I ADABOOST = adaptive boosting

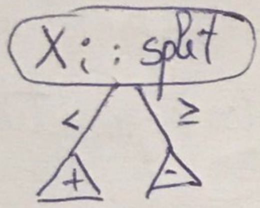
→ algoritm de clasificare, la fel ca ID3, Bayes, KNN

→ Underfitting : - nu învață destul de bine datele
- nu are acuratețe bună pe datele de antrenament

→ Boosting = alg iterativ, la care la fiecare iterație, ne focusăm pe o anumită zonă din datele de antrenament

• Compași de decizie

- avem doar variabile continue în cazul AdaBoost

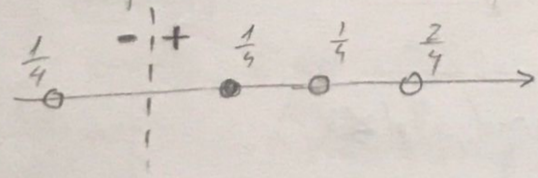


$$\Rightarrow \begin{cases} X_i < split \Rightarrow + \\ X_i \geq split \Rightarrow - \end{cases} \Rightarrow \begin{cases} sgn(split - X_i) \\ sgn(X_i - split) \end{cases}$$

// $sgn(split - X_i) \Rightarrow X_i < split \Rightarrow$ partea stângă / de desupt este considerată pe moment zona pozitivă.

// $sgn(X_i - split) \Rightarrow split < X_i \Rightarrow$ partea dreaptă / de deasupra este considerată pe moment zona pozitivă

• Eroare



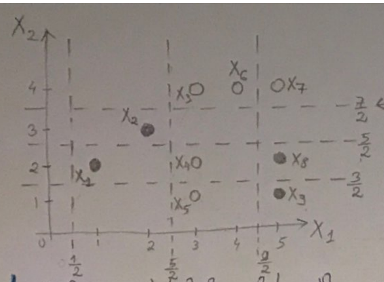
$\Rightarrow$ Eroarea - fără ponderi $= \frac{2}{4}$

- cu ponderi $= 1 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} = \frac{3}{4}$ → fiecare pct clasificat greșit * ponderea sa

• Algoritm

- Inițial, se folosește distribuția normală:

$D_1(i) = \frac{1}{m}$, $i = \overline{1,m}$, m = nr total de puncte de clasificat



primul split ales (eroarea cea mai mică)

↳ Realizăm tabelul split-urilor, în care completăm erorile fără ponderi realizate de acestea.

! CUM LUĂM SPLIT-URILE? → La mijlocul distanței dintre două puncte etichetate diferit. + pt una din axe, luăm și un split exterior, ai pe o parte a split-ului să se afle toate punctele.

split	$\frac{1}{2}$	$\frac{5}{2}$	$\frac{9}{2}$
sgn(split − X_1)	$\frac{4}{9}$	$\boxed{\frac{2}{9}}$	$\frac{6}{9}$
sgn(−split + X_1)	$\frac{5}{9}$	$\frac{7}{9}$	$\frac{3}{9}$
split	$\frac{3}{2}$	$\frac{5}{2}$	$\frac{7}{2}$
sgn(split − X_2)	$\frac{4}{9}$	$\frac{3}{9}$	$\boxed{\frac{2}{9}}$
sgn(X_2 − split)	$\frac{5}{9}$	$\frac{6}{9}$	$\frac{7}{9}$

sgn(split − X_i) ⇒ X_i < split
sgn(X_i − split) → split < X_i

$\sum_i \Delta_t(i)$, unde $i \in \{$indicele punctelor etichetate greșit$\}$
"suma ponderilor punctelor etich greșit"

! $Error_{\Delta_t}(X_i < split) = 1 - Error_{\Delta_t}(split \leq X_i)$
$Err_{\Delta_t}(X < split)$ / $Err_{\Delta_t}(X \geq split)$

↳ Alegem compasul de decizie cu eroarea cea mai mică.
În caz de egalitate, alegem unul random.
↳ în cazul nostru vom alege $\boxed{sgn\left(\frac{7}{2} - X_2\right) = h_1(x)}$

↳ Notăm eroarea: $\boxed{\varepsilon_1 = Error_{\Delta_1}(sgn(split - X_i))} = \frac{2}{9}$

↳ Calculăm ponderea: $\boxed{\alpha_1 = \frac{1}{2}\ln\frac{1-\varepsilon_1}{\varepsilon_1}} = \frac{1}{2}\ln\frac{1-\frac{2}{9}}{\frac{2}{9}} = \frac{1}{2}\ln\frac{7}{2}$

↳ Scriem structura următoarei distribuții:

MI $\boxed{\Delta_2 = \frac{1}{Z_1}\cdot\Delta_1(i)\cdot\left(e^{-\alpha_1}\right)^{y_1\cdot h_1(x_i)}}$

$$= \boxed{\begin{cases} \frac{1}{Z_1}\cdot\Delta_1(i)\cdot e^{-\alpha_1}, & \text{pentru pct etichetate corect} \\ \frac{1}{Z_1}\cdot\Delta_1(i)\cdot e^{\alpha_1}, & \text{pentru pct etichetate greșit} \end{cases}} \; !$$

⇒ Cazul nostru: $\Delta_2 = \begin{cases} \frac{1}{Z_1}\cdot\frac{1}{9}\cdot e^{-\frac{1}{2}\ln\frac{7}{2}}, & i \in \{1,2,3,6,7,8,9\} \\ \frac{1}{Z_1}\cdot\frac{1}{9}\cdot e^{\frac{1}{2}\ln\frac{7}{2}}, & i \in \{4,5\} \end{cases}$

OBS: y_t = etichetă $\in \{-1, 1\}$
$y_t \cdot h_t(x_i) = \begin{cases} 1, & \text{dacă pct este etichetat greșit} \\ -1, & \text{altfel} \end{cases}$

↳ Calculăm Z_1

$\boxed{Z_1 = \Delta_1(i)\cdot\left(\text{nr noduri corecte}\cdot e^{-\alpha_1} + \text{nr noduri greșite}\cdot e^{\alpha_1}\right)}$

$$\Rightarrow Z_1 = \frac{1}{9}\cdot\left(7\cdot e^{-\frac{1}{2}\ln\frac{7}{2}} + 2\cdot e^{\frac{1}{2}\ln\frac{7}{2}}\right)$$
$$= \frac{1}{9}\left(7\cdot\sqrt{e^{\ln\frac{2}{7}}} + 2\cdot\sqrt{e^{\ln\frac{7}{2}}}\right)$$
$$= \frac{1}{9}\left(7\cdot\underbrace{\sqrt{\frac{2}{7}}}_{e^{-\alpha_1}} + 2\cdot\underbrace{\sqrt{\frac{7}{2}}}_{e^{\alpha_1}}\right) = \frac{7}{9}\sqrt{\frac{2}{7}} + \frac{2}{9}\sqrt{\frac{7}{2}} = 0,8315$$

// Apoi înlocuim în Δ_1 și aflăm valorile pentru ambele ramuri

$$\boxed{Z_t = \sum_{i=1}^{m} D_t(i) \cdot e^{-\alpha_t y_i h_t(x)}}$$

$e^{-\alpha_t}$: nodurii etichetate corect

e^{α_t} nodurii etichetate greșit.

$$\boxed{H_t(x) = \sum_{j=1}^{t} \alpha_j \cdot h_j(x)}$$

pentru a stabili semnul unui punct după t iterații

- Condiții de oprire
  - → am atins nr maxim de iterații
  - → $\varepsilon_t = 0$ sau $\varepsilon_t = \frac{1}{2}$

M2 etichetate corect: $i \in \{1,2,3,6,7,8,9\}$:

$$D_1(1)\cdot a + D_1(2)a + D_1(3)a + D_1(6)a + D_1(7)a + D_1(8)a + D_1(9)a = \frac{1}{2}$$

$$\Leftrightarrow \frac{1}{9}a + \frac{1}{9}a + \frac{1}{9}a + \frac{1}{9}a + \frac{1}{9}a + \frac{1}{9}a + \frac{1}{9}a = \frac{1}{2}$$

$$\Leftrightarrow \frac{7}{9}a = \frac{1}{2} \Rightarrow a = \frac{1}{2}\cdot\frac{9}{7} = \frac{9}{14}$$

etichetate greșit: $i \in \{4,5\}$

$$D_1(4)b + D_1(5)b = \frac{1}{2}$$

$$\Leftrightarrow \frac{1}{9}b + \frac{1}{9}b = \frac{1}{2} \Leftrightarrow \frac{2}{9}b = \frac{1}{2} \Rightarrow b = \frac{1}{2}\cdot\frac{9}{2}$$

$$\Rightarrow b = \frac{9}{4}$$

$$\Rightarrow D_2(i) = \begin{cases} \frac{1}{9}\cdot\frac{9}{14}, & i \in \{1,2,3,6,7,8,9\} \\ \frac{1}{9}\cdot\frac{9}{4}, & i \in \{4,5\} \end{cases}$$

$$\Leftrightarrow D_2(i) = \begin{cases} \frac{1}{14}, & i \in \{1,2,3,6,7,8,9\} \\ \frac{1}{4}, & i \in \{4,5\} \end{cases}$$

M3 $D_{t+1}(i) = \begin{cases} D_t(i)/2\varepsilon_t, & \text{nodurii etichetate corect} \\ D_t(i)/2(1-\varepsilon_t), & \text{nodurii etichetate greșit} \end{cases}$

t	ε_t	α_t	$D_t(1)$	$D_t(2)$	......	$D_t(m)$	$Err_{D_t}(H_t)$

ponderi

t	α_t	x_1	x_2	...	x_m
	$H_t(x_i)$			...	

semnul deci de $h_t(x)$ pt x_i

semnul deci de $H_t(x_i)$ pt x_i

, unde $\boxed{H_t(x_i) = \sum_{j=1}^{t} \alpha_t \cdot h_t(x_i)}$

$$\boxed{\frac{a}{b\sqrt{c}} = \frac{b\sqrt{c}}{a} \cdot \left(\frac{a}{b\sqrt{c}}\right)^2}$$

Granițe de decizie:

imp.

