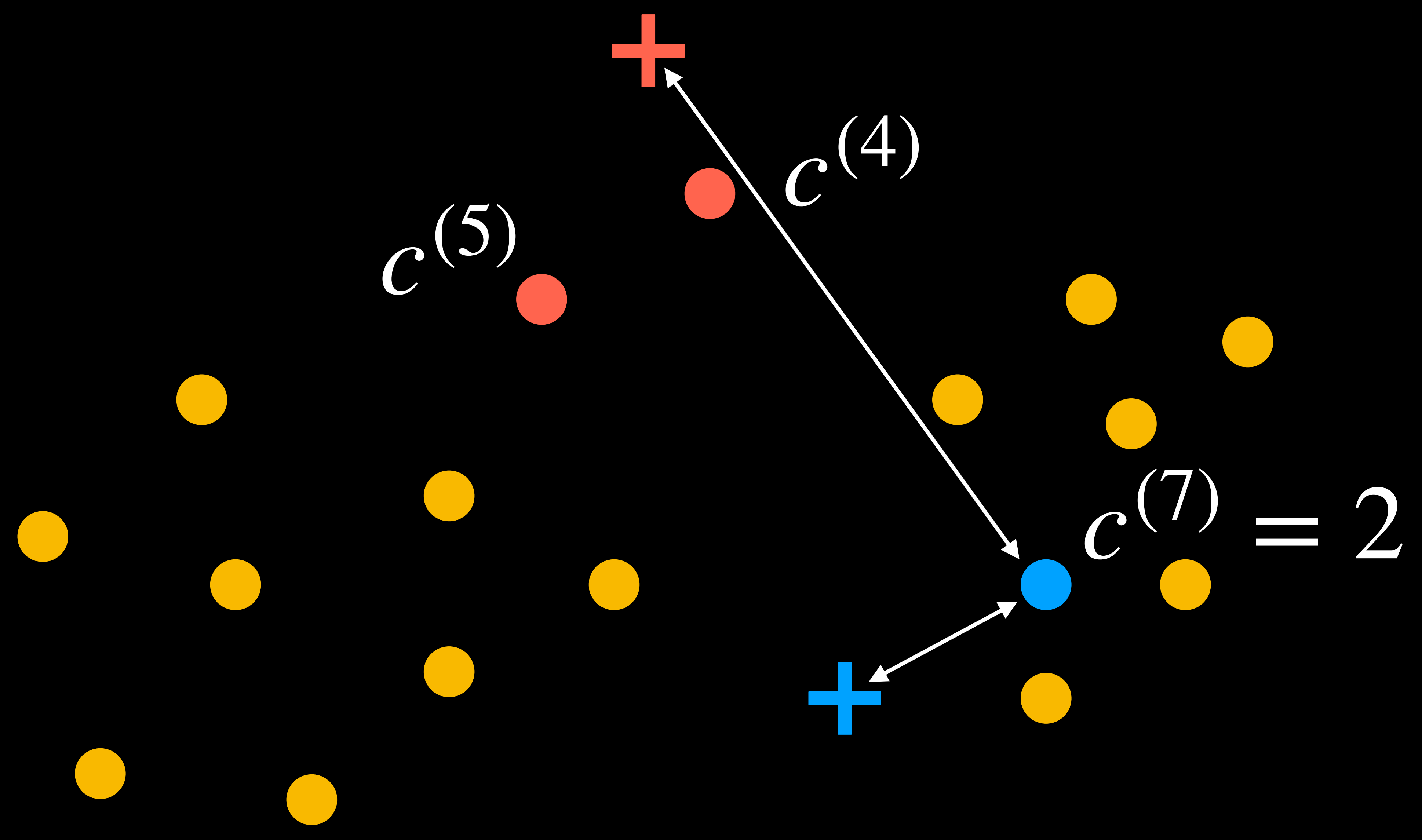
x_1	x_2
1.2	1.2
3.2	5.4
4.3	6.4
3.2	5.4
…	…



Assign centroids to data based on minimum distance

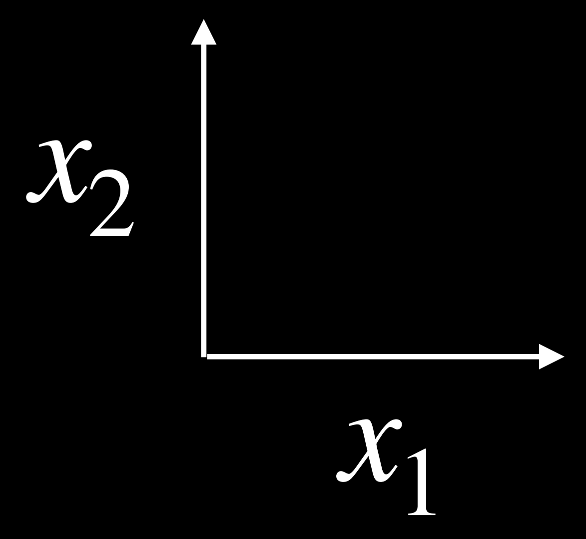
$$c^{(i)} := \arg\min_{j} \|x^{(i)} - \mu_j\|^2$$

K-Means Clustering



Dataset

x_1	x_2
1.2	1.2
3.2	5.4
4.3	6.4
3.2	5.4
...	...



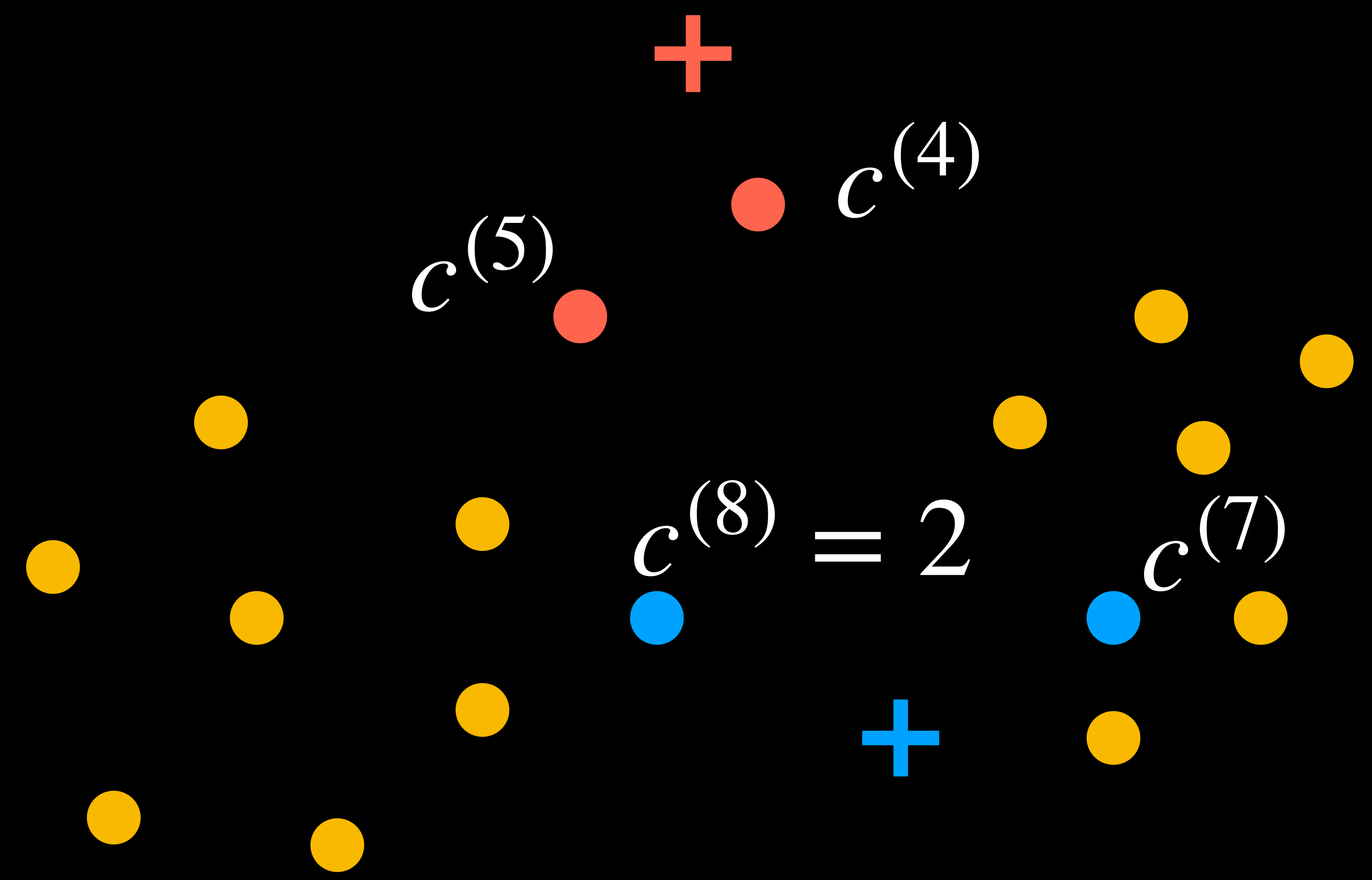
Assign centroid to data based on minimum distance

