TEORIE ML 2

SĂPTĂMÂNA 9

→ $\varepsilon_t = err_{D_t}(h_t) = Pr_{x \sim D_t}[y \neq h_t(x)] \Leftrightarrow$ eroarea ponderată la antrenare (luăm în calcul probabilitățile instanțelor clasificate greșit, nu nr. lor);

→ AdaBoost se oprește când găsim $\varepsilon_t = 0$ sau $\varepsilon_t = \frac{1}{2}$ pentru o ipoteză $h_t(x)$;

→ la fiecare iterație, alegem cea mai bună ipoteză slabă (care are eroarea ponderată la antrenare minimă);

→ $\varepsilon_t = \frac{1}{2} - \gamma_t$;

→ $S = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ este setul de date de antrenament; $x_i \in X$, $y_i \in \{-1, +1\}$;

→ $h : X \to \{-1, +1\} \Leftrightarrow$ ipotezele; T este nr. de iterații;

→ $D_1(i) = \frac{1}{m}$, $\forall i = \overline{1, m}$;

→ la fiecare iterație t $(t = \overline{1, T})$, producem ipoteza h_t cu eroarea ponderată la antrenare ε_t, facem update la distribuție $D_{t+1}(i) = \frac{1}{Z_t} D_t(i) \cdot e^{-\alpha_t y_i h_t(x_i)}$, $\forall i = \overline{1, m}$ (unde Z_t este un normalizator pentru ca suma probabilităților să fie 1, $e^{-\alpha_t y_i h_t(x_i)}$ este $e^{-\alpha_t}$, dacă $y_i = h_t(x_i)$ și e^{α_t}, altfel și

$\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$;

→ la sfârșit, calculăm $H_T = sign\left(\sum_{t=1}^{T} \alpha_t h_t\right)$

⟺ votul majoritar;

→ normalizator ⟺ factor de normalizare;

→ AdaBoost poate produce rezultate diferite atunci când are posibilitatea să aleagă între 2 sau mai multe (cele mai bune) ipoteze slabe;

→ $err_{D_{t+1}}(h_t) = \frac{1}{2}$ ⟹ ipoteza h_t nu poate fi reselectată și la iterația $t+1$ (poate fi reselectată la o iterație ulterioară);

→ $Z_t = e^{-\alpha_t} \cdot (1-\varepsilon_t) + e^{\alpha_t} \cdot \varepsilon_t \Rightarrow$
$Z_t = 2\sqrt{\varepsilon_t(1-\varepsilon_t)}$ $(0 < Z_t < 1)$;

→ $D_{t+1}(i) = \begin{cases} \frac{D_t(i)}{2\varepsilon_t}, & i \in M = \{i \mid y_i \neq h_t(x_i)\}; \\ \frac{D_t(i)}{2(1-\varepsilon_t)}, & i \in C = \{i \mid y_i = h_t(x_i)\} \end{cases}$

→ $\varepsilon_i > \varepsilon_j \Rightarrow \alpha_i < \alpha_j$ (antimonotonie);

→ $err_{D_{t+1}}(h_t) = \frac{1}{Z_t} \cdot e^{\alpha_t} \cdot \varepsilon_t$, unde
$err_{D_{t+1}}(h_t) = Pr_{D_{t+1}}(\{x_i \mid h_t(x_i) \neq y_i\})$;

↳ $err_{D_{t+1}}(h_t) = \frac{1}{2}$;

→ $\sum_{i \in M} D_t(i) = \varepsilon_t$, $M = \{ i \mid y_i \neq h_t(x_i) \}$;

→ atunci când lucrăm cu AdaBoost, ~~este~~ este neapărat necesar să ținem cont și de compuși exteriori (clasifică toate instanțele cu același semn);

↳ slides DT, de la slide-ul 107;

→ condiții de învățabilitate γ-slabă (condiții de convergență);

→ $D_{T+1}(i) = \left(m \cdot \prod_{t=1}^{T} Z_t\right)^{-1} \cdot e^{-y_i f(x_i)}$, unde $f(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$ (margine algebrică);

→ $err_S(H_T) \le \prod_{t=1}^{T} Z_t$; $\left(\gamma_t = \frac{1}{2} - \varepsilon_t\right)$;

→ $\prod_{t=1}^{T} Z_t \le e^{-2\sum_{t=1}^{T} \gamma_t^2}$

→ AdaBoost este un algoritm de optimizare secvențială (la fiecare iterație, își optimizează Z_t în funcție de α_t);

↳ slides DT, de la slide-ul 83;

→ garanție de învățare empirică γ-slabă ⟺ dacă există un $\gamma > 0$ astfel încât $\gamma \le \gamma_t$, * pentru $t = \overline{1,T}$, atunci după un anumit nr. de iterații, eroarea la antrenare este strict mai mică decât un ε oarecare fixat ($\varepsilon \le \frac{1}{m}$ ⟹ eroare la antrenare 0);

→ condiția de învățabilitate γ-slabă ⟺ $\text{eros}(H_T) \le e^{-2\sum_{t=1}^{T}\gamma_t^2}$;

SĂPTĂMÂNA 10

→ AdaBoost produce overfitting atunci când datele de antrenament și cele de test sunt foarte mixate

→ învățare supervizată ⟺ clasificare, regresie;

→ învățare nesupervizată ⟺ clusterizare;

→ regresie ⟺ aproximare de funcții de probabilitate;

→ învățare nesupervizată ⟺ nu mai avem etichete pentru instanțele de antrenament;

→ clusterizare (clustering) ⟺ grupare;

→ funcție/măsură de distanță ⟹ suplinește folosirea etichetelor;

→ similaritate ⟺ măsură ce ne arată cât de similare sunt 2 clustere;

→ clusterizare [→ ierarhică / ↓ neierarhică (algoritmul K-means și algoritmul EM);

→ similaritatea este inversul distanței ⟺ cu cât distanța este mai mică, cu atât similaritatea este mai mare;

→ „single-linkage" (nearest-neighbour) $\Leftrightarrow$
$d(A,B) = \min\{d(x,y) \mid x \in A, y \in B\}$;
→ „complete-linkage" (furthest-neighbour) $\Leftrightarrow$
$d(A,B) = \max\{d(x,y) \mid x \in A, y \in B\}$;
→ „average-linkage" $\Leftrightarrow$ $d(A,B) = \frac{1}{|A|\,|B|} \sum_{\substack{x \in A \\ y \in B}} d(x,y)$;

→ A, B sunt 2 clustere;
→ clusterizare ierarhică → bottom-up; top-down;

→ în arbore, înălțimile arcelor sunt proporționale cu similaritatea (cu cât similaritatea este mai mare, cu atât înălțimea este mai mică);
↳ dendrograme $\Leftrightarrow$ arborii pentru clusterizare ierarhică;
→ în general ~~se~~ dendrogramele se fac bottom-up;
→ inițial considerăm că fiecare instanță este în propriul său cluster (se formează clustere singleton); apoi încercăm să le combinăm 2 câte 2; ne uităm la vecinii fiecărui cluster până când obținem un cluster care le înglobează pe toate;
↳ algoritmul bottom-up;

→ trasând o paralelă cu axa orizontală unde distanța dintre clustere este maximă le obținem pe cele pe care le-am obține natural;

↳ carte ML, pag. 755;

→ în mod implicit lucrăm cu distanța euclidiană, dacă nu este specificată alta;

→ putem lucra cu varianta aplatizată a unei ierarhii (în plan);

↳ carte ML, pag. 758;

→ „single-linkage" are tendința să formeze clustere alungite („elongate", înșiruite); „complete-linkage" și „average-linkage" au tendința să formeze clustere de formă mai degrabă sferică;

→ rezultatele (ierarhiile) obținute nu sunt determinate neapărat în mod unic (maximul de similaritate poate fi atins de mai multe perechi);

↳ book-overview, pag. 32;

→ dacă atunci când folosim „single-linkage" și „complete-linkage" se obțin dendrograme identice, nu rezultă în mod neapărat că vom obține aceeași folosind „average-linkage" dendrogramă;

↳ carte ML, pag. 761;

→ $d(X \cup Y, Z) = \dfrac{|X|\, d(X,Z) + |Y|\, d(Y,Z)}{|X| + |Y|}$;

→ putem lucra și cu matrice de distanțe;

↳ slides Cluster, de la slide-ul 4;

→ clusterizare top-down (divizivă) ⟺ inițial pornim de la un cluster cu toate instanțele și le divizăm folosindu-ne de o funcție de coeziune ($coh(A) = \dfrac{C_{|A|}^{2}}{\sum_{x,y \in A} d(x,y)}$ de exemplu) și un de split (identificăm clusterul cu coeziunea cea mai mică și îl spargem în 2 subclustere); facem aceste diviziuni până ajungem la clustere singleton;

↳ slides Cluster, de la slide-ul 12;

→ nr. maxim de niveluri dintr-o dendrogramă (văzută ca arbore în sensul teoriei grafurilor) este $n-1$, iar cel minim este $\lceil \log_2 n \rceil$, unde n este nr. de instanțe de clusterizat;

→ există o anumită corespondență între clusterizarea ierarhică cu similaritate de tip ⟶ „single-linkage" și aflarea arborelui de cost minim dintr-un graf;

„complete-linkage" și aflarea unei clici dintr-un graf;

→ clusterizare partițională ⇔ clusterizare neierarhizată;
→ pe parcursul ~~aplicării~~ execuției algoritmului K-means K este fixat ($K \geq 2$); K ne spune câte clustere să formăm (clusterele sunt disjuncte);
↳ formăm K-partiție;
→ $\mu : \mathcal{P}(\mathbb{R}^m) \to \mathbb{R}^m$ ⇔ funcție de calcul al mediei;
→ centroid ⇔ centru $\overset{\text{presumptiv}}{\text{de greutate}}$;
↳ se alg inițial $f_1, \ldots, f_k$ din $\mathbb{R}^m$;
→ $X = \{x_1, \ldots, x_n\} \subseteq \mathbb{R}^m$ set de date;
→ d ⇔ măsură de distanță din $\mathbb{R}^m$;
→ la fiecare pas formăm K clustere și asignăm instanțele care sunt cele mai apropiate de centroidul respectiv;
→ rezultatul lui K-means nu este neapărat unic;
→ după ce alcătuim clusterele, punem centroizii în poziții mai bune ($f_j = \mu(c_j)$, c_j este un cluster);
↳ centroizii chiar devin centre de greutate;
→ criteriul de oprire neatisfăcut ⇔ continuăm să reorganizăm clusterele;

↳ carte ML, pag. 769;

→ condiții de oprire ⟺ până când componența clusterelor nu se mai modifică de la o iterație la alta sau până când pozițiile centroizilor nu se mai schimbă de la o iterație la alta (cele 2 criterii nu sunt neapărat independente simultan);

→ outlier ⟺ punct singuratic;

→ K-means poate să producă rezultate diferite în funcție de inițializări diferite;

→ funcție de evaluare a „calității" clusterelor (funcție de „coeziune"/„distorsiune"/„eroare totală") ⟺ „suma celor mai mici pătrate" $(J_K(C, \mu) = \sum \|x_i - \mu_{C(x_i)}\|^2$, unde C este K-partiție, μ este K-configurație de centroizi și $\mu_{C(x_i)}$ este centroidul cel mai apropiat de x_i);

→ rezultatul lui K-means poate să difere în funcție de măsura de distanță folosită (chiar dacă avem aceeași inițializare);

→ K-means este sensibil la prezența outlierelor;

↳ carte ML, pag. 775;

→ dacă pe parcursul execuției lui K-means un cluster este vid, centroidul respectiv nu-și schimbă poziția la următoarea iterație;

→ înainte de K-means, este bine să se aplice un algoritm de identificare al outlierilor;
→ atunci când lucrăm cu distanța euclidiană separatorii dintre clustere la K-means sunt de tip liniar;

SĂPTĂMÂNILE 11-14

→ algoritmul EM (Expectation - Maximization) ⟺ permite crearea unor modele probabiliste care pe de o parte, depind de un set de parametri θ, iar pe de altă parte, includ pe lângă variabilele obișnuite („observabile" sau „vizibile") x și variabile necunoscute z („neobservabile", „ascunse" sau „latente");
→ în general, în astfel de situații/modele, nu se poate face în mod direct o estimare a parametrilor modelului (θ) în așa fel încât să se garanteze atingerea maximului verosimilității datelor observabile x;
→ algoritmul EM procedează în manieră iterativă ⟺ constituie o modalitate convenabilă de estimare a parametrilor θ;
→ log-verosimilitatea datelor observabile (x) ⟺ $\log P(x|\theta)$;
→ log-verosimilitatea datelor complete (observabile, x și neobservabile, z) ⟺ $\log P(x, z|\theta)$;

→ log-verosimilitatea datelor observabile (X) se poate exprima în funcție de datele neobservabile (Z) ⇔ $l(\theta) = \log P(X|\theta) = \log\left(\sum_Z P(X,Z|\theta)\right)$;

→ q ⇔ o funcție/distribuție de probabilitate definită peste variabilele ascunse/neobservabile Z;