0

	t	α_t	R_1	R_2	R_3	R_4	R_5	R_6
$h_1(x) \to$	1	0,42	+	+	+	−	−	−
$h_2(x) \to$	2	0,64	+	+	−	+	+	−
$h_3(x) \to$	3	0,92	+	−	−	+	−	−
$H_3(R) \longrightarrow$			+	+	−	+	−	−

$$\Rightarrow H_3(R_1) = \underset{sigm}{}(0,42+0,64+0,92) = +$$

$$H_3(R_2) = sigm(0,42+0,64-0,92) = sigm(1,06-0,92) = +$$

$$H_3(R_3) = sigm(0,42-0,64-0,92) = -$$

$$H_3(R_4) = sigm(-0,42+0,64+0,92) = +$$

$$H_3(R_5) = sigm(-0,42+0,64-0,92) = -$$

$$H_3(R_6) = sigm(-0,42-0,64-0,92) = -$$

II CLUSTERIZARE IERARHICĂ

• **Produs scalar**: $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = x_1y_1 + x_2y_2 + \dots + x_my_m$

• **Normă**: $\left\| \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \right\|_p = \left(|x_1|^p + |x_2|^p + \dots + (x_m)^p\right)^{\frac{1}{p}}$

$p \geq 1$

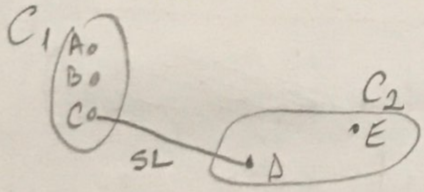
// Dacă nu se specifică, se folosește norma euclidiană, $p=2$

• **Distanța indusă de normă**: $d_p(x, y) = \|x - y\|_p$

• **Unghiul dintre 2 vectori**: $\cos(x, y) = \dfrac{xy}{\|x\|\|y\|}$

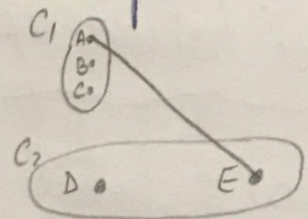
• **Clusterizare ierarhică**: // lucrăm cu clustere

1. Single-linkage: $\boxed{d_{sl}(C_1, C_2) = \min\{d(x,y) / x \in C_1, y \in C_2\}}$



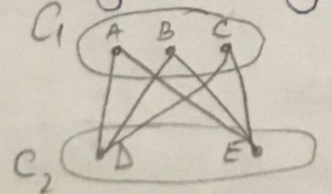
$d_{sl}(C_1, C_2) = d(C, D)$

2. Complete-linkage: $\boxed{d_{cl}(C_1, C_2) = \max\{d(x,y) / x \in C_1, y \in C_2\}}$



$d_{cl}(C_1, C_2) = d(A, E)$

3. Average-linkage: $\boxed{d_{al}(C_1, C_2) = \text{avg}\{d(x,y) / x \in C_1, y \in C_2\}}$



$$d_{al}(C_1, C_2) = \frac{d(A,D) + d(A,E) + d(B,D) + d(B,E) + d(C,D) + d(C,E)}{6}$$

5. Matricea lui Ward:

↳ aici lucrăm doar cu distanța euclidiană

• Centroid: $\boxed{\mu_{cluster} = \frac{\sum_{x \in cluster} x}{|Cluster|}}$

EX: $Cluster = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \end{bmatrix} \right\}$

$\Rightarrow \mu_{cluster} = \frac{\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 5 \\ 6 \end{bmatrix}}{3} = \begin{bmatrix} \frac{1+3+5}{3} \\ \frac{2+4+6}{3} \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

$$\boxed{\mu_{A \cup B} = \frac{|A|\mu_A + |B|\mu_B}{|A| + |B|}}$$

$$\boxed{\frac{\sum_{x \in C_1} x + \sum_{y \in C_2} y}{|C_1| + |C_2|}}$$

$$\boxed{d_{ward}(C_1, C_2) = \sum_{x \in C_1 \cup C_2} d^2(x, \mu_{C_1 \cup C_2}) - \sum_{y \in C_1} d^2(y, \mu_{C_1}) - \sum_{z \in C_2} d^2(z, \mu_{C_2})}$$

↳ $\boxed{d_{ward}(C_1, C_2) = \frac{|C_1||C_2|}{|C_1| + |C_2|} \cdot \|\mu_{C_1} - \mu_{C_2}\|^2}$

• **Algoritmul pentru formarea dendogramei bottom up**

La o iterație: - calculăm distanțele între oricare 2 clustere

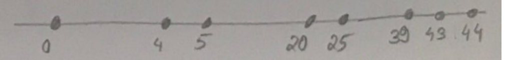
- luăm distanța minimă și combinăm clusterele corespunzătoare

↳ se oprește când avem un nr de clustere specificat sau am unit toate clusterele

Folosim convenții în caz de egalitate

EXEMPLE

1. Single linkage



I

d	0	4	5	20	25	39	43	44
0	0							
4	4	0						
5	5	1	0					
20	20	16	15	0				
25	25	21	20	5	0			
39	...	35	34	...	14	0		
43	...	39	38	...	...	4	0	
44	...	40	39	...	...	...	1	0

// luăm distanța minimă între oricare 2 clustere

// la prima iterație fiecare punct aparține unui cluster

$\Rightarrow$ distanța minimă din tabel = 1

$\Rightarrow C_1 = \{4, 5\}$, $R(C_1) = 1$

II Desenăm din nou tabelul și calculăm distanțele

d	0	4/5	20	25	39	43	44
0	0						
4/5	4	0					
20	20	15	0				
25	25	20	5	0			
39	39	34	19	14	0		
43	43	38	23	18	4	0	
44	44	39	24	19	5	1	0

// ⓘ = valoare aleasă din cele 2 valori din tabelul anterior

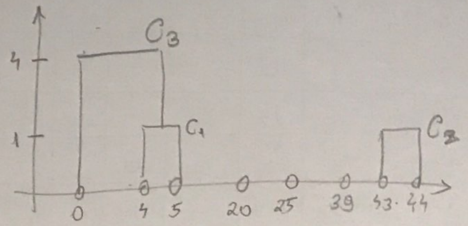
$\Rightarrow$ distanța minimă $= 1 \Rightarrow C_2 = \{43, 44\}$

$R(C_2) = 1$

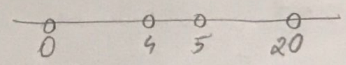
III // tabel $\rightarrow$ dist min $= 4 \Rightarrow C_3 = C_1 \cup \{0\}$

$R(C_3) = 4$

$\Rightarrow$ Dendrogramă :



2. Matricea lui Ward



d	0	4	5	20
0	0			
4	8	0		
5	12,5	0,5	0	
20	200	128	112,5	0

$d_{ward}(0,4) = \frac{n_0 \cdot n_4}{n_0 + n_4} \cdot d^2(\mu_0, \mu_4)$

$= \frac{1 \cdot 1}{1+1} \cdot d^2(0,4)$

$= \frac{1}{2} \cdot (4-0)^2 = \frac{1}{2} \cdot 16 = 8$

$d_{ward}(0,5) = \frac{1}{2}(5-0)^2 = \frac{1}{2} \cdot 25 = 12{,}5$

$d_{ward}(0,20) = \frac{1}{2}(20-0)^2 = \frac{1}{2} \cdot 400 = 200$

$d_{ward}(4,5) = \frac{1}{2}(5-4)^2 = \frac{1}{2} \cdot 1 = 0{,}5$

$d_{ward}(4,20) = \frac{1}{2} \cdot 16^2 = \frac{1}{2} \cdot 256 = 128$

$d_{ward}(5,20) = \frac{1}{2} \cdot 15^2 = \frac{1}{2} \cdot 225 = 112{,}5$

$\Rightarrow C_1 = \{4, 5\}$

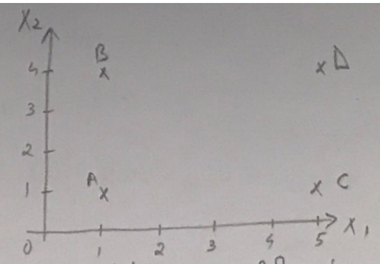
$R(C_1) = 0{,}5$

4

ALTE EXEMPLE

Ex 1:

	X_1	X_2
A	1	1
B	1	4
C	5	1
D	5	4



a) Aplicați alg de clusterizare ierarhică bottom up, folosind:

- distanța euclidiană pt vectori
- distanța single linkage pentru clustere

t=1

d_2	A	B	C	D
A	0			
B	3	0		
C	4	5	0	
D	5	4	3	0

$d_{SL}(A,B) = \min\{d_2(A,B)\}$

$= d_2(A,B) = \sqrt{(1-1)^2 + (1-4)^2}$

$= \sqrt{0 + 3^2} = 3$

$d_{SL}(A,C) = d_2(A,C) = \sqrt{(1-5)^2 + (1-1)^2}$

$= 4$

$d_{SL}(A,D) = d_2(A,D) = \sqrt{(1-5)^2 + (1-4)^2}$

$= \sqrt{16+9} = 5$

$d_{SL}(B,C) = d_2(B,C) = \sqrt{(1-5)^2 + (4-1)^2} = 5$

$d_{SL}(B,D) = d_2(B,D) = \sqrt{(1-5)^2 + (4-4)^2} = 4$

$d_{SL}(C,D) = d_2(C,D) = 3$

Convenție: În caz de distanțe egale între clustere, vom alege să unim clusterele cu punctele în fct de ordinea alfabetică

$\Rightarrow$ dist minimă $= 3 = d_{SL}(A,B) \rightarrow C_1 = \{A, B\}$, $R(C_1) = 3$

t=2

d	A|B	C	D
A|B	0		
C	4	0	
D	4	3	0

$d_{SL}(A|B, C) \overset{\text{def}}{=} \min\{d_{SL}(A,C), d_{SL}(B,C)\}$

$= d_2(A,C) = 4$

$d_{SL}(A|B, D) \overset{\text{def}}{=} \min\{d_{SL}(A,D), d_{SL}(B,D)\} = 4$

$d_{SL}(C,D) = d_2(C,D) = 3$

$\Rightarrow C_2 = \{C, D\}$, $R(C_2) = 3$

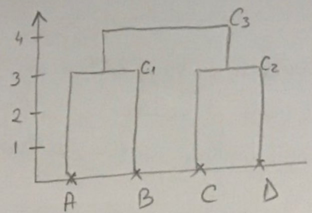
5

d	AB	C|D
AB	0	
C|D	4	0

$d_{SL}(AB, C|D) = \min\{d_{SL}(AB,C), d_{SL}(AB,D)\}$
$= d_{SL}(AB,C) = 4$

$\Rightarrow C_3 = C_1 \cup C_2 = \{A,B,C,D\},\ h(C_3)=4$

Dendrograma:



b) —//— (folosind complete-linkage pt clustere)

d	A	B	C	D
A	0			
B	3	0		
C	4	5	0	
D	5	4	3	0

$d_{CL}(A,B) = \max\{d_2(A,B)\}$
$= d_2(A,B) = 3$

$d_{CL}(A,C) = \max\{d_2(A,C)\} = 4$

....

$\Rightarrow C_1 = \{A,B\},\ h(C_1)=3$

d	A|B	C	D
A|B	0		
C	5	0	
D	5	3	0

$d_{CL}(A|B, C) = \max\{d_{CL}(A,C), d_{CL}(B,C)\}$
$= d_{CL}(B,C) = 5$

$d_{CL}(A|B, D) = \max\{d_{CL}(A,D), d_{CL}(B,D)\}$
$= d_{CL}(A,D) = 5$

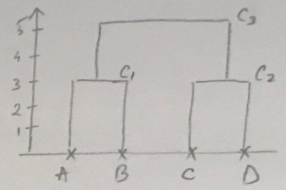
$\Rightarrow C_2 = \{C,D\},\ h(C_2)=3$

d	AB	C|D
AB	0	
C|D	5	0

$d_{CL}(AB, C|D) = \max\{d_{CL}(AB,C), d_{CL}(AB,D)\}$
$= d_{CL}(AB,C) = 5$

$\Rightarrow C_3 = C_1 \cup C_2 = \{A,B,C,D\},\ h(C_3)=5$

Dendrograma:



c) —//— (folosind average-linkage pt clustere)

d	A	B	C	D
A	0			
B	3	0		
C	4	5	0	
D	5	4	3	0

$d_{AL}(A,B) = avg\{d_2(A,B)\} = d_2(A,B) = 3$

$d_{AL}(A,C) = \frac{d_2(A,C)}{1} = d_2(A,C) = 4$

$d_{AL}(A,D) = \frac{d_2(A,D)}{1} = d_2(A,D) = 5$

$d_{AL}(B,C) = \frac{d_2(B,C)}{1} = 5$

....

