# Mathematics of Image and Data Analysis
# Math 5467

## $k$-means clustering

Instructor: Jeff Calder
Email: jcalder@umn.edu

http://www-users.math.umn.edu/~jwcalder/5467

# Announcements

- Projects due Friday

# Last time

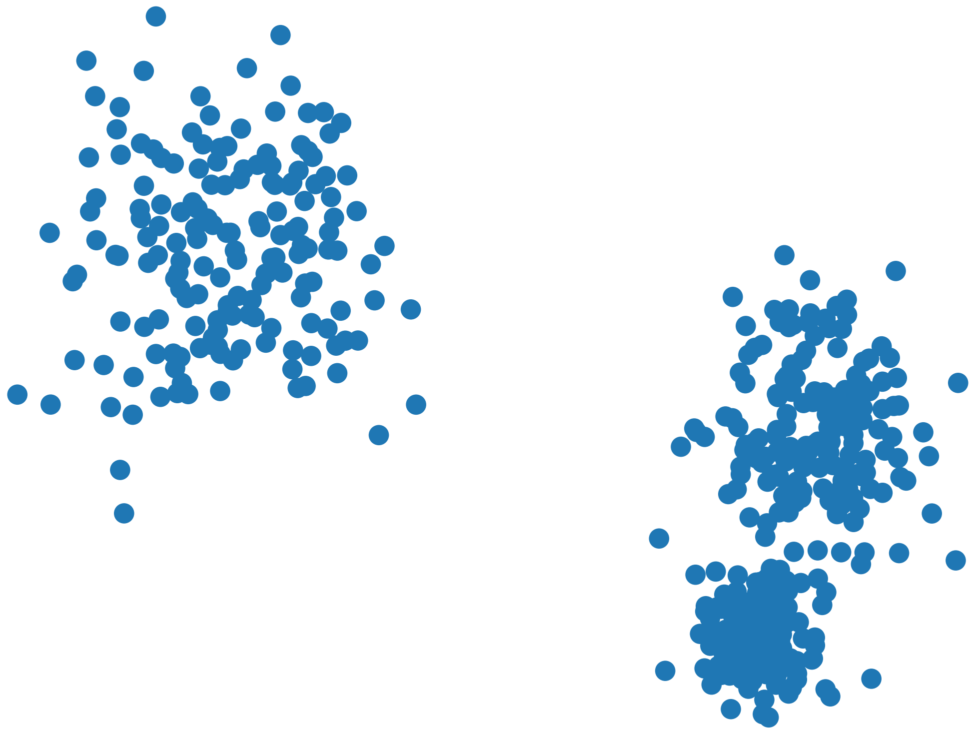
- Principal Component Analysis (PCA) Applications

# Today

$k$**-means clustering**:

- Theory
- Clustering MNIST digits

# Clustering

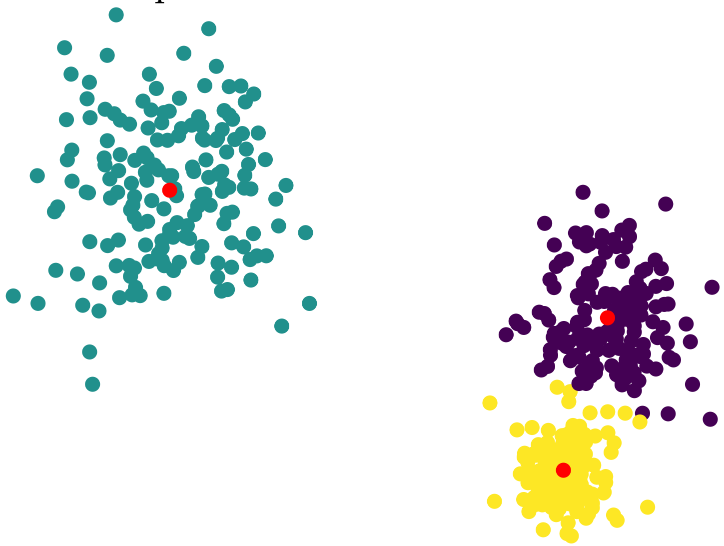


# $k$-means clustering

Let $x_1, x_2, \ldots, x_m$ be datapoints in $\mathbb{R}^n$. The $k$-means algorithm is guided by the task of minimizing the energy over the choice of cluster centers $c_i$

$$E(c_1, c_2, \ldots, c_k) = \sum_{i=1}^{m} \min_{1 \le j \le k} \|x_i - c_j\|^2.$$

Minimizing $E$ is an NP-hard problem.



