Q1

$$D = \{ x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)} \} = \left\{ \binom{0}{1}, \binom{2}{0}, \binom{1}{1}, \binom{3}{*} \right\}$$

$$\theta = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \sigma_1^2 \\ \sigma_2^2 \end{bmatrix}, \quad \theta_0 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad x_4 = \binom{3}{x_{42}}$$

$$Q(\theta, \theta^{old}) = \int_z \ln p(x, z \mid \theta)\, p(z \mid x, \theta^{old})\, dz$$

$$p(z \mid x, \theta^{old}) = \frac{p(x_{41}, x_{42} \mid \theta^{old})}{\int p(x_{41}, x_{42} \mid \theta^{old})\, dx_{42}}$$

$$= \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{9}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{42}^2}{2}}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{9}{2}} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{42}^2}{2}}\, dx_{42}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{42}^2}{2}}$$

$$\Rightarrow \quad Q(\theta, \theta^{old}) = \int \ln(x, z \mid \theta) \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{42}^2}{2}}\, dx_{42}$$

$$= \sum_{k=1}^{3} \ln p(x_k \mid \theta) + \int_{-\infty}^{\infty} \ln p\left(\binom{3}{x_{42}} \Big| \theta\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{42}^2}{2}}\, dx_{42}$$

$$= \sum_{k=1}^{3} \ln p(x_k \mid \theta) + \left[\ln\left(\frac{1}{2\pi\sigma_1\sigma_2}\right) - \frac{(3-\mu_1)^2}{2\sigma_1^2} \right] \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{42}^2}{2}}\, dx_{42} + \int_{-\infty}^{\infty} -\frac{(x_{42}-\mu_2)^2}{2\sigma_2^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{42}^2}{2}}\, dx_{42}$$

$$= \sum_{k=1}^{3} \ln p(x_k \mid \theta) + \ln\left(\frac{1}{2\pi\sigma_1\sigma_2}\right) - \frac{(3-\mu_1)^2}{2\sigma_1^2} - \frac{1}{2\sigma_2^2} \int_{-\infty}^{\infty} x_{42}^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{42}^2}{2}}\, dx_{42} + \frac{2\mu_2}{2\sigma_2^2} \int_{-\infty}^{\infty} x_{42} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{42}^2}{2}}\, dx_{42}$$

$$+ \frac{\mu_2^2}{2\sigma_2^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{42}^2}{2}}\, dx_{42}$$

$$= \sum_{k=1}^{3} \ln p(x_k \mid \theta) + \ln\left(\frac{1}{2\pi\sigma_1\sigma_2}\right) - \frac{(3-\mu_1)^2}{2\sigma_1^2} - \frac{1}{2\sigma_2^2} - \frac{\mu_2^2}{2\sigma_2^2}$$

$$= \sum_{k=1}^{3} \ln \frac{1}{2\pi\sigma_1\sigma_2} - \left[\frac{(x_{k1}-\mu_1)^2}{2\sigma_1^2} + \frac{(x_{k2}-\mu_2)^2}{2\sigma_2^2} \right] + \ln\left(\frac{1}{2\pi\sigma_1\sigma_2}\right) - \frac{(3-\mu_1)^2}{2\sigma_1^2} - \frac{1}{2\sigma_2^2} - \frac{\mu_2^2}{2\sigma_2^2}$$

$$= 4\left(\ln \frac{1}{2\pi\sigma_1\sigma_2}\right) - \frac{\mu_1^2}{2\sigma_1^2} - \frac{(1-\mu_2)^2}{2\sigma_2^2} - \frac{(2-\mu_1)^2}{2\sigma_1^2} - \frac{(\mu_2)^2}{2\sigma_2^2} - \frac{(1-\mu_1^2)}{2\sigma_1^2} - \frac{(1-\mu_2)^2}{2\sigma_2^2}$$

$$- \frac{1}{2\sigma_2^2} - \frac{\mu_2^2}{2\sigma_2^2}$$

$$\frac{\partial Q(\theta, \theta^{old})}{\partial \mu_2} = \frac{(1-\mu_2)}{\sigma_2^2} - \frac{\mu_2}{\sigma_2^2} + \frac{(1-\mu_2)}{\sigma_2^2} - \frac{\mu_2}{\sigma_2^2} = 0$$

$$1 - \mu_2 - \mu_2 + 1 - \mu_2 - \mu_2 = 0, \quad \boxed{\mu_2 = \frac{1}{2}}$$

first improved estimate of μ_2

Q2

(9.10) $p(z) = \prod_{k=1}^{K} \pi_k^{z_k}$, (9.11) $p(x|z) = \prod_{k=1}^{K} N(x|\mu_k, \Sigma_k)^{z_k}$