$\Rightarrow C_1 = \{A,B\},\ h(C_1)=3$

d	A|B	C	D
A|B	0		
C	4,5	0	
D	4,5	3	0

$d_{AL}(A|B, C) \stackrel{def}{=} \frac{d_2(A,C)+d_2(B,C)}{2} = \frac{4+5}{2} = \frac{9}{2}$

$d_{AL}(A|B, D) \stackrel{def}{=} \frac{d_2(A,D)+d_2(B,D)}{2} = \frac{5+4}{2} = \frac{9}{2}$

$d_{AL}(C,D) \stackrel{def}{=} \frac{d_2(C,D)}{1} = d_2(C,D) = 3$

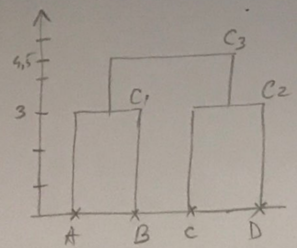
$\Rightarrow C_2 = \{C,D\},\ h(C_2)=3$

d	AB	C|D
AB	0	
C|D	4,5	0

$d_{AL}(AB, C|D) \stackrel{def}{=} \frac{d_{AL}(AB,C)+d_{AL}(AB,D)}{2}$
$= \frac{4,5+4,5}{2} = 4,5$

$\Rightarrow C_3 = C_1 \cup C_2 = \{A,B,C,D\},\ h(C_3)=4,5$

Dendrograma:



d) —//— (folosind metrica lui Ward pt clustere)

d	A	B	C	D
A	0			
B	4,5	0		
C	8	12,5	0	
D	12,5	8	4,5	0

$$d_{ward}(A,B) = \frac{d_A \cdot d_B}{d_A + d_B} \cdot \cancel{d_2}\|\mu_A - \mu_B\|^2 \quad \cancel{\left\| \begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}1\\4\end{bmatrix} \right\|^2}$$

$$\boxed{d_A = |A|}$$

$$= \frac{1 \cdot 1}{1+1} \cdot \cancel{d_2(A,B)} = \left\| \begin{bmatrix}0\\-3\end{bmatrix} \right\|^2$$

$$= \frac{1}{2} \cdot \cancel{\sqrt{(1-1)^2+(1-4)^2} = \frac{3}{2}} \left(\sqrt{0^2+(-3)^2}\right)^2 = \frac{1}{2} \cdot 3^2 = 4,5$$

$$d_{ward}(A,B) = \underbrace{\|A - \mu_{AB}\|^2 + \|B - \mu_{AB}\|^2}_{\sum_{x \in A \cup B} \|x - \mu_{AB}\|^2} - \|A - \mu_A\|^2 - \|B - \mu_B\|^2$$

$$= \left\| \begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}1\\2,5\end{bmatrix} \right\|^2 + \left\| \begin{bmatrix}1\\4\end{bmatrix} - \begin{bmatrix}1\\2,5\end{bmatrix} \right\|^2 + \left\| \begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix} \right\|^2 + \left\| \begin{bmatrix}1\\4\end{bmatrix} - \begin{bmatrix}1\\4\end{bmatrix} \right\|^2$$

$$\mu_{AB} = \frac{A+B}{|A|+|B|} = \frac{\begin{bmatrix}1\\1\end{bmatrix} + \begin{bmatrix}1\\4\end{bmatrix}}{2} = \begin{bmatrix}1\\2,5\end{bmatrix} \qquad \left\| \begin{bmatrix}x_1\\x_2\end{bmatrix} \right\|_p = (x_1^p + x_2^p)^{\frac{1}{p}}$$

$$\mu_A = A, \quad \mu_B = B$$

$$= \left\| \begin{bmatrix}0\\-1,5\end{bmatrix} \right\|^2 + \left\| \begin{bmatrix}0\\1,5\end{bmatrix} \right\|^2 + \left\| \begin{bmatrix}0\\0\end{bmatrix} \right\|^2 + \left\| \begin{bmatrix}0\\0\end{bmatrix} \right\|^2$$

$$= \sqrt{0^2+\left(-\tfrac{3}{2}\right)^2}^2 + \sqrt{0^2+\left(\tfrac{3}{2}\right)^2}^2 + \sqrt{0^2+0^2}^2 + \sqrt{0^2+0^2}^2$$

$$= \left(\sqrt{\tfrac{9}{4}}\right)^2 + \left(\sqrt{\tfrac{9}{4}}\right)^2 = \left(\tfrac{3}{2}\right)^2 + \left(\tfrac{3}{2}\right)^2 = \frac{9}{4} + \frac{9}{4} = \frac{18}{4} = \frac{9}{2} = 4,5$$

$$d_{ward}(A,C) = \frac{|A| \cdot |C|}{|A|+|C|} \cdot \left\| \begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}5\\1\end{bmatrix} \right\|^2 = \frac{1}{2} \cdot \left((-4)^2 + (0)^2\right) = \frac{1}{2} \cdot 16 = 8$$

$$d_{ward}(A,D) = \frac{|A||D|}{|A|+|D|} \cdot \left\| \begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}5\\4\end{bmatrix} \right\|^2 = \frac{1}{2}\left((-4)^2 + (-3)^2\right) = \frac{1}{2} \cdot 25 = 12,5$$

$$d_{ward}(B,C) = \frac{|B||C|}{|B|+|C|} \cdot \left\| \begin{bmatrix}1\\4\end{bmatrix} - \begin{bmatrix}5\\1\end{bmatrix} \right\|^2 = \frac{1}{2}\left((-4)^2 + 3^2\right) = \frac{1}{2} \cdot 25 = 12,5$$

$$d_{ward}(B,D) = \frac{|B||D|}{|B|+|D|} \left\| \begin{bmatrix}1\\4\end{bmatrix} - \begin{bmatrix}5\\4\end{bmatrix} \right\|^2 = \frac{1}{2} \cdot (-4)^2 = \frac{16}{2} = 8$$

$$d_{ward}(C,D) = \frac{|C||D|}{|C|+|D|} \left\| \begin{bmatrix}5\\1\end{bmatrix} - \begin{bmatrix}5\\4\end{bmatrix} \right\|^2 = \frac{1}{2}\left(0^2 + (-3)^2\right) = \frac{9}{2} = 4,5$$

$$\Rightarrow C_1 = \{A, B\}, \; R(C_1) = 4,5$$

d	A|B	C	D
A|B	0		
C	12,1	0	
D	12,1	4,5	0

$$d_{ward}(A|B, C) = \frac{|C_1||C|}{|C_1|+|C|} \cdot \|\mu_{A|B} - \mu_C\|$$

$$\mu_{AB} = \frac{A+B}{2} = \frac{\begin{bmatrix}1\\1\end{bmatrix} + \begin{bmatrix}1\\4\end{bmatrix}}{2} = \begin{bmatrix}1\\2,5\end{bmatrix}$$

$$\dashrightarrow = \frac{2 \cdot 1}{2+1} \cdot \left\| \begin{bmatrix}1\\2,5\end{bmatrix} - \begin{bmatrix}5\\1\end{bmatrix} \right\|^2 = \frac{2}{3} \cdot \left((-4)^2 + 1.5^2\right)$$

$$= \frac{2}{3} \cdot 16 + \frac{\cancel{2}}{\cancel{3}} \cdot \frac{\cancel{9}\,3}{4\,2} = \frac{32}{3} + \frac{3}{2} = 10,6 + 1,5 = 12,1$$

$$d_{ward}(A|B, D) = \frac{|C_1||D|}{|C_1|+|D|} \left\| \begin{bmatrix}1\\2,5\end{bmatrix} - \begin{bmatrix}5\\4\end{bmatrix} \right\|^2 = \frac{2}{3} \cdot \left((-4)^2 + (-1,5)^2\right)$$

$$= \frac{2}{3} \cdot 16 + \frac{2}{3} \cdot \frac{9}{4} = 12,1$$

$$\Rightarrow C_2 = \{C, D\}, \; R(C_2) = 4,5$$

d	AB	C|D
AB	0	
C|D	16	0

$$d_{ward}(AB, C|D) = \frac{|C_1||C_2|}{|C_1|+|C_2|} \cdot \|\mu_{AB} - \mu_{CD}\|^2$$

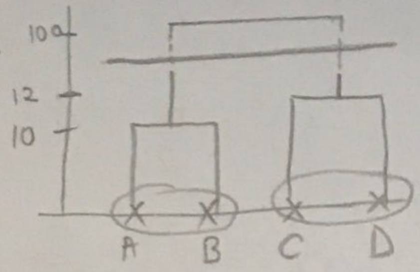
$$\mu_{CD} = \frac{\begin{bmatrix}5\\1\end{bmatrix} + \begin{bmatrix}5\\4\end{bmatrix}}{2} = \begin{bmatrix}5\\2,5\end{bmatrix}$$

$$\dashrightarrow = \frac{4}{4} \cdot \left\| \begin{bmatrix}1\\2,5\end{bmatrix} - \begin{bmatrix}5\\2,5\end{bmatrix} \right\|^2 = (-4)^2 = 16$$

$$\Rightarrow C_3 = C_1 \cup C_2 = \{A, B, C, D\}, \; R(C_3) = 16$$

Numar natural de clustere

↳ parcurgând dendograma de jos în sus, o tăiem pe aceasta cu o linie orizontală, acolo unde distanța crește foarte mult.