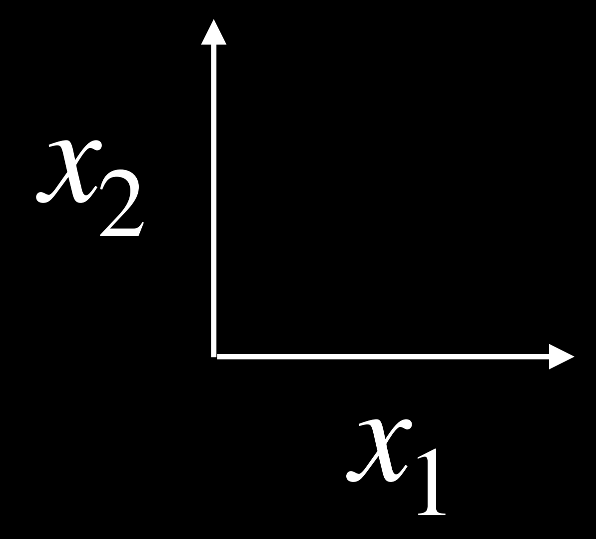
$$c^{(i)} := \arg\min_j \|x^{(i)} - \mu_j\|^2$$

K-Means Clustering

Dataset

x_1	x_2	c
1.2	1.2	1
3.2	5.4	1
4.3	6.4	2
3.2	5.4	1
...	...	...





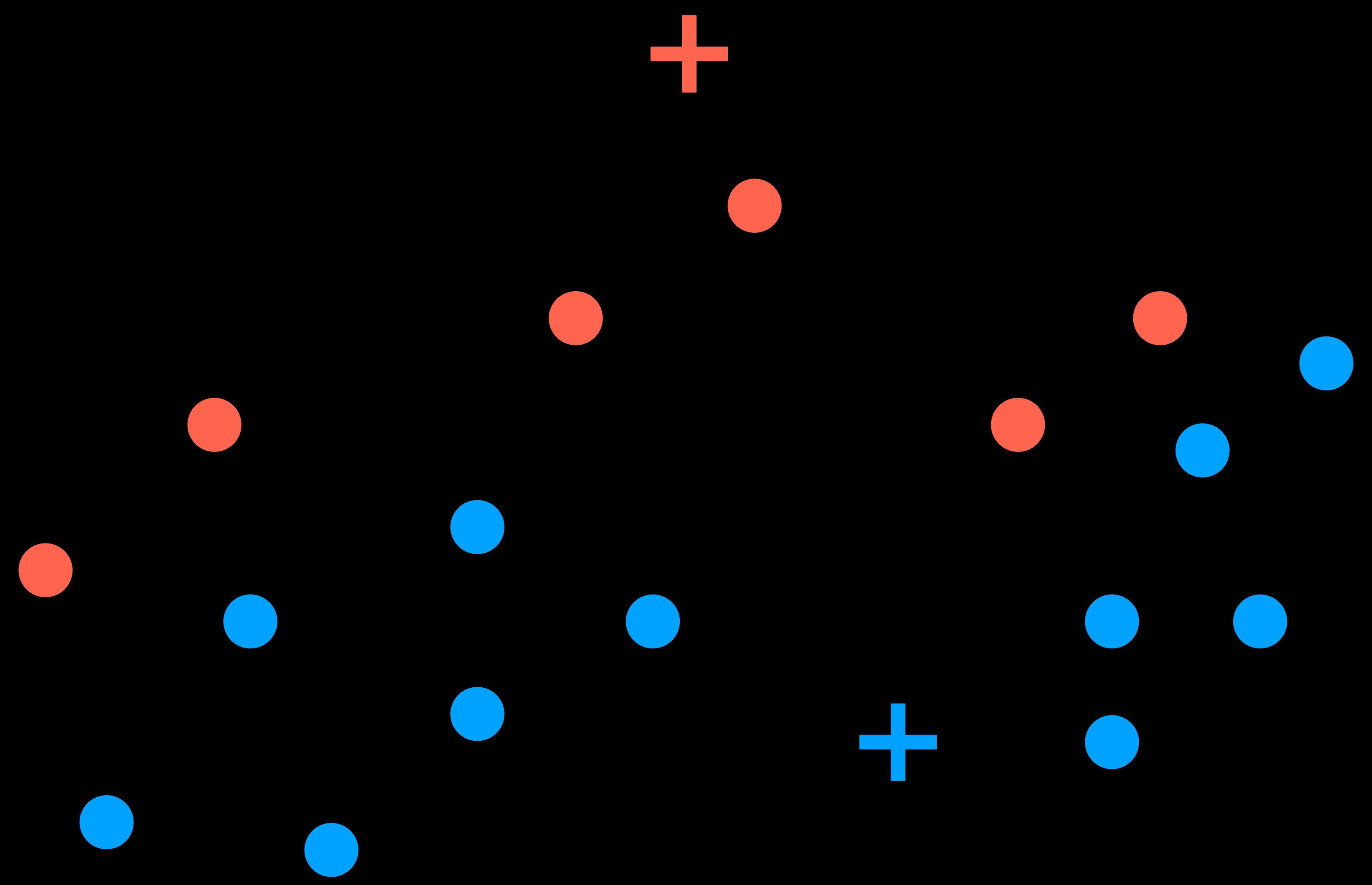
Assign centroids to data based on minimum distance

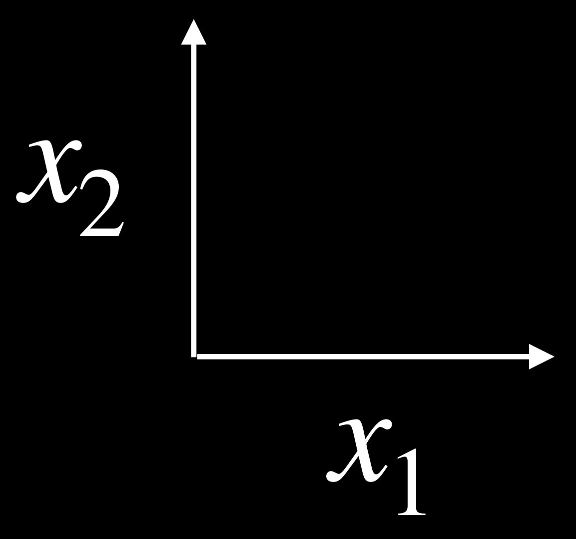
$$c^{(i)} := \arg\min_j \|x^{(i)} - \mu_j\|^2$$

K-Means Clustering

Dataset

x_1	x_2	c
1.2	1.2	1
3.2	5.4	1
4.3	6.4	2
3.2	5.4	1
...	...	...





