1.

Denote that $\bar{X}_k = (\bar{X}_{1k}, \bar{X}_{2k}, \ldots, \bar{X}_{pk})$, which is the mean vector associated with the K cluster.

$$\frac{1}{2}\sum_{k=1}^{k}\sum_{c(i)=k}\sum_{c(i')=k}\|X_i - X_{i'}\|^2$$

$$= \frac{1}{2}\sum_{k=1}^{k}\sum_{c(i)=k}\sum_{c(i')=k}\|(X_i - \bar{X}_k) - (X_{i'} - \bar{X}_k)\|^2$$

$$= \frac{1}{2}\sum_{k=1}^{k}\sum_{c(i)=k}\sum_{c(i')=k}\left(\|X_i - \bar{X}_k\|^2 + \|X_{i'} - \bar{X}_k\|^2 - 2\langle X_i - \bar{X}_k, X_{i'} - \bar{X}_k\rangle\right)$$

$$= \frac{1}{2}\sum_{k=1}^{k}\Big(N_k\big(\sum_{c(i)=k}\|X_i - \bar{X}_k\|^2 + \sum_{c(i')=k}\|X_{i'} - \bar{X}_k\|^2\big)$$

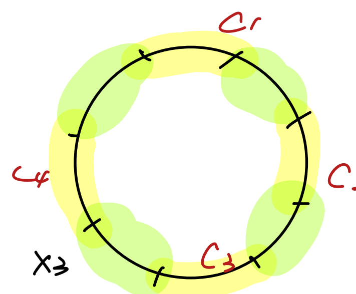
$$- 2\langle \sum_{c(i)=k}(X_i - \bar{X}_k), \sum_{c(i')=k}(X_{i'} - \bar{X}_k)\rangle$$

$$= \sum_{k=1}^{k} N_k \sum_{c(i)=k}\|X_i - \bar{X}_k\|^2, \text{ where } N_k = \sum_{i=1}^{N} I(C(i)=k)$$

2.

(1) within point scatter

$$\Rightarrow W(C) = \frac{1}{2}\sum_{k=1}^{k}\sum_{c(i)=k}\sum_{c(i')=k}\|X_i - X_{i'}\|^2$$



C denote the cluster

$$= \frac{1}{2}\sum_{k=1}^{k}\left[\|X_1 - X_1\|^2 + \|X_1 - X_2\|^2 + \|X_2 - X_1\|^2 + \|X_2 - X_2\|^2\right]$$

$$= \sum_{k=1}^{k}(\|X_1 - X_2\|^2)$$

$$\|X_1 - X_2\|^2 = \|[\cos(\tfrac{3}{2}\pi - \varepsilon), \sin(\tfrac{3}{2}\pi - \varepsilon)] - [\cos(\tfrac{3}{2}\pi + \varepsilon), \sin(\tfrac{3}{2}\pi + \varepsilon)]\|^2$$

$$= (2 \sin\tfrac{3}{2}\pi \sin\varepsilon)^2 + (-2\sin\varepsilon \cos\tfrac{3}{2}\pi)^2$$

$\because \cos\frac{3}{2}\pi = 0, \quad \sin\frac{3}{2}\pi = -1$

$$\Rightarrow \|X_1 - X_2\|^2 = 4\sin^2\varepsilon$$

$$\Rightarrow W(C) = \sum_{k=1}^{4} 4\sin^2\varepsilon = 16\sin^2\varepsilon \quad \#$$

Denote C' as the cluster

$$W(C') = \sum_{k=1}^{4} \|X_3 - X_1\|^2$$

$$= \sum_{k=1}^{4} \|[\cos(\pi+\varepsilon), \sin(\pi+\varepsilon)] - [\cos(\tfrac{3}{2}\pi - \varepsilon), \sin(\tfrac{3}{2}\pi - \varepsilon)]\|^2$$

$$= \sum_{k=1}^{4} [(-\cos\varepsilon - \sin\varepsilon)^2 + (-\sin\varepsilon - \cos\varepsilon)^2]$$

$$= 8(\cos\varepsilon + \sin\varepsilon)^2 \quad \#$$

Since other cluster will have larger within point scatter, so, we choose only to compare $W(C), W(C')$