Ex: 

$\Rightarrow$ nr nat de clustere $= 2$

Demonstrații

• → Notatie : $\Delta(X,Y) = d_{ward}(X,Y)$

→ Dem că $\Delta(X,Y) = \sum_{i\in X\cup Y} \|x_i - \mu_{X\cup Y}\|^2 - \sum_{j\in X} \|x_j - \mu_X\|^2 - \sum_{k\in Y}\|x_k - \mu_Y\|^2$

$$= \frac{m_X m_Y}{m_X + m_Y} \|\mu_X - \mu_Y\|^2, \text{ unde } \begin{matrix} m_X = |X| \\ m_Y = |Y| \end{matrix}$$

$$\Delta(X,Y) = \sum_{i\in X\cup Y} \|x_i - \mu_{X\cup Y}\|^2 - \sum_{j\in X}\|x_j - \mu_X\|^2 - \sum_{k\in Y}\|x_k - \mu_Y\|^2$$

$$= \sum_{i\in X\cup Y}(x_i - \mu_{X\cup Y})^2 - \sum_{j\in X}(x_j - \mu_X)^2 - \sum_{k\in Y}(x_k - \mu_Y)^2$$

$$= \cancel{\sum_{i\in X\cup Y} x_i^2} - 2\mu_{X\cup Y}\sum_{i\in X\cup Y} x_i + \sum_{i\in X\cup Y}\mu_{X\cup Y}^2 - \cancel{\sum_{j\in X} x_j^2} + 2\mu_A \sum_{j\in X} x_j$$

$$- \sum_{j\in X}\mu_X^2 - \cancel{\sum_{k\in Y} x_k^2} + 2\mu_B\sum_{k\in Y} x_k - \sum_{k\in Y}\mu_Y^2$$

$$= -\frac{2(m_X\mu_X + m_Y\mu_Y)^2}{m_X+m_Y} + \frac{(m_X\mu_X + m_Y\mu_Y)^2}{m_X + m_Y} + 2\mu_X m_X \mu_X + 2\mu_Y m_Y \mu_Y - m_X\mu_X^2 - m_Y\mu_Y^2$$

$$= -\frac{(m_X\mu_X + m_Y\mu_Y)^2}{m_X+m_Y} + 2m_X\mu_X^2 + 2m_Y\mu_Y^2 - m_X\mu_X^2 - m_Y\mu_Y^2$$

$$= \frac{-(m_X\mu_X + m_Y\mu_Y)^2 + (m_X\mu_X^2 + m_Y\mu_Y^2)(m_X+m_Y)}{m_X+m_Y}$$

$$= \frac{m_X m_Y}{m_X+m_Y}(\mu_X^2 - 2\mu_X\mu_Y + \mu_Y^2) = \frac{m_X m_Y}{m_X+m_Y}\|\mu_X - \mu_Y\|^2$$

III K-means

• Algoritm

↳ Input: $\underbrace{x_1, x_2, \dots x_m}_{\text{puncte}} \in \mathbb{R}^d$ și $\underbrace{K}_{\text{nr de clustere dorite}} \in \mathbb{N}$

↳ Output: O anumită K-partiție pt $\{x_1, \dots, x_m\}$
// descompunere în mulțimi disjuncte

↳ Procedură:

1. Inițializare:
   - se fixează în mod arbitrar $\underbrace{\mu_1^0, \mu_2^0, \dots, \mu_K^0}_{\text{centroizii inițiali}}$

   $\boxed{\mu_j^t, \; t = \text{nr iterației}, \; j = \text{nr clusterului}}$

   - se asignează fiecare instanță x_i la centroidul cel mai apropiat (pe baza distanței), formând astfel clusterele $C_1^0, C_2^0, \dots, C_K^0$

2. Corpul iterativ:

   Pas I: Se calculează noile poziții ale centroizilor:

   $$\boxed{\mu_j^t = \frac{1}{|C_j^{t-1}|} \cdot \sum_{x_i \in C_j^{t-1}} x_i}, \text{ pt } j = \overline{1, K}$$

   Pas II: Se reasignează instanțele $x_1, x_2, \dots x_m$ în funcție de distanță $\Rightarrow C_1^{t+1}, C_2^{t+1}, \dots C_K^{t+1}$

   // La alg K-means se folosește doar distanța euclidiană

3. Oprire - atunci când poz centroizilor nu se schimbă
- componența clusterelor nu se mai schimbă de la o iterație la alta

• Exemplu

	X_1	X_2
A	-1	0
B	1	0
C	0	1
D	3	0
E	3	1

$\mu_1^0 = \begin{bmatrix}-1\\0\end{bmatrix}, \quad \mu_2^0 = \begin{bmatrix}3\\1\end{bmatrix}$

a) Aplicați 2-means ($K=2$).

$\boxed{\text{Metoda I}}$ (analitică)!

t=0:

	A	B	C	D	E
μ_1^0	0	2	$\sqrt{2}$	4	$\sqrt{17}$
μ_2^0	$\sqrt{17}$	$\sqrt{5}$	3	1	0

$\|\mu_1^0 - A\| = \left\|\begin{bmatrix}-1\\0\end{bmatrix} - \begin{bmatrix}-1\\0\end{bmatrix}\right\| = \left\|\begin{bmatrix}0\\0\end{bmatrix}\right\| = 0$

$\|\mu_1^0 - B\| = \left\|\begin{bmatrix}-1\\0\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix}\right\| = \left\|\begin{bmatrix}-2\\0\end{bmatrix}\right\| = \sqrt{(-2)^2} = 2$

$\|\mu_1^0 - C\| = \left\|\begin{bmatrix}-1\\0\end{bmatrix} - \begin{bmatrix}0\\1\end{bmatrix}\right\| = \left\|\begin{bmatrix}-1\\-1\end{bmatrix}\right\| = \sqrt{(-1)^2+(-1)^2} = \sqrt{2}$

$\|\mu_1^0 - D\| = \left\|\begin{bmatrix}-1\\0\end{bmatrix} - \begin{bmatrix}3\\0\end{bmatrix}\right\| = \left\|\begin{bmatrix}-4\\0\end{bmatrix}\right\| = \sqrt{(-4)^2} = 4$

$\|\mu_1^0 - E\| = \left\|\begin{bmatrix}-1\\0\end{bmatrix} - \begin{bmatrix}3\\1\end{bmatrix}\right\| = \left\|\begin{bmatrix}-4\\-1\end{bmatrix}\right\| = \sqrt{(-4)^2+(-1)^2} = \sqrt{17}$

$\|\mu_2^0 - A\| = \left\|\begin{bmatrix}3\\1\end{bmatrix} - \begin{bmatrix}-1\\0\end{bmatrix}\right\| = \left\|\begin{bmatrix}4\\1\end{bmatrix}\right\| = \sqrt{4^2+1^2} = \sqrt{17}$

$\|\mu_2^0 - B\| = \left\|\begin{bmatrix}3\\1\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix}\right\| = \left\|\begin{bmatrix}2\\1\end{bmatrix}\right\| = \sqrt{2^2+1^2} = \sqrt{5}$

$\|\mu_2^0 - C\| = \left\|\begin{bmatrix}3\\1\end{bmatrix} - \begin{bmatrix}0\\1\end{bmatrix}\right\| = \left\|\begin{bmatrix}3\\0\end{bmatrix}\right\| = \sqrt{3^2} = 3$

$\|\mu_2^0 - D\| = \left\|\begin{bmatrix}3\\1\end{bmatrix} - \begin{bmatrix}3\\0\end{bmatrix}\right\| = \left\|\begin{bmatrix}0\\1\end{bmatrix}\right\| = \sqrt{1^2} = 1$

$\|\mu_2^0 - E\| = \left\|\begin{bmatrix}3\\1\end{bmatrix} - \begin{bmatrix}3\\1\end{bmatrix}\right\| = \left\|\begin{bmatrix}0\\0\end{bmatrix}\right\| = 0$

$\Rightarrow \cancel{C_1^0 = \{A,B,C\}} \quad C_1^0 = \{A, B, C\}$

$\cancel{C_2^0 = \{B,D,E\}} \quad C_2^0 = \{D, E\}$

t=1 $\mu_1^1 = \dfrac{A+C+B}{3} = \dfrac{\begin{bmatrix}-1\\0\end{bmatrix}+\begin{bmatrix}0\\1\end{bmatrix}+\begin{bmatrix}1\\0\end{bmatrix}}{3} = \dfrac{\begin{bmatrix}0\\1\end{bmatrix}}{3} = \begin{bmatrix}0\\\frac{1}{3}\end{bmatrix}$

$\mu_2^1 = \dfrac{D+E}{2} = \dfrac{\begin{bmatrix}3\\0\end{bmatrix}+\begin{bmatrix}3\\1\end{bmatrix}}{2} = \dfrac{\begin{bmatrix}6\\1\end{bmatrix}}{2} = \begin{bmatrix}3\\0{,}5\end{bmatrix}$

	A	B	C	D	E
μ_1^1	1,05	1,05	0,66	3,01	3,07
μ_2^1	4,03	2,06	3,04	0,5	0,5

$\|\mu_1^1 - A\| = \left\|\begin{bmatrix}0\\\frac{1}{3}\end{bmatrix} - \begin{bmatrix}-1\\0\end{bmatrix}\right\| = \left\|\begin{bmatrix}1\\\frac{1}{3}\end{bmatrix}\right\| = \sqrt{1^2+\left(\frac{1}{3}\right)^2} = \sqrt{1+\frac{1}{9}} = 1{,}05$

$\|\mu_1^1 - B\| = \left\|\begin{bmatrix}0\\\frac{1}{3}\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix}\right\| = \left\|\begin{bmatrix}-1\\\frac{1}{3}\end{bmatrix}\right\| = \sqrt{(-1)^2+\left(\frac{1}{3}\right)^2} = \sqrt{1+\frac{1}{9}} = 1{,}05$

$\|\mu_1^1 - C\| = \left\|\begin{bmatrix}0\\\frac{1}{3}\end{bmatrix} - \begin{bmatrix}0\\1\end{bmatrix}\right\| = \left\|\begin{bmatrix}0\\-\frac{2}{3}\end{bmatrix}\right\| = \sqrt{\left(-\frac{2}{3}\right)^2} = \sqrt{\frac{4}{9}} = 0{,}66$

$\|\mu_1^1 - D\| = \left\|\begin{bmatrix}0\\\frac{1}{3}\end{bmatrix} - \begin{bmatrix}3\\0\end{bmatrix}\right\| = \left\|\begin{bmatrix}-3\\\frac{1}{3}\end{bmatrix}\right\| = \sqrt{(-3)^2+\frac{1}{3^2}} = \sqrt{9+\frac{1}{9}} = 3{,}01$

$\|\mu_1 - E\| = \left\|\begin{bmatrix}0\\\frac{1}{3}\end{bmatrix} - \begin{bmatrix}3\\1\end{bmatrix}\right\| = \left\|\begin{bmatrix}-3\\-\frac{2}{3}\end{bmatrix}\right\| = \sqrt{(-3)^2+\left(-\frac{2}{3}\right)^2} = \sqrt{9+\frac{4}{9}} = 3{,}07$

$$\|\mu_2' - A\| = \left\| \begin{bmatrix} 3 \\ 0,5 \end{bmatrix} - \begin{bmatrix} -1 \\ 0 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 4 \\ 0,5 \end{bmatrix} \right\| = \sqrt{16 + (0,5)^2} = 4,03$$

$$\|\mu_2' - B\| = \left\| \begin{bmatrix} 3 \\ 0,5 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 2 \\ 0,5 \end{bmatrix} \right\| = \sqrt{4 + (0,5)^2} = 2,06$$

$$\|\mu_2' - C\| = \left\| \begin{bmatrix} 3 \\ 0,5 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 3 \\ -0,5 \end{bmatrix} \right\| = \sqrt{9 + (-0,5)^2} = 3,04$$

$$\|\mu_2' - D\| = \left\| \begin{bmatrix} 3 \\ 0,5 \end{bmatrix} - \begin{bmatrix} 3 \\ 0 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 0 \\ 0,5 \end{bmatrix} \right\| = \sqrt{(0,5)^2} = 0,5$$