→ $\log P(X|\theta) \geq \sum_Z q(Z) \log\left(\frac{P(X,Z|\theta)}{q(Z)}\right)$, $\forall X$ (fixat), $\forall \theta$ și $\forall q$ distribuție probabilistă peste variabilele neobservabile Z;

↳ orice funcție de forma $F(q,\theta) = \sum_Z q(Z) \log\left(\frac{P(X,Z|\theta)}{q(Z)}\right)$ constituie o margine inferioară pentru funcția de log-verosimilitate a datelor incomplete/observabile, $l(\theta) = \log P(X|\theta)$;

→ q ⇔ în F este o variabilă de tip funcțional;

→ expresia funcției F este de fapt o medie, $E_{q(Z)}\left[\log\frac{P(X,Z|\theta)}{q(Z)}\right]$, atunci când X, q și parametrul θ se consideră fixați, iar Z este lăsat să varieze;

→ inegalitatea lui Jensen ⇔ considerând X o variabilă aleatoare (reală), dacă φ este o funcție (reală) convexă, atunci $\varphi(E[X]) \leq E[\varphi(X)]$ și dacă φ este o funcție concavă, atunci $\varphi(E[X]) \geq E[\varphi(X)]$;

→ entropia relativă (divergența Kullback-Leibler) ⇔ $KL(q(Z) \| P(Z|X,\theta)) = -\sum_Z q(Z) \log\left(\frac{P(Z|X,\theta)}{q(Z)}\right)$;

→ $\log P(X|\theta) = F(q(Z),\theta) + KL(q(Z) \| P(Z|X,\theta))$;

↳ diferența dintre funcția obiectiv $l(\theta) = \log P(X|\theta)$ și marginea sa inferioară $F(q(Z),\theta)$ este $KL(q(Z) \| P(Z|X,\theta))$;

→ în loc să calculeze maximul funcției de log-verosimilitate $\log P(X \mid \theta)$ în raport cu θ, algoritmul EM va maximiza marginea sa inferioară, $F(q(z), \theta)$, în raport cu ambele argumente, $q(z)$ și θ;

→ pentru a căuta maximul (un maxim local al) marginii inferioare $F(q(z), \theta)$, algoritmul EM aplică metoda creșterii pe coordonate ⟺ după ce inițial ~~z~~ se fixează $\theta^{(0)}$ eventual aleatoriu, se maximizează iterativ funcția $F(q(z), \theta)$ în mod alternativ (mai întâi în raport cu distribuția $q(z)$ și apoi în raport cu parametrul θ);

↳ $\begin{cases} \cdot \text{ pasul E} \Leftrightarrow q^{(t)}(z) = \underset{q(z)}{\arg\max}\, F(q(z), \theta^{(t)}); \\ \cdot \text{ pasul M} \Leftrightarrow \theta^{(t+1)} = \underset{\theta}{\arg\max} (F(q^{(t)}(z), \theta)); \end{cases}$

→ $P(B \cap C \mid A) = P(B \mid A) \cdot P(C \mid A, B)$;

→ $KL(p \,\|\, q) \geq 0, \forall p, q \Rightarrow KL(q(z) \,\|\, P(z \mid X, \theta)) \geq 0$;

→ dacă $\theta^{(t)}$ este valoarea obținută pentru parametrul sau parametrii θ la iterația t a algoritmului EM și considerând această valoare fixată, atunci maximul lui F în raport cu argumentul / distribuția $q(z)$ este atins pentru distribuția $P(z \mid X, \theta^{(t)})$ și valoarea maximului este

$$\max_{q(z)} F(q(z), \theta^{(t)}) = E_{P(z \mid X, \theta^{(t)})} [\log P(z \mid X, \theta^{(t)})] + H(P(z \mid X, \theta^{(t)}));$$

→ $KL(p\|q) \geq 0$ și $KL(p\|q) = 0 \Leftrightarrow p = q$;

→ $E[X-Y] = E[X] - E[Y]$;

→ dacă notăm $Q(\theta|\theta^{(t)}) = E_{P(z|x,\theta^{(t)})}[\log P(x,z|\theta)]$

și $G_t(\theta) = F(P(z|x,\theta^{(t)}),\theta)$, atunci

$G_t(\theta^{(t)}) = \log(P|\theta^{(t)}) = Q(\theta^{(t)}|\theta^{(t)}) + H(P(z|x,\theta^{(t)}))$

și $G_t(\theta) = Q(\theta|\theta^{(t)}) + H[P(z|x,\theta^{(t)})]$;

↳ $H[P(z|x,\theta^{(t)})]$ nu depinde de $\theta \Rightarrow$

$\underset{\theta}{\arg\max}\, G_t(\theta) = \underset{\theta}{\arg\max}\, Q(\theta|\theta^{(t)})$;

→ $\theta^{(t+1)} = \underset{\theta}{\arg\max}\, F(P(z|x,\theta^{(t)}),\theta) = \underset{\theta}{\arg\max}\, G_t(\theta) =$

$= \underset{\theta}{\arg\max}\, Q(\theta|\theta^{(t)})$;

→ reformularea algoritmului EM (ca uzuală):

- pasul E' $\Leftrightarrow$ calculează $Q(\theta|\theta^{(t)}) =$
  $= E_{P(z|x,\theta^{(t)})}[\log P(x,z|\theta)]$;
- pasul M' $\Leftrightarrow$ calculează $\theta^{(t+1)} =$
  $= \underset{\theta}{\arg\max}\, Q(\theta|\theta^{(t)})$;

→ $Q(\theta|\theta^{(t)}) \Leftrightarrow$ media log-verosimilității datelor complete (observabile și neobservabile) în raport cu distribuția condițională $P(z|x,\theta^{(t)})$;

→ dacă considerăm iterațiile $t = 0, 1, \ldots$ și $\theta^{(t+1)} = \underset{\theta}{\arg\max}\, Q(\theta|\theta^{(t)})$ ($\theta^{(0)}$ ales în mod arbitrar), atunci pentru orice t fixat (arbitrar) și pentru orice θ astfel încât $Q(\theta|\theta^{(t)}) \geq Q(\theta^{(t)}|\theta^{(t)})$

are loc inegalitatea $\log P(X|\theta) - \log P(X|\theta^{(t)}) \geq$
$\geq Q(\theta|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t)})$;

↳ orice îmbunătățire a valorii funcției auxiliare $Q(\theta|\theta^{(t)})$ conduce la o îmbunătățire cel puțin la fel de mare a valorii funcției obiectiv $l(\theta)$; dacă înlocuim θ cu $\theta^{(t+1)} = \arg\max_{\theta} Q(\theta|\theta^{(t)})$, atunci $\log P(X|\theta^{(t+1)}) \geq \log P(X|\theta^{(t)})$ (în final, vom avea $l(\theta^{(0)}) \leq l(\theta^{(t)}) \leq l(\theta^{(t+1)}) \leq \ldots$);

→ versiunea „generalizată" a algoritmului EM ⟺ la pasul M la iterația t a algoritmului EM, este suficient ca în loc să se ia $\theta^{(t+1)} = \arg\max_{\theta} Q(\theta|\theta^{(t)})$, să se aleagă $\theta^{(t+1)}$ astfel încât $Q(\theta^{(t+1)}|\theta^{(t)}) > Q(\theta^{(t)}|\theta^{(t)})$;

→ inegalitatea lui Gibbs ⟺ $CH(\theta|\theta^{(t)}) \geq CH(\theta^{(t)}|\theta^{(t)})$;

→ $E[X] = \sum_x x \cdot p(x)$ și $E[g(x)] = \sum_x g(x) p(x)$ (pentru variabile discrete);

→ teorema lui Bayes (3 variabile) ⟺

$$P(R|H,S) = \frac{P(H|R,S)P(R|S)}{P(H|R,S)P(R|S) + P(H|\neg R,S)P(\neg R|S)};$$

→ se dă $X = \{x_1, \ldots, x_n\}$ (date observabile) generat independent cu ajutorul distribuțiilor/ipotezelor $h_1, \ldots, h_m$ și $Z = \{z_1, \ldots, z_n\}$ (date neobservabile) și se cere determinarea lui $\hat{h}$ care maximizează (local) $P(X|h)$;

CamScanner

→ algoritmul EM pentru distribuții Bernoulli:

- inițializare ⇔ atribuire de valori arbitrare pentru parametrii π, p și q ($\pi^{(0)}$, $p^{(0)}$ și $q^{(0)}$ în intervalul $(0,1)$);
- pentru $t = 0, \dots, T-1$ (T fixat) (sau până când log-verosimilitatea datelor observabile nu mai crește semnificativ sau până când $|\pi^{(t)} - \pi^{(t+1)}| < \varepsilon$, $|p^{(t)} - p^{(t+1)}| < \varepsilon$ și $|q^{(t)} - q^{(t+1)}| < \varepsilon$, ε fixat), execută pasul E și pasul M:
- pasul E ⇔ pentru $i = \overline{1,n}$, calculează
$$\mu_i^{(t)} = \frac{p^{(t)x_i}(1-p^{(t)})^{1-x_i}\,\pi^{(t)}}{(p^{(t)x_i}(1-p^{(t)})^{1-x_i})\cdot \pi^{(t)} + q^{(t)x_i}(1-q^{(t)})^{1-x_i}(1-\pi^{(t)})};$$
- pasul M ⇔ calculează
$$p^{(t+1)} = \frac{\delta Q(\theta|\theta^{(t)})}{\delta p} = \frac{\sum_{i=1}^{n}\mu_i^{(t)}x_i}{\sum_{i=1}^{n}\mu_i^{(t)}},$$
$$q^{(t+1)} = \frac{\sum_{i=1}^{n}(1-\mu_i^{(t)})x_i}{\sum_{i=1}^{n}(1-\mu_i^{(t)})} \text{ și } \pi^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}\mu_i^{(t)};$$
- returnează $\pi^{(t+1)}$, $p^{(t+1)}$ și $q^{(t+1)}$;

→ la AdaBoost, e suficient să punem un prag exterior pe una din axe (coincide sau după toate punctele);

→ distribuție binomială ⇔ $X \sim B(n,p)$, $E[X] = np$, $Var(X) = np(1-p)$, unde n este nr. de repetări și p este probabilitatea;

→ clusterizare ⇔ se agregă după distanța minimă;

→ metoda lui Ward ⟺ $\Delta(X,Y) = \sum_{x_i \in X \cup Y} \|x_i - \mu_{X \cup Y}\|^2 - \sum_{x_i \in X} \|x_i - \mu_X\|^2 - \sum_{x_i \in Y} \|x_i - \mu_Y\|^2$ ($\|x_i - \mu_{X \cup Y}\|^2$ este norma la pătrat ⟺ distanța dintre x_i și centroidul $X \cup Y$, adică media aritmetică a tuturor punctelor din $X \cup Y$ ⟺ $\|x_i - \mu_{X \cup Y}\|^2 = \left(\sqrt{(x_{ix} - \mu_{X \cup Y_x})^2 + (x_{iy} - \mu_{X \cup Y_y})^2}\right)^2$);

↳ pentru matricea inițială de clusterizare, luăm distanța dintre 2 puncte, o ridicăm la pătrat și împărțim la 2;

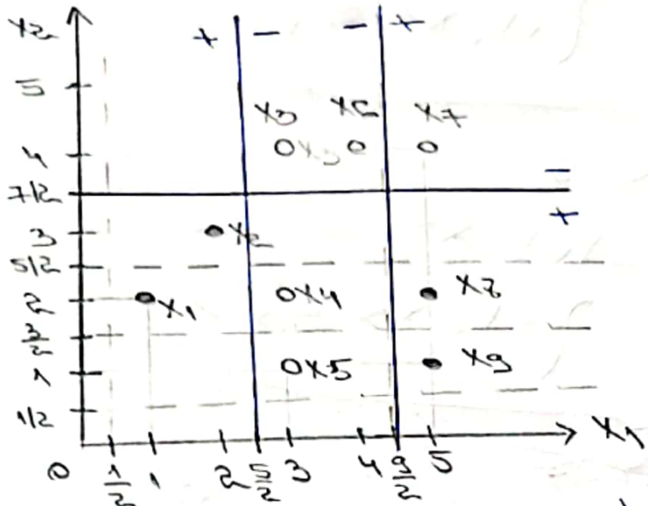
→ $\mu_X = \frac{1}{n_X} \sum_{x_i \in X} x_i$, unde n_X este nr. de elemente din clusterul X;

→ $\Delta(X,Y) = \frac{n_X n_Y}{n_X + n_Y} \|\mu_X - \mu_Y\|^2$;

EXERCIȚII TEST 2

Ex. 1

algoritmul AdaBoost; $T=3$;



- pentru $T=1$, vom avea separatorii $\frac{1}{2}, \frac{5}{2}, \frac{9}{2}$ pentru x_1 ($\frac{1}{2}$ prag exterior) și $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}$ pentru x_2 ($\frac{1}{2}$ prag exterior); inițial x_i au probabilitatea $\frac{1}{9}$ $\forall i=\overline{1,9}$;

-

s	$\frac{1}{2}$	$\frac{5}{2}$	$\frac{9}{2}$
$err_{D_1}(x_1<s)$	$\frac{4}{9}$	$\frac{2}{9}$	$\frac{6}{9}=\frac{2}{3}$
$err_{D_1}(x_1\geq s)$	$\frac{5}{9}$	$\frac{7}{9}$	$\frac{1}{3}$