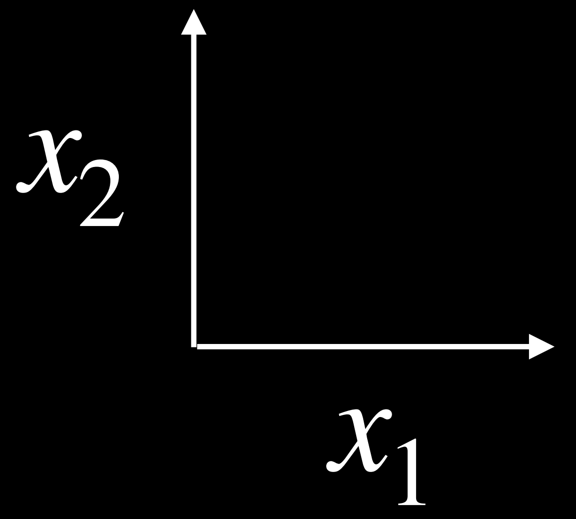
Assign centroid to data based on minimum distance

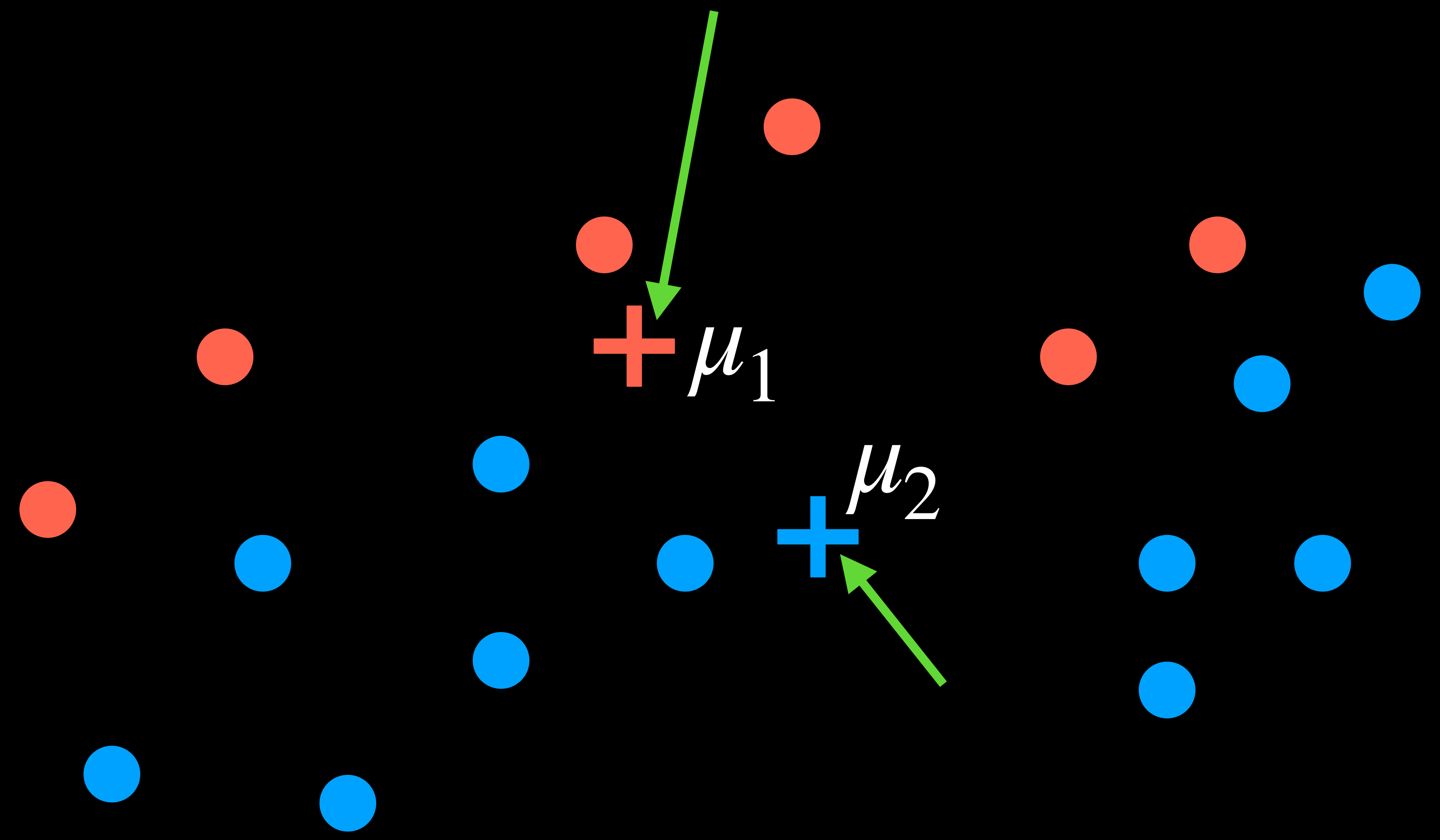
$$c^{(i)} := \arg\min_j \|x^{(i)} - \mu_j\|^2$$

K-Means Clustering

Dataset

x_1	x_2	c
1.2	1.2	1
3.2	5.4	1
4.3	6.4	2
3.2	5.4	1
...	...	...