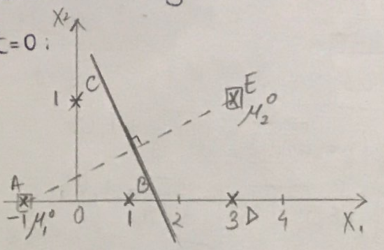
$$\|\mu_2' - E\| = \left\| \begin{bmatrix} 3 \\ 0,5 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 0 \\ -0,5 \end{bmatrix} \right\| = \sqrt{(-0,5)^2} = 0,5$$

$(0,5)^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4} = 0,25$

$$\left.\begin{aligned} \Rightarrow C_1' &= \{A, B, C\} = C_1^0 \\ C_2' &= \{D, E\} = C_2^0 \end{aligned}\right\} \Rightarrow \text{algoritmul se oprește}$$

$\boxed{\text{Metoda II}}$ (geometrică)

t=0:



// facem desenul

// marcăm centroizii

// trasăm mediatoarea/mediatoarele între centroizi (1NN)

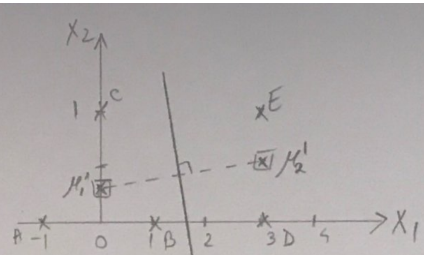
// vedem cum sunt împărțite instanțele

$$\Rightarrow C_1^0 = \{A, B, C\}$$
$$C_2^0 = \{D, E\}$$

t=0: // calculăm centroizii noi

$$\mu_1' = \frac{A+B+C}{3} = \begin{bmatrix} 0 \\ \frac{1}{3} \end{bmatrix}$$

$$\mu_2' = \frac{D+E}{2} = \begin{bmatrix} 3 \\ \frac{1}{2} \end{bmatrix}$$

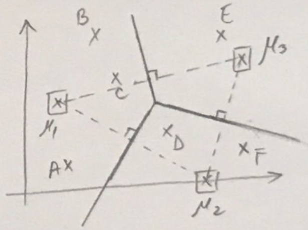


// repetăm pașii

// avem aceleași condiții de oprire

$$\left.\begin{aligned} \Rightarrow C_1' &= \{A, B, C\} = C_1^0 \\ C_2' &= \{D, E\} = C_2^0 \end{aligned}\right\} \Rightarrow \text{alg se oprește}$$

// Dacă avem 3 centroizi =>



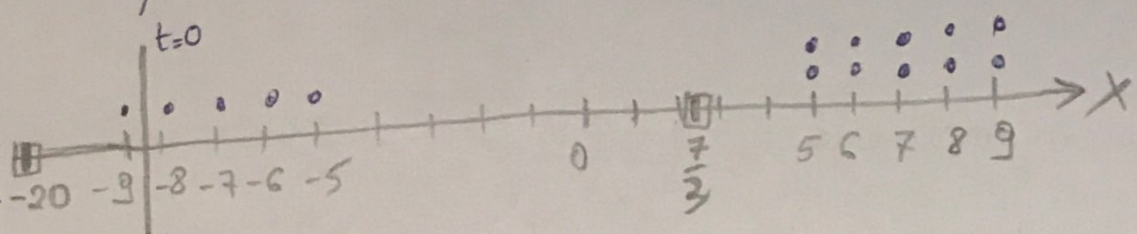
$$\Rightarrow C_1 = \{A, B, C\}$$
$$C_2 = \{D, F\}$$
$$C_3 = \{E\}$$

IV K-means - alg optimizare

// demonstratie criteriu J

Ex 1

X
-9
-8
-7
-6
-5
5
5
6
6
7
7
8
8
9
9



Știm: $\mu_1^0 = -20$

$\mu_2^0 = \frac{7}{3} \simeq 2,33$

Demonstrați în manieră analitică că:

$J(C^0, \mu^0) \geq J(C^1, \mu^1)$ după rularea 2-means

t=0: // facem desenul

// marcăm centroizii

// trasăm mediatoarea și vedem clusterele

$\frac{-20+2,33}{2} = -8,8$

$\Rightarrow C_1^0 = \{-9\}$

$C_2^0 = \{-8,-7,-6,-5,5,5,6,6,7,7,8,8,9,9\}$

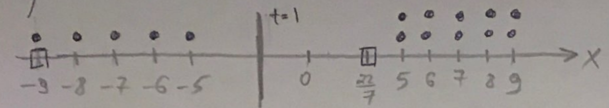
t=1: // calculăm noii centroizi și apoi repetăm pașii de mai sus

$\mu_1^1 = \frac{-9}{1} = -1$

$\mu_2^1 = \frac{\cancel{-8}\cancel{-7}\cancel{-6}\cancel{-5}+\cancel{5}+5+\cancel{6}+6+\cancel{7}+7+\cancel{8}+8+9+9}{14}$

$= \frac{11+7+17+9}{14} = \frac{28+26}{14} = \frac{54}{14} = \frac{22}{7} = 3,14$

// refacem desenul și marcăm centroizii



// trasăm mediana : $\frac{-9+3,14}{2} = -2,93$

$\Rightarrow C_1^1 = \{-9,-8,-7,-6,-5\}$

$C_2^1 = \{5,5,6,6,7,7,8,8,9,9\}$

// Am aplicat până aici 2-means.

Vom demonstra : $J(C^0,\mu^0) \geq J(C^1,\mu^1)$, arătând

că $J(C^0,\mu^0) \boxed{\geq} J(C^0,\mu^1) \boxed{\geq} J(C^1,\mu^1)$

inegalitatea 1 inegalitatea 2

• Inegalitatea 1: ← split-ul de la t=0

$J(C^0,\mu^0) = \|-9-(-20)\|^2 + \|-8-\frac{7}{3}\|^2 + \dots + \|9-\frac{7}{3}\|^2$

$J(C^0,\mu^1) = \|-9-(-9)\|^2 + \|-8-(-9)\|^2 + \dots + \|9-\frac{22}{7}\|^2$

$$\boxed{J(C^{t_1},\mu^{t_2}) = \sum_{i=1}^{m} \|x_i - \mu^{t_2}_{C_j^{t_1}}\|^2, \text{ unde } x_i \in C_j^{t_1}}$$

Notăm cu: $f(x) = \|-9-x\|^2$

$g(x) = \|-8-x\|^2 + \|-7-x\|^2 + \dots + \|9-x\|^2$

$\Rightarrow J(C^0,\mu^0) = f(-20) + g(\frac{7}{3})$

$J(C^0,\mu^1) = f(-9) + g(\frac{22}{7})$

$f(x) = \|9-x\|^2 = (\sqrt{(9-x)^2})^2 = (9-x)^2 = 81 \mp 18x + x^2$

$= \underbrace{1}_{a} \cdot x^2 - \underbrace{18}_{b} x + \underbrace{81}_{c}$ // funcție de gradul 2

$a = 1 > 0 \Rightarrow$ funcția are minim $\Rightarrow x_{min} = \frac{-b}{2a}$

$\Rightarrow x_{min} = -\frac{(+18)}{2\cdot 1} = -\frac{18}{2} = -9$

$g(x) = (-8-x)^2 + (-7-x)^2 + \dots + (9-x)^2 + (9-x)^2$

$= 64 + 16x + x^2 + 49 + 14x + x^2 + \dots + 81 - 18x + x^2$

$= x^2 \underbrace{(14)}_{a} \mp \underbrace{(10 \mp 12 \mp 14 \mp 16 \mp 18)}_{b=-88} x + 3\cdot 64 + \dots$

$a = 14 > 0 \Rightarrow$ funcția are minim $\Rightarrow x_{min} = \frac{-+b}{2a}$

$x_{min} = \frac{-+88}{28} = +\frac{22}{7}$

$J(C^0,\mu^0) = f(-20) + g(\frac{7}{3})$

$J(C^0,\mu^1) = f(-9) + g(\frac{22}{7})$

x_{min} pt f $\quad x_{min}$ pt g

$\Rightarrow J(C^0,\mu^0) > J(C^0,\mu^1)$ $\boxed{A}$

• Inegalitatea 2 ← split-ul t=1

$J(C^0,\mu^1) = \|-9-(-9)\|^2 + \dots + \|-5-(+\frac{22}{7})\|^2 + \|5-\frac{22}{7}\|^2 + \dots + \|9-\frac{22}{7}\|^2$

$J(C^1,\mu^1) = \|-9-(-9)\|^2 + \dots + \|-5-(-9)\|^2 + \|5-\frac{22}{7}\|^2 + \dots \|9-\frac{22}{7}\|^2$

// comparăm termen cu termen

$\Rightarrow J(C^0,\mu^1) > J(C^1,\mu^1)$ $\boxed{B}$

Din $\boxed{A} + \boxed{B} \Rightarrow J(C^0, \mu^0) > J(C^1, \mu^1)$

// am terminat demonstrația analitică

// dacă se cerea demonstrație numerică, doar calculam $J(C^0, \mu^0)$ și $J(C^1, \mu^1)$ și comparam valorile.

$\boxed{\text{EX 2}}$

X_1	X_2
1	1
1	4
5	1
5	4

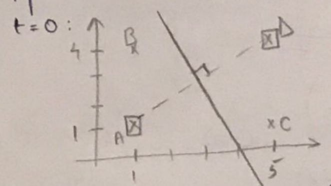
Considerăm că aplicăm alg 2-means.