Let $\varepsilon > 0$ and ε be sufficient small $\Rightarrow \varepsilon \to 0$

$$\Rightarrow \left.\begin{array}{l} W(C) = 16\sin^2\varepsilon \to 0 \\ W(C') = 8(\cos\varepsilon + \sin\varepsilon)^2 \to 8 \end{array}\right\} W(C) < W(C')$$

⇒ The global minimun of the minimum of the within point scatter is achieved by pairs of points on the circle with angle differe 2ε.

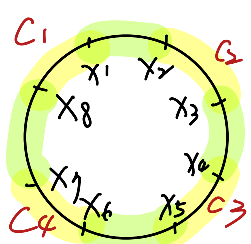
(2)

```python
import numpy as np
from sklearn.cluster import KMeans

# Set the number of data points and the value of epsilon
n_data_points = 8
epsilon = 0.01

# Generate data points on the unit circle
angles = []
angles+=[m * np.pi / 2 + epsilon for m in range(1, 5)]
angles+=[m * np.pi / 2 - epsilon for m in range(1, 5)]

data_points = np.array([(np.cos(angle), np.sin(angle)) for angle in angles])

# Set the number of clusters and the number of iterations
K = 4
n_iterations = 10

# Run K-means clustering with random initialization 10 times and find the minimum within-point scatter
min_within_point_scatter = float('inf')
for i in range(n_iterations):
    kmeans = KMeans(n_clusters=K, init='random', n_init=1, random_state=None)
    kmeans.fit(data_points)

    within_point_scatter = kmeans.inertia_
    min_within_point_scatter = min(min_within_point_scatter, within_point_scatter)

print(f"Minimum within-point scatter achieved in the simulation: {min_within_point_scatter}")
```

✓ 0.0s

```
Minimum within-point scatter achieved in the simulation: 0.0007999733336888611
```

(3)



Applying Lloyd's Algorithm, if the four red dots above happen to be the four initial center and each of the are exactly the mean of the closet 2 pts, then by the algorithm to assign the each data pt to the closet center, we have

$C_1 = \{x_1, x_8\}$, $C_2 = \{x_2, x_3\}$, $C_3 = \{x_4, x_5\}$, $C_4 = \{x_6, x_9\}$

And after that, since all center are already the mean of the data pts it contains, so it reaches local minimum. However, by Q1, we know that this is not the global optimal.

(4) For H-W algorithm, if centroid are initiated as $\{(\cos\theta, \sin\theta),\ \theta = \frac{n\pi}{4},\ n = 1, 3, 5, 7\}$, then, since each of the data points will first be randomly assigned to a centroid (not assigned to the closet one), and after that, they will be reassigned to the centroid with the smallest total sum of squared distance, and then the centroid will be recalculated by the mean of its data points.

And this makes the "local minimum for Lloyd's algorithm", but not likely to be the "local minimum for H-W's algorithm".

3,

We have the Gaussian 2-mixture model:

$$p_x = \frac{1}{2}N(\mu_1, 1) + \frac{1}{2}N(\mu_2, 1)$$

$$= \frac{1}{2}\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2}}\right] + \frac{1}{2}\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2}}\right],$$

$$p(z) = \frac{1}{2}, \quad p(z, x) = p(z)\,p(x|z)$$

$$\Rightarrow p(z, x) = \begin{cases} \frac{1}{2}\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2}}\right), & z = 1 \\ \frac{1}{2}\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2}}\right), & z = 2 \end{cases}$$

$$p(z_i = 1 | x, \theta) = \frac{p(z_i = 1, x | \theta)}{p(x|\theta)}$$

$$= \frac{1}{1 + \exp\left(-\frac{1}{2}[(x-\mu_2)^2 - (x-\mu_1)^2]\right)}$$

$$= \frac{1}{1 + \exp\{-(\mu_1 - \mu_2)(x - \frac{1}{2}(\mu_1 + \mu_2))\}}$$

Then, we have

① $$p(z_i = 1 | x_i, \theta) = \frac{1}{1 + e^{-(\mu_1 - \mu_2)(x_i - \frac{1}{2}(\mu_1 + \mu_2))}}$$

