I Pen-and-paper

1. - Expectation

$$P(x_n \mid c_k) = \mathcal{N}(x_n \mid \mu_k, \Sigma_k) = \tag{1}$$
$$= \frac{1}{(2\pi)^{\frac{D}{2}} \cdot |\Sigma_k|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(x-\mu_k)^\top \cdot (\Sigma_k)^{-1} \cdot (x-\mu_k)}$$

$$P(x_1 \mid c_1) = \mathcal{N}(x_1 \mid \mu_1, \Sigma_1) =$$
$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)^\top \cdot \left(\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)}$$
$$= 0.066$$

$$P(x_1 \mid c_2) = \mathcal{N}(x_1 \mid \mu_2, \Sigma_2) =$$
$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^\top \cdot \left(\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)}$$
$$= 0.023$$

$$P(x_2 \mid c_1) = \mathcal{N}(x_2 \mid \mu_1, \Sigma_1) =$$
$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)^\top \cdot \left(\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)}$$
$$= 0.009$$

$$P(x_2 \mid c_2) = \mathcal{N}(x_2 \mid \mu_2, \Sigma_2) =$$
$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^\top \cdot \left(\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)}$$
$$= 0.048$$

$$P(x_3 \mid c_1) = \mathcal{N}(x_3 \mid \mu_1, \Sigma_1) =$$

$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)^{\top} \cdot \left(\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)}$$

$$= 0.034$$

$$P(x_3 \mid c_2) = \mathcal{N}(x_3 \mid \mu_2, \Sigma_2) =$$

$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^{\top} \cdot \left(\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)}$$

$$= 0.062$$

$$P(x_n, c_k) = P(x_n \mid c_k) \cdot \pi_k = \mathcal{N}(x_n \mid \mu_k, \Sigma_k) \cdot \pi_k \tag{2}$$

$$P(x_1, c_1) = \mathcal{N}(x_1 \mid \mu_1, \Sigma_1) \cdot \pi_1 = 0.066 \times 0.5 = 0.033$$

$$P(x_1, c_2) = \mathcal{N}(x_1 \mid \mu_2, \Sigma_2) \cdot \pi_2 = 0.023 \times 0.5 = 0.011$$

$$P(x_2, c_1) = \mathcal{N}(x_2 \mid \mu_1, \Sigma_1) \cdot \pi_1 = 0.009 \times 0.5 = 0.004$$

$$P(x_2, c_2) = \mathcal{N}(x_2 \mid \mu_2, \Sigma_2) \cdot \pi_2 = 0.048 \times 0.5 = 0.024$$

$$P(x_3, c_1) = \mathcal{N}(x_3 \mid \mu_1, \Sigma_1) \cdot \pi_1 = 0.034 \times 0.5 = 0.017$$

$$P(x_3, c_2) = \mathcal{N}(x_3 \mid \mu_2, \Sigma_2) \cdot \pi_2 = 0.062 \times 0.5 = 0.031$$

$$P(x_n) = \sum_{k=1}^{3} P(x_n, c_k) = \sum_{k=1}^{3} (\pi_k \cdot \mathcal{N}(x_n \mid \mu_k, \Sigma_k)) \tag{3}$$

$$P(x_1) = \sum_{k=1}^{3} (\pi_k \cdot \mathcal{N}(x_1 \mid \mu_k, \Sigma_k)) = 0.033 + 0.011 = 0.044$$

$$P(x_2) = \sum_{k=1}^{3} (\pi_k \cdot \mathcal{N}(x_2 \mid \mu_k, \Sigma_k)) = 0.004 + 0.024 = 0.029$$

$$P(x_3) = \sum_{k=1}^{3}(\pi_k \cdot \mathcal{N}(x_3 \mid \mu_k, \Sigma_k)) = 0.017 + 0.031 = 0.048$$