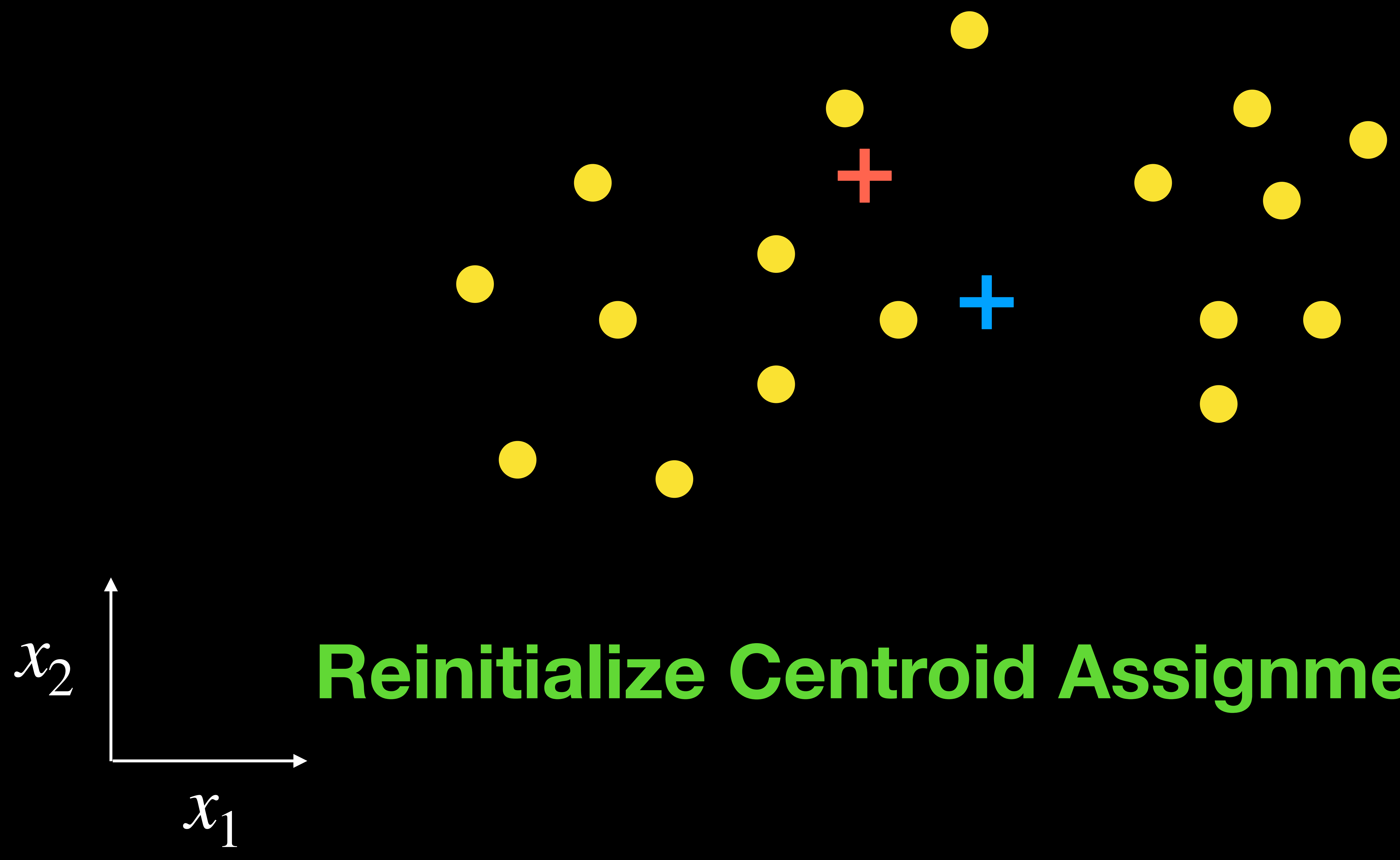
Update Centroids to Being Means of Their Clusters

$$\mu_j := \frac{\sum_{i=1}^{n} \mathbf{1}\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{n} \mathbf{1}\{c^{(i)} = j\}}$$

K-Means Clustering

Dataset

x_1	x_2	c
1.2	1.2	-
3.2	5.4	-
4.3	6.4	-
3.2	5.4	-
...	...	...