$\mu_1^0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ \mu_2^1 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$

Dem analitic $J(C^0, \mu^0) \geq J(C^1, \mu^1)$

Aplicăm K-means:

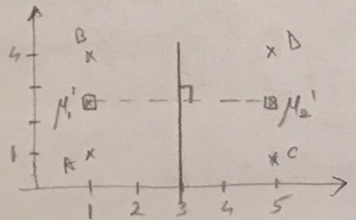
t=0:



$\Rightarrow C_1 = \{A, B\}$
$C_2 = \{C, D\}$

t=1: $\mu_1^1 = \frac{A+B}{2} = \frac{\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 4 \end{bmatrix}}{2} = \begin{bmatrix} 1 \\ 2,5 \end{bmatrix}$

$\mu_2^1 = \frac{C+D}{2} = \frac{\begin{bmatrix} 5 \\ 1 \end{bmatrix} + \begin{bmatrix} 5 \\ 4 \end{bmatrix}}{2} = \begin{bmatrix} 5 \\ 2,5 \end{bmatrix}$



$\Rightarrow C_1 = \{A, B\}$
$C_2 = \{C, D\}$

// Am terminat de aplicat 2-means

Vom demonstra $J(C^0, \mu^0) \geq J(C^1, \mu^1)$, arătând că

$$\underbrace{J(C^0, \mu^0) \geq J(C^0, \mu^1)}_{\text{inegalitatea 1}} \geq J(C^1, \mu^1)$$

(inegalitatea 2: $J(C^0, \mu^1) \geq J(C^1, \mu^1)$)

• Inegalitatea 1

$$J(C^0, \mu^0) = \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 1 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 5 \\ 1 \end{bmatrix} - \begin{bmatrix} 5 \\ 4 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 5 \\ 4 \end{bmatrix} - \begin{bmatrix} 5 \\ 4 \end{bmatrix} \right\|^2$$

$$J(C^0, \mu^1) = \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 2,5 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 1 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 \\ 2,5 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 5 \\ 1 \end{bmatrix} - \begin{bmatrix} 5 \\ 2,5 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 5 \\ 4 \end{bmatrix} - \begin{bmatrix} 5 \\ 2,5 \end{bmatrix} \right\|^2$$

↳ split-ul t=0

Notăm cu: $f(x) = \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - x \right\|^2 + \left\| \begin{bmatrix} 1 \\ 4 \end{bmatrix} - x \right\|^2$

$g(y) = \left\| \begin{bmatrix} 5 \\ 1 \end{bmatrix} - y \right\|^2 + \left\| \begin{bmatrix} 5 \\ 4 \end{bmatrix} - y \right\|^2$

$\Rightarrow J(C^0, \mu^0) = f\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) + g\left(\begin{bmatrix} 5 \\ 4 \end{bmatrix}\right)$

$J(C^0, \mu^1) = f\left(\begin{bmatrix} 1 \\ 2,5 \end{bmatrix}\right) + g\left(\begin{bmatrix} 5 \\ 2,5 \end{bmatrix}\right)$

$$f(x) = \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 1 \\ 4 \end{bmatrix} - \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|^2$$

$$= (1-x_1)^2 + (1-x_2)^2 + (1-x_1)^2 + (4-x_2)^2$$

$$= 1 - 2x_1 + x_1^2 + 1 - 2x_2 + x_2^2 + 1 - 2x_1 + x_1^2 + 16 - 8x_2 + x_2^2$$

$$= (2x_1^2 + (-2-2)x_1 + 2) + (2x_2^2 + (-2-8)x_2 + 17)$$

$$= \underbrace{(2x_1^2 + (-4)x_1 + 2)}_{f_1(x_1)} + \underbrace{(2x_2^2 + (-10)x_2 + 17)}_{f_2(x_2)}$$

$$\begin{cases} f_1(x_1) = \underset{a}{2}x_1^2 + \underset{b}{(-4)}x_1 + \underset{c}{2} \\ \quad a = 2 > 1 \Rightarrow f_1 \text{ are minim} \Rightarrow x_{1\,min} = \frac{-b}{2a} = \frac{-(-4)}{4} = 1 \\ f_2(x_2) = \underset{a}{2}x_2^2 + \underset{b}{(-10)}x_2 + \underset{c}{17} \\ \quad a = 2 > 1 \Rightarrow f_2 \text{ are minim} \Rightarrow x_{2\,min} = \frac{-b}{2a} = \frac{-(-10)}{4} = \frac{10}{4} = 2{,}5 \end{cases}$$

$$g(y) = \left\| \begin{bmatrix} 5 \\ 1 \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 5 \\ 4 \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\|^2$$

$$= (5-y_1)^2 + (1-y_2)^2 + (5-y_1)^2 + (4-y_2)^2$$

$$= 2(25 - 10y_1 + y_1^2) + 1 - 2y_2 + y_2^2 + 16 - 8y_2 + y_2^2$$

$$= \underbrace{(2y_1^2 + (-20)y_1 + 50)}_{g_1(y_1)} + \underbrace{(2y_2^2 + (-10)y_2 + 17)}_{g_2(y_2)}$$

$$\begin{cases} g_1(y_1) = \underset{a}{2}y_1^2 + \underset{b}{(-20)}y_1 + \underset{c}{50} \\ \quad a = 2 > 1 \Rightarrow g_1 \text{ are min} \Rightarrow y_{1\,min} = \frac{-b}{2a} = \frac{-(-20)}{4} = 5 \\ g_2(y_2) = \underset{a}{2}y_2^2 + \underset{b}{(-10)}y_2 + \underset{c}{17} \\ \quad a = 2 > 1 \Rightarrow g_2 \text{ are min} \Rightarrow y_{2\,min} = \frac{-b}{2a} = \frac{-(-10)}{4} = \frac{10}{4} = 2{,}5 \end{cases}$$

$$\left.\begin{aligned} J(C^0, \mu^0) &= f_1(1) + f_2(1) + g_1(5) + g_2(4) \\ &\quad\; \shortparallel \qquad\; \vee \qquad\;\; \shortparallel \qquad\;\; \vee \\ J(C^0, \mu^1) &= f_1(1) + f_2(2{,}5) + g_1(5) + g_2(2{,}5) \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow J(C^0, \mu^0) > J(C^0, \mu^1) \quad \boxed{A}$$

- Inegalitatea 2

$$J(C^0, \mu^1) = \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 2{,}5 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 1 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 \\ 2{,}5 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 5 \\ 1 \end{bmatrix} - \begin{bmatrix} 5 \\ 2{,}5 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 5 \\ 4 \end{bmatrix} - \begin{bmatrix} 5 \\ 2{,}5 \end{bmatrix} \right\|^2$$

$$J(C^1, \mu^1) = \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 2{,}5 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 1 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 \\ 2{,}5 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 5 \\ 1 \end{bmatrix} - \begin{bmatrix} 5 \\ 2{,}5 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 5 \\ 4 \end{bmatrix} - \begin{bmatrix} 5 \\ 2{,}5 \end{bmatrix} \right\|^2$$

$$\Rightarrow J(C^0, \mu^1) = J(C^1, \mu^1) \quad \boxed{B}$$

Din $\boxed{A} + \boxed{B} \Rightarrow J(C^0, \mu^0) > J(C^1, \mu^1)$

V ML - funcția de verosimilitate

- **Introducere:**

Sunt mai multe moduri de a scrie o probabilitate: $D = (0,0,0,1)$ Bernoulli

V1: $P(X=0) = \frac{3}{4}$
$P(X=1) = \frac{1}{4}$

V2: $\begin{pmatrix} 0 & 1 \\ 1-\theta & \theta \end{pmatrix}$, $X \sim Bernoulli(\theta)$
$\boxed{\theta_{MLE} = \frac{1}{4}}$

V3: $\boxed{P(X=x) = \begin{cases} \frac{3}{4}, & x=0 \\ \frac{1}{4}, & x=1 \end{cases} = \left(\frac{3}{4}\right)^{1\{x=0\}} \cdot \left(\frac{1}{4}\right)^{1\{x=1\}}}$!

$\boxed{1\{a\} = \begin{cases} 0, & a = False \\ 1, & a = True \end{cases}}$

- **Date continue:**

$D = (1,2,3) \Rightarrow$ nu putem spune ce distribuție trebuie să folosim, dar încercăm să exprimăm datele cu ajutorul distribuției normale: $\mathcal{N}(\mu, \sigma^2) \xrightarrow{MLE} \mu_{MLE} = \dots\dots$
$\sigma^2_{MLE} = \dots\dots$

$D = (1,2,3) \Rightarrow \pi_1 \mathcal{N}(\mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(\mu_2, \sigma_2^2)$!
$\pi_1 + \pi_2 = 1$

- $\boxed{\theta_{MLE} \overset{def}{=} \underset{\theta \in [0,1]}{\arg\max} \mathcal{L}_D(\theta) = \underset{\theta \in [0,1]}{\arg\max} \ell_D(\theta)}$

funcția de verosimilitate; funcția de logverosimilitate

EXERCIȚIU: $D = (0,0,0,1)$, $\theta = ?$, $X_i \sim Bernoulli(\theta)$

- $\mathcal{L}_D(\theta) = P(D \mid \theta)$
$= P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4 \mid \theta)$

Știm că datele sunt independente și avem mai multe variante de a continua calculul:

V_1: $\overset{indep}{=} P(X_1=0|\theta)P(X_2=0|\theta)P(X_3=0|\theta)P(X_4=1|\theta)$

↳ prob ca $X_1=0$, știind θ

$$=(1-\theta)(1-\theta)(1-\theta)\cdot\theta=(1-\theta)^3\theta$$

V_2: $\overset{indep}{=}\prod_{i=1}^{4} P(X_i=x_i|\theta)$ ← scris mai general

$$=\prod_{i=1}^{4}(1-\theta)^{1\{x_i=0\}}\cdot\theta^{1\{x_i=1\}}$$

$$=(1-\theta)^{\sum_{i=1}^{4}1\{x_i=0\}}\cdot\theta^{\sum_{i=1}^{4}1\{x_i=1\}}$$ → nr de 1 din Δ

$$=(1-\theta)^{m_0}\cdot\theta^{m_1}$$ → nr de zerouri din Δ

- $\ell_\Delta(\theta)\overset{def}{=}\ln\mathcal{L}_\Delta(\theta)=m_0\ln(1-\theta)+m_1\ln\theta$
- $\ell'_\Delta(\theta)=m_0\cdot\frac{1}{1-\theta}\cdot(-1)+m_1\cdot\frac{1}{\theta}$

$$=\frac{m_1}{\theta}-\frac{m_0}{1-\theta}$$

$$\ell'_\Delta(\theta)=0\iff\frac{m_1}{\theta}-\frac{m_0}{1-\theta}=0\Rightarrow\frac{m_1}{\theta}=\frac{m_0}{1-\theta}$$

$$\iff\frac{m_1}{m_0}=\frac{\theta}{1-\theta}\iff\frac{m_1}{m_0+m_1}=\frac{\theta}{1-\cancel{\theta}+\cancel{\theta}}$$

$$\iff\boxed{\frac{m_1}{m_0+m_1}=\theta_{MLE}}\Rightarrow\boxed{\theta_{MLE}=\frac{1}{4}}$$