$$\gamma_{kn} = P(c_k \mid x_n) = \frac{P(x_n, c_k)}{P(x_n)} \tag{4}$$

$$\gamma_{11} = \frac{P(x_1, c_1)}{P(x_1)} = \frac{0.033}{0.044} = 0.743$$

$$\gamma_{21} = \frac{P(x_1, c_2)}{P(x_1)} = \frac{0.011}{0.044} = 0.257$$

$$\gamma_{12} = \frac{P(x_2, c_1)}{P(x_2)} = \frac{0.004}{0.029} = 0.156$$

$$\gamma_{22} = \frac{P(x_2, c_2)}{P(x_2)} = \frac{0.024}{0.029} = 0.844$$

$$\gamma_{13} = \frac{P(x_3, c_1)}{P(x_3)} = \frac{0.017}{0.048} = 0.353$$

$$\gamma_{23} = \frac{P(x_3, c_2)}{P(x_3)} = \frac{0.031}{0.048} = 0.647$$

- Maximization

$$N_k = \sum_{n=1}^{2} \gamma_{kn} \tag{5}$$

$$N_1 = \sum_{n=1}^{3} \gamma_{1n} = 0.743 + 0.156 + 0.353 = 1.252$$

$$N_2 = \sum_{n=1}^{3} \gamma_{2n} = 0.257 + 0.844 + 0.647 = 1.748$$

$$\mu_k = \frac{1}{N_k} \cdot \sum_{n=1}^{3} \gamma_{kn} \cdot x_n \tag{6}$$

$$\begin{aligned}
\mu_1 &= \frac{1}{N_1} \cdot \sum_{n=1}^{3} \gamma_{1n} \cdot x_n \\
&= \frac{1}{1.252} \cdot \left(0.743 \cdot \begin{bmatrix}1\\2\end{bmatrix} + 0.156 \cdot \begin{bmatrix}-1\\1\end{bmatrix} + 0.353 \cdot \begin{bmatrix}1\\0\end{bmatrix}\right) \\
&= \begin{bmatrix}0.751\\1.311\end{bmatrix}
\end{aligned}$$

$$
\begin{aligned}
\mu_2 &= \frac{1}{N_2} \cdot \sum_{n=1}^{3} \gamma_{2n} \cdot x_n \\
&= \frac{1}{1.748} \cdot \left(0.257 \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0.844 \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix} + 0.647 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \\
&= \begin{bmatrix} 0.034 \\ 0.777 \end{bmatrix}
\end{aligned}
$$

$$
\Sigma_k = \frac{1}{N_k} \cdot \sum_{n=1}^{3} \gamma_{kn} \cdot (x_n - \mu_k) \cdot (x_n - \mu_k)^\top \tag{7}
$$

$$
\begin{aligned}
\Sigma_1 &= \frac{1}{N_1} \cdot \sum_{n=1}^{3} \gamma_{1n} \cdot (x_n - \mu_1) \cdot (x_n - \mu_1)^\top \\
&= \frac{1}{1.252} \cdot \Bigg(0.743 \cdot \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.751 \\ 1.311 \end{bmatrix}\right) \cdot \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.751 \\ 1.311 \end{bmatrix}\right)^\top \\
&\qquad + 0.156 \cdot \left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0.751 \\ 1.311 \end{bmatrix}\right) \cdot \left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0.751 \\ 1.311 \end{bmatrix}\right)^\top \\
&\qquad + 0.353 \cdot \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.751 \\ 1.311 \end{bmatrix}\right) \cdot \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.751 \\ 1.311 \end{bmatrix}\right)^\top \Bigg) \\
&= \begin{bmatrix} 0.436 & 0.078 \\ 0.078 & 0.778 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\Sigma_2 &= \frac{1}{N_2} \cdot \sum_{n=1}^{3} \gamma_{2n} \cdot (x_n - \mu_2) \cdot (x_n - \mu_2)^\top \\
&= \frac{1}{1.748} \cdot \Bigg(0.257 \cdot \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.034 \\ 0.777 \end{bmatrix}\right) \cdot \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.034 \\ 0.777 \end{bmatrix}\right)^\top \\
&\qquad + 0.844 \cdot \left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0.034 \\ 0.777 \end{bmatrix}\right) \cdot \left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0.034 \\ 0.777 \end{bmatrix}\right)^\top \\
&\qquad + 0.647 \cdot \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.034 \\ 0.777 \end{bmatrix}\right) \cdot \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.034 \\ 0.777 \end{bmatrix}\right)^\top \Bigg) \\
&= \begin{bmatrix} 0.999 & -0.215 \\ -0.215 & 0.467 \end{bmatrix}
\end{aligned}
$$

