Machine Learning homework 11 solution
Clustering

Wiktor Jurasz - M.Nr. 03709419

January 26, 2019

1 Problem 1

$$\mathbb{E}[x] = \mathbb{E}(\sum_k \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)) = \sum_k \mathbb{E}(\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)) = \sum_k \pi_k \mu_k \tag{1}$$

$$\begin{aligned} \mathbb{E}[xx^T] &= \mathbb{E}(\sum_k \pi_k \mathcal{N}(xx^T|\mu_k, \Sigma_k)) = \sum_k \mathbb{E}(\pi_k \mathcal{N}(xx^T|\mu_k, \Sigma_k)) = \sum_k \mathbb{E}[\pi_k]\mathbb{E}[\mathcal{N}(x|\mu_k, \Sigma_k)] = \\ &= \sum_k \mathbb{E}[\pi_k](Cov[x] + \mathbb{E}[x]\mathbb{E}[x]^T) = \sum_k \pi_k(\Sigma_k + \mu_k \mu_k^T) \end{aligned} \tag{2}$$

$$Cov[x] = \sum_k \pi_k(\Sigma_k + \mu_k\mu_k^T) - \sum_k \pi_k\mu_k(\sum_k \pi_k\mu_k)^T \tag{3}$$

2 Problem 2

2.1 Step E

$$\gamma(z_{ik}) = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j)} \tag{4}$$

$$\forall k \in (1..K) : \Sigma_k = \sigma^2 I = \Sigma \tag{5}$$

$$\gamma(z_{ik}) = \frac{\pi_k \exp(\frac{-||x_i-\mu_k||_2^2}{2\Sigma})}{\sum_{j=1}^K \pi_j \exp(\frac{-||x_i-\mu_j||_2^2}{2\Sigma})} \tag{6}$$

$$d = \min_k ||x_i - \mu_k||_2^2 \tag{7}$$

$$\gamma(z_{ik}) = \frac{\pi_k \exp(\frac{d-||x_i-\mu_k||_2^2}{2\Sigma})}{\sum_{j=1}^K \pi_j \exp(\frac{d-||x_i-\mu_j||_2^2}{2\Sigma})} \tag{8}$$

$$\forall ||x_i - \mu_k|| \neq d, \lim_{\Sigma \to 0} \exp(\frac{d - ||x_i - \mu_k||_2^2}{2\Sigma}) = 0 \tag{9}$$

$$\forall ||x_i - \mu_k|| = d, \lim_{\Sigma \to 0} \exp(\frac{d - ||x_i - \mu_k||_2^2}{2\Sigma}) = 1 \tag{10}$$

$$\lim_{\Sigma \to 0} \sum_{j=1}^K \exp(\frac{d - ||x_i - \mu_k||_2^2}{2\Sigma}) = 1 \tag{11}$$

$$(9) \ \& \ (10) \ \& \ (11) \Rightarrow \gamma(z_{ik}) = \begin{cases} 1, & \text{if } k = \text{argmin}_j ||x_i - \mu_j||_2^2 \\ 0, & \text{otherwise} \end{cases} = z_{ik} \tag{12}$$

Q.E.D

2.2 Step M

$$P(X,Z|\theta) = P(X|Z,\theta)P(Z|\theta) = \prod_{n=1}^{N}\prod_{k=1}^{K}\pi_k^{z_{nk}}\mathcal{N}(x_n|\mu_k,\Sigma_k)^{z_{nk}} \tag{13}$$

$$log(P(X,Z|\theta)) = \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}(log(\pi_k) + log(\mathcal{N}(x_n|\mu_k,\Sigma_k))) = \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}(log(\pi_k) - \frac{1}{2}log(2\pi\Sigma_k) - \frac{(x_n-\mu_k)^2}{2\Sigma_k}) \tag{14}$$

$$\forall k \in (1..K) : \Sigma_k = \sigma^2 I = \Sigma \tag{15}$$

$$log(P(X,Z|\theta)) = \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}log(\pi_k) - \sum_{n=1}^{N}\sum_{k=1} z_{nk}\frac{1}{2}log(2\pi\Sigma) - \sum_{n=1}^{N}\sum_{k=1} z_{nk}\frac{(x_n-\mu_k)^2}{2\Sigma} \tag{16}$$

