1.

- Observations: $x = \left\{ \begin{pmatrix} 1 \\ 0.6 \\ 0.1 \end{pmatrix}, \begin{pmatrix} 0 \\ -0.4 \\ 0.8 \end{pmatrix}, \begin{pmatrix} 0 \\ 0.2 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 1 \\ 0.4 \\ -0.1 \end{pmatrix} \right\}$

Most of our calculations were made using programming functions as a helpful resource.

- $y_1 \perp\!\!\!\perp \{y_2, y_3\}$ — y_1: Bernoulli; $\{y_2, y_3\}$: Gaussian

- $p_1 = P(y_1 = 1) = 0.3$
- $\pi_1 = 0.5$
- $N_1\left(\mu_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}\right)$

- $p_2 = P(y_1 = 1) = 0.7$
- $\pi_2 = 0.5$
- $N_2\left(\mu_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1.5 & 1 \\ 1 & 1.5 \end{pmatrix}\right)$

E-Step:

Cluster 1

- $p(x_1 | C=1) = p_1 \times N(x_1 | \mu_1, \Sigma_1) = p_1 \times \left(\frac{1}{\sqrt{2\pi \cdot |\Sigma_1|}} \cdot \exp\left(-\frac{1}{2} \cdot (x_1 - \mu_1)^T \cdot \Sigma_1^{-1} \cdot (x_1 - \mu_1)\right)\right) = 0.01997$
- $p(x_2 | C=1) = (1 - p_1) \times N(x_2 | \mu_1, \Sigma_1) = 0.03503$
- $p(x_3 | C=1) = (1 - p_1) \times N(x_3 | \mu_1, \Sigma_1) = 0.04786$
- $p(x_4 | C=1) = p_1 \times N(x_4 | \mu_1, \Sigma_1) = 0.01771$

Xn variable will only use y_2, y_3 (normally distrabuted)

Cluster 2

- $p(x_1 | C=2) = p_2 \times N(x_1 | \mu_2, \Sigma_2) = p_2 \times \left(\frac{1}{\sqrt{2\pi \cdot |\Sigma_2|}} \cdot \exp\left(-\frac{1}{2} \cdot (x_1 - \mu_2)^T \cdot \Sigma_2^{-1} \cdot (x_1 - \mu_2)\right)\right) = 0.08373$
- $p(x_2 | C=2) = (1 - p_2) \times N(x_2 | \mu_2, \Sigma_2) = 0.02046$
- $p(x_3 | C=2) = (1 - p_2) \times N(x_3 | \mu_2, \Sigma_2) = 0.03887$
- $p(x_4 | C=2) = p_2 \times N(x_4 | \mu_1, \Sigma_1) = 0.08715$

Joint Probabilities

Cluster 1

- $p(C=1, x_1) = \pi_1 \times N(x_1 | \mu_1, \Sigma_1) = 0.00997$
- $p(C=1, x_2) = 0.01752$
- $p(C=1, x_3) = 0.02393$
- $p(C=1, x_4) = 0.00886$

Cluster 2

- $p(C=2, x_1) = \pi_2 \times N(x_1 | \mu_2, \Sigma_2) = 0.04187$
- $p(C=2, x_2) = 0.01023$
- $p(C=2, x_3) = 0.01944$
- $p(C=2, x_4) = 0.04358$

Cluster 1

- $\gamma(c_{11}) = p(c=1 \mid x_1) = \dfrac{p(c=1, x_1)}{p(x_1)^{*}} = 0.19259$
- $\gamma(c_{21}) = 0.63135$
- $\gamma(c_{31}) = 0.55181$
- $\gamma(c_{41}) = 0.16892$

$$\boxed{{}^{*}\; p(x_n) = \sum_{k=1}^{K} p(c_k=1, x_n)}$$

Cluster 2

- $\gamma(c_{12}) = p(c=2 \mid x_1) = \dfrac{p(c=2, x_1)}{p(x_1)^{*}} = 0.80741$
- $\gamma(c_{22}) = 0.36865$
- $\gamma(c_{32}) = 0.44819$
- $\gamma(c_{42}) = 0.83108$

M-Step

- $N_k = \sum_{n=1}^{N} \gamma(c_{nk})$
- $\mu_k = \dfrac{1}{N_k} \cdot \sum_{n=1}^{N} \gamma(c_{nk}) \cdot x_n$
- $\Sigma_k = \dfrac{1}{N_k} \cdot \sum_{n=1}^{N} \gamma(c_{nk}) \cdot (x_n - \mu_k) \cdot (x_n - \mu_k)^T$
- $\pi_k = p(c_k = 1) = \dfrac{N_k}{N}$

only the X_n attributes y_2, y_3 are included in this calculation, given that y_1 follows Bernoulli distribution, and not a Gaussian distribution.

only the X_n attributes y_1 are included in this calculation, given that y_2, y_3 follow a Gaussian distribution, and not a Bernoulli distribution.

Bernoulli Probability of Sucess Update

- $p_k = \dfrac{\sum_{n=1}^{N} \gamma_{nk} \cdot x_n}{N_k}$

Cluster 1 | $N_1 = 1.54467$; $\mu_1 = \begin{pmatrix} 0.02651 \\ 0.50713 \end{pmatrix}$; $\Sigma_1 = \begin{pmatrix} 0.14136 & -0.10541 \\ -0.10541 & 0.09605 \end{pmatrix}$; $\pi_1 = 0.38617$

Cluster 2 | $N_2 = 2.45533$; $\mu_2 = \begin{pmatrix} 0.30914 \\ 0.21042 \end{pmatrix}$; $\Sigma_2 = \begin{pmatrix} 0.10829 & -0.08865 \\ -0.08865 & 0.10412 \end{pmatrix}$; $\pi_2 = 0.61383$

Cluster 1 | $p_1 = P(y_1 = 1) = 0.23404$

Cluster | $p_2 = P(y_1 = 1) = 0.66732$

2. $X_{new} = \begin{pmatrix} 1 \\ 0.3 \\ 0.7 \end{pmatrix}$

In this exercise, we will be using the updated values from Exercise 1!!