-

s	$1/2$	$3/2$	$5/2$	$7/2$
$err_{D_1}(x_2<s)$	$\frac{4}{9}$	$\frac{4}{9}$	$\frac{3}{9}=\frac{1}{3}$	$\frac{2}{9}$
$err_{D_1}(x_2\geq s)$	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{2}{3}$	$\frac{7}{9}$

- $x_1<s \Rightarrow -1$, $x\geq s$ și $+1$, $x<s$ și pentru $x_1\geq s \Rightarrow +1$, $x\geq s$ și -1, $x<s$ (+, − pentru $x_1 < s$ și −, + pentru $x_1\geq s$);
- eroarea ponderată minimă la antrenare este $\frac{2}{9}=\varepsilon$

și se obține pentru $x_2 < \frac{7}{2}$ și $x_1 < \frac{5}{2}$;

↳ alegem $x_2 < \frac{7}{2} \Rightarrow h_1 = sign\left(\frac{7}{2} - x_2\right)$;

- $\underbrace{h = sign(s - x) = \begin{cases} -1, & x \geq s; \\ +1, & x < s; \end{cases}}_{x < s}$

- $\underbrace{h = sign(x - s) = \begin{cases} +1, & x \geq s; \\ -1, & x < s; \end{cases}}_{x \geq s}$

- x_4 și x_5 sunt greșit clasificați;
- $\gamma_1 = \frac{1}{2} - \varepsilon_1 = \frac{1}{2} - \frac{2}{9} = \frac{9-4}{18} = \frac{5}{18}$;
- $M = \{x_4; x_5\}$, $C = \{x_1; x_2; x_3; x_6; x_7; x_8; x_9\}$;
- $D_1(i) = \frac{1}{9}$, $\forall i = \overline{1,9}$;
- $D_2(i) = \begin{cases} \frac{D_1(i)}{2\varepsilon_1}, & i \in M; \\ \frac{D_1(i)}{2(1-\varepsilon_1)}, & i \in C; \end{cases}$

- $\frac{D_1(i)}{2\varepsilon_1} = \frac{\frac{1}{9}}{2 \cdot \frac{2}{9}} = \frac{1}{\cancel{9}} \cdot \frac{\cancel{9}}{4} = \frac{1}{4}$;
- $\frac{D_1(i)}{2(1-\varepsilon_1)} = \frac{\frac{1}{9}}{2\left(1 - \frac{2}{9}\right)} = \frac{\frac{1}{9}}{2 \cdot \frac{7}{9}} = \frac{1}{\cancel{9}} \cdot \frac{\cancel{9}}{14} = \frac{1}{14}$;

$D_2(i) = \begin{cases} \frac{1}{4}, & i \in M; \\ \frac{1}{14}, & i \in C; \end{cases}$

- pentru iterația $T=2$, vom avea separatorii $\frac{1}{2}$, $\frac{5}{2}$, $\frac{9}{2}$ pentru x_1 și $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{2}$, $\frac{7}{2}$;

s	$1/2$	$5/2$	$9/2$
$err_{D_2}(x_1 < s)$	$\frac{4}{14} = \frac{2}{7}$	$\frac{2}{14} = \frac{1}{7}$	$\frac{4}{14} + \frac{2}{4} = \frac{11}{14}$
$err_{D_2}(x_1 \geq s)$	$\frac{5}{7}$	$\frac{6}{7}$	$\frac{3}{14}$

s	$1/2$	$3/2$	$5/2$	$7/2$
$err_{D_2}(x_2 < s)$	$\frac{4}{14} = \frac{2}{7}$	$\frac{1}{4} + \frac{3}{14} = \frac{13}{28}$	$\frac{2}{4} + \frac{1}{14} = \frac{16}{28}$	$\frac{2}{4} = \frac{1}{2}$
$err_{D_2}(x_2 \geq s)$	$\frac{5}{7}$	$\frac{15}{28}$	$\frac{12}{28}$	$\frac{1}{2}$

- eroarea ponderată minimă la antrenare este $\varepsilon_2 = \frac{1}{7}$ și se obține pentru $x_1 < \frac{5}{2}$;

$\hookrightarrow h_2 = sign\left(\frac{5}{2} - x_1\right)$;

- $\gamma_2 = \frac{1}{2} - \varepsilon_2 = \frac{1}{2} - \frac{1}{7} = \frac{7-2}{14} = \frac{5}{14}$;

- ~~$D_3(i)$ =~~ instanțele greșit clasificate sunt x_7 și x_9;

- $D_3(i) = \begin{cases} \frac{D_2(i)}{2\varepsilon_2}, & i \in \{8; 9\}; \\ \frac{D_2(i)}{2(1-\varepsilon_2)}, & i \notin \{8; 9\}; \end{cases}$

- $\frac{D_2(i)}{2\varepsilon_2} = \frac{\frac{1}{14}}{2 \cdot \frac{1}{7}} = \frac{1}{14} \cdot \frac{7}{2} = \frac{1}{4}$;

- $\frac{D_2(i)}{2(1-\varepsilon_2)} = \frac{D_2(i)}{2\left(1-\frac{1}{7}\right)} = \frac{D_2(i)}{2 \cdot \frac{6}{7}} =$

$$= \begin{cases} \dfrac{\frac{1}{14}}{2\cdot\frac{6}{7}} = \dfrac{1}{14}\cdot\dfrac{7}{12} = \dfrac{1}{24}; \\ \dfrac{\frac{1}{4}}{2\cdot\frac{6}{7}} = \dfrac{1}{4}\cdot\dfrac{7}{12} = \dfrac{7}{48}; \end{cases}$$

$$\hookrightarrow D_3(i) = \begin{cases} \frac{1}{4}, & i \in \{8; 9\}; \\ \frac{7}{48}, & i \in \{4; 5\}; \\ \frac{1}{24}, & i \notin \{4; 5; 8; 9\}; \end{cases}$$

- pentru $T=3$, calculăm mediile ponderate la antrenare;

θ	$1/2$	$5/2$	$9/2$
$err_{D_3}(x_1 < \theta)$	$\frac{7}{12}$	$\frac{2}{4} = \frac{1}{2}$	$\frac{21}{24}$
$err_{D_3}(x_1 \geq \theta)$	$\frac{5}{12}$	$\frac{1}{2}$	$\frac{3}{24} = \frac{1}{8}$

- $\frac{2}{24} + \frac{2}{4} = \frac{1}{12} + \frac{1}{2} = \frac{1+6}{12} = \frac{7}{12};$
- $\frac{2}{24} + \frac{2\cdot 7}{48} + \frac{2}{4} = \cancel{\frac{1}{12} + \frac{1}{2} + \frac{1}{2} = \frac{1}{12} + \frac{1}{2} = \frac{1}{12}}$

$= \frac{1}{12} + \frac{7}{24} + \frac{1}{2} = \frac{2+7+12}{24} = \frac{21}{24};$

θ	$1/2$	$3/2$	$5/2$	$7/2$
$err_{D_3}(x_2 < \theta)$	$\frac{7}{12}$	$\frac{23}{48}$	$\frac{1}{3}$	$\frac{7}{24}$
$err_{D_3}(x_2 \geq \theta)$	$\frac{5}{12}$	$\frac{25}{48}$	$\frac{2}{3}$	$\frac{17}{24}$

- $\frac{2}{24} + \frac{2}{4} = \frac{7}{12}$;
- $\frac{7}{48} + \frac{2}{24} + \frac{1}{4} = \frac{7}{48} + \frac{1}{12} + \frac{1}{4} = \frac{7+4+12}{48} = \frac{23}{48}$;
- $\frac{2\cdot 7}{48} + \frac{1}{24} = \frac{7}{24} + \frac{1}{24} = \frac{8}{24} = \frac{1}{3}$;
- $2\cdot\frac{7}{48} = \frac{7}{24}$;
- eroarea ponderată minimă la antrenare este $\varepsilon_3 = \frac{1}{8}$ și se obține pentru $x_1 \geq \frac{9}{2}$;

  $\hookrightarrow \underline{h_3 = \text{sign}\left(x_1 - \frac{9}{2}\right)}$;
- $\gamma_3 = \frac{1}{2} - \varepsilon_3 = \frac{1}{2} - \frac{1}{8} = \frac{4-1}{8} = \frac{3}{8}$;
- instanțele greșit clasificate sunt x_1, x_2;
- $D_4(i) = \begin{cases} \frac{D_3(i)}{2\varepsilon_3}, & i \in \{1;2\}; \\ \frac{D_3(i)}{2(1-\varepsilon_3)}, & i \notin \{1;2\}; \end{cases}$
- $\frac{D_3(i)}{2\varepsilon_3} = \frac{\frac{1}{24}}{2\cdot\frac{1}{8}} = \frac{1}{24}\cdot\frac{4}{1} = \frac{1}{6}$; $\frac{\frac{1}{24}}{2\cdot\frac{3}{8}} = \frac{1}{24}\cdot\frac{4}{3} = \frac{1}{18}$;
- $\frac{D_3(i)}{2(1-\varepsilon_3)} = \frac{D_3(i)}{2\gamma_3} = \begin{cases} \frac{\frac{7}{48}}{2\cdot\frac{3}{8}} = \frac{7}{48}\cdot\frac{4}{3} = \frac{7}{36}; \\ \frac{\frac{1}{4}}{2\cdot\frac{3}{8}} = \frac{1}{\cancel{4}}\cdot\frac{\cancel{4}}{3} = \frac{1}{3}; \end{cases}$

$\hookrightarrow D_4(i) = \begin{cases} \frac{1}{6}, & i \in \{1;2\}; \\ \frac{1}{18}, & i \in \{3;6;7\}; \\ 7/36, & i \in \{4;5\}; \\ 1/3, & i \in \{8;9\}; \end{cases}$

- $\alpha_1 = \frac{1}{2} \ln \frac{1-\varepsilon_1}{\varepsilon_1} = \frac{1}{2} \ln \frac{1-\frac{2}{9}}{\frac{2}{9}} = \frac{1}{2} \ln \frac{\frac{7}{9}}{\frac{2}{9}} =$

$= \ln \left(\frac{7}{\cancel{9}} \cdot \frac{\cancel{9}}{2} \right)^{1/2} = \ln \sqrt{\frac{7}{2}}$;

- $\alpha_2 = \frac{1}{2} \ln \frac{1-\varepsilon_2}{\varepsilon_2} = \frac{1}{2} \ln \frac{1-\frac{1}{7}}{\frac{1}{7}} = \frac{1}{2} \ln \frac{\frac{6}{7}}{\frac{1}{7}} =$

$= \ln \left(\frac{6}{\cancel{7}} \cdot \frac{\cancel{7}}{1} \right)^{1/2} = \ln \sqrt{6}$;

- $\alpha_3 = \frac{1}{2} \ln \frac{1-\varepsilon_3}{\varepsilon_3} = \frac{1}{2} \ln \frac{1-\frac{1}{8}}{\frac{1}{8}} = \frac{1}{2} \ln \frac{\frac{7}{8}}{\frac{1}{8}} =$

$= \ln \left(\frac{7}{\cancel{8}} \cdot \frac{\cancel{8}}{1} \right)^{1/2} = \ln \sqrt{7}$;

- $H_3 = \operatorname{sign}\left(\sum_{i=1}^{3} \alpha_i h_i \right) = \operatorname{sign}\Big(\ln \sqrt{\frac{7}{2}} \operatorname{sign}\left(\frac{7}{2} - x_2\right) + \ln\sqrt{7} \underbrace{\operatorname{sign}\left(x_1 - \frac{9}{2}\right)}_{-1} +$

$+ \ln\sqrt{6} \underbrace{\operatorname{sign}\left(\frac{3}{2} - x_1\right)}_{-1} \Big) = \operatorname{sign}\left(\ln\sqrt{\frac{7}{2}} - \ln\sqrt{6} - \ln\sqrt{7} \right) =$

$= \operatorname{sign}\left(\ln \left(\sqrt{\frac{7}{2}} \cdot \frac{1}{\sqrt{6}} \cdot \frac{1}{\sqrt{7}} \right) \right) =$

$= \operatorname{sign}\left(\ln \left(\frac{\cancel{\sqrt{7}}}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}\cdot\sqrt{3}} \cdot \frac{1}{\cancel{\sqrt{7}}} \right) \right) =$

$= \operatorname{sign}\left(\ln \frac{1}{2\sqrt{3}} \right) = \operatorname{sign}\left(\underbrace{\ln 1}_{0} - \ln 2\sqrt{3} \right) =$

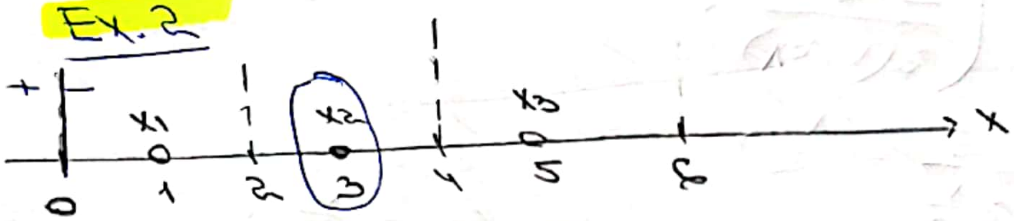
$= \operatorname{sign}\Big(\underbrace{- \underbrace{\ln 2\sqrt{3}}_{>0}}_{<0} \Big) = -1$;