$$
\pi_k = \frac{N_k}{N} \tag{8}
$$

$$
\pi_1 = \frac{N_k}{N} = \frac{1.252}{3} = 0.417
$$

$$\pi_2 = \frac{N_k}{N} = \frac{1.748}{3} = 0.583$$

2. (a)

$$P(x_n \mid c_k) = \mathcal{N}(x_n \mid \mu_k, \Sigma_k) = \tag{9}$$
$$= \frac{1}{(2\pi)^{\frac{D}{2}} \cdot |\Sigma_k|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(x-\mu_k)^\top \cdot (\Sigma_k)^{-1} \cdot (x-\mu_k)}$$

$$P(x_1 \mid c_1) = \mathcal{N}(x_1 \mid \mu_1, \Sigma_1) =$$
$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 0.436 & 0.078 \\ 0.078 & 0.778 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}0.751\\1.311\end{bmatrix}\right)^\top \cdot \left(\begin{bmatrix}0.436 & 0.078\\0.078 & 0.778\end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}0.751\\1.311\end{bmatrix}\right)}$$
$$= 0.196$$

$$P(x_1 \mid c_2) = \mathcal{N}(x_1 \mid \mu_2, \Sigma_2) =$$
$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 0.999 & -0.215 \\ -0.215 & 0.467 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}0.034\\0.777\end{bmatrix}\right)^\top \cdot \left(\begin{bmatrix}0.999 & -0.215\\-0.215 & 0.467\end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}0.034\\0.777\end{bmatrix}\right)}$$
$$= 0.014$$

$$P(x_2 \mid c_1) = \mathcal{N}(x_2 \mid \mu_1, \Sigma_1) =$$
$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 0.436 & 0.078 \\ 0.078 & 0.778 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix}-1\\1\end{bmatrix} - \begin{bmatrix}0.751\\1.311\end{bmatrix}\right)^\top \cdot \left(\begin{bmatrix}0.436 & 0.078\\0.078 & 0.778\end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix}-1\\1\end{bmatrix} - \begin{bmatrix}0.751\\1.311\end{bmatrix}\right)}$$
$$= 0.008$$

$$P(x_2 \mid c_2) = \mathcal{N}(x_2 \mid \mu_2, \Sigma_2) =$$
$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 0.999 & -0.215 \\ -0.215 & 0.467 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix}-1\\1\end{bmatrix} - \begin{bmatrix}0.034\\0.777\end{bmatrix}\right)^\top \cdot \left(\begin{bmatrix}0.999 & -0.215\\-0.215 & 0.467\end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix}-1\\1\end{bmatrix} - \begin{bmatrix}0.034\\0.777\end{bmatrix}\right)}$$
$$= 0.144$$

$$P(x_3 \mid c_1) = \mathcal{N}(x_3 \mid \mu_1, \Sigma_1) =$$

$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 0.436 & 0.078 \\ 0.078 & 0.778 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.751\\1.311\end{bmatrix}\right)^{\top} \cdot \left(\begin{bmatrix}0.436 & 0.078\\0.078 & 0.778\end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.751\\1.311\end{bmatrix}\right)}$$

$$= 0.077$$

$$P(x_3 \mid c_2) = \mathcal{N}(x_3 \mid \mu_2, \Sigma_2) =$$

$$= \frac{1}{(2\pi)^{\frac{2}{2}} \cdot \begin{vmatrix} 0.999 & -0.215 \\ -0.215 & 0.467 \end{vmatrix}^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.034\\0.777\end{bmatrix}\right)^{\top} \cdot \left(\begin{bmatrix}0.999 & -0.215\\-0.215 & 0.467\end{bmatrix}\right)^{-1} \cdot \left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0.034\\0.777\end{bmatrix}\right)}$$

$$= 0.105$$

$$c_i = \arg\max_{c_i} P(c_i \mid x) = \arg\max_{c_i} \frac{P(c_i)P(x \mid c_i)}{P(x)} = \arg\max_{c_i} P(c_i)P(x \mid c_i)$$

$$\begin{aligned}
\text{cluster}(x_1) &= \arg\max_{c_i}\{P(c_i)P(x_1 \mid c_i)\} \\
&= \arg\max_{c_i}\{\pi_1 P(x_1 \mid c_1), \pi_2(P(x_1 \mid c_2))\} \\
&= \arg\max_{c_i}\{0.417 \times 0.196, 0.583 \times 0.014\} \\
&= \arg\max_{c_i}\{0.082, 0.008\} \\
&= c_1
\end{aligned}$$