Since Σ is known and fixed

$$log(P(X,Z|\theta)) \propto \Sigma(\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}log(\pi_k) - \sum_{n=1}^{N}\sum_{k=1} z_{nk}\frac{1}{2}log(2\pi\Sigma)) - \sum_{n=1}^{N}\sum_{k=1} z_{nk}(x_n-\mu_k)^2 \tag{17}$$

$$lim_{\Sigma\to 0}\Sigma(\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}log(\pi_k) - \sum_{n=1}^{N}\sum_{k=1} z_{nk}\frac{1}{2}log(2\pi\Sigma)) - \sum_{n=1}^{N}\sum_{k=1} z_{nk}(x_n-\mu_k)^2 = -\sum_{n=1}^{N}\sum_{k=1} z_{nk}(x_n-\mu_k)^2 = -J(X,Z,\mu) \tag{18}$$

Maximizing above function is equivalent to minimizing J in second step of Loyd's algorithm

Q.E.D

3 Problem 3

$$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

Answer: **D** - No covariance and equal variances

$$\begin{bmatrix} 3 & 0 \\ 0 & 0.5 \end{bmatrix}$$

Answer: **B** - No covariance and unequal variances

$$\begin{bmatrix} 0.9 & -0.8 \\ -0.8 & 1.2 \end{bmatrix}$$

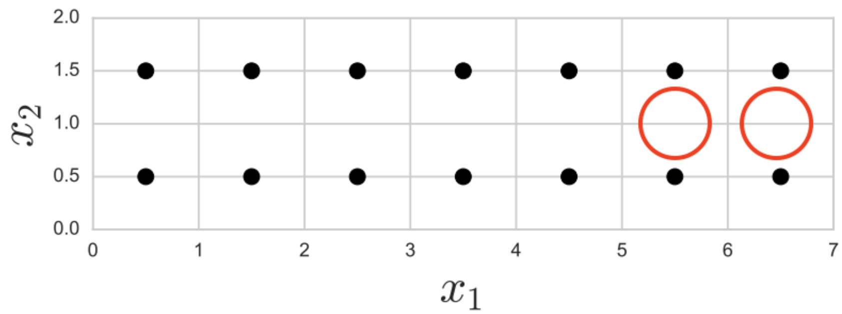
Answer: **C** - Has covariance and variances are smaller then A (also unequal variances but it's hard to see from plot)

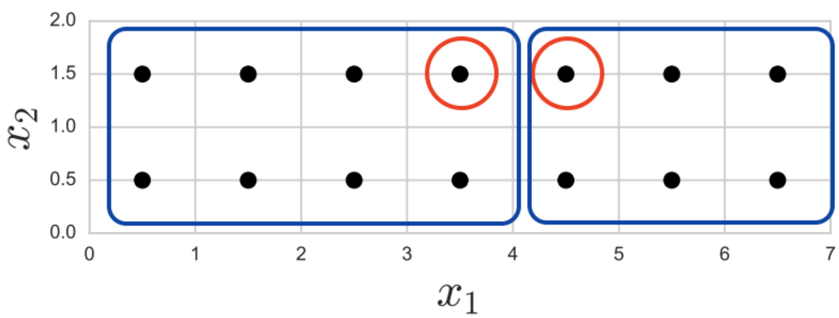
$$\begin{bmatrix} 2 & -1.7 \\ -1.7 & 2 \end{bmatrix}$$

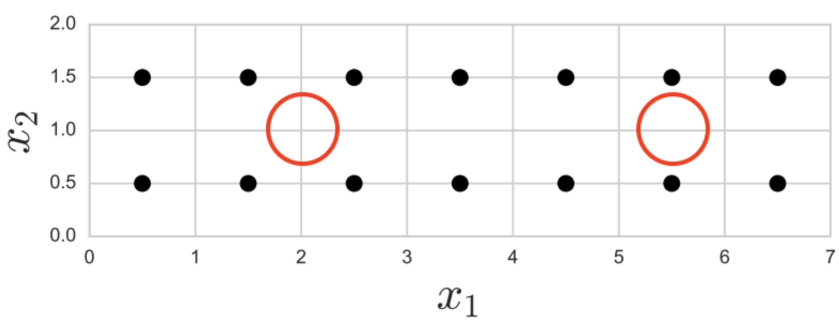
Answer: **A** - Has covariance and equal variances (bigger then C)