- $err_s(H_T) = \begin{cases} \frac{2}{9}, & T=1 \ (x_4, x_5); \\ \frac{2}{9}, & T=2 \ (x_8, x_9); \\ 0, & T=3; \end{cases}$

T	α_T	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	$\ln\sqrt{7/2}$	+1	+1	−1	+1	+1	−1	−1	+1	+1
2	$\ln\sqrt{6}$	+1	+1	−1	−1	−1	−1	−1	−1	−1
3	$\ln\sqrt{7}$	−1	−1	−1	−1	−1	−1	+1	+1	+1
	$H_T(x_i)$	+1	+1	−1	−1	−1	−1	−1	+1	+1

- $H_3(x_i) = \text{sign}\left(\ln\sqrt{\frac{7}{2}}\ \text{sign}\left(\frac{7}{2} - x_{i,2}\right) + \ln\sqrt{6}\ \text{sign}\left(\frac{5}{2} - x_{i,1}\right) + \ln\sqrt{7}\ \text{sign}\left(x_{i,1} - \frac{9}{2}\right)\right);$

Ex. 2



$x_1 = 1,\ x_2 = 3,\ x_3 = 5;$

a) valoarea ponderilor inițial;
primul compas decizional;
probabilitățile instanțelor după $T = 1$;

- avem 3 instanțe de antrenament $\Rightarrow$
$D_1(i) = \frac{1}{3},\ \forall i = \overline{1,3};$
- separatorii sunt 0, 2, 4 (0 sunt separatori exteriori);
-

s	0	2	4
$err_{D_1}(x < s)$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$
$err_{D_1}(x \geq s)$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$

• eroarea ponderată minimă la antrenare este $\varepsilon_1 = \frac{1}{3}$ și se obține pentru $x<0$, $x \ge 2$ și $x<4$;

↳ alegem $x<0 \Rightarrow h_1 = sign(0-x) = sign(-x)$;

• $\alpha_1 = \frac{1}{2}\ln\frac{1-\varepsilon_1}{\varepsilon_1} = \frac{1}{2}\ln\frac{1-\frac{1}{3}}{\frac{1}{3}} = \frac{1}{2}\ln\frac{\frac{2}{3}}{\frac{1}{3}} = \ln\left(\frac{2}{\cancel{3}}\cdot\frac{\cancel{3}}{1}\right)^{1/2} = \ln\sqrt{2}$;

• $err_S(H_1) = \varepsilon_1 = \frac{1}{3}$ (instanța greșită este x_2);

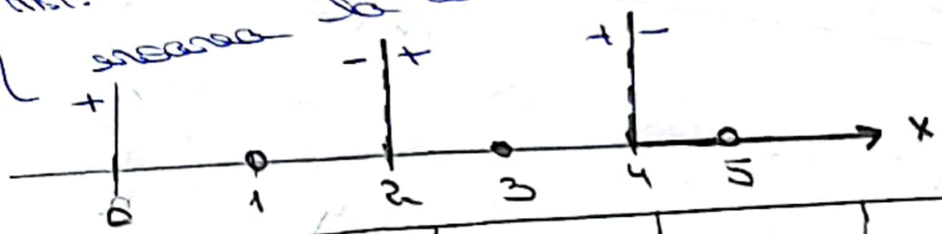
• $D_2(i) = \begin{cases} \frac{D_1(i)}{2\varepsilon_1}, & i \in \{2\}; \\ \frac{D_1(i)}{2(1-\varepsilon_1)}, & i \in \{1;3\}; \end{cases}$

• $\frac{D_1(i)}{2\varepsilon_1} = \frac{\cancel{1/3}}{2\cancel{1/3}} = \frac{1}{2}$;

• $\frac{D_1(i)}{2(1-\varepsilon_1)} = \frac{\frac{1}{3}}{2(1-\frac{1}{3})} = \frac{1}{\cancel{3}}\cdot\frac{1}{2}\cdot\frac{\cancel{3}}{2} = \frac{1}{4}$;

↳ $D_2(i) = \begin{cases} \frac{1}{2}, & i \in \{2\}; \\ \frac{1}{4}, & i \in \{1;3\}; \end{cases}$

b) { nr. minim de iterații în care se atinge eroarea la antrenare 0;



s	0	2	4
$err_{D_2}(x<s)$	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{4}$
$err_{D_2}(x \ge s)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{4}$

• eroarea minimă ponderată la antrenare este $\varepsilon_2 = \frac{1}{4}$ și se obține pentru $x \geq 2$ și $x < 4$;

↳ alegem $x \geq 2 \Rightarrow h_2 = \text{sign}(x-2)$;

• $\alpha_2 = \frac{1}{2}\ln\frac{1-\varepsilon_2}{\varepsilon_2} = \frac{1}{2}\ln\frac{1-\frac{1}{4}}{\frac{1}{4}} = \frac{1}{2}\ln\frac{\frac{3}{4}}{\frac{1}{4}} =$

$= \ln\left(\frac{3}{\cancel{4}}\cdot\frac{\cancel{4}}{1}\right)^{\frac{1}{2}} = \ln\sqrt{3}$;

• $D_3(i) = \begin{cases} \frac{D_2(i)}{2\varepsilon_2}, & i\in\{3\}; \\ \frac{D_2(i)}{2(1-\varepsilon_2)}, & i\in\{1;2\}; \end{cases}$

↳ instanța clasificată greșit este x_3;

• $\frac{D_2(i)}{2\varepsilon_2} = \frac{\frac{1}{4}}{2\cdot\frac{1}{4}} = \frac{1}{2}\cdot\frac{2}{1} = \frac{1}{2}$;

• $\frac{D_2(1)}{2(1-\varepsilon_2)} = \frac{\frac{1}{4}}{2(1-\frac{1}{4})} = \frac{\frac{1}{4}}{2\cdot\frac{3}{4}} = \frac{1}{\cancel{4}}\cdot\frac{\cancel{4}}{6} = \frac{1}{6}$;

• $\frac{D_2(2)}{2(1-\varepsilon_2)} = \frac{\frac{1}{2}}{2\cdot\frac{3}{4}} = \frac{1}{\cancel{4}}\cdot\frac{\cancel{4}}{3} = \frac{1}{3}$;

↳ $D_3(i) = \begin{cases} \frac{1}{2}, & i=3; \\ \frac{1}{3}, & i=2; \\ \frac{1}{6}, & i=1; \end{cases}$

•

α_T	x_1	x_2	x_3
$\ln\sqrt{2}$	-1	-1	-1
$\ln\sqrt{3}$	-1	$+1$	$+1$
$H_2(x_i)$	-1	$+1$	$+1$

• $H_2(x_2) = sign(\ln\sqrt{2}\cdot(-1) + \ln\sqrt{3}\cdot(+1)) =$

$= sign(-\ln\sqrt{2} + \ln\sqrt{3}) =$

$= sign(\underbrace{\ln \underbrace{\frac{\sqrt{3}}{\sqrt{2}}}_{>1}}_{>0}) = +1;$

↳ $err_S(H_2) = \frac{1}{3};$

a	0	2	4
$err_{D_3}(x<a)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$
$err_{D_3}(x\geq a)$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{5}{6}$

• eroarea ponderată minimă la antrenare este $\varepsilon_3 = \frac{1}{6}$ și se obține pentru $x<4$;

↳ $h_3 = sign(4-x)$;

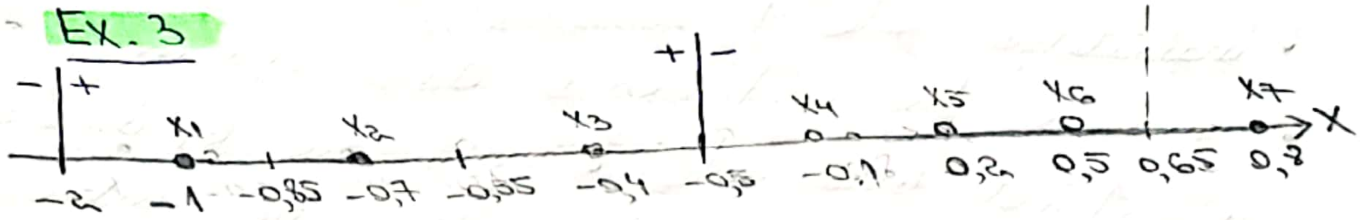
• $\alpha_3 = \frac{1}{2}\ln\frac{1-\varepsilon_3}{\varepsilon_3} = \frac{1}{2}\ln\frac{1-\frac{1}{6}}{\frac{1}{6}} = \frac{1}{2}\ln\frac{\frac{5}{6}}{\frac{1}{6}} =$

$= \ln\left(\frac{5}{\cancel{6}}\cdot\frac{\cancel{6}}{1}\right)^{\frac{1}{2}} = \ln\sqrt{5};$

α_t	x_1	x_2	x_3
$\ln\sqrt{2}$	-1	-1	-1
$\ln\sqrt{3}$	-1	$+1$	$+1$
$\ln\sqrt{5}$	$+1$	$+1$	-1
$H_3(x_i)$	-1	$+1$	-1

↳ $err_S(H_3) = 0 \Rightarrow$ pentru $T=3$, eroarea la antrenare este 0;

Ex. 3



a) separatorul pentru h_1;

- separatorii decizionali pentru setul de date de antrenament sunt ~~x~~ -2 (separator exterior), $-0,5$ și $0,65$;

-

s	-2	$-0,5$	$0,65$
$err_{D_1}(x < s)$	$\frac{4}{7}$	$\frac{1}{7}$	$\frac{4}{7}$
$err_{D_1}(x \geq s)$	$\frac{3}{7}$	$\frac{6}{7}$	$\frac{3}{7}$

- $D_1(i) = \frac{1}{7}, \forall i = \overline{1,7}$;
- eroarea ponderată minimă la antrenare este $\frac{1}{7} = \varepsilon_1$ și se obține pentru $x < -0,5$;

  ↳ $\underline{h_1 = sign(-0,5 - x)}$;

b) $\left\{ \begin{array}{l} \varepsilon_1;\ \alpha_1;\ \text{acuratețea la antrenare dacă oprim algoritmul acum;} \end{array} \right.$

- $\varepsilon_1 = err_{D_1}(x_1 < -0,5)$;

  ↳ $\underline{\varepsilon_1 = \frac{1}{7}}$;

- $\alpha_1 = \frac{1}{2} \ln \frac{1-\varepsilon_1}{\varepsilon_1} = \frac{1}{2} \ln \frac{1-\frac{1}{7}}{\frac{1}{7}} = \frac{1}{2} \ln \frac{\frac{6}{7}}{\frac{1}{7}} =$

  $= \ln\left(\frac{6}{\cancel{7}} \cdot \frac{\cancel{7}}{1}\right)^{\frac{1}{2}}$;

  ↳ $\underline{\alpha_1 = \ln \sqrt{6}}$;

- acuratețea este 1 - eroarea ;
- eroarea la antrenare dacă oprim algoritmul aici este egală cu ε_1 deoarece $H_1(x_i)$ este compus doar din h_1, adică $H_1(x_i) = \cancel{\frac{1}{7}} \, sign(d_1 h_1) \Rightarrow err_S(H_1) = \varepsilon_1 = \frac{1}{7}$;

↳ acuratețea la antrenare este $1 - err_S(H_1) = 1 - \frac{1}{7} = \frac{6}{7}$;

c) valorile noilor probabilități $D_2(i)$;
- instanța clasificată greșit este x_7;
- $D_2(i) = \begin{cases} \frac{D_1(i)}{2\varepsilon_1}, & i = 7; \\ \frac{D_1(i)}{2(1-\varepsilon_1)}, & i \neq 7; \end{cases}$

- $\frac{D_1(7)}{2\varepsilon_1} = \frac{\cancel{\frac{1}{7}}}{2 \cdot \cancel{\frac{1}{7}}} = \frac{1}{2}$;
- $\frac{D_1(i)}{2(1-\varepsilon_1)} = \frac{\frac{1}{7}}{2(1-\frac{1}{7})} = \frac{\frac{1}{7}}{2 \cdot \frac{6}{7}} = \frac{1}{\cancel{7}} \cdot \frac{\cancel{7}}{12} = \frac{1}{12}$;

↳ $D_2(i) = \begin{cases} \frac{1}{2}, & i = 7; \\ \frac{1}{12}, & i \neq 7; \end{cases}$

d) •

s	-2	$-0,5$	$0,65$
$err_{D_2}(X < s)$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
$err_{D_2}(X \geq s)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

- $\frac{3}{12} + \frac{1}{2} = \frac{1}{4} + \frac{1}{2} = \frac{1+2}{4} = \frac{3}{4}$;
- $\frac{3}{12} + \frac{1}{2} = \frac{3}{4}$;
- eroarea ponderată minimă la antrenare este $\varepsilon_2 = \frac{1}{4}$ și se obține la $x \geq -2$, $x \geq 0{,}65$;