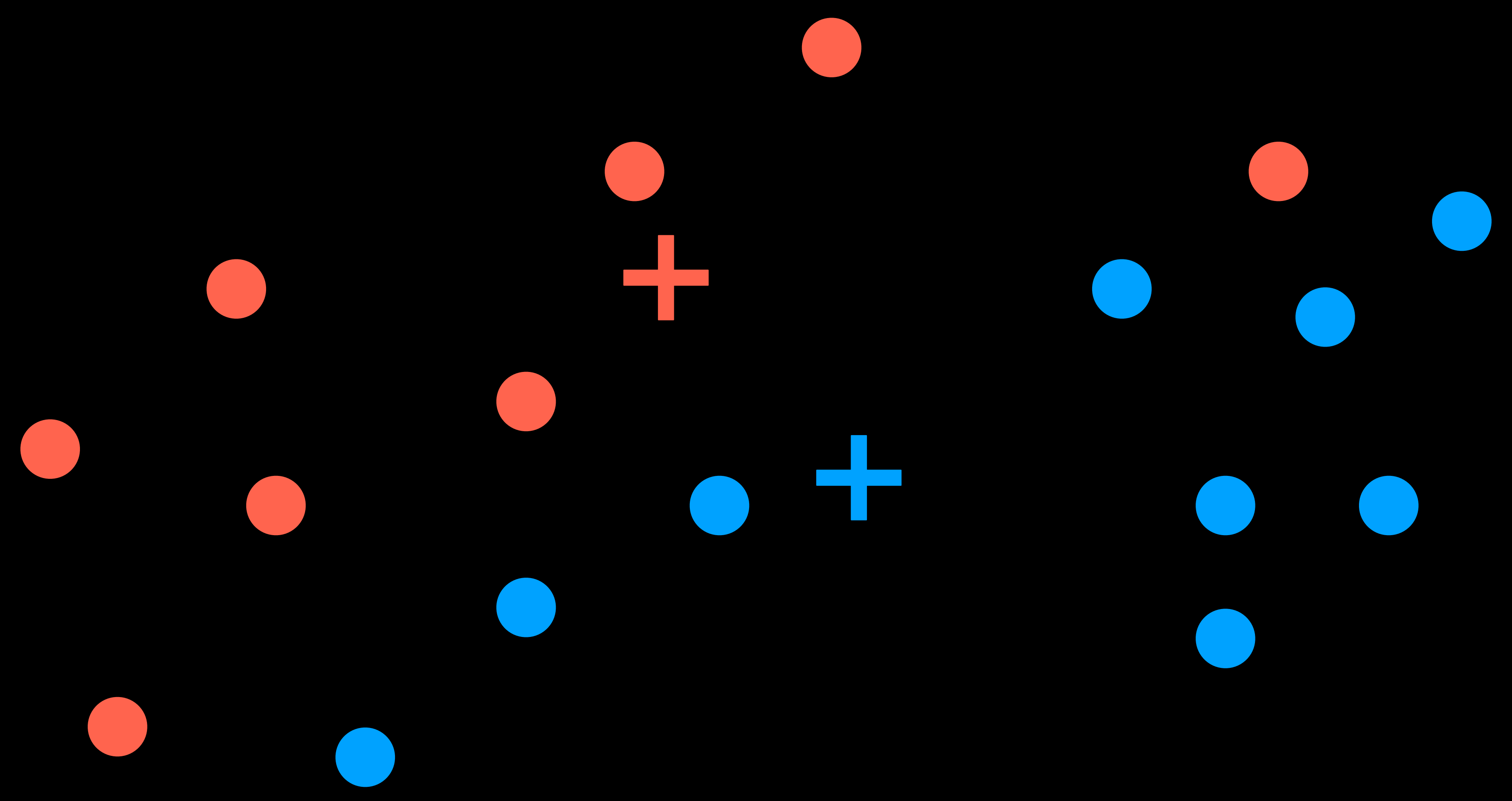


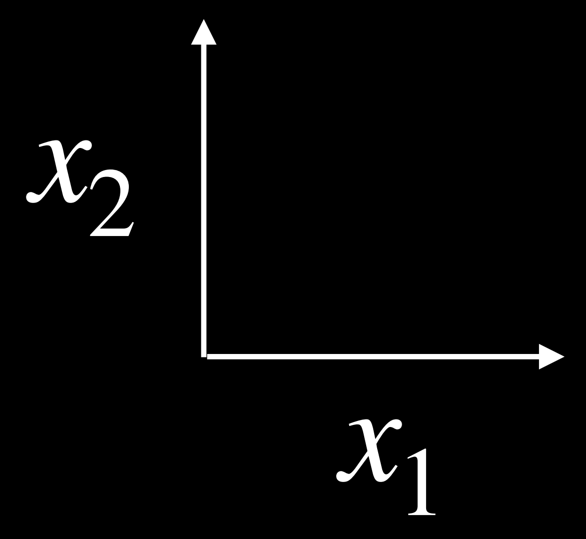
Reinitialize Centroid Assignments

K-Means Clustering

Dataset

x_1	x_2	c
1.2	1.2	1
3.2	5.4	2
4.3	6.4	2
3.2	5.4	1
…	…	…





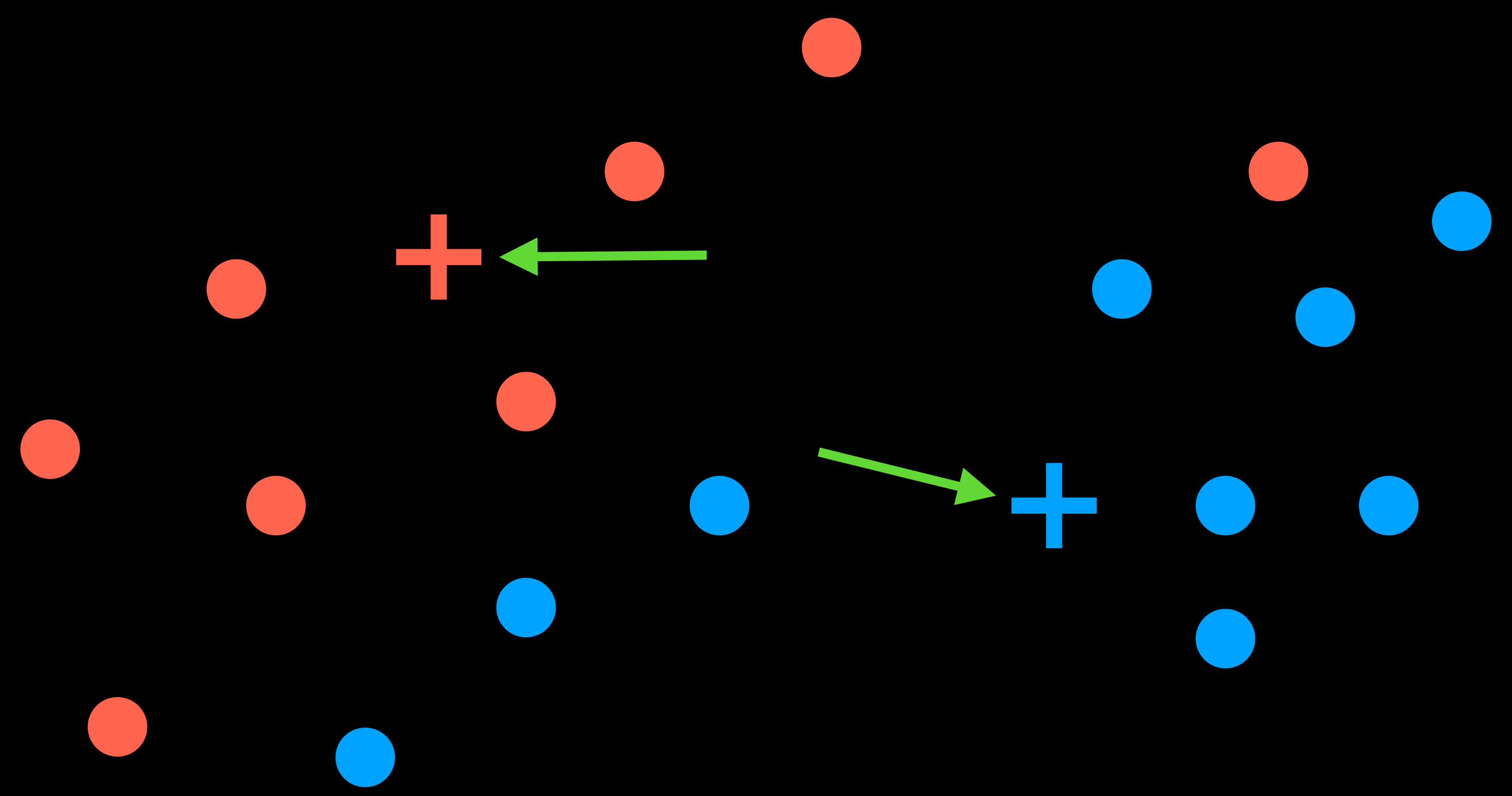
Assign Centroid to Data Based on Minimum Distance

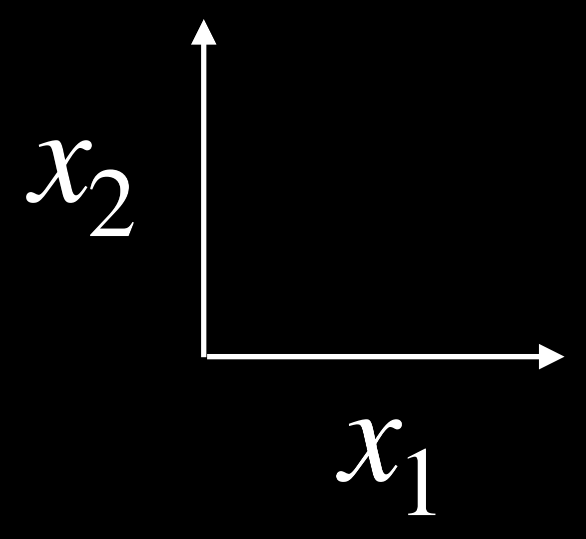
$$c^{(i)} := \arg\min_j \|x^{(i)} - \mu_j\|^2$$

K-Means Clustering

Dataset

x_1	x_2	c
1.2	1.2	1
3.2	5.4	2
4.3	6.4	2
3.2	5.4	1
...	...	...