# $k$-means clustering

**$k$-means algorithm:** We start with some randomized initial values for the means $c_1^0, c_2^0, \ldots, c_k^0$, and iterate the steps below until convergence.

1. Update the clusters

$$\Omega_j^t = \left\{ x_i \,:\, \|x_i - c_j^t\|^2 = \min_{1 \le \ell \le k} \|x_i - c_\ell^t\|^2 \right\}. \tag{1}$$

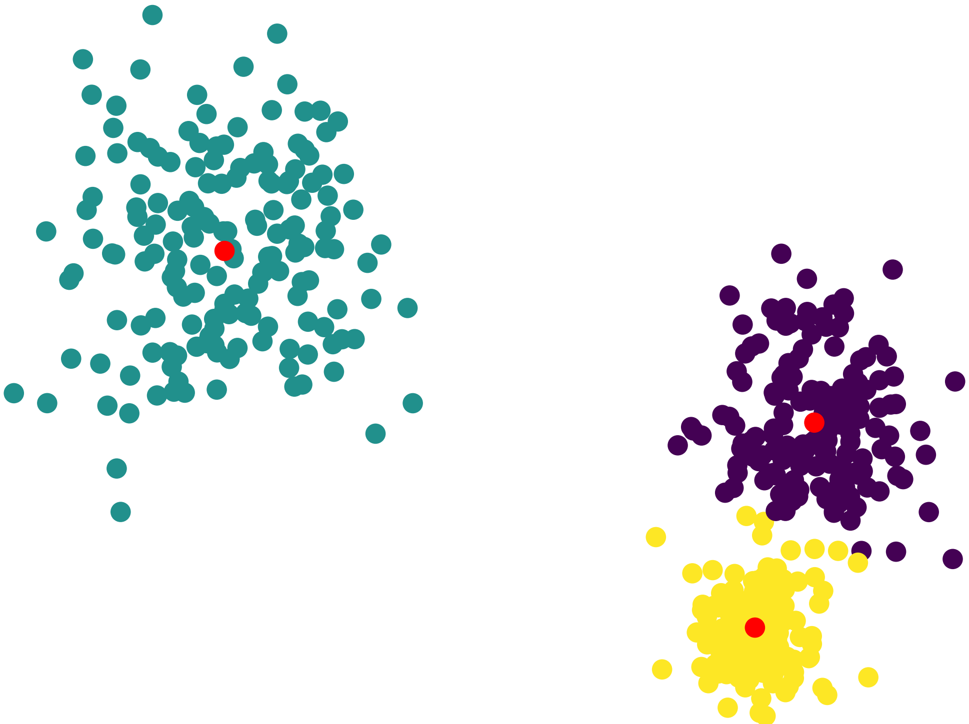
2. Update the cluster centers (means)

$$c_j^{t+1} = \frac{1}{\#\Omega_j^t} \sum_{x \in \Omega_j^t} x. \tag{2}$$

The algorithm converges when $c_j^{t+1} = c_j^t$ for all $j$.