  ↳ alegem $x \geq -2 \Rightarrow \underline{h_2 = \text{sign}(x+2)}$;

d) { exemplele de antrenament cu ponderile cele mai mici după a doua iterație;
- instanțele de antrenament greșit clasificate sunt x_4, x_5, x_6;
- exemplele de antrenament cu ponderile cele mai mici sunt cele corect clasificate;

  ↳ $\underline{x_1, x_2, x_3}$ (deoarece la iterația anterioară probabilitățile cele mai mici erau pentru $i \neq 7$);

e) acuratețea la antrenare pentru iterația 2;
- $\alpha_2 = \frac{1}{2}\ln\frac{1-\varepsilon_2}{\varepsilon_2} = \frac{1}{2}\ln\frac{1-\frac{1}{4}}{\frac{1}{4}} = \frac{1}{2}\ln\frac{\frac{3}{4}}{\frac{1}{4}} = \ln\left(\frac{3}{\cancel{4}}\cdot\frac{\cancel{4}}{1}\right)^{\frac{1}{2}} = \ln\sqrt{3}$;
-

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
α_t							
$\ln\sqrt{6}$	+1	+1	+1	−1	−1	−1	−1
$\ln\sqrt{3}$	+1	+1	+1	+1	+1	+1	+1
$H_2(x_i)$	+1	+1	+1	−1	−1	−1	−1

• $H_2(x_4) = \text{sign}\left(\ln\sqrt{6}\cdot(-1) + \ln\sqrt{3}\cdot(+1)\right) =$

$= \text{sign}\left(-\ln\sqrt{6} + \ln\sqrt{3}\right) = \text{sign}\left(\ln\frac{1}{\sqrt{2}}\right) = -1$;

$= \text{sign}\left(\ln\frac{\sqrt{3}}{\sqrt{2}\cdot\sqrt{3}}\right) = \text{sign}\left(\underbrace{\ln\underbrace{\frac{1}{\sqrt{2}}}_{<1}}_{<0}\right)$

• $H_2(x_7) = \cancel{\text{sign}(}\ H_2(x_4)$;

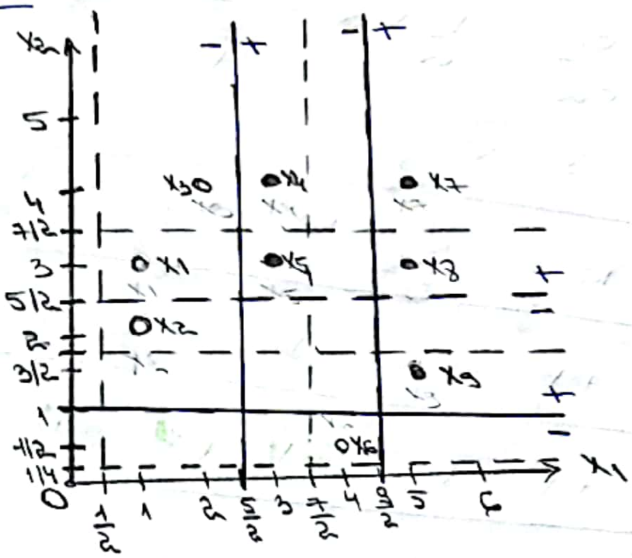
• acuratețea la antrenare $\Leftrightarrow$ nr. instanțelor corect clasificate din total $\Leftrightarrow 1 - \text{err}_S(H_2)$;

• $\text{err}_S(H_2) = \frac{1}{7}$;

↳ acuratețea la antrenare este $1 - \frac{1}{7} = \frac{6}{7}$

$\Rightarrow$ nu se îmbunătățește față de iterația 1;

Ex. 4



a) primul compas de decizie h_1;

- ~~xxxx~~ pragurile pentru $x_1 \Leftrightarrow \frac{1}{2}$ (exterior), $\frac{5}{2}$, $\frac{7}{2}$, $\frac{9}{2}$;
- pragurile pentru $x_2 \Leftrightarrow \frac{1}{4}$ (exterior), 1, $\frac{7}{4}$, $\frac{5}{2}$, $\frac{7}{2}$;
- $D_1(i) = \frac{1}{9}, \forall i = \overline{1,9}$;
-

s	$1/2$	$5/2$	$7/2$	$9/2$
$err_{D_1}(x_1 < s)$	$\frac{5}{9}$	$\frac{8}{9}$	$\frac{6}{9} = \frac{2}{3}$	$\frac{7}{9}$
$err_{D_1}(x_1 \geq s)$	$\frac{4}{9}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{2}{9}$

-

s	$1/4$	1	$7/4$	$5/2$	$7/2$
$err_{D_1}(x_2 < s)$	$\frac{5}{9}$	$\frac{6}{9} = \frac{2}{3}$	$\frac{5}{9}$	$\frac{6}{9} = \frac{2}{3}$	$\frac{5}{9}$
$err_{D_1}(x_2 \geq s)$	$\frac{4}{9}$	$\frac{1}{3}$	$\frac{4}{9}$	$\frac{1}{3}$	$\frac{4}{9}$

- eroarea ponderată minimă este $\varepsilon_1 = \frac{1}{9} \neq \frac{1}{2}$; se obține pentru $x_1 \geq \frac{5}{2}$;

  ↳ $h_1 = sign\left(x_1 - \frac{5}{2}\right)$;

b) probabilitățile instanțelor după prima iterație;
instanțele cu probabilitatea cea mai mare;

- instanța clasificată greșit de h_1 este x_6;
- $D_2(i) = \begin{cases} \frac{D_1(i)}{2\varepsilon_1}, & i = 6; \\ \frac{D_1(i)}{2(1-\varepsilon_1)}, & i \neq 6; \end{cases}$
- $\frac{D_1(6)}{2\varepsilon_1} = \frac{\cancel{\frac{1}{9}}}{2 \cdot \cancel{\frac{1}{9}}} = \frac{1}{2}$;
- $\frac{D_1(i)}{2(1-\varepsilon_1)} = \frac{\frac{1}{9}}{2\left(1-\frac{1}{9}\right)} = \frac{\frac{1}{9}}{2 \cdot \frac{8}{9}} = \frac{1}{\cancel{9}} \cdot \frac{\cancel{9}}{16} = \frac{1}{16}$;

↳ $D_2(i) = \begin{cases} \frac{1}{2}, & i = 6; \\ \frac{1}{16}, & i \neq 6; \end{cases}$

↳ probabilitatea cea mai mare o are x_6 (instanța incorect clasificată);

c) • $err_{D_2}(x_1 \leq \frac{5}{2}) = D_2(6) = \frac{1}{2}$;

↳ eroarea ponderată la antrenarea predicției de h_1 după prima iterație este $\frac{1}{2}$ (proprietatea $err_{D_{t+1}}(h_t) = \frac{1}{2}$);

d) • $\alpha_1 = \frac{1}{2} \ln \frac{1-\varepsilon_1}{\varepsilon_1} = \frac{1}{2} \ln \frac{1-\frac{1}{9}}{\frac{1}{9}} = \frac{1}{2} \ln \frac{\frac{8}{9}}{\frac{1}{9}} = \ln\left(\frac{8}{9} \cdot \frac{9}{1}\right)^{\frac{1}{2}} = \ln \sqrt{8} = \ln 2\sqrt{2}$;

• $err_S(h_1) = \varepsilon_1 = \frac{1}{9}$ ($H_1 = sign\, h_1 \alpha_1$);

•

s	$1/2$	$5/2$	$7/2$	$9/2$
$err_{D_2}(x_1 < s)$	$\frac{5}{16}$	$\frac{1}{2}$	$\frac{6}{16} = \frac{3}{8}$	$\frac{7}{8}$
$err_{D_2}(x_1 \geq s)$	$\frac{11}{16}$	$\frac{1}{2}$	$\frac{5}{8}$	$\frac{1}{8}$

• $\frac{3}{16} + \frac{1}{2} + \frac{3}{16} = \frac{6}{16} + \frac{8}{16} = \frac{14}{16} = \frac{7}{8}$;

•

s	$1/4$	1	$7/4$	$5/2$	$7/2$
$err_{D_2}(x_2 < s)$	$\frac{5}{16}$	$\frac{13}{16}$	$\frac{3}{4}$	$\frac{13}{16}$	$\frac{3}{4}$
$err_{D_2}(x_2 \geq s)$	$\frac{11}{16}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{4}$

• eroarea ponderată minimă la antrenare este $\varepsilon_2 = \frac{1}{8}$ şi se obţine pentru $x_1 \geq \frac{3}{2}$;

↳ $h_2 = sign\left(x_1 - \frac{3}{2}\right)$;

2) instanţe clasificate greşit de H_2;

• $\alpha_2 = \frac{1}{2} \ln \frac{1-\varepsilon_2}{\varepsilon_2} = \frac{1}{2} \ln \frac{1-\frac{1}{8}}{\frac{1}{8}} = \frac{1}{2} \ln \frac{\frac{7}{8}}{\frac{1}{8}} =$
$= \ln\left(\frac{7}{8} \cdot \frac{8}{1}\right)^{\frac{1}{2}} = \ln \sqrt{7}$;

•

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
α_1	−1	−1	−1	+1	+1	+1	+1	+1	+1
α_2	−1	−1	−1	−1	−1	−1	+1	+1	+1
$H_2(x_i)$	−1	−1	−1	+1	+1	+1	+1	+1	+1

• $H_2(x_4) = sign(\alpha_1 \cdot (+1) + \alpha_2 \cdot (-1)) = sign\left(\ln \frac{\sqrt{8}}{\sqrt{7}}\right) =$
$= sign(\ln \sqrt{8} - \ln \sqrt{7}) =$
$= sign\left(\underbrace{\ln \underbrace{\sqrt{\frac{8}{7}}}_{>1}}_{>0}\right) = +1$;

↳ x_6 este clasificat greşit $\left(eroare(H_2) = \frac{1}{9}\right)$;

3) a treia iteraţie;

• h_2 clasifică greşit instanţele x_4 şi x_5;

• $D_3(i) = \begin{cases} \frac{D_2(i)}{2\varepsilon_2}, & i \in \{4;5\}; \\ \frac{D_2(i)}{2(1-\varepsilon_2)}, & i \notin \{4;5\}; \end{cases}$

• $\frac{D_2(i)}{2\varepsilon_2} = \frac{\frac{1}{16}}{2 \cdot \frac{1}{8}} = \frac{1}{16} \cdot \frac{4}{1} = \frac{1}{4}$;

- $\dfrac{D_2(i)}{2(1-\varepsilon_2)} = \begin{cases} \dfrac{\frac{1}{2}}{2(1-\frac{1}{8})} = \dfrac{\frac{1}{2}}{2\cdot\frac{7}{8}} = \dfrac{1}{\cancel{2}}\cdot\dfrac{\cancel{8}^{2}}{7} = \dfrac{2}{7}; \\[2ex] \dfrac{\frac{1}{16}}{2\cdot\frac{7}{8}} = \dfrac{1}{\cancel{16}_{4}}\cdot\dfrac{\cancel{8}}{7} = \dfrac{1}{28}; \end{cases}$

$\hookrightarrow D_3(i) = \begin{cases} \frac{1}{4}, & i \in \{4;5\}; \\ \frac{2}{7}, & i = 6; \\ \frac{1}{28}, & i \notin \{4;5;6\}; \end{cases}$

-

s	$1/2$	$5/2$	$7/2$	$9/2$
$\Pr_{D_3}(X_1 < s)$	$\frac{17}{28}$	$\frac{5}{7}$	$\frac{3}{14}$	$\frac{1}{2}$
$\Pr_{D_3}(X_1 \geq s)$	$\frac{11}{28}$	$\frac{2}{7}$	$\frac{9}{14}$	$\frac{1}{2}$

- $\frac{2}{4} + \frac{3}{28} = \frac{1}{2} + \frac{3}{28} = \frac{14+3}{28} = \frac{17}{28};$
- $\frac{3}{28} + \frac{3}{28} = \frac{6}{28} = \frac{3}{14}$
- $\frac{2}{4} = \frac{1}{2};$
-

s	$1/4$	1	$7/4$	$5/2$	$7/2$
$\Pr_{D_3}(X_2 < s)$	$\frac{17}{28}$	$\frac{25}{28}$	$\frac{6}{7}$	$\frac{25}{28}$	$\frac{9}{14}$
$\Pr_{D_3}(X_2 \geq s)$	$\frac{11}{28}$	$\frac{3}{28}$	$\frac{1}{7}$	$\frac{3}{28}$	$\frac{5}{14}$

- $\frac{2}{7} + \frac{2}{4} + \frac{3}{28} = \frac{2}{7} + \frac{17}{28} = \frac{25}{28}$
- $\frac{2}{7} + \frac{2}{28} + \frac{2}{4} = \frac{2}{7} + \frac{1}{14} + \frac{1}{2} = \frac{4+1+7}{14} = \frac{12}{14} = \frac{6}{7};$
- $\frac{1}{28} + \frac{2}{7} + \frac{1}{4} + \frac{2}{28} = \frac{1+8+7+2}{28} = \frac{18}{28} = \frac{9}{14};$

• eroarea ponderată minimă la antrenare este $\varepsilon_3 = \frac{3}{28}$ și se obține pentru $x_2 \geq 1$ și $x_2 \geq \frac{5}{2}$;