$\sum_z p(z)p(x|z) = \sum_z \prod_{k=1}^{K} \pi_k^{z_k} N(x|\mu_k, \Sigma_k)$, $\sum_k z_k = 1$, $z_k \in \{0, 1\}$

$z \Rightarrow \begin{bmatrix} 1 \\ 0 \\ \vdots \end{bmatrix} \cdots \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}$

$\sum_z P(x|z)p(z) = \pi_1^{z} N(x|\mu_1, \Sigma_1)^{z_1} + \cdots + \pi_k^{z_k} N(x|\mu_k, \Sigma_k)^{z_k}$

$= \sum_{k=1}^{K} \pi_k N(x|\mu, \Sigma_k) = p(x)$ (9.7) ✓

Q3

(9.49) $E[x] = \sum_{k=1}^{K} \pi_k \mu_k$, (9.50) $\mathrm{cov}[x] = \sum_{k=1}^{K} \pi_k \{\Sigma_k + \mu_k \mu_k^T\} - E[x]E[x]^T$

$p(x) = \sum_{k=1}^{K} \pi_k p(x|k)$

$E[x] = \sum_x x \sum_{k=1}^{K} \pi_k P(x|k) = \sum_{k=1}^{K} \pi_k \sum_x x\, p(x|k) = \sum_{k=1}^{K} \pi_k E[x|k] = \sum_{k=1}^{K} \pi_k \mu_k$ ✓

$\mathrm{cov}[x] = E((x-E(x))(x-E(x))^T = E(xx^T) - E(x)E(x)^T$

$\Rightarrow E[xx^T] - E[x]E[x]^T$

$= \sum_x x \cdot x^T p(x) - E[x]E[x]^T$

$= \sum_x xx^T \sum_{k=1}^{K} \pi_k\, p(x|k) - E[x]E[x]^T$

$= \sum_{k=1}^{K} \pi_k \sum_x xx^T p(x|k) - E[x]E[x]^T$

$= \sum_{k=1}^{K} \pi_k E(xx^T|x) - E[x]E[x]^T$

$= \sum_{k=1}^{K} \pi_k (\Sigma_k + \mu_k \mu_k^T) - E[x]E[x]^T$ ✓

Q4

(9.52) $P(x|z, \mu) = \prod_{k=1}^{K} p(x|\mu_k)^{z_k}$

(9.53) $P(z|\pi) = \prod_{k=1}^{K} \pi_k^{z_k}$

(9.47) $P(x|\mu, \pi) = \sum_{k=1}^{K} \pi_k p(x|\mu_k)$

$$p(x|\mu,\pi) = \sum_z p(x|z,\mu)\,p(z|\pi) = \sum_z \prod_{k=1}^{K} p(x|\mu_k)^{z_k} \cdot \pi_k^{z_k}$$
$$= p(x|\mu_1)^{z_1} \cdot \pi_1^{z_1} + \dots + p(x|\mu_k)^{z_k} \cdot \pi_k^{z_k}$$
$$= \sum_{k=1}^{K} p(x|\mu_k) \cdot \pi_k \checkmark$$

Q5

$$(9.55)\ E_z[\ln p(X,Z|\mu,\pi)] = \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk})\left\{\ln \pi_k + \sum_{i=1}^{D}[x_{ni}\ln\mu_{ki} + (1-x_{ni})\ln(1-\mu_{ki})]\right\}$$