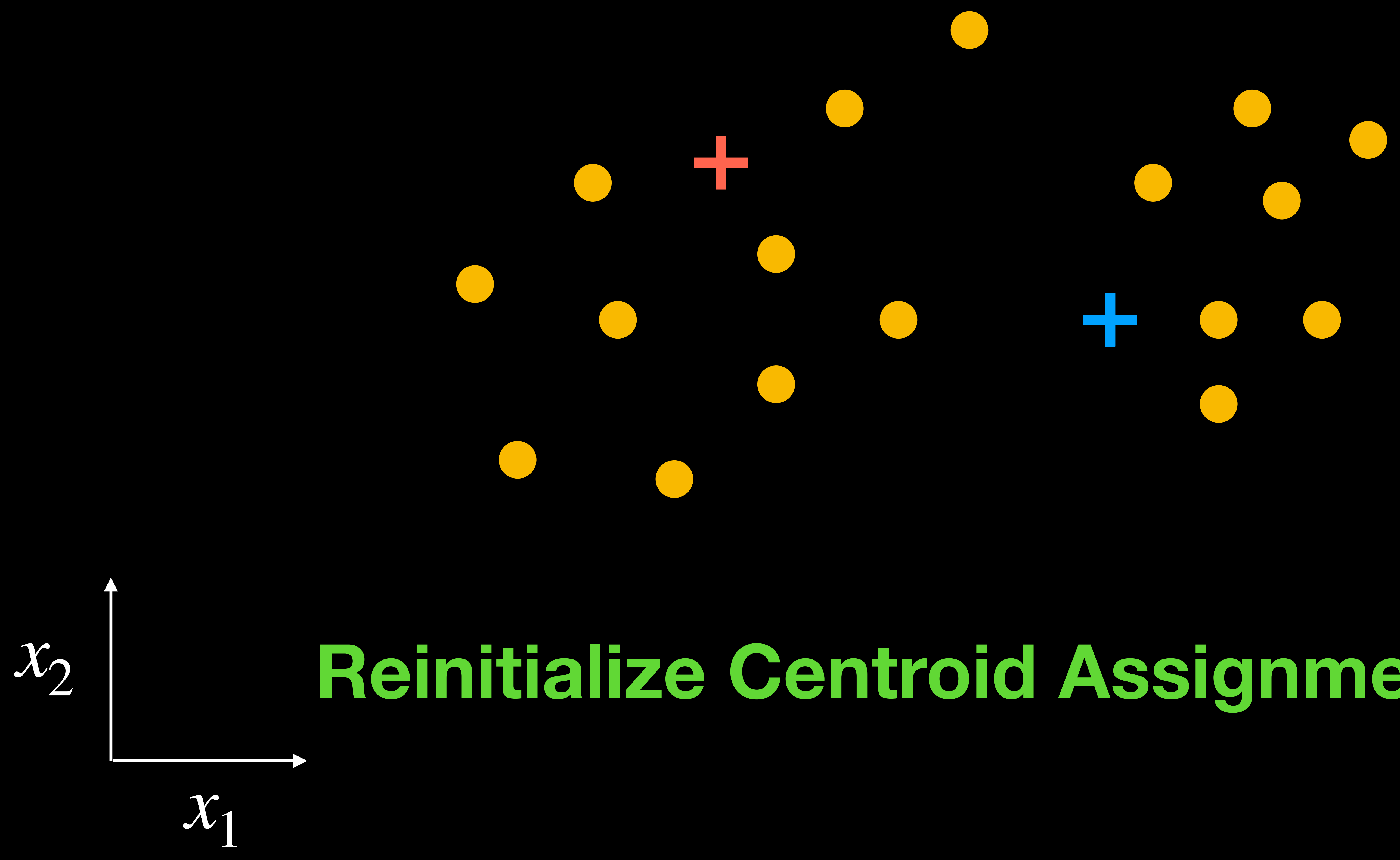
Update Centroids to Being Means of Their Clusters

$$\mu_j := \frac{\sum_{i=1}^{n} \mathbf{1}\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{n} \mathbf{1}\{c^{(i)} = j\}}$$

K-Means Clustering

Dataset

x_1	x_2	c
1.2	1.2	-
3.2	5.4	-
4.3	6.4	-
3.2	5.4	-
...	...	...

