3. **(Algoritmul EM/GMM, cazul $\pi_1 = \pi_2 = 1/2$, $\sigma_1^2 = \sigma_2^2$: executarea unei iteraţii)**

Fie un model de mixtură gaussiană (engl., Gaussian mixture model, GMM) cu două componente având varianţe cunoscute şi probabilităţi *a priori* egale pentru selecţia celor două distribuţii:

$$\frac{1}{2}N(x;\mu_1,\sigma) + \frac{1}{2}N(x;\mu_2,\sigma),\ x \in \mathbb{R}. \quad \Rightarrow P(z_{ij}) = \frac{1}{2}$$

În continuare se va considera că $n = 2$, $\sigma = 1/2$, $x_1 = 0.5$ şi $x_2 = 2$, iar valorile iniţiale pentru μ_1 şi μ_2 sunt 1 şi respectiv 2.

Executaţi în mod manual o iteraţie a algoritmului EM (varianta din cartea *Machine Learning* a lui Tom Mitchell, pag. 193; vezi problema 4) pe aceste date, astfel:

a. (Pasul E) Calculaţi mai întâi $P(z_{ij} = 1|X_i, \mu)$, probabilităţile a posteriori de apartenenţă a datelor observate (x_1 şi x_2) la cele două componente ale mixturii. Am folosit notaţia $\mu = (\mu_1, \mu_2)$.

b. (Pasul M) Re-calculaţi valorile parametrilor μ_1 şi μ_2 în funcţie de probabilităţile calculate la punctul precedent.

c. Care credeţi că va fi tendinţa de „mişcare" a mediilor la următoarele iteraţii? Faceţi comparaţia cu cazul în care s-ar fi lucrat cu $\sigma = 1$ sau cu $\sigma = 1/4$. Cunoaşteţi un rezultat teoretic care explică acest „comportament" al algoritmului EM pentru rezolvare de mixturi de distribuţii gaussiene (GMM) pentru valori din ce în ce mai mici ale lui σ? Daca da, formulaţi rezultatul respectiv.

***Indicaţie*: În vederea efectuării calculelor, pentru convenienţă puteţi considera valorile distribuţiei normale/gaussiene standard $N(x;\mu = 0, \sigma = 1)$ în punctele 0, 0.5, 1, 1.5, 2, 3, 4, 5 şi 6 ca fiind respectiv 0.4, 0.35, 0.24, 0.13, 0.05, 0.04, 0.000134, 0.000001 şi 0.000000.... Legătura dinspre o distribuţie gaussiană nestandard şi distribuţia gaussiană standard se face cu ajutorul transformării $x \to \frac{x-\mu}{\sigma}$.**

a) $$P(z_{ij}=1 \mid X_i,\mu) \overset{TB}{=} \frac{P(z_{ij}=1)\cdot P(X_i=x_i,\ \mu=\mu_j)}{P(x_i,\mu_j)\cdot P(z_{ij}=1) + P(x_i,\mu_{1-j})\cdot P(z_{ij}=1)} =$$

$$= \frac{P(X=x_i,\mu=\mu_j)\cdot \cancel{\frac{1}{2}}}{\cancel{\frac{1}{2}}P(X=x_i,\mu=\mu_j) + \cancel{\frac{1}{2}}P(X=x_i,\mu=\mu_{1-j})} = \frac{P(X=x_i,\mu=\mu_j)}{P(X=x_i,\mu=\mu_j)+P(X=x_i,\mu=\mu_{1-j})}$$

$$\Rightarrow P(z_{11}=1 \mid x_1,\mu_1) = \frac{N(0.5;1;\frac{1}{2})}{N(0.5;1;\frac{1}{2})+N(0.5;2;\frac{1}{2})} = \frac{N(1;0;1)}{N(1;0;1)+N(3;0;1)} = \frac{0.24}{0.24+0.04} = 0.857$$

(Aplicăm substituția $x_{nou} = \frac{x_{vechi}-\mu}{\sigma}$)

$$P(z_{12}=1 \mid x_1,\mu_2) = \frac{N(0.5;2;\frac{1}{2})}{N(0.5;1;\frac{1}{2})+N(0.5;2;\frac{1}{2})} = \frac{N(3;0;1)}{N(1;0;1)+N(3;0;1)} = \frac{0.04}{0.24+0.04} = 0.142$$

$$P(z_{21}=1 \mid x_2,\mu_1) = \frac{N(2;1;\frac{1}{2})}{N(2;1;\frac{1}{2})+N(2;2;\frac{1}{2})} = \frac{N(2;0;1)}{N(2;0;1)+N(0;0;1)} = \frac{0.05}{0.05+0.4} = 0.111$$

$$P(z_{22}=1 \mid x_2,\mu_2) = \frac{N(2;2;\frac{1}{2})}{N(2;1;\frac{1}{2})+N(2;2;\frac{1}{2})} = \frac{N(0;0;1)}{N(2;0;1)+N(0;0;1)} = \frac{0.4}{0.05+0.4} = 0.888$$

9.

b) $$\mu_j = \frac{\sum E[z_{ij}]\cdot x_i}{\sum E[z_{ij}]} = \frac{\sum P(z_{ij}=1 \mid x_i,\mu_j)\cdot x_i}{\sum P(z_{ij}=1 \mid x_i,\mu_j)}$$

$$\Rightarrow \mu_1 = \frac{P(z_{11}=1 \mid x_1,\mu_1)\cdot x_1 + P(z_{21}=1 \mid x_2,\mu_1)\cdot x_2}{P(z_{11}=1 \mid x_1,\mu_1)+P(z_{21}=1 \mid x_2,\mu_1)} = \frac{0.857\cdot 0.5+0.111\cdot 2}{0.857+0.111}$$

$$= \frac{0.4285+0.222}{0.968} = 0.67$$

$$\mu_2 = \frac{P(z_{12}=1 \mid x_1,\mu_2)\cdot x_1 + P(z_{22}=1 \mid x_2,\mu_2)\cdot x_2}{P(z_{12}=1 \mid x_1,\mu_2)+P(z_{22}=1 \mid x_2,\mu_2)} = \frac{0.142\cdot 0.5+0.888\cdot 2}{0.142+0.888}$$

$$= \frac{0.071+1.776}{1.03} = 1.79$$

c) Se observă că tendința de „mișcare" este de apropiere a centroizilor, așadar μ_1 tinde spre ~~dreapta~~ x_1 / stânga, iar μ_2 spre stânga. tot

?

μ_1 s-a mutat acum de la $\mu_1^{(0)}=1$ la $\mu_1^{(1)}=0{,}67$ | acum $\mu_2^{(0)}=2$, $\mu_2^{(1)}=1{,}79$

Se obs. că $\mu_1^{(1)}$ și $\mu_2^{(1)}$ sunt relativ simetrice față de $x_1=0{,}5$ și $x_2=2$.

La iterațiile următoare, μ_1 va continua să meargă spre $x_1=0{,}5$, iar μ_2 se va „retrage" spre $x_2=2$.

(Se poate verif. printr-o implementare relativ simplă.)

Pt. $\sigma=1$, în pr. 16 din culegere am văzut că μ_1 și μ_2 se depl. ambele spre centrul intervalului [0,5; ...

Pt $\sigma=1/4$, tendința de depl. $\mu_1\to x_1$ și $\mu_2\to x_2$ este chiar mai accentuată

Pr. 40.b spune că $\sigma\to 0 \Rightarrow$ EM $\to$ K-means, deci $\mu_1\to x_1$ și $\mu_2\to x_2$

10.