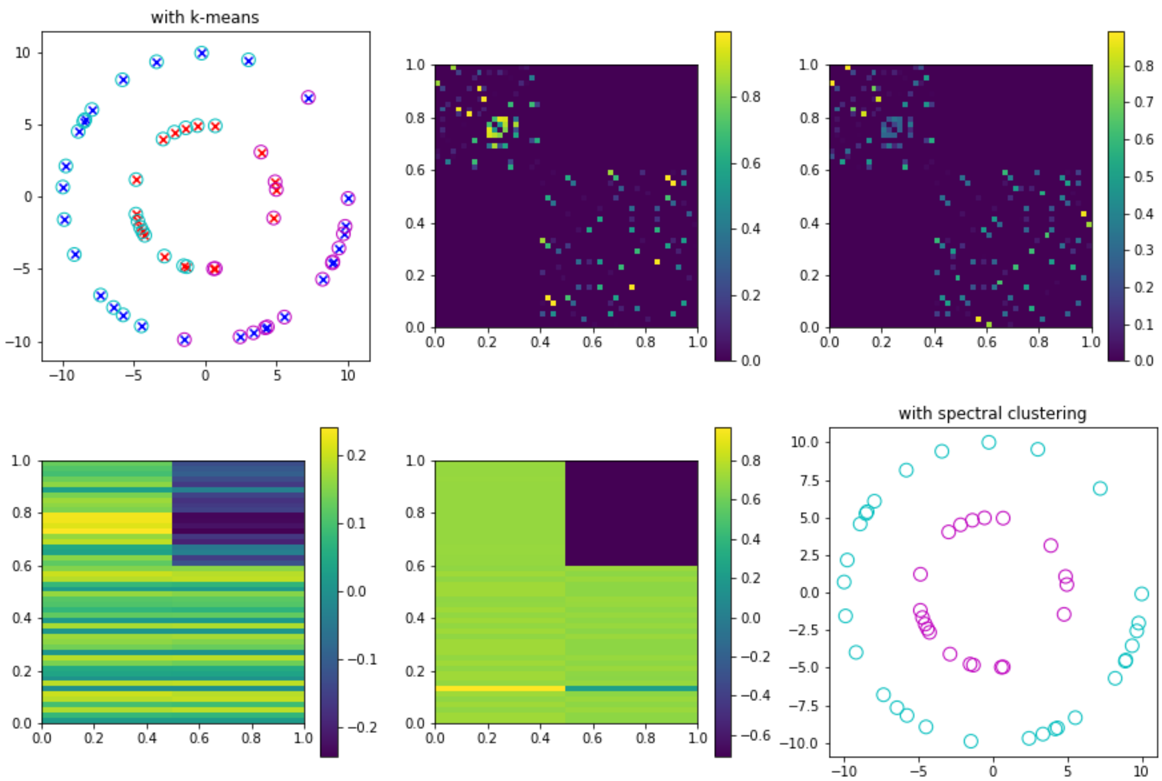
$$(9.59)\ \mu_k = \bar{x}_k$$

$$\frac{\partial E_z}{\partial \mu_k} = \sum_{n=1}^{N} \gamma(z_{nk})\left(\frac{x_{ni}}{\mu_{ki}} - \frac{(1-x_{ni})}{(1-\mu_{ki})}\right) = 0$$
$$\sum_{n=1}^{N} \gamma(z_{nk})(x_n(1-\mu_k) - \mu_k(1-x_n)) = 0$$
$$\sum_{n=1}^{N} \gamma(z_{nk})x_n = \sum_{n=1}^{N}\gamma(z_{nk})\mu_k$$
$$\mu_k = \sum_{n=1}^{N}\gamma(z_{nk})x_n \cdot \frac{1}{\sum_{n=1}^{N}\gamma(z_{nk})} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})x_n = \bar{x}_k \checkmark$$

Q6.

$$AA^T v = \lambda v$$
$$A^T A A^T v = A^T \lambda v = \lambda A^T v$$
$$\Rightarrow (A^T A)(A^T v) = \lambda(A^T v), \text{ set } B = A^T v \Rightarrow (A^T A)B = \lambda B$$

$A^T v$ is an eigenvector of $A^T A$


with k-means
with spectral clustering

In [15]:
```
fig.savefig('pic')
```

Q7

The first plot shows the attempt to perform classification using k-means. However, it fails to show a clear classification as the first plot has mixed points in the outer and inner rings. The second plot shows the affinity matrix of data points with the zero diagonal entries. The third plot shows the normalized Laplacian matrix. The fourth plot shows the sorted eigenvectors of matrix L. The fifth plot shows the result of normalization with rows of X. Then, the last plot shows the successfully classified data points into an outer and inner rings after spectral clustering.

Q8 ¶

sigsq represents the distance among data points within the same cluster. This means that it also represents the level of similarity among data points. If sigsq is too small, the affinity among points also becomes too small, close to 0, and that the data points do not have to be really similar to be in the same cluster. On the other hand, if sigsq is too large, affinity among points becomes too large, close to 1, and data points have to be almost the same to be in the same cluster. A sigsq value that is either too small or too large cannot produce desired clustering result.

Q9

There is not a sigsq value that produces desired clustering result for random data. The spectral clustering seems to choose the large gap between outer and inner rings as the split, and thus produce results that have mixed points for both outer and inner rings. We can see that spectral clustering only works for certain data sets and sigsq values.

In [ ]:
```

```