- $$\boxed{\mathcal{N}(x|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$$

medie varianță

EXERCIȚII

a) $\Delta=(x_1,x_2,x_3)=(1,2,3)$ ← variabile aleatoare continue

X_1, X_2, X_3 – VA iid, $X_i\sim\mathcal{N}(\mu,\sigma^2=3^2)$

$\mu_{MLE}=?$

$$\mu_{MLE}=\arg\max_{\mu\in R}\mathcal{L}_\Delta(\mu)=\arg\max_{\mu\in R}\ell_\Delta(\mu)$$

$$\mathcal{L}_\Delta(\mu)=P(\Delta|\mu)=P(X_1=x_1,X_2=x_1,X_3=x_3|\mu)$$

$$=P(X_1=x_1|\mu)P(X_2=x_2|\mu)P(X_3=x_3|\mu)$$

$$=P(X_1=1|\mu)P(X_2=2|\mu)P(X_3=3|\mu)$$

$$=\frac{1}{\sqrt{2\pi}\cdot3}\cdot e^{-\frac{1}{2}\left(\frac{1-\mu}{3}\right)^2}\cdot\frac{1}{\sqrt{2\pi}\cdot3}\cdot e^{-\frac{1}{2}\left(\frac{2-\mu}{3}\right)^2}\cdot\frac{1}{\sqrt{2\pi}\cdot3}\cdot e^{-\frac{1}{2}\left(\frac{3-\mu}{3}\right)^2}$$

$$=\left(\frac{1}{\sqrt{2\pi}\cdot3}\right)^3\cdot e^{-\frac{1}{2}\cdot\frac{1}{3^2}\left((1-\mu)^2+(2-\mu)^2+(3-\mu)^2\right)}$$

$$\overset{indep}{=}P(X_1=x_1|\mu)P(X_2=x_2|\mu)P(X_3=x_3|\mu)$$

$$=\prod_{i=1}^{3}P(X_i=x_i|\mu)$$

$$=\prod_{i=1}^{3}\frac{1}{\sqrt{2\pi}\cdot3}\cdot e^{-\frac{1}{2}\left(\frac{x_i-\mu}{3}\right)^2}$$

$$\ell_\Delta(\mu)=\ln\mathcal{L}_\Delta(\mu)=\sum_{i=1}^{3}\left(-\ln(\sqrt{2\pi}\cdot3)-\frac{1}{2}\left(\frac{x_i-\mu}{3}\right)^2\right)$$

$$= -3 \cdot \ln(3\cdot\sqrt{2\pi}) - \frac{1}{2}\cdot\frac{1}{9}\cdot\sum_{i=1}^{3}(x_i-\mu)^2$$

$$= \underbrace{-3\ln(3\sqrt{2\pi})}_{\text{constantă}} - \frac{1}{18}\sum_{i=1}^{3}(x_i-\mu)^2$$

$$\mu_{MLE} = \underset{\mu\in R}{\arg\max} -\frac{1}{18}\sum_{i=1}^{3}(x_i-\mu)^2$$

$$= \underset{\mu\in R}{\arg\min}\sum_{i=1}^{3}(x_i-\mu)^2$$

$$= \underset{\mu\in R}{\arg\min}\sum_{i=1}^{3}(x_i^2-2x_i\mu+\mu^2)$$ fct de gradul 2

$$= \underset{\mu\in R}{\arg\min}\ \underbrace{3}_{a}\mu^2+\mu\Big(\underbrace{-2\sum_{i=1}^{3}x_i}_{b}\Big)+\underbrace{\sum_{i=1}^{3}x_i^2}_{c}$$

$$= \frac{-b}{2a} = \frac{-(-\cancel{2}\sum_{i=1}^{3}x_i)}{\cancel{2}\cdot 3} = \frac{\sum_{i=1}^{3}x_i}{3} = \frac{1+2+3}{3}$$

$$\Rightarrow \boxed{\mu_{MLE}=2}$$

b) $\Delta = (x_1,x_2,x_3) = (1,2,3)$

X_1, X_2, X_3 - VA iid: $X_i \sim N(\mu,\sigma^2)$

$\mu_{MLE}, \sigma^2_{MLE} = ?$

$$(\mu_{MLE},\sigma^2_{MLE}) = \underset{\substack{\mu\in R\\ \sigma>0}}{\arg\max}\ L_\Delta(\mu,\sigma) = \underset{\substack{\mu\in R\\ \sigma>0}}{\arg\max}\ \ell_\Delta(\mu,\sigma)$$

$$L_\Delta(\mu,\sigma) \overset{def}{=} P(\Delta|\mu,\sigma)$$

$$= \cancel{P(X_1=x_1|\mu,\sigma)P(X_2=x_2|\mu,\sigma)P(X_3=x_3|\mu,\sigma)}$$

$$\overset{indep}{=} \prod_{i=1}^{3}P(X_i=x_i|\mu,\sigma)$$

$$= \prod_{i=1}^{3}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$\ell_\Delta(\mu,\sigma) \overset{def}{=} \ln L_\Delta(\mu,\sigma) = \sum_{i=1}^{3}\left(-\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma}\left(\frac{x_i-\mu}{\cancel{\sigma}}\right)^2\right)$$

$$\frac{\partial \ell_\Delta}{\partial\mu} = \sum_{i=1}^{3}\cancel{2}(x_i-\mu)(-1)\cdot\left(-\frac{1}{\cancel{2}\sigma^2}\right)$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{3}(x_i-\mu)$$

$$=0 \iff \sum_{i=1}^{3}(x_i-\mu)=0 \Rightarrow \sum_{i=1}^{3}x_i-3\mu=0$$

$$\Rightarrow \mu = \frac{\sum_{i=1}^{3}x_i}{3} \Rightarrow \boxed{\mu_{MLE} = \frac{1+2+3}{3} = 2}$$

$$\frac{\partial \ell_\Delta}{\partial\sigma} = \sum_{i=1}^{3} -\frac{1}{\sigma} - \frac{1}{\cancel{2}}(x_i-\mu)^2\cdot(-1)\cdot\cancel{2}\cdot\sigma^{-3}$$

$$\ln'(\sigma\sqrt{2\pi}) = \frac{1}{\sigma\cdot\cancel{\sqrt{2\pi}}}\cdot\cancel{\sqrt{2\pi}} = \frac{1}{\sigma}$$

$$\left(\frac{1}{\sigma^2}\right)' = (\sigma^{-2})' = -2\cdot\sigma^{-3}$$

$$= \sum_{i=1}^{3} -\frac{1}{\sigma} + \sigma^{-3}(x_i-\mu)^2$$

$$= -\frac{3}{\sigma} + \cancel{3}\sigma^{-3}\sum_{i=1}^{3}(x_i-\mu)^2$$

$$= -\frac{3}{\sigma} + \frac{\sum_{i=1}^{3}(x_i-\mu)^2}{\sigma^3}$$

$$= 0 \Leftrightarrow -\frac{3}{\sigma} + \frac{\sum_{i=1}^{3}(x_i-\mu)^2}{\sigma^3} = 0$$

$$\Leftrightarrow \frac{\sum_{i=1}^{3}(x_i-\mu)^2}{\sigma^3} = \frac{3}{\sigma}$$

$$\Leftrightarrow \frac{\sum_{i=1}^{3}(x_i-\mu_{MLE})^2}{3} = \sigma^2_{MLE} \Rightarrow \sigma^2_{MLE} = \frac{(1-2)^2+(2-2)^2+(3-2)^2}{3}$$

$$\sigma^2_{MLE} = (-1)^2+0^2+1^2$$

$$\boxed{\sigma^2_{MLE} = \frac{2}{3}}$$

$$\boxed{\begin{aligned} \mu_{MLE} &= \frac{\sum_{i=1}^{m} x_i}{m} \\ \sigma^2_{MLE} &= \frac{\sum_{i=1}^{m}(x_i-\mu_{MLE})^2}{m} \end{aligned}}$$

!

VI EM/GMM

→ expectation maximization

→ Gauss mixture model

• <u>Mixture</u> $\pi_1 \underbrace{\mathcal{N}(\mu_1, \sigma_1^2)} + \pi_2 \underbrace{\mathcal{N}(\mu_2, \sigma_2^2)}$, $\pi_1 + \pi_2 = 1$

gaussiene

// ex cu 2 gaussiene ↑, dar o mixtură de distribuții poate avea n gaussiene

$$= \begin{cases} (X_1, Z_1), (X_2, Z_2) - \text{VA iid} & \text{// date complete} \\ Z_i \sim \text{Categorială}(\pi_1, \pi_2) - \text{VA latente / ascunse} \\ \quad X_i \mid Z_i = 1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \\ \quad X_i \mid Z_i = 2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \\ \quad X_i - \text{VA observabile} \end{cases}$$

// altă definiție

• <u>EXERCITIU</u> - <u>calculare μ_{1MLE} și μ_{2MLE}</u> - DERIVARE

$D = (1, 3, 10, 12)$

Datele au fost produse de urm mixtură:

$$X \sim \frac{1}{4}\mathcal{N}(x; \mu_1, \sigma_1^2 = 1^2) + \frac{3}{4}\mathcal{N}(x; \mu_2, \sigma_2^2 = 4^2)$$

$\pi_1 = \frac{1}{4}, \pi_2 = \frac{3}{4}$, $\pi_1 + \pi_2 = 1$ ✓

$\mu_{1;MLE}, \mu_{2;MLE} = ?$

$$\left[\begin{array}{l} P(Z_i=1)=\frac{1}{4} \\ P(Z_i=2)=\frac{3}{4} \end{array}\right\} \text{prob cu care se alege o gaussiană}$$

$P(Z_i=z/\pi_1,\pi_2,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2) = P(Z_i=1)$ (notație)

$$P(X_i/\underbrace{Z_i=1}_{Z_{i1}=1}) = \mathcal{N}(x_i/\mu_1,\sigma_1^2=1^2)$$

$$P(X_i/\underbrace{Z_i=2}_{Z_{i2}=1}) = \mathcal{N}(x_i/\mu_2,\sigma_2^2=4^2)$$

$$P(X_i=x_i) = \frac{1}{4}\mathcal{N}(x_i/\mu_1,\sigma_1^2=1^2)+\frac{3}{4}\mathcal{N}(x_i/\mu_2,\sigma_2^2=4^2)$$

$$P(\underbrace{Z_i=1}_{Z_{i1}=1}/X_i=x_i) \overset{FB}{=} \frac{P(X_i=x_i/Z_i=1)P(Z_i=1)}{P(X_i=x_i)}$$

$$= \frac{P(X_i=x_i/Z_i=1)P(Z_i=1)}{P(X_i=x_i/Z_i=1)P(Z_i=1)+P(X_i=x_i/Z_i=2)P(Z_i=2)}$$

$$P(\underbrace{Z_i=2}_{Z_{i2}=1}/X_i=x_i) = \frac{P(X_i=x_i/Z_i=2)P(Z_i=2)}{P(X=x_i)}$$

! la toate aceste ar trebui să trecem și parametrii $\pi_1,\pi_2,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2$ la partea marginală

$$\left[\begin{array}{l} Z_i=(Z_{i1},Z_{i2}) \Rightarrow \\ Z_{i1}+Z_{i2}=1 \end{array}\right.$$

X	Z_{i1}	Z_{i2}
1	Z_{11}	Z_{12}
3	Z_{21}	Z_{22}
10	Z_{31}	Z_{32}
12	Z_{41}	Z_{42}

$$\boxed{\begin{array}{l} Z_i=1 \Leftrightarrow Z_{i1}=1 \\ Z_i=2 \Leftrightarrow Z_{i2}=1 \end{array}}$$

NOTAȚII

$Z_i=1$
$Z_i=2$ SAU $Z_{ij} \overset{not}{=} 1\{Z_i=j\}$

„x_i a fost generat de gaussiana $\mathcal{N}_j$?"

$$\left[\begin{array}{l} \mu=(\mu_1,\mu_2) \\ \mu^{(t)}=(\mu_1^{(t)},\mu_2^{(t)}) \end{array}\right.$$

Presupunem că suntem la iterația $t+1$:

Pasul E

$$\boxed{E_{Z/X,\mu^{(t)}}[\underbrace{Z_{ij}}_{\in\{0,1\}}] = 0\cdot P(Z_{ij}=0/X_i=x_i,\mu^{(t)}) + 1\cdot P(Z_{ij}=1/X_i=x_i,\mu^{(t)})}$$

$$\underbrace{E_{Z/X,\mu^{(t)}}[Z_{i1}]}_{\gamma_{i1}} = P(Z_{i1}=1/X_i=x_i,\mu^{(t)})$$

$$= \frac{P(X_i=x_i/Z_{i1}=1,\mu^{(t)})}{P(X_i=x_i)}$$

$\boxed{\gamma_{i2}=1-\gamma_{i1}}$

$$= \frac{P(X_i=x_i/Z_{i1}=1,\mu^{(t)})\cdot P(Z_{i1}=1)}{P(X_i=x_i/Z_{i1}=1,\mu^{(t)})P(Z_{i1}=1)+P(X_i=x_i/Z_{i2}=2,\mu^{(t)})P(Z_{i2}=2)}$$

$$= \frac{\frac{1}{4}\mathcal{N}(x_i/\mu_1^{(t)},\sigma_1^2=1^2)}{\frac{1}{4}\mathcal{N}(x_i/\mu_1^{(t)},1^2)+\frac{3}{4}\mathcal{N}(x_i,\mu_2^{(t)},4^2)}$$

$$\boxed{\mathcal{N}(x/\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$$

$$\boxed{\mathcal{N}(x_i/\mu,\sigma^2)=\frac{1}{\sigma}\cdot\mathcal{N}\left(\frac{x_i-\mu}{\sigma}/0,1\right)}$$

valorile distribuție normale cu $\mu=0$ și $\sigma=1$ sunt date în enunțul problemei

! Pasul E presupune calculul variabilelor γ_{ij}, $i\in\{1,2,3,4\}$, $j\in\{1,2\}$

Pasul M

$$\mathcal{L}_{D_c}(\mu_1,\mu_2) = P(\overbrace{D_c}^{\text{notat } Y \text{ în carte}}/\mu_1,\mu_2)$$

$$= P((X_1=x_1, Z_{11}=z_{11}, Z_{12}=z_{12}), \ldots, (X_4=x_4, Z_{41}=z_{41}, Z_{42}=z_{42})/\mu_1,\mu_2)$$

$\overset{\text{independență}}{=} \prod_{i=1}^{4} P(\underbrace{x_i, z_{i1}, z_{i2}}_{y_i} / \mu_1, \mu_2)$ form. de înmulțire

$= \prod_{i=1}^{4} P(z_{i1}, z_{i2} / \mu_1, \mu_2) P(x_i / z_{i1}, z_{i2}, \mu_1, \mu_2)$

$P(z_{i1}, z_{i2} / \mu_1, \mu_2)$ - nu depind de gaussiene

$= P(z_{i1}, z_{i2}) = \begin{cases} \frac{1}{4}, z_{i1} = 1 \ \& \ z_{i2} = 0 \\ \frac{3}{4}, z_{i2} = 1 \ \& \ z_{i1} = 0 \end{cases}$

$= \left(\frac{1}{4}\right)^{1\{z_{i1}=1\}} \cdot \left(\frac{3}{4}\right)^{1\{z_{i2}=1\}} = \left(\frac{1}{4}\right)^{z_{i1}} \cdot \left(\frac{3}{4}\right)^{z_{i2}}$

$\mathcal{L}_{D_c}(\mu_1, \mu_2) = \prod_{i=1}^{4} \left(\frac{1}{4}\right)^{z_{i1}} \cdot \left(\frac{3}{4}\right)^{z_{i2}} \cdot \mathcal{N}(x_i / \mu_1, 1^2)^{z_{i1}} \cdot \mathcal{N}(x_i / \mu_2, 4^2)^{z_{i2}}$

$\ell_{D_c}(\mu_1, \mu_2) = \ln \mathcal{L}_{D_c}(\mu_1, \mu_2)$

$= \sum_{i=1}^{4} \Big(z_{i1} \cdot \ln \frac{1}{4} + z_{i2} \cdot \ln \frac{3}{4} + z_{i1} \cdot \ln(\mathcal{N}(x_i; \mu_1, 1^2))$
$+ z_{i2} \cdot \ln(\mathcal{N}(x_i; \mu_2, 4^2))\Big)$

$= \sum_{i=1}^{4} \Big(z_{i1} \ln \frac{1}{4} + z_{i2} \ln \frac{3}{4} + z_{i1} \cdot \left(-\ln \sqrt{2\pi} - \frac{1}{2}\left(\frac{x_i - \mu_1}{1}\right)^2\right)$
$+ z_{i2}\left(-\ln 4\sqrt{2\pi} - \frac{1}{2}\left(\frac{x_i - \mu_2}{4}\right)^2\right)\Big)$

Obs: Variabilele aleatoare z_{ij} nu pot fi derivate! De aceea definim funcția auxiliară Q:

$Q(\mu, \mu^{(t)}) = E_{z/x, \mu^{(t)}}[\ell_{D_c}(\mu_1, \mu_2)]$

$= \sum_{i=1}^{4} \Big(E[z_{i1}]\left(\ln \frac{1}{4} - \ln \sqrt{2\pi} - \frac{1}{2}\left(\frac{x_i - \mu_1}{1}\right)^2\right)$
$+ E[z_{i2}]\left(\ln \frac{3}{4} - \ln 4\sqrt{2\pi} - \frac{1}{2}\left(\frac{x_i - \mu_2}{4}\right)^2\right)\Big)$

$\left.\begin{array}{l} E[z_{i1}] = \gamma_{i1} \\ E[z_{i2}] = \gamma_{i2} \end{array}\right\} \to$ calculate la pasul E

$= \sum_{i=1}^{4} \Big(\gamma_{i1}\left(\ln \frac{1}{4} - \ln \sqrt{2\pi} - \frac{1}{2}\left(\frac{x_i - \mu_1}{1}\right)^2\right)$
$+ \gamma_{i2}\left(\ln \frac{3}{4} - \ln 4\sqrt{2\pi} - \frac{1}{2}\left(\frac{x_i - \mu_2}{4}\right)^2\right)\Big)$

$\boxed{\frac{\partial Q}{\partial \mu_1} = \sum_{i=1}^{4} \gamma_{i1}\left(-\frac{1}{2} \cdot 2(x_i - \mu_1) \cdot (-1)\right)}$

$\begin{bmatrix} \mu_1^{(t+1)} \\ \mu_2^{(t+1)} \end{bmatrix} = \underset{\mu \in \mathbb{R}^2}{\arg\max}\ Q(\mu, \mu^{(t)})$

$= \begin{bmatrix} \underset{\mu_1 \in \mathbb{R}}{\arg\max} \sum_{i=1}^{4} \Big(\gamma_{i1}\Big(\overbrace{\ln \frac{1}{4} - \ln \sqrt{2\pi}}^{\text{constante}} - \frac{1}{2}\left(\frac{x_i - \mu_1}{1}\right)^2\Big)\Big) \\ \underset{\mu_2 \in \mathbb{R}}{\arg\max} \sum_{i=1}^{4} \Big(\gamma_{i2}\Big(\underbrace{\ln \frac{3}{4} - \ln 4\sqrt{2\pi}}_{\text{constante}} - \frac{1}{2}\left(\frac{x_i - \mu_2}{4}\right)^2\Big)\Big) \end{bmatrix}$

$= \begin{bmatrix} \underset{\mu_1 \in \mathbb{R}}{\arg\max} \sum_{i=1}^{4} \gamma_{i1}\left(-\frac{1}{2}\left(\frac{x_i - \mu_1}{1}\right)^2\right) \\ \underset{\mu_2 \in \mathbb{R}}{\arg\max} \sum_{i=1}^{4} \gamma_{i2}\left(-\frac{1}{2}\left(\frac{x_i - \mu_2}{4}\right)^2\right) \end{bmatrix}$

$= \begin{bmatrix} \underset{\mu_1 \in \mathbb{R}}{\arg\min} \sum_{i=1}^{4} \gamma_{i1}(x_i - \mu_1)^2 \\ \underset{\mu_2 \in \mathbb{R}}{\arg\min} \sum_{i=1}^{4} \gamma_{i2}\left(\frac{x_i - \mu_2}{4}\right)^2 \end{bmatrix}$ constantă

$= \begin{bmatrix} \underset{\mu_1 \in \mathbb{R}}{\arg\min} \sum_{i=1}^{4} (\gamma_{i1} x_i^2 - 2\gamma_{i1} \cdot x_i \cdot \mu_1 + \gamma_{i1} \mu_1^2) \\ \underset{\mu_2 \in \mathbb{R}}{\arg\min} \sum_{i=1}^{4} (\gamma_{i2} x_i^2 - 2\gamma_{i2} x_i \cdot \mu_2 + \gamma_{i2} \mu_2^2) \end{bmatrix}$

$$= \begin{bmatrix} \underset{\mu_1 \in \mathbb{R}}{\operatorname{argmin}} \underbrace{\sum_{i=1}^{4} \gamma_{i1} x_i^2}_{c} - \underbrace{\sum_{i=1}^{4} 2\gamma_{i1} x_i \mu_1}_{b} + \underbrace{\sum_{i=1}^{4} \gamma_{i1} \mu_1^2}_{a} \\ \underset{\mu_2 \in \mathbb{R}}{\operatorname{argmin}} \underbrace{\sum_{i=1}^{4} \gamma_{i2} x_i^2}_{c} - \underbrace{\sum_{i=1}^{4} 2\gamma_{i2} x_i \mu_2}_{b} + \underbrace{\sum_{i=1}^{4} \gamma_{i2} \mu_2^2}_{a} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{-b}{2a} \\ \frac{-b}{2a} \end{bmatrix} = \begin{bmatrix} \frac{-(-\cancel{2} \sum_{i=1}^{4} \gamma_{i1} \cdot x_i)}{\cancel{2} \cdot \sum_{i=1}^{4} \gamma_{i1}} \\ \frac{-(-2 \sum_{i=1}^{4} \gamma_{i2} x_i)}{2 \sum_{i=1}^{4} \gamma_{i2}} \end{bmatrix} = \begin{bmatrix} \frac{\sum_{i=1}^{4} \gamma_{i1} x_i}{\sum \gamma_{i1}} \\ \frac{\sum \gamma_{i2} x_i}{\sum \gamma_{i2}} \end{bmatrix} \quad !$$

$\boxed{\text{ALGORITM GM/EMM}}$ – RULARE DE MÂNĂ

Initialization: $\mu_1^{(0)}, \mu_2^{(0)}$

for $t=1$: max Iterations do

    E step: Compute $\gamma_{ij} = E_{Z/X,\mu_1,\mu_2}[Z_{ij}]$, $\forall i \in \overline{1,m}, j \in \overline{1,K}$

    M step: $\mu_1^{(t)} \leftarrow \frac{\sum_{i=1}^{m} \gamma_{i1} x_i}{\sum_{i=1}^{m} \gamma_{i1}}$, $\mu_2^{(t)} \leftarrow \frac{\sum_{i=1}^{m} \gamma_{i2} x_i}{\sum_{i=1}^{m} \gamma_{i2}}$

end for

! La implementare nu este necesară demonstrarea formulelor prin derivare! Trebuie doar să facem calculele.