② $$p(z_i = 2 | x_i, \theta) = \frac{1}{1 + e^{-(\mu_2 - \mu_1)(x_i - \frac{1}{2}(\mu_1 + \mu_2))}}$$

(2)

$$Q(\beta,\alpha) := \mathbb{E}_{p(z^n \mid X^n, \alpha)}[\log p(z^n, X^n \mid \beta)]$$

$$= \mathbb{E}_{p(z^n \mid X^n, \alpha)}[\log \prod_{i=1}^{n} p(z_i, X_i \mid \beta)]$$

$$= \mathbb{E}_{p(z^n \mid X^n, \alpha)}[\sum_{i=1}^{n} \log p(z_i, X_i \mid \beta)]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{2} p(z_i = j \mid X_i, \alpha) \log p(z_i = j, X_i \mid \beta)$$

$$= \sum_{i=1}^{n}\Big[\frac{1}{1+\exp\{-(\alpha_1-\alpha_2)(X_i - \frac{1}{2}(\alpha_1+\alpha_2))\}} \times [-\log 2 - \frac{1}{2}\log 2\pi - \frac{1}{2}(X_i-\beta_1)^2] + \frac{1}{1+\exp\{-(\alpha_2-\alpha_1)(X_i - \frac{1}{2}(\alpha_1+\alpha_2))\}} \times [-\log 2 - \frac{1}{2}\log 2\pi - \frac{1}{2}(X_i-\beta_2)^2] \Big]$$

//

(3) Goal: argmax $Q(\beta,\alpha)$ in terms of X^n and α

Let $\beta_1 = \mu_1^T$, $\beta_2 = \mu_2^T$

$$\begin{cases} \frac{d}{d\beta_1} Q(\beta,\alpha) = \sum_{i=1}^{n} p(z_i = 1 \mid X_i, \alpha)(X_i - \beta_1) = 0 & \text{— ①} \\ \frac{d}{d\beta_2} Q(\beta,\alpha) = \sum_{i=1}^{n} p(z_i = 2 \mid X_i, \alpha)(X_i - \beta_2) = 0 & \text{— ②} \end{cases}$$

By ①

$$\hat{B}_1 = M_1^T = \frac{\sum_{i=1}^{n} p(z_i = 1 \mid x_i, \alpha)\, x_i}{\sum_{i=1}^{n} p(z_i = 1 \mid x_i, \alpha)}$$

$$= \frac{\sum \frac{x_i}{1 + \exp\{-(\alpha_1 - \alpha_2)(x_i - \frac{1}{2}(\alpha_1 + \alpha_2))\}}}{\sum \frac{1}{1 + \exp\{-(\alpha_1 - \alpha_2)(x_i - \frac{1}{2}(\alpha_1 + \alpha_2))\}}}$$

By ②

$$\hat{B}_2 = M_2^T = \frac{\sum_{i=1}^{n} p(z_i = 2 \mid x_i, \alpha)\, x_i}{\sum_{i=1}^{n} p(z_i = 2 \mid x_i, \alpha)}$$

$$= \frac{\sum \frac{x_i}{1 + \exp\{-(\alpha_2 - \alpha_1)(x_i - \frac{1}{2}(\alpha_1 + \alpha_2))\}}}{\sum \frac{1}{1 + \exp\{-(\alpha_2 - \alpha_1)(x_i - \frac{1}{2}(\alpha_1 + \alpha_2))\}}}$$