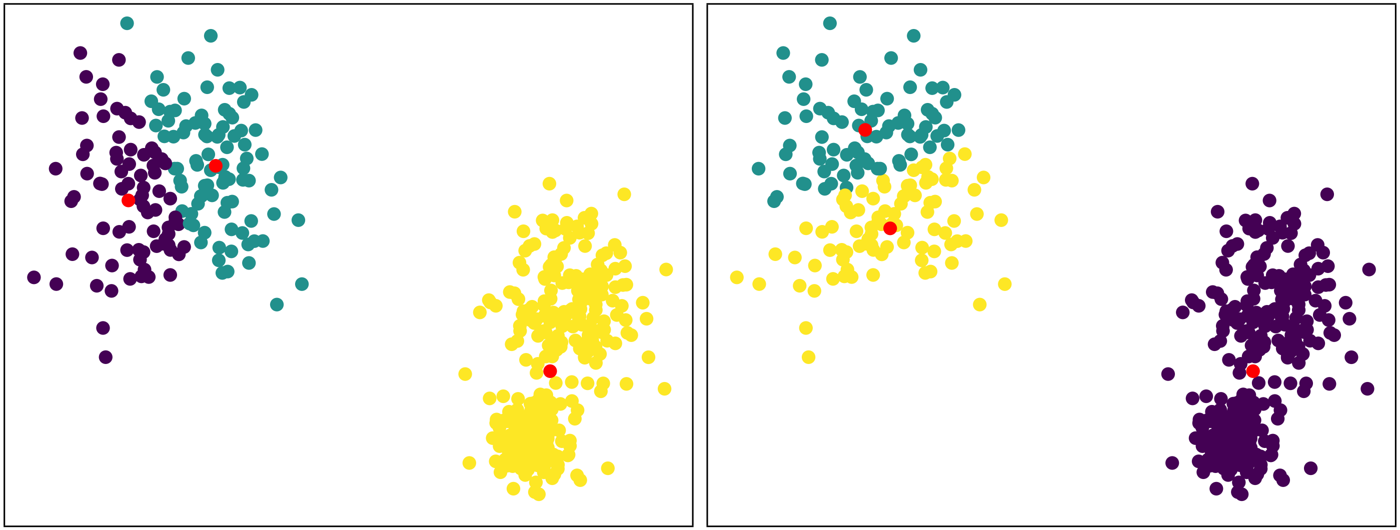
Code Demo

## $k$-means clustering result



# Poor clustering by $k$-means

Clustering depends on the random initialization.



## $k$-means clustering in Python (.ipynb)

# Lemma on centroids

**Lemma 1.** *Let $y_1, y_2, \ldots, y_m$ be points in $\mathbb{R}^n$, and define the function $f : \mathbb{R}^n \to \mathbb{R}$ by*

$$f(x) = \sum_{i=1}^{m} \|y_i - x\|^2.$$

*Then the unique minimizer of $f$ is the centroid*

$$\nabla \|Ax\|^2 = 2A^2x$$

$$c = \frac{1}{m} \sum_{i=1}^{m} y_i.$$

*In particular, $f(c) < f(x)$ if $x \neq c$.*

Proof:

$$\nabla_x f(x) = \sum_{i=1}^{m} \nabla_x \|y_i - x\|^2$$

$$= -\sum_{i=1}^{m} 2(y_i - x) = 0$$

$$\sum_{i=1}^{m} (y_i - x) = 0$$

$$\sum_{i=1}^{m} y_i = \sum_{i=1}^{m} x = mx$$

$$x = \frac{1}{m} \sum_{i=1}^{m} y_i \qquad \blacksquare$$

# Convergence of $k$-means

Recall the $k$-means enegy

$$E(c_1, c_2, \ldots, c_k) = \sum_{i=1}^{m} \min_{1 \le j \le k} \|x_i - c_j\|^2. \tag{3}$$

**Theorem 2.** *The $k$-means algorithm descends on the energy (3), that is*

$$E(c_1^{t+1}, c_2^{t+1}, \ldots, c_k^{t+1}) \le E(c_1^t, c_2^t, \ldots, c_k^t). \tag{4}$$

*Furthermore, we have equality in (4) if and only if $c_j^{t+1} = c_j^t$ for $j = 1, \ldots, k$, and hence the $k$-means algorithm converges in a finite number of iterations.*

**Note:**

- $k$-means does **not** in general find a global minimum of $E$.
- It is useful because it is fast, guaranteed to converge, and often finds good clustering.

## Proof:

rewrite energy over clusters

$$E(c_1^t, \dots, c_k^t) = \sum_{i=1}^{m} \min_{1 \le j \le k} \| x_i - c_j^t \|^2$$

$$= \sum_{j=1}^{k} \sum_{x \in \Omega_j^t} \| x - c_j^t \|^2$$

cluster center update

Lemma →

$$\ge \sum_{j=1}^{k} \sum_{x \in \Omega_j^t} \| x - c_j^{t+1} \|^2$$

$$\ge \sum_{i=1}^{m} \min_{1 \le j \le k} \| x_i - c_j^{t+1} \|^2$$

$$= E(c_1^{t+1}, c_2^{t+1}, \dots, c_k^{t+1})$$

equality iff $c_j^{t+1} = c_j^t \ \forall j$