Cluster 1 | $\bullet\ p(x_{new} \mid C=1) = p_1 \times N(x_{new} \mid \mu_1, \Sigma_1) = 0.00634$

Cluster 2 | $\bullet\ p(x_{new} \mid C=2) = 0.04567$

Cluster 1 | $p(C=1, x_{new}) = \pi_1 \cdot N(x_{new} \mid \mu_1, \Sigma_1) = 0.00245$

Cluster 2 | $p(C=2, x_{new}) = \pi_2 \cdot N(x_{new} \mid \mu_2, \Sigma_2) = 0.02803$

$\longrightarrow P(X_{new}) = P(C=1, X_{new}) + P(C=2, X_{new}) = 0.03048$

Cluster Memberships (Posteriors) for X_{new}:

Cluster 1 | $\gamma(C_{new1}) = \dfrac{P(C=1, x_{new})}{P(X_{new})} = 0.08029$

Cluster 2 | $\gamma(C_{new2}) = \dfrac{P(C=2, X_{new})}{P(X_{new})} = 0.91971$

3. We will be using the updated values!

- $p(x_1|C=1) = p_1 \times N(x_1|\mu_1,\Sigma_1) = 0.23147$
- $p(x_1|C=2) = p_2 \times N(x_1|\mu_2,\Sigma_2) = 0.94954$

$\longrightarrow X_1$ belongs to $C=2$

- $p(x_2|C=1) = (1-p_1) \times N(x_2|\mu_1,\Sigma_1) = 1.26633$
- $p(x_2|C=2) = (1-p_2) \times N(x_2|\mu_2,\Sigma_2) = 0.08874$

$\longrightarrow X_2$ belongs to $C=1$

- $p(x_3|C=1) = (1-p_1) \times N(x_3|\mu_1,\Sigma_1) = 1.43811$
- $p(x_3|C=2) = (1-p_2) \times N(x_3|\mu_2,\Sigma_2) = 0.45417$

$\longrightarrow X_3$ belongs to $C=1$

- $p(x_4|C=1) = p_1 \times N(x_4|\mu_1,\Sigma_1) = 0.02077$
- $p(x_4|C=2) = p_2 \times N(x_4|\mu_1,\Sigma_1) = 0.72331$

$\longrightarrow X_4$ belongs to $C=2$

$C_1 = \{X_2, X_3\}, \quad C_2 = \{X_1, X_4\}$

Since both clusters have the same number of observations, we will calculate the Silhouette for both of them.

$$\boxed{\text{Silhouette:} \quad s(x) = \begin{cases} 1 - \dfrac{a(x)}{b(x)}, & b(x) > a(x) \\ \dfrac{b(x)}{a(x)} - 1, & \text{c.c} \end{cases}}$$

- a equals to medium distance from X_i to the points in the same cluster
- b equals to minimum medium distance from X_i to the points of another cluster, in this case, we only have two cluster.

- Important to know we will be using manhattan distance as our distance measurement

$$a(x_1) = d(x_1,x_4) = \overbrace{|1-1| + |0.6-0.4| + |0.1-(-0.1)|}^{\text{Manhattan Distance}} = 0.40000$$

$a(x_2) = d(x_2,x_3) = 0.90000$

$a(x_3) = d(x_3,x_2) = 0.90000$

$a(x_4) = d(x_4,x_1) = 0.40000$

$$b(x_1) = \frac{d(x_1,x_2) + d(x_1,x_3)}{2} = 2.25000$$

$$b(x_2) = \frac{d(x_2,x_1) + d(x_2,x_4)}{2} = 2.70000$$

$$b(x_3) = \frac{d(x_3,x_1) + d(x_3,x_4)}{2} = 1.80000$$

$$b(x_4) = \frac{d(x_4,x_2) + d(x_4,x_3)}{2} = 2.25000$$

Using the Silhouette function: $\longrightarrow$

- $s(x_1) = 0.82222$
- $s(x_2) = 0.66667$
- $s(x_3) = 0.50000$
- $s(x_4) = 0.82222$

The silhoutte of a cluster corresponds to the mean of the silhouttes of the observations in that cluster.

- $s(C_1) = \frac{s(x_2) + s(x_3)}{2} = 0.58333$

- $s(C_2) = \frac{s(x_1) + s(x_4)}{2} = 0.82222$

4.- The purity can be defined as a measurement that indicates the clustering's quality:

- A purity score closer to 1 reveals that our clustering is highly adequate.
- Since we are given a purity value of 0.75, we can conclude that 75% of the observations were correctly assigned to their clusters.
- Thus, 25% of our observations were assigned to the wrong cluster.
- Since we have 4 observations, only one of them (25%) was incorrectly classified.
- Given that we don't know if that observation belongs to either cluster 1 or cluster 2, in the worst case scenario, it doesn't belong to neither of them.
- Therefore, we would need at most 3 classes - 1 for each cluster and a new class for the observation that was incorrectly classified.