$$\begin{aligned}
\text{cluster}(x_2) &= \arg\max_{c_i}\{P(c_i)P(x_2 \mid c_i)\} \\
&= \arg\max_{c_i}\{\pi_1 P(x_2 \mid c_1), \pi_2(P(x_2 \mid c_2))\} \\
&= \arg\max_{c_i}\{0.417 \times 0.008, 0.583 \times 0.144\} \\
&= \arg\max_{c_i}\{0.003, 0.084\} \\
&= c_2
\end{aligned}$$

$$
\begin{aligned}
\text{cluster}(x_3) &= \arg\max_{c_i}\{P(c_i)P(x_3 \mid c_i)\} \\
&= \arg\max_{c_i}\{\pi_1 P(x_3 \mid c_1), \pi_2(P(x_3 \mid c_2))\} \\
&= \arg\max_{c_i}\{0.417 \times 0.077, 0.583 \times 0.105\} \\
&= \arg\max_{c_i}\{0.032, 0.061\} \\
&= c_2
\end{aligned}
$$

(b) The cluster c_2 is the largest of the two, as its covariance matrix has greater values.

$$
\begin{aligned}
\text{silhouette}(x_2) &= \frac{d_2(X_2, X_1) - d_2(X_2, X_3)}{\max\{d_2(X_2, X_3), d_2(X_2, X_1)\}} \\
&= \frac{d_2\left(\begin{bmatrix}-1\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}\right) - d_2\left(\begin{bmatrix}-1\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}\right)}{\max\left\{d_2\left(\begin{bmatrix}-1\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}\right), d_2\left(\begin{bmatrix}-1\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}\right)\right\}} \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
\text{silhouette}(x_3) &= \frac{d_2(X_3, X_1) - d_2(X_3, X_2)}{\max\{d_2(X_3, X_2), d_2(X_3, X_1)\}} \\
&= \frac{d_2\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}\right) - d_2\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}-1\\1\end{bmatrix}\right)}{\max\left\{d_2\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}-1\\1\end{bmatrix}\right), d_2\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}\right)\right\}} \\
&= -0.106
\end{aligned}
$$

$$
\begin{aligned}
\text{silhouette}(c_2) &= 0.5\text{silhouette}(x_2) + 0.5\text{silhouette}(x_3) \\
&= 0.5 \times 0 + 0.5 \times (-0.106) \\
&= -0.053
\end{aligned}
$$

II Programming

1. - Seed = 0
     - Silhouette: 0.114
     - Purity: 0.767
   - Seed = 1
     - Silhouette: 0.114
     - Purity: 0.763
   - Seed = 2
     - Silhouette: 0.114
     - Purity: 0.767
2. The non-determinism comes from the random initialization of the centroids.
3.