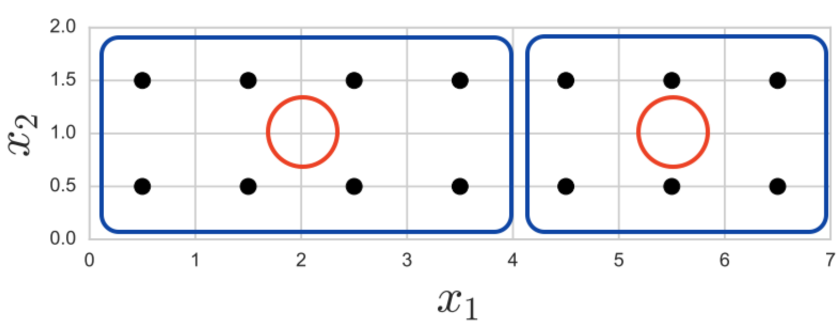
4 Problem 4

4.1



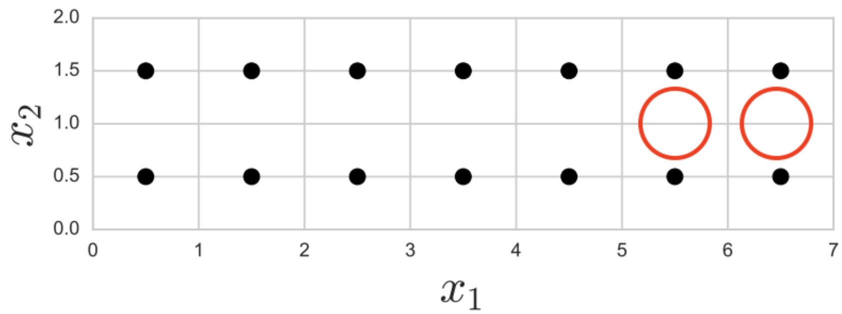


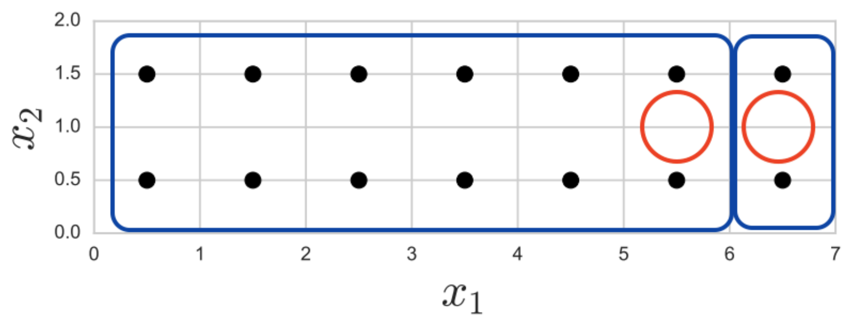


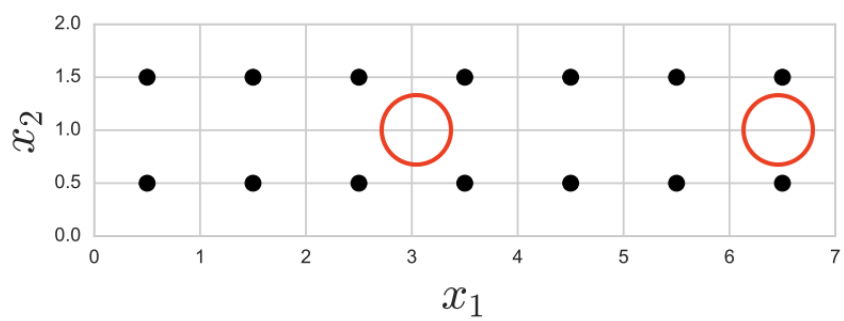


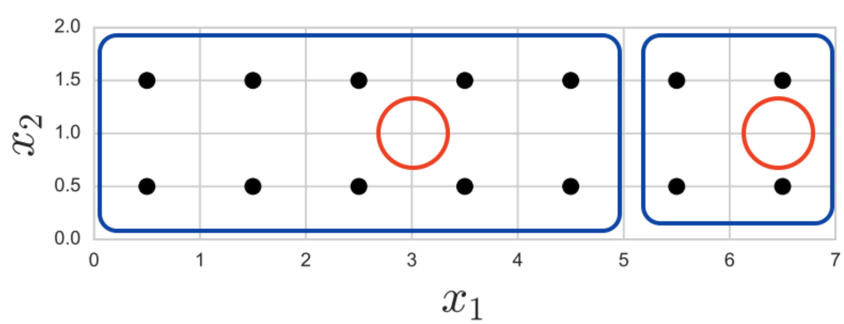
It took 2 iterations to converge

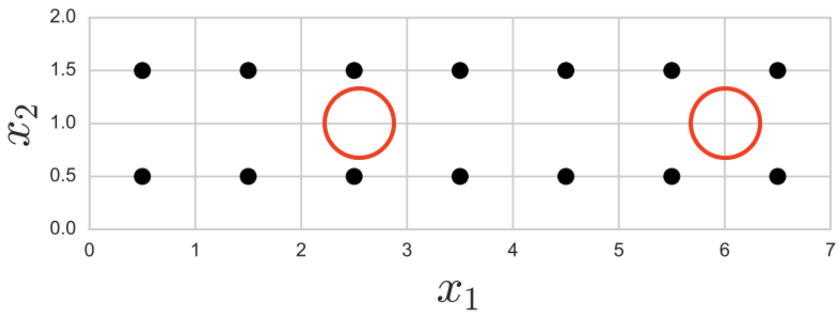
4.2

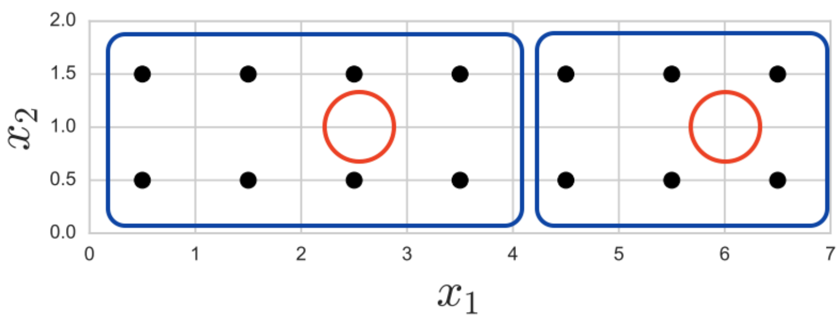

2.0
1.5
1.0
0.5
0.0
x_2
0
1
2
3
4
5
6
7
x_1