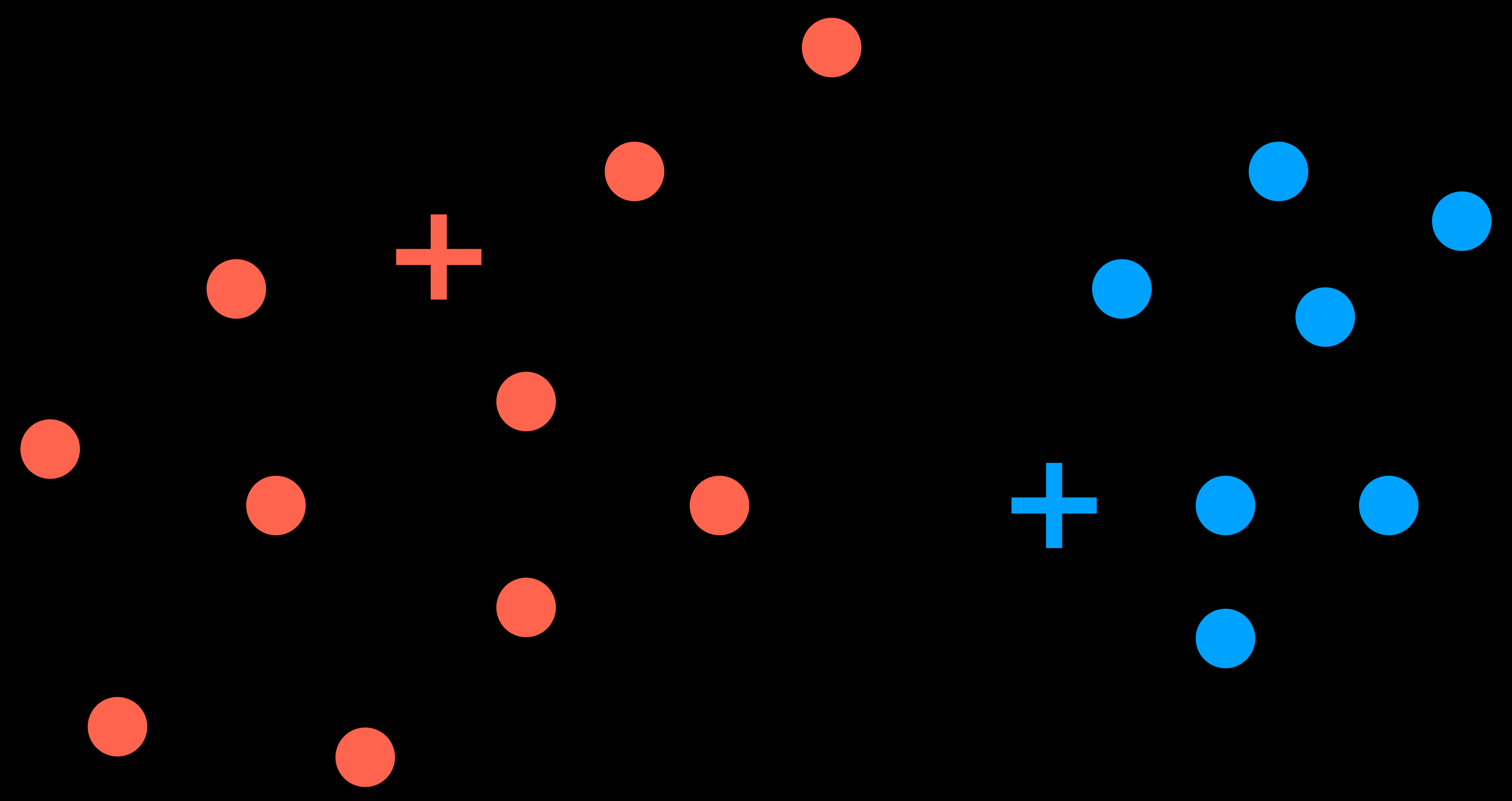


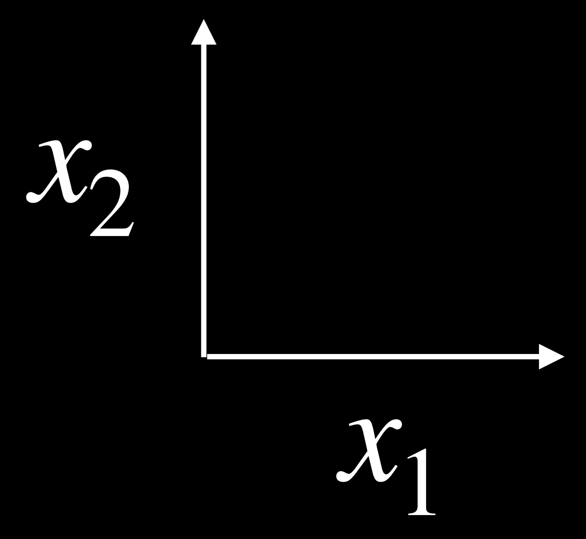
Reinitialize Centroid Assignments

K-Means Clustering

Dataset

x_1	x_2	c
1.2	1.2	1
3.2	5.4	2
4.3	6.4	1
3.2	5.4	2
…	…	…





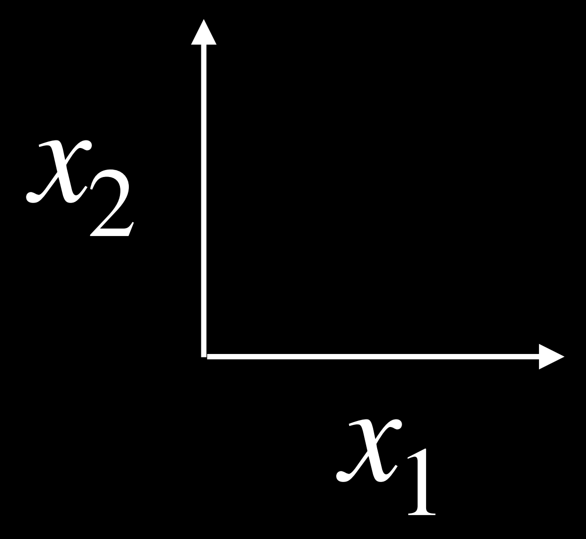
Assign Centroid to Data Based on Minimum Distance

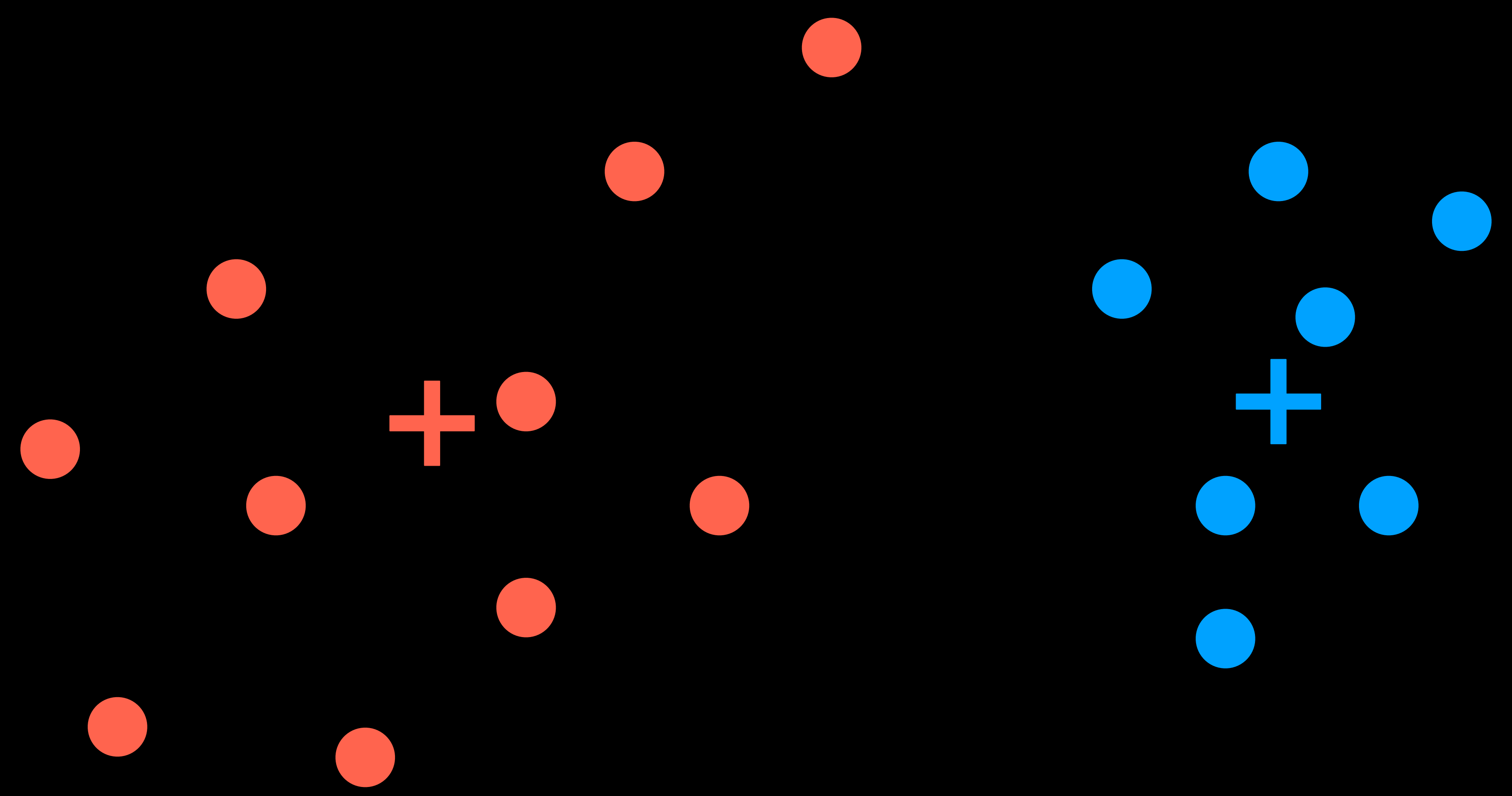
$$c^{(i)} := \arg\min_j \|x^{(i)} - \mu_j\|^2$$

K-Means Clustering

Dataset

x_1	x_2	c
1.2	1.2	1
3.2	5.4	2
4.3	6.4	1
3.2	5.4	2
...	...	...