↳ $h_3 = \text{sign}(x_2 - 1)$ și $h_3' = \text{sign}(x_2 - \frac{5}{2})$;

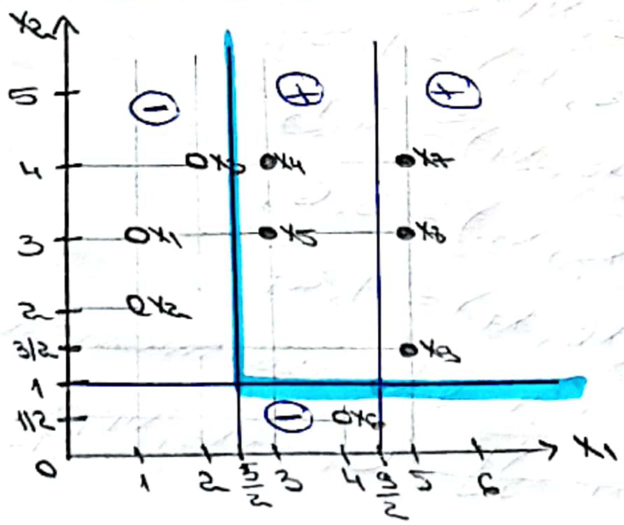
• $\alpha_3 = \frac{1}{2} \ln \frac{1-\varepsilon_3}{\varepsilon_3} = \frac{1}{2} \ln \frac{1-\frac{3}{28}}{\frac{3}{28}} = \frac{1}{2} \ln \frac{\frac{25}{28}}{\frac{3}{28}} =$

$= \ln\left(\frac{25}{28} \cdot \frac{28}{3}\right)^{\frac{1}{2}} = \ln \frac{5}{\sqrt{3}}$; $\alpha_3' = \alpha_3$;

•

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
α_1	-1	-1	-1	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$
α_2	-1	-1	-1	-1	-1	-1	$+1$	$+1$	$+1$
α_3	$+1$	$+1$	$+1$	$+1$	$+1$	-1	$+1$	$+1$	$+1$
α_3'	$+1$	-1	$+1$	$+1$	$+1$	-1	$+1$	$+1$	-1
$H_3(x_i)$	-1	-1	-1	$+1$	$+1$	-1	$+1$	$+1$	$+1$
$H_3'(x_i)$	-1	-1	-1	$+1$	$+1$	-1	$+1$	$+1$	$+1$

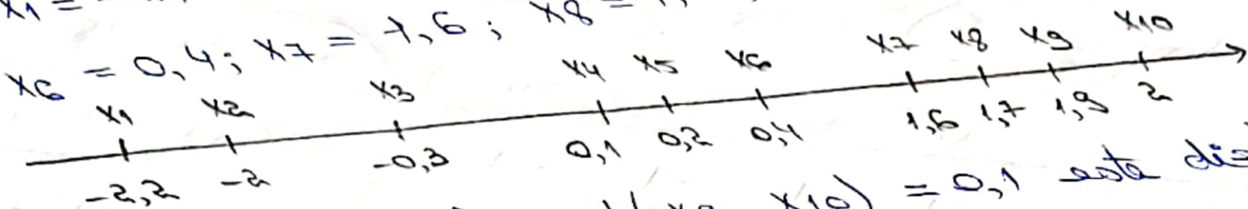
↳ $\text{eroare}(H_3) = \text{eroare}(H_3') = 0$ (h_3 și h_3' obțin același rezultat α_3 și ε_3 și $\text{eroare}(H_3)$, deși clasifică diferit instanțele);



Ex. 5

$S = \{-2,2;\ -2,0;\ -0,3;\ 0,1;\ 0,2;\ 0,4;\ 1,6;\ 1,7;\ 1,9;\ 2,0\}$;
dendrograma corespunzătoare; „single-linkage";

- $x_1 = -2,2$; $x_2 = -2$; $x_3 = -0,3$; $x_4 = 0,1$; $x_5 = 0,2$; $x_6 = 0,4$; $x_7 = 1,6$; $x_8 = 1,7$; $x_9 = 1,9$; $x_{10} = 2$;



- $d(x_4, x_5) = d(x_7, x_8) = d(x_9, x_{10}) = 0,1$ este distanța minimă $\Rightarrow C_1 = \{x_4; x_5\}$, $C_2 = \{x_7; x_8\}$, $C_3 = \{x_9; x_{10}\}$;
- $d(x_1, x_2) = 0,2$;
- $d(C_1, x_6) \stackrel{SL}{=} d(x_5, x_6) = 0,2$;
- $d(C_2, C_3) \stackrel{SL}{=} d(x_8, x_9) = 0,2$;
- $d(x_2, x_3) = 1,7$; $d(x_3, C_1) \stackrel{SL}{=} d(x_3, x_4) = 0,4$;

↳ $d(x_1, x_2) = d(C_1, x_6) = d(C_2, C_3) = 0,2$ este cea minimă $\Rightarrow C_4 = \{x_1; x_2\}$, $C_5 = \{x_4; x_5; x_6\}$ și $C_6 = \{x_7; x_8; x_9; x_{10}\}$;

- $d(C_4, x_3) \stackrel{SL}{=} d(x_2, x_3) = 1,7$;
- $d(x_3, C_5) \stackrel{SL}{=} d(x_3, x_4) = 0,4$;
- $d(x_3, C_6) \stackrel{SL}{=} d(x_3, x_7) = 1,9$;
- $d(C_4, C_5) \stackrel{SL}{=} d(x_2, x_4) = 2,1$;
- $d(C_4, C_6) \stackrel{SL}{=} d(x_2, x_7) = 3,6$;
- $d(C_5, C_6) \stackrel{SL}{=} d(x_6, x_7) = 1,2$;

↳ $d(x_3, C_5) = 0,4$ este cea minimă $\Rightarrow$ ~~$C_7 = \{x_3, x_1, x_2\}$~~, $C_7 = \{x_3; x_4; x_5; x_6\}$;

• $d(C_4, C_7) \overset{SL}{=} d(x_2, x_3) = 1,7$;

• ~~$d(C_5, C_7) \overset{SL}{=} d(x_3, x_4) = 0,4$;~~

• $d(C_6, C_7) \overset{SL}{=} d(x_6, x_7) = 1,2$;

↳ $d(C_6, C_7) = 1,2$ este cea minimă $\Rightarrow C_8 = C_7 \cup C_6$

$\Rightarrow C_8 = \{x_3; x_4; x_5; x_6; x_7; x_8; x_9\}$;

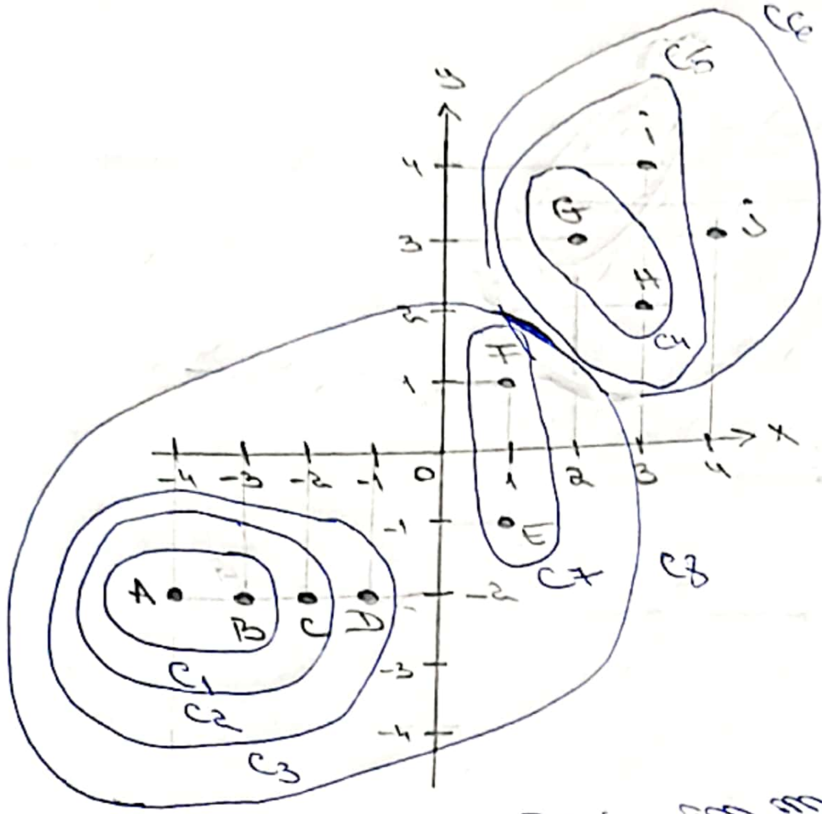
• cea mai mare distanță dintre 2 clustere este între C_7 și C_8 ⇒ tăiem dendrograma între C_7 și C_8 cu o dreaptă paralelă cu axa de coordonate;

↳ $\{x_1; x_2\}$, $\{x_3; x_4; x_5; x_6\}$, $\{x_7; x_8; x_9; x_{10}\}$;

EX. 6

$\begin{cases} A(-4;-2), B(-3;-2), C(-2;-2), D(-1;-2), E(1;-1), \\ F(1;1), G(2;3), H(3;2), I(3;4), J(4;3); \end{cases}$

a) reprezentarea grafică a datelor;



b) { clusterizarea ierarhică în manieră bottom-up, cu „single-linkage", distanța euclidiană;

- $SL \Leftrightarrow d(C_1, C_2) = \min\{d(x,y) \mid x \in C_1, y \in C_2\}$;
- dacă la o iterație avem aceeași distanță între 2 clustere, prioritatea la alcătuirea noului cluster este dată de ordinea alfabetică;
- $d(A,B) = d(B,C) = d(C,D) = 1$;
- $d(A,C) = d(B,D) = 2$;
- $d(G,I) = d(H,J) = \sqrt{2}$;
- $d(E,F) = 2$;

↳ $d(A,B) = d(B,C) = d(C,D) = 1$ este minimă $\Rightarrow C_1 = \{A;B\}$, $R(C_1) = 1$; $c(C_1) = 1$.

- $d(C_1, C) \overset{SL}{=} d(B,C) = 1$;

↳ $d(C_1, C) = d(C,D) = 1$ este minimă $\Rightarrow$ $C_2 = C_1 \cup \{C\} = \{A;B;C\}$, $R(C_2) = \frac{1+1+2}{3}$;, $c(C_2) = \frac{3}{4} = 0{,}75$;
$R(C_2) = \frac{4}{3} = 1{,}(3)$;

- h este direct proporțională cu media aritmetică a distanțelor dintre punctele din clusterul respectiv; coeziunea e este ~~...~~ inversul mediei: $\overline{h-1}$;

$\hookrightarrow h(C_2) = \frac{d(A,B)+d(B,C)+d(A,C)}{3}$;

- $d(C_2, D) \overset{SL}{=} d(C, D) = 1$;

$\hookrightarrow d(C_2, D)$ este minimă $\Rightarrow C_3 = C_2 \cup \{D\}$, $C_3 = \{A; B; C; D\}$ și $h(C_3) = \frac{1+1+2+2+1+3}{6}$,

$h(C_3) = \frac{10}{6} = 1,(6)$; $c(C_3) = \frac{6}{10} = 0,6$;

- $d(C_3, E) \overset{SL}{=} d(D,E) = \sqrt{5}$; $d(G,H) = d(G,J)$ este minimă $\Rightarrow$ $= \sqrt{2}$;

$\hookrightarrow d(G,H) = d(H,J) = \sqrt{2}$

$C_4 = \{G; H\}$ și $h(C_4) = \sqrt{2} = 1,414$,

$c(C_4) = \frac{1}{\sqrt{2}} = 0,707$;

- $d(C_4, I) \overset{SL}{=} d(G, I) = \sqrt{2}$;
- $d(C_4, J) \overset{SL}{=} d(H, J) = \sqrt{2}$;

$\hookrightarrow d(C_4, I) = d(C_4, J) = \sqrt{2}$ este minimă $\Rightarrow$

$C_5 = C_4 \cup \{I\} = \{G; H; I\}$ și $h(C_5) = \frac{2\sqrt{2}+2}{3}$,

$h(C_5) = 1,609$, $c(C_5) = \frac{3}{2\sqrt{2}+2} = 0,621$;

- $d(C_5, J) \overset{SL}{=} d(\cancel{H}\, I, J) = \sqrt{2}$;
- $d(C_5, F) \overset{SL}{=} d(G, F) = \sqrt{5}$;

$\hookrightarrow d(C_5, J) = \sqrt{2}$ este minimă $\Rightarrow C_6 = C_5 \cup \{J\}$,

$C_6 = \{G; H; I; J\}$ și $h(C_6) = \frac{4\sqrt{2}+4}{6} = \frac{2\sqrt{2}+2}{3} = 1,609$

$c(C_6) = \frac{3}{2\sqrt{2}+2} = 0,621$;

• $d(C_3,E) \overset{SL}{=} d(D,E) = \sqrt{5}$;

• $d(C_6, \cancel{E}F) \overset{SL}{=} d(G,F) = \sqrt{5}$;

• $d(E,F) = 2$;