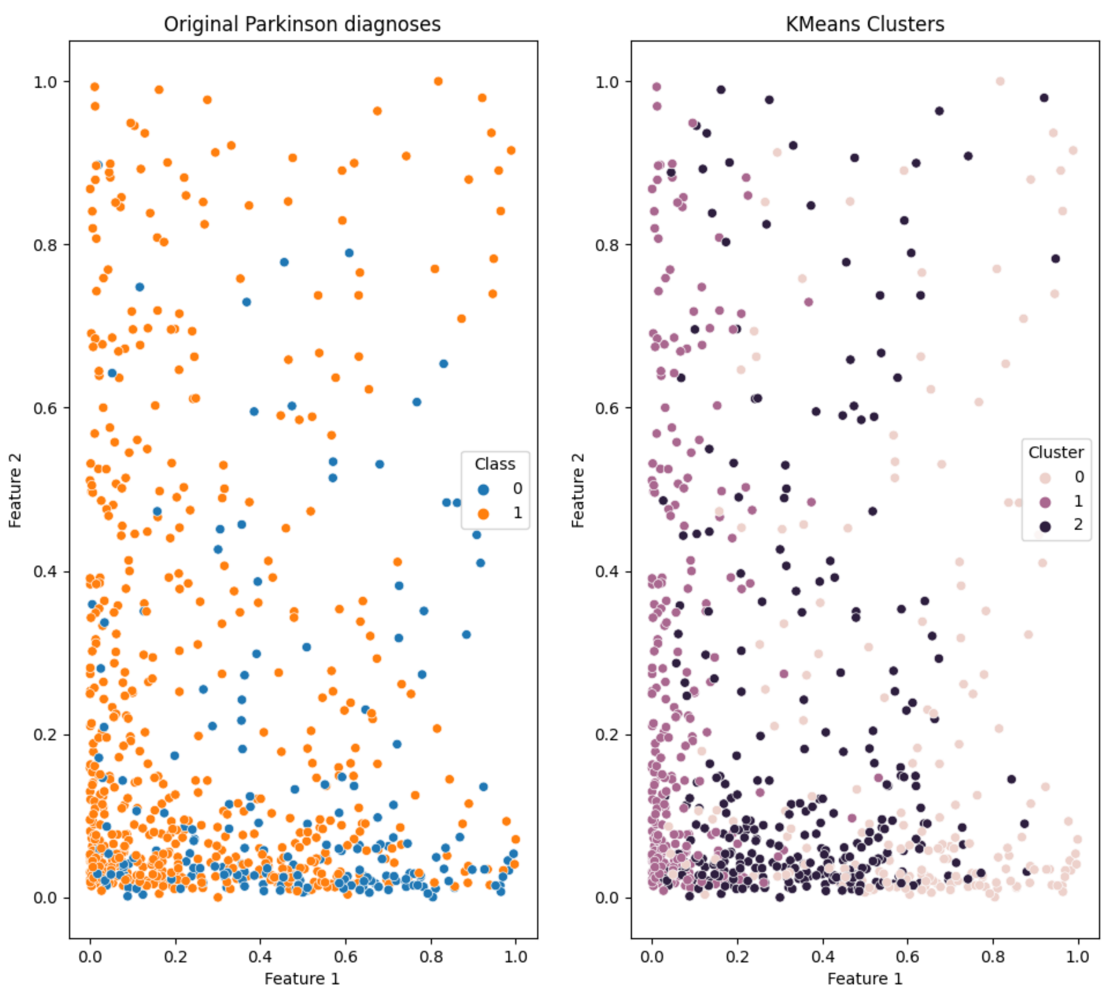

Figure 1: Scatter Plot

4. The number of principal components required to explain more than 80% of the variance is 31.

III Appendix

```
from scipy.io.arff import loadarff
from sklearn import cluster
from sklearn import metrics
from sklearn.decomposition import PCA
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns


def purity_score(y_true, y_pred):
    confusion_matrix = metrics.cluster.contingency_matrix(y_true, y_pred)
    return np.sum(np.amax(confusion_matrix, axis=0)) / np.sum(confusion_matrix)


if __name__ == "__main__":
    data = loadarff("pd_speech.arff")
    df = pd.DataFrame(data[0])
    df["class"] = df["class"].astype(int)
    X, y = df.drop("class", axis=1), df["class"]
    X = MinMaxScaler().fit_transform(X)

    LABELS = None
    for seed in (0, 1, 2):
        kmeans = cluster.KMeans(n_clusters=3, random_state=seed).fit(X)
        print("Seed {} Silhouette: {}".format(
            seed, metrics.silhouette_score(X, kmeans.labels_)))
        print("Seed {} Purity: {}".format(seed,
                                          purity_score(y, kmeans.labels_)))
        if (seed == 0):
            LABELS = kmeans.labels_

    X_variance = X[:, np.argsort(np.var(X, axis=0))[::-1][:2]]
    f1, f2 = X_variance[:, 0], X_variance[:, 1]

    _, ax = plt.subplots(1, 2)
    sns.scatterplot(x=f1, y=f2, hue=y, ax=ax[0])
    ax[0].set_title("Original Parkinson diagnoses")
    ax[0].legend(loc="center right", title="Class")
    ax[0].set_xlabel("Feature 1")
    ax[0].set_ylabel("Feature 2")

    sns.scatterplot(x=f1, y=f2, hue=LABELS, ax=ax[1])
    ax[1].set_title("KMeans Clusters")
    ax[1].legend(loc="center right", title="Cluster")
    ax[1].set_xlabel("Feature 1")
    ax[1].set_ylabel("Feature 2")

    plt.show()

    pca = PCA(n_components=0.8, svd_solver='full')
    pca.fit(X)
    print("{} principal components are necessary.".format(pca.n_components_))
```