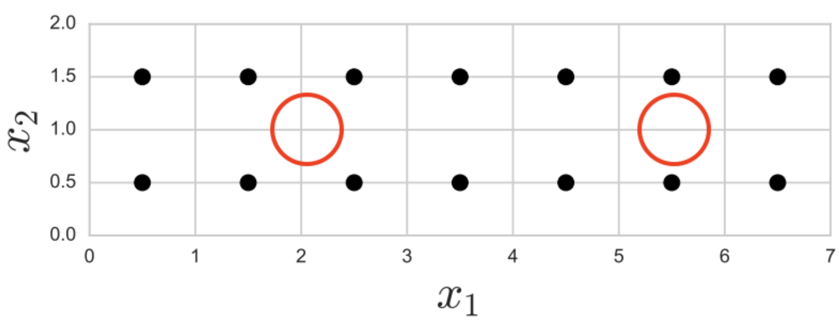

2.0
1.5
1.0
0.5
0.0
x_2
0
1
2
3
4
5
6
7
x_1

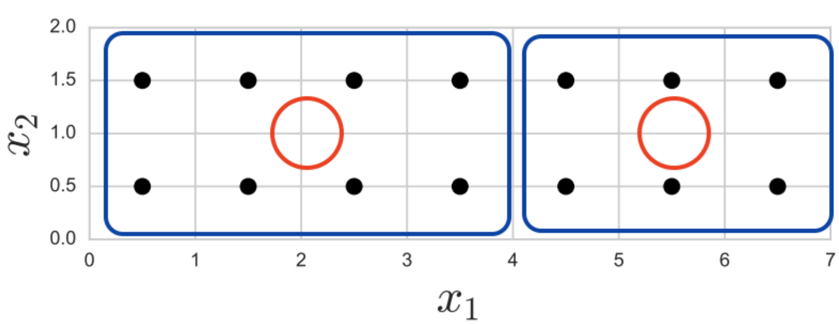

2.0
1.5
1.0
0.5
0.0
x_2
0
1
2
3
4
5
6
7
x_1


2.0
1.5
1.0
0.5
0.0
x_2
0
1
2
3
4
5
6
7
x_1


2.0
1.5
1.0
0.5
0.0
0 1 2 3 4 5 6 7
x_2
x_1


2.0
1.5
1.0
0.5
0.0
0 1 2 3 4 5 6 7
x_2
x_1


2.0
1.5
1.0
0.5
0.0
0 1 2 3 4 5 6 7
x_2
x_1


2.0
1.5
1.0
0.5
0.0
0 1 2 3 4 5 6 7
x_2
x_1