↳ $d(E,F) = 2$ este minimă $\Rightarrow C_7 = \{E;F\}$ și

$h(C_7) = 2$, $c(C_7) = \frac{1}{2} = 0,5$;

• $d(C_3,C_7) \overset{SL}{=} d(D,E) = \sqrt{5}$;

• $d(C_6,C_7) \overset{SL}{=} d(G,F) = \sqrt{5}$;

↳ $d(C_3,C_7) = d(C_6,C_7) = \sqrt{5}$ este minimă și

$\Rightarrow C_8 = C_3 \cup C_7 = \{A;B;C;D;E;F\}$

$h(C_8) = \frac{10+2+\sqrt{5}+\sqrt{13}+\sqrt{18}+5+\sqrt{34}+\sqrt{26}+\sqrt{17}+\sqrt{10}}{15}$

$h(C_8) = 3,02$ și $c(C_8) = \frac{15}{45,2996} = 0,331$;

• $C_9 = C_8 \cup C_6$ și $h(C_9) = 4,559$, $c(C_9) = 0,219$;

c) clusterizare ierarhică „complete-linkage";

• „complete-linkage" (CL) $\Leftrightarrow d(C_1,C_2) = \max\{d(x,y) \mid x \in C_1, y \in C_2\}$;

• $d(A,B) = d(B,C) = d(C,D) = 1$ este minimă $\Rightarrow$

$C_1 = \{A;B\}$ și $h(C_1) = 1$, $c(C_1) = 1$;

• $d(C_1,C) \overset{CL}{=} d(A,C) = 2$;

↳ $d(C,D) = 1$ este minimă $\Rightarrow C_2 = \{C;D\}$

și $h(C_2) = 1$, $c(C_2) = 1$;

• $d(G,i) = d(G,H) = d(H,j) = d(i,j) = \sqrt{2}$;

• $d(C_1,C_2) \overset{CL}{=} d(A,D) = 3$;

↳ $d(G,H) = d(G,i) = d(H,j) = d(i,j) = \sqrt{2}$

este minimă $\Rightarrow C_3 = \{G;H\}$ și $h(C_3) = \sqrt{2}$,

$h(C_3) = 1,414$, $c(C_3) = \frac{1}{\sqrt{2}} = 0,707$;

- $d(C_3, i) \overset{CL}{=} d(H, i) = 2$;
- $d(C_3, j) \overset{CL}{=} d(G, j) = 2$;

  ↳ $d(i, j) = \sqrt{2}$ este minimă $\Rightarrow C_4 = \{i; j\}$ și $h(C_4) = \sqrt{2} = 1{,}414$, $c(C_4) = \frac{1}{\sqrt{2}} = 0{,}707$;

- $d(C_3, C_4) \overset{CL}{=} d(G, j) = 2$;

  ↳ $d(E, F) = d(C_3, C_4) = 2 \Rightarrow C_5 = \{E; F\}$ și $h(C_5) = 2$, $c(C_5) = \frac{1}{2} = 0{,}5$;

- $d(C_2, C_5) \overset{CL}{=} d(C, F) = \sqrt{10}$;
- $d(C_3, C_5) \overset{CL}{=} d(G, E) = \sqrt{17}$;

  ↳ $d(C_3, C_4) = 2$ este minimă $\Rightarrow C_6 = C_3 \cup C_4$, $C_6 = \{G; H; i; j\}$ și $h(C_6) = \frac{(\sqrt{2}+\sqrt{2}) \cdot 2 + 4}{6}$,

  $h(C_6) = \frac{2\sqrt{2}+2}{3} = 1{,}609$, $c(C_6) = \frac{3}{2\sqrt{2}+2} = 0{,}621$;

- ~~$d(C_1, C_2)$~~ $d(C_5, C_6) \overset{CL}{=} d(i, E) = \sqrt{29}$;

  ↳ $d(C_1, C_2) = 3$ este minimă $\Rightarrow C_7 = C_1 \cup C_2$, $C_7 = \{A; B, C; D\}$ și $h(C_7) = \frac{10}{6} = 1,(6)$,

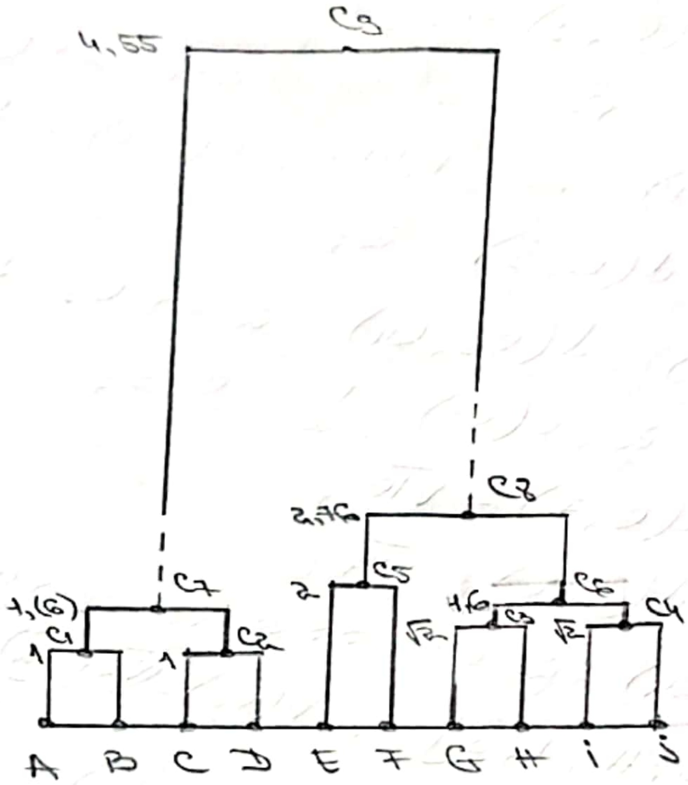
  $c(C_6) = \frac{6}{10} = 0{,}6$;

- $d(C_5, C_7) \overset{CL}{=} d(A, F) = \sqrt{34}$;

  ↳ $d(C_5, C_6) = \sqrt{29}$ este minimă $\Rightarrow C_8 = C_5 \cup C_6$, $C_8 = \{G; H; i; j; E; F\}$ și $h(C_8) = 2{,}76$,

  $c(C_8) = 0{,}36$;

- $C_9 = C_7 \cup C_8$ și $h(C_9) = 4{,}55$, $c(C_9) = 0{,}21$;



Ex. 8

"single – linkage", "average – linkage";

$$d(X \cup Y, Z) = \frac{|X|\, d(X,Z) + |Y|\, d(Y,Z)}{|X| + |Y|};$$

	A	B	C	D	E	F
A	0					
B	0,12	0				
C	0,51	0,25	0			
D	0,84	0,16	0,14	0		
E	0,28	0,77	0,70	0,45	0	
F	0,34	0,61	0,93	0,20	0,67	0

- "single – linkage" ⇔ $d(C_1, C_2) = \min\{d(A,B) \mid A \in C_1, B \in C_2\}$ (SL);

• $d(A,B) = 0{,}12$ este minimă $\Rightarrow C_1 = \{A; B\}$ și

$h(C_1) = 0{,}12$ (și pentru SL și AL);

• $d(AB,C) = \min\{d(A,C); d(B,C)\} = 0{,}25$;

• $d(AB,D) = \min\{d(A,D); d(B,D)\} = 0{,}16$;

• $d(AB,E) = \min\{d(A,E); d(B,E)\} = 0{,}28$;

• $d(AB,F) = \min\{d(A,F); d(B,F)\} = 0{,}34$;

• „average-linkage" $\Leftrightarrow d(C_1,C_2) = \dfrac{\sum d(A,B)}{|C_1|\cdot|C_2|}$, unde

$A \in C_1$, $B \in C_2$ (AL);

• $d(AB,C) = \dfrac{d(A,C)+d(B,C)}{2} = \dfrac{0{,}51+0{,}25}{2} = 0{,}38$;

• $d(AB,D) = \dfrac{d(A,D)+d(B,D)}{2} = \dfrac{0{,}84+0{,}16}{2} = \dfrac{1}{2} = 0{,}5$;

• $d(AB,E) = \dfrac{d(A,E)+d(B,E)}{2} = \dfrac{0{,}28+0{,}77}{2} = \dfrac{1{,}05}{2} = 0{,}525$

• $d(AB,F) = \dfrac{d(A,F)+d(B,F)}{2} = \dfrac{0{,}34+0{,}61}{2} = 0{,}475$;

SL	AB	C	D	E	F
AB	0				
C	0,25	0			
D	0,16	0,14	0		
E	0,28	0,70	0,45	0	
F	0,34	0,93	0,20	0,67	0

AL	AB	C	D	E	F
AB	0				
C	0,38	0			
D	0,50	0,14	0		
E	0,525	0,70	0,45	0	
F	0,475	0,93	0,20	0,67	0

• $d(C,D) = 0{,}14$ este ca minimă și pentru SL $\Rightarrow C_2 = \{C; D\}$ și $h(C_2) = 0{,}14$;

și pentru AL

• $d(AB,CD) = \min\{d(AB,C); d(AB,D)\} = 0{,}16$;

• $d(CD,E) = \min\{d(C,E); d(D,E)\} = 0{,}45$;

• $d(CD,F) = \min\{d(C,F); d(D,F)\} = 0{,}20$;

• $d(AB,CD) = d(AB,\{C\} \cup \{D\}) = \dfrac{2d(AB,C) + 2d(AB,D)}{4} =$

$= \dfrac{d(AB,C) + d(AB,D)}{2} = \dfrac{0,38 + 0,50}{2} = 0,44;$

• $d(CD,E) = \dfrac{d(C,E) + d(D,E)}{2} = \dfrac{0,70 + 0,45}{2} = 0,575;$

• $d(CD,F) = \dfrac{d(C,F) + d(D,F)}{2} = \dfrac{0,93 + 0,20}{2} = 0,565;$

AL	AB	CD	E	F
AB	0			
CD	0,44	0		
E	0,525	0,575	0	
F	0,475	0,565	0,67	0

SL	AB	CD	E	F
AB	0			
CD	0,16	0		
E	0,28	0,45	0	
F	0,34	0,20	0,67	0

• $d(AB,CD) = 0,16$ pentru SL şi $d(AB,CD) = 0,44$ pt. AL sunt minime $\Rightarrow C_3 = \{A; B; C; D\}$ şi $h(C_3) = 0,16$ pentru SL şi $h(C_3) = 0,44$ pentru AL;

• $d(ABCD, E) = \min\{d(AB,E); d(CD,E)\} = 0,28;$

• $d(ABCD, F) = \min\{d(AB,F); d(CD,F)\} = 0,20;$

• $d(ABCD,E) = d(\{A;B\} \cup \{C;D\}, E) =$

$= \dfrac{2d(AB,E) + 2d(CD,E)}{4} = \dfrac{\cancel{2}(0,525 + 0,575)}{\cancel{4}\,2} = 0,55;$

• $d(ABCD,F) = d(\{A;B\} \cup \{C;D\}, F) =$

$= \dfrac{2d(AB,F) + 2d(CD,F)}{4} = \dfrac{\cancel{2}(0,475 + 0,565)}{\cancel{4}\,2} = 0,52;$

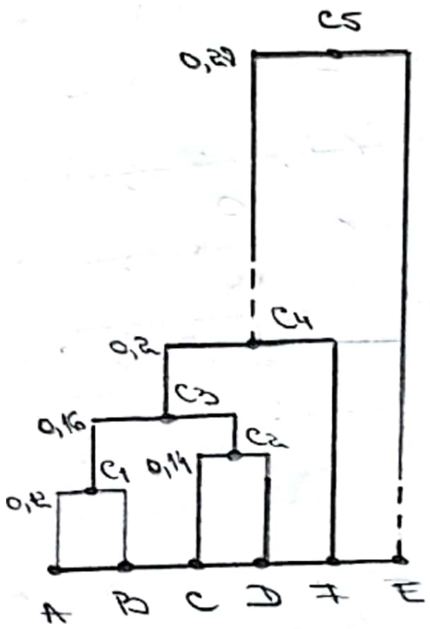
AL	ABCD	E	F
ABCD	0		
E	0,55	0	
F	0,52	0,67	0

SL	ABCD	E	F
ABCD	0		
E	0,28	0	
F	0,20	0,67	0

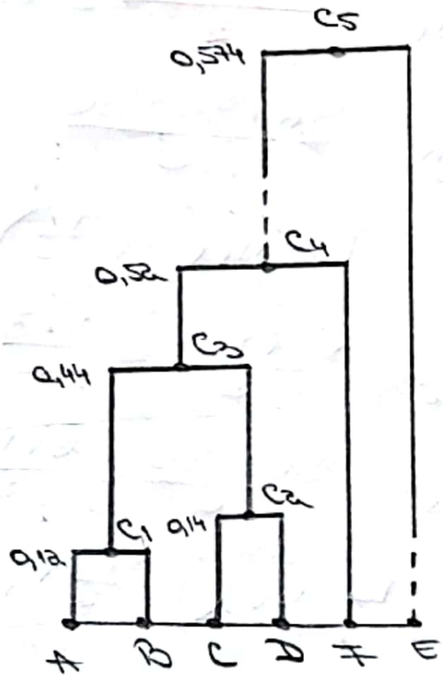
- $d(ABCD, F) = 0,52$ pentru AL și $d(ABCD, F) = 0,20$ pentru SL sunt minime $\Rightarrow C_4 = \{A; B; C; D; F\}$ și $h(C_4) = 0,52$ (AL); și $h(C_4) = 0,20$ (SL),
- $d(ABCDF, E) = \min\{d(ABCD, E); d(F, E)\} = 0,28$;
- $d(ABCDF, E) = d(\{A; B; C; D\} \cup \{F\}, E) = \dfrac{4d(ABCD, E) + d(F, E)}{5} = \dfrac{4 \cdot 0,55 + 0,67}{5} = 0,574$;

SL	ABCDF	E
ABCDF	0	
E	0,28	0

AL	ABCDF	E
ABCDF	0	
E	0,574	0



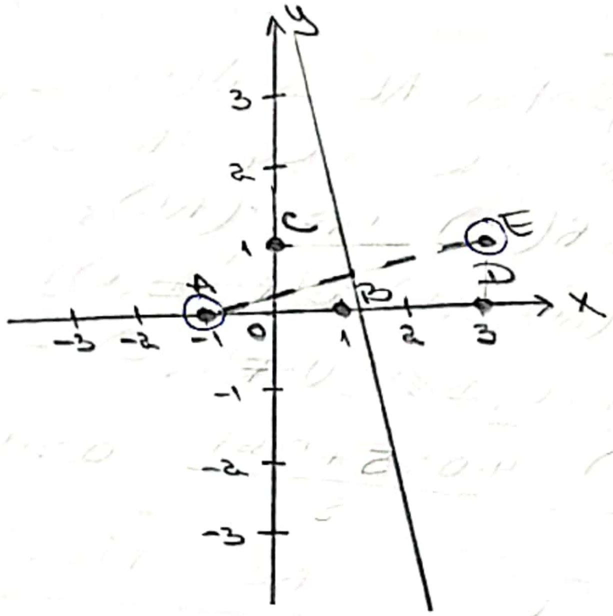
(SL)



(AL)

Ex. 9

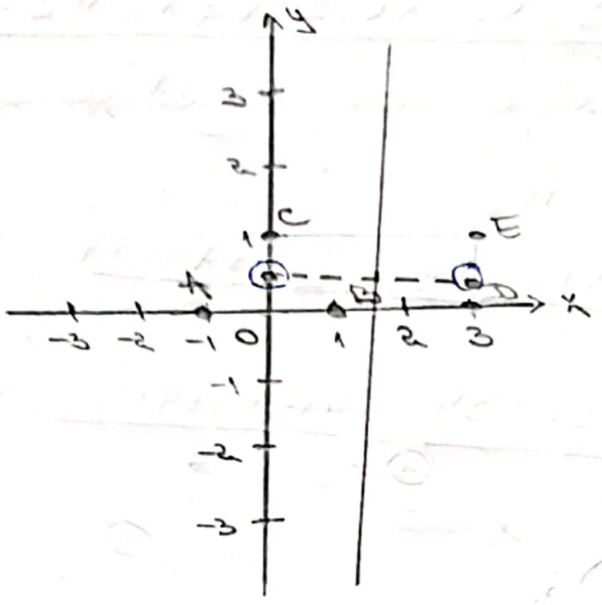
$\begin{cases} A(-1;0),\ B(1;0),\ C(0;1),\ D(3;0),\ E(3;1); \\ 2\text{-means};\ M_1(-1;0),\ M_2(3;1);\ 2 \text{ iterații}; \end{cases}$



- $C_1 = \{A; B; C\}$, $C_2 = \{E; D\}$;
- $x_{M_1} = \dfrac{x_A + x_B + x_C}{3} = \dfrac{-\cancel{1} + \cancel{1} + 0}{3} = 0$;
- $y_{M_1} = \dfrac{y_A + y_B + y_C}{3} = \dfrac{0 + 0 + 1}{3} = \dfrac{1}{3}$;

  $\hookrightarrow M_1\left(0; \frac{1}{3}\right)$;
- $x_{M_2} = \dfrac{x_E + x_D}{2} = \dfrac{3+3}{2} = \dfrac{6}{2} = 3$;
- $y_{M_2} = \dfrac{y_E + y_D}{2} = \dfrac{0+1}{2} = \dfrac{1}{2}$;

  $\hookrightarrow M_2\left(3; \frac{1}{2}\right)$;

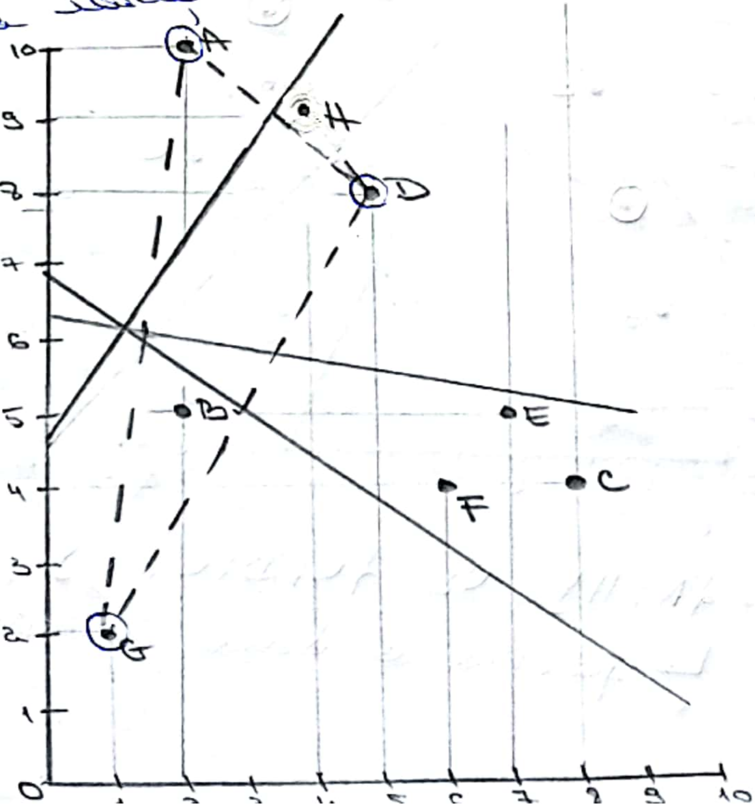


↳ $C_1 = \{A; B; C\}$ și $C_2 = \{D; E\}$;

Ex. 10

$A(2;10)$, $B(2;5)$, $C(8;4)$, $D(5;8)$, $E(7;5)$, $F(6;4)$, $G(1;2)$, $H(4;9)$; 3-means; $\mu_1 = A$, $\mu_2 = D$; $\mu_3 = G$; distanța euclidiană;

a) prima iterație:



↳ $C_1=\{A\}$, $C_2=\{C;D;E;F;H\}$, $C_3=\{G;B\}$.

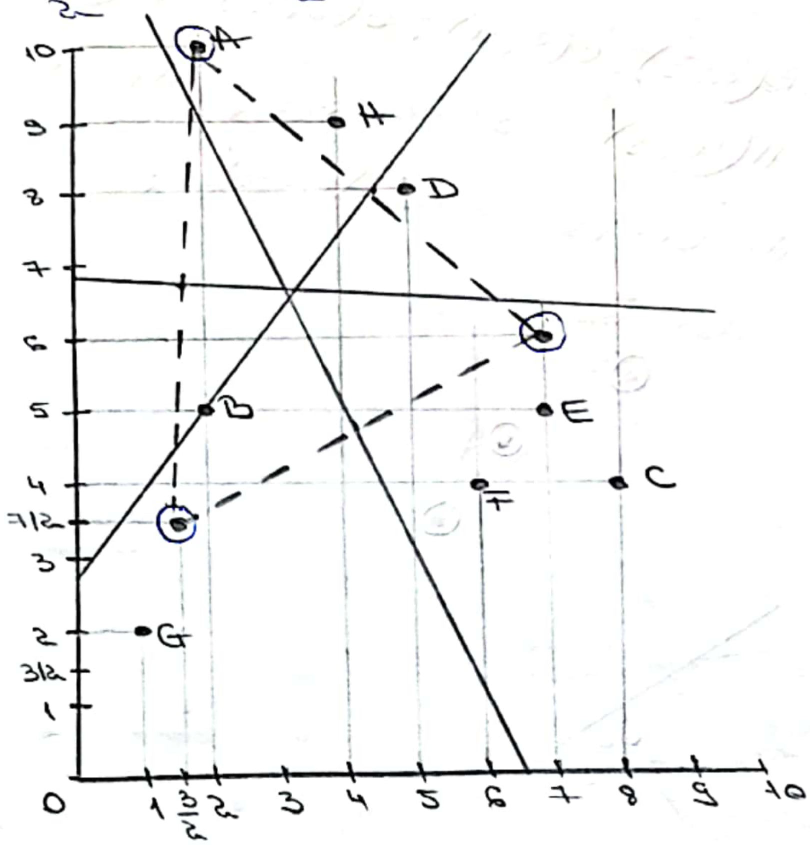
b) nr. iterații pentru care algoritmul converge;

- $C_1=\{A\} \Rightarrow \mu_1=A$;
- $x_{\mu_2}=\dfrac{x_C+x_D+x_E+x_F+x_H}{5}=\dfrac{8+5+7+6+4}{5}=\dfrac{30}{5}=6$;
- $y_{\mu_2}=\dfrac{y_C+y_D+y_E+y_F+y_H}{5}=\dfrac{4+8+5+4+9}{5}=\dfrac{30}{5}=6$;

↳ $\mu_2(6;6)$;

- $x_{\mu_3}=\dfrac{x_G+x_B}{2}=\dfrac{2+1}{2}=\dfrac{3}{2}$;
- $y_{\mu_3}=\dfrac{y_G+y_B}{2}=\dfrac{5+2}{2}=\dfrac{7}{2}$;

$\Rightarrow \mu_3\left(\dfrac{3}{2};\dfrac{7}{2}\right)$;



↳ $C_1=\{A;H\}$, $C_2=\{C;D;E;F\}$, $C_3=\{G;B\}$;

↳ pentru a doua iterație;

EXERCIȚII SESIUNE

Ex. 1

demonstrație $\log P(X|\theta) \geq \sum_{z} q(z) \log\left(\frac{P(X,z|\theta)}{q(z)}\right)$;

- log cu baza supraunitară $\Leftrightarrow$ funcție concavă (1);
- inegalitatea lui Jensen $\Leftrightarrow$ considerând X o variabilă aleatoare (uniformă), dacă φ este o funcție concavă, atunci $\varphi(E[X]) \geq E[\varphi(X)]$ (2);
- (1), (2) $\Rightarrow \log(E[X]) \geq E[\log(X)]$ (3);
- $l(\theta) = \log P(X|\theta) = \log\left(\sum_{z} P(X,z|\theta)\right) =$

$= \log\left(\sum_{z} q(z) \frac{P(X,z|\theta)}{q(z)}\right) = \log\left(E_{q(z)}\left[\frac{P(X,z|\theta)}{q(z)}\right]\right)$ (4)

- în (3), $X \rightarrow \frac{P(X,z|\theta)}{q(z)} \Rightarrow \log\left(E_{q(z)}\left[\frac{P(X,z|\theta)}{q(z)}\right]\right) \geq$

$\geq E_{q(z)}\left[\log\left(\frac{P(X,z|\theta)}{q(z)}\right)\right]$ (5);

- (5), (4) $\Rightarrow l(\theta) \geq \sum_{z} q(z) \log \frac{P(X,z|\theta)}{q(z)}$; $l(\theta) = \log P(X|\theta)$;

↳ adevărat;

Ex. 2

demonstrație $\log P(X|\theta) = F(q(z),\theta) + KL(q(z)\,||\,P(z|X,\theta))$;

- $F(q(z),\theta) = \sum_{z} q(z) \log\left(\frac{P(X,z|\theta)}{q(z)}\right) =$

$= \sum_{z} q(z) \log\left(\frac{P(z|X,\theta) \cdot P(X|\theta)}{q(z)}\right)$ (1);

↳ $P(B \cap C|A) = P(B|A) \cdot P(C|A,B) \Leftrightarrow$

$P(X,z|\theta) = P(X|\theta) \cdot P(z|X,\theta)$;

- $\log a \cdot b = \log a + \log b$ (2);

• (1), (2) $\Rightarrow F(q(z), \theta) = \sum_z q(z)\left(\log \frac{P(z|X,\theta)}{q(z)} + \log P(X|\theta)\right) =$

$= \sum_z \left(q(z) \log \frac{P(z|X,\theta)}{q(z)} + q(z) \log P(X|\theta)\right)$ (3);

• $\sum_i a + \sum_i b = \sum_i (a+b)$ (4);

• (3), (4) $\Rightarrow F(q(z), \theta) = \sum_z q(z) \log \frac{P(z|X,\theta)}{q(z)} + \sum_z q(z) \log P(X|\theta) =$

$= \sum_z q(z) \log \frac{P(z|X,\theta)}{q(z)} + P(X|\theta) \sum_z q(z)$ (5);

• $KL(q(z) \| P(