Outline

- Kmeans Clustering
- Graph Clustering
Supervised versus Unsupervised Learning

Supervised Learning:
- Learning from labeled observations
- Classification, regression, ...

Unsupervised Learning:
- Learning from unlabeled observations
- Discover hidden patterns
- Clustering (today)
Clustering

- Given \( \{x_1, x_2, \ldots, x_n\} \) and \( K \) (number of clusters)
- Output \( A(x_i) \in \{1, 2, \ldots, K\} \) (cluster membership)
Can we split the data into two clusters?
Can we split the data into **two clusters**?
Clustering is Subjective

- Non-trivial to say one clustering is better than the other
- Each algorithm has two parts:
  - Define the **objective function**
  - Design an algorithm to **minimize this objective function**
K-means Objective Function

- Partition dataset into $C_1, C_2, \ldots, C_K$ to minimize the following objective:

$$ J = \sum_{k=1}^{K} \sum_{x \in C_k} \|x - m_k\|_2^2, $$

where $m_k$ is the mean of $C_k$. 

Multiple ways to minimize this objective

Hierarchical Agglomerative Clustering

Kmeans Algorithm (Today)
K-means Objective Function

- Partition dataset into $C_1, C_2, \ldots, C_K$ to minimize the following objective:

$$J = \sum_{k=1}^{K} \sum_{x \in C_k} \|x - m_k\|_2^2,$$

where $m_k$ is the mean of $C_k$.

- Multiple ways to minimize this objective
  - Hierarchical Agglomerative Clustering
  - Kmeans Algorithm (Today)
  - …
K-means Algorithm
K-means Algorithm

- Re-write objective:

\[ J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \left\| x_n - m_k \right\|_2^2, \]

where \( r_{nk} \in \{0, 1\} \) is an indicator variable

\[ r_{nk} = 1 \text{ if and only if } x_n \in C_k \]

- Alternative optimization between \( \{ r_{nk} \} \) and \( \{ m_k \} \)
  - Fix \( \{ m_k \} \) and update \( \{ r_{nk} \} \)
  - Fix \( \{ r_{nk} \} \) and update \( \{ m_k \} \)
K-means Algorithm

- Step 0: Initialize \( \{m_k\} \) to some values
K-means Algorithm

- Step 0: Initialize \( \{m_k\} \) to some values
- Step 1: Fix \( \{m_k\} \) and minimize over \( \{r_{nk}\} \):

\[
\begin{aligned}
r_{nk} &= \begin{cases} 
1 & \text{if } k = \arg \min_j \|x_n - m_j\|^2_2 \\
0 & \text{otherwise}
\end{cases}
\end{aligned}
\]
K-means Algorithm

- Step 0: Initialize \( \{m_k\} \) to some values
- Step 1: Fix \( \{m_k\} \) and minimize over \( \{r_{nk}\} \):
  
  \[
  r_{nk} = \begin{cases} 
  1 & \text{if } k = \arg \min_j \|x_n - m_j\|_2^2 \\
  0 & \text{otherwise}
  \end{cases}
  \]

- Step 2: Fix \( \{r_{nk}\} \) and minimize over \( \{m_k\} \):
  
  \[
  m_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}}
  \]
K-means Algorithm

- Step 0: Initialize \( \{m_k\} \) to some values
- Step 1: Fix \( \{m_k\} \) and minimize over \( \{r_{nk}\} \):
  \[
  r_{nk} = \begin{cases} 
  1 & \text{if } k = \arg \min_j \|x_n - m_j\|^2_2 \\
  0 & \text{otherwise}
  \end{cases}
  \]
- Step 2: Fix \( \{r_{nk}\} \) and minimize over \( \{m_k\} \):
  \[
  m_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}}
  \]
- Step 3: Return to step 1 unless stopping criterion is met
K-means Algorithm

Equivalent to the following procedure:

- Step 0: Initialize centers \( \{ m_k \} \) to some values
- Step 1: Assign each \( x_n \) to the nearest center:
  \[
  A(x_n) = \arg \min_j \| x_n - m_j \|_2^2
  \]
  Update clusters:
  \[
  C_k = \{ x_n : A(x_n) = k \} \quad \forall k = 1, \ldots, K
  \]
- Step 2: Calculate mean of each cluster \( C_k \):
  \[
  m_k = \frac{1}{|C_k|} \sum_{x_n \in C_k} x_n
  \]
- Step 3: Return to step 1 unless stopping criterion is met
More on K-means Algorithm

- Always \textbf{decrease} the objective function for each update
- Objective function will keep unchanged when step 1 doesn’t change cluster assignment $\Rightarrow$ Converged
More on K-means Algorithm

- Always decrease the objective function for each update
- Objective function will keep unchanged when step 1 doesn’t change cluster assignment ⇒ Converged
- May not converge to global minimum
  
  Sensitive to initial values

Kmeans++: A better way to initialize the clusters
More on K-means Algorithm

- Always decrease the objective function for each update
- Objective function will keep unchanged when step 1 doesn’t change cluster assignment ⇒ Converged
- May not converge to global minimum
  - Sensitive to initial values
- Kmeans++: A better way to initialize the clusters
Coming up

- Clustering

Questions?