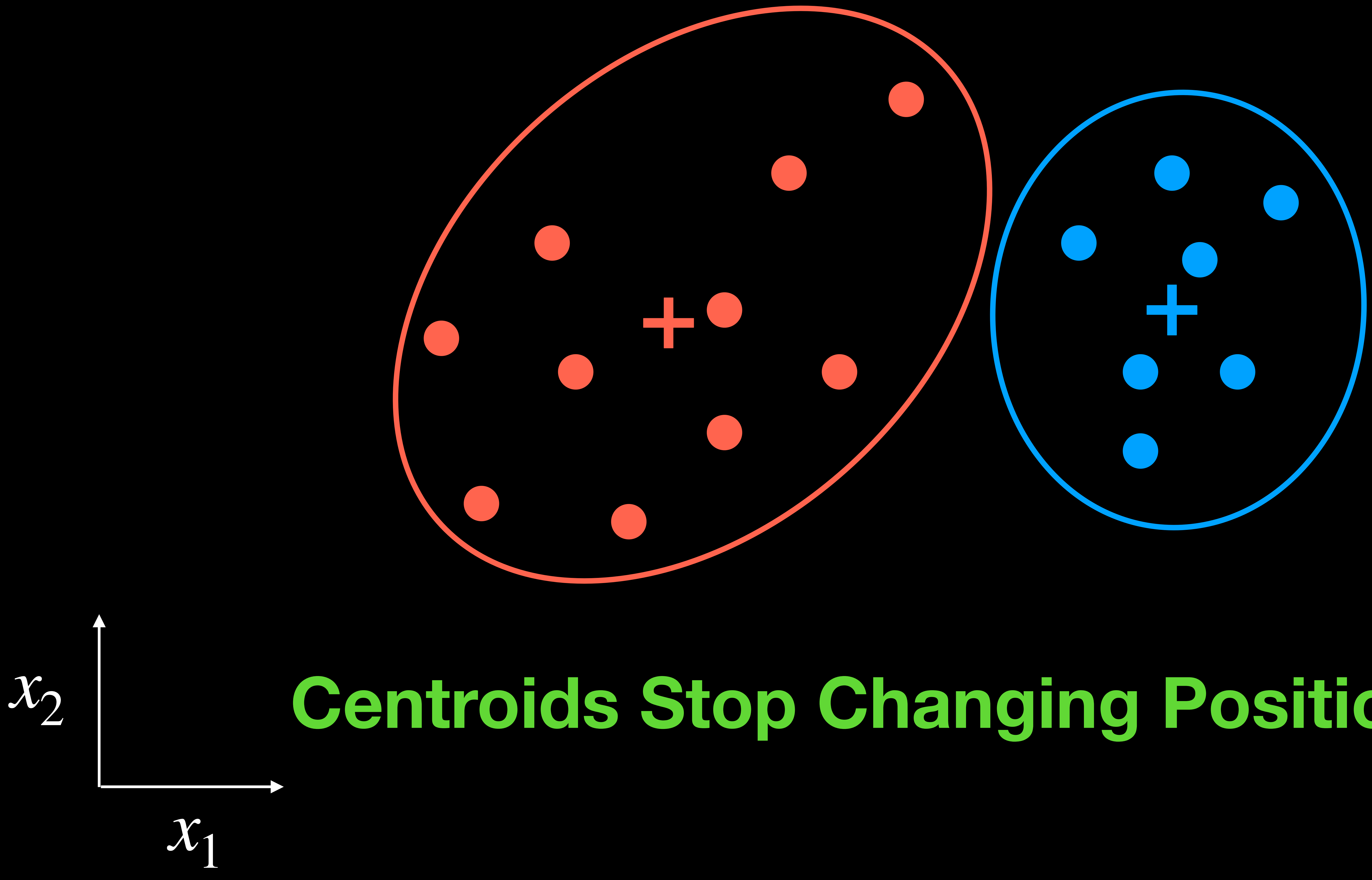


Update Centroids to Being Means of Their Clusters

K-Means Clustering

Dataset

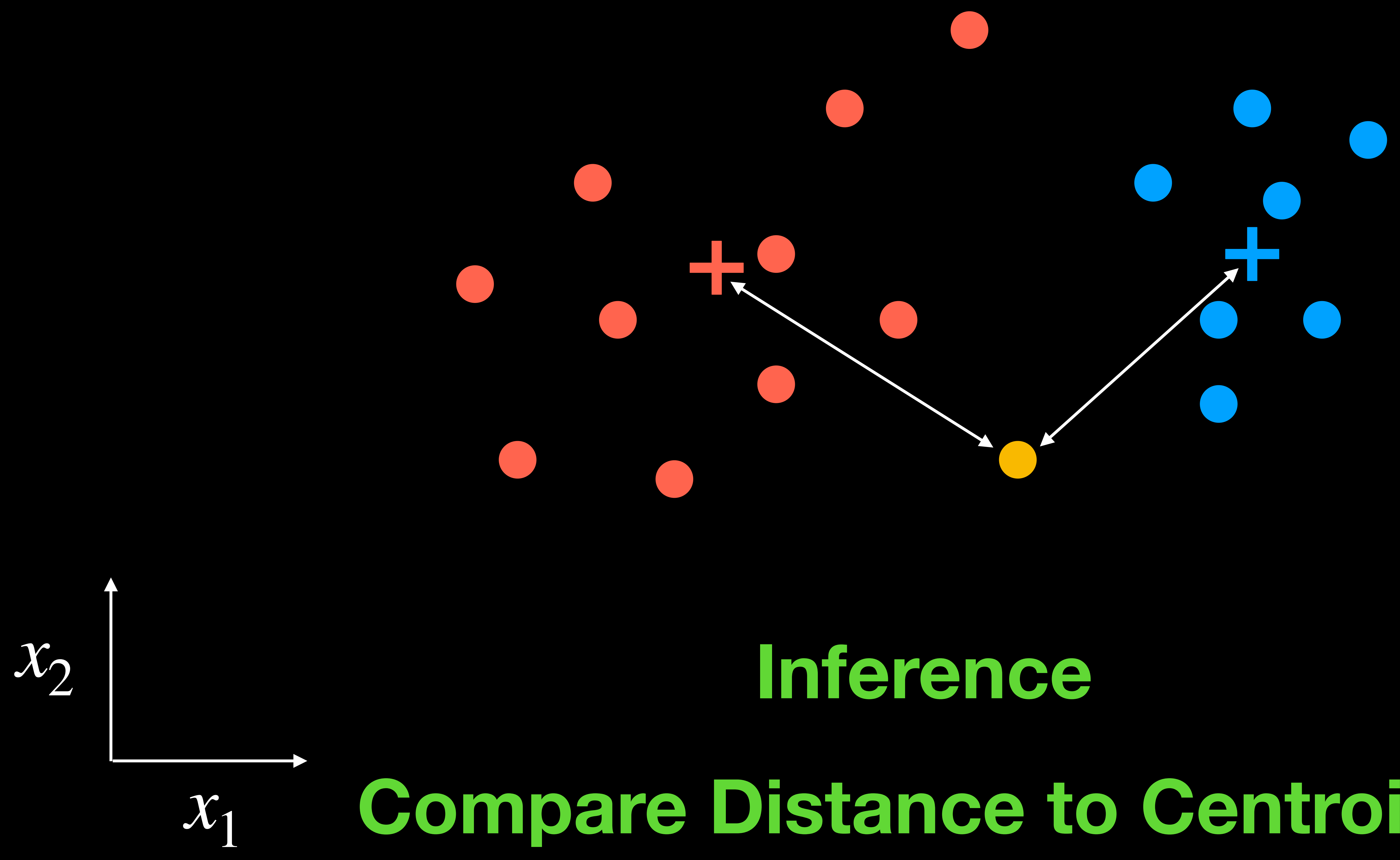
x_1	x_2	c
1.2	1.2	1
3.2	5.4	2
4.3	6.4	1
3.2	5.4	2
...	...	...



K-Means Clustering

Dataset

x_1	x_2
1.2	1.2
3.2	5.4
4.3	6.4
3.2	5.4
...	...



Inference

Compare Distance to Centroids

K-Means Clustering - Algorithm

Dataset

x_1	x_2
1.2	1.2
3.2	5.4
4.3	6.4
3.2	5.4
...	...

Initialize cluster centroids: $\mu_1, \mu_2, \ldots, \mu_k \in \mathbb{R}^d$

Repeat until convergence:

For every i, set:

$$c^{(i)} := \arg\min_j \|x^{(i)} - \mu_j\|^2$$

For every j, set:

$$\mu_j := \frac{\sum_{i=1}^{n} \mathbf{1}\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{n} \mathbf{1}\{c^{(i)} = j\}}$$

K-Means Clustering - Loss Function

Loss that is being minimized

$$J(c, \mu) = \sum_{i=1}^{n} \left\| x^{(i)} - \mu_{c(i)} \right\|^2$$

- K-means is **coordinate descent** on $J(c, \mu)$: **distortion function**
- $J(c, \mu)$ is generally non-convex and susceptible to local minima
- Problem can be fixed by trying different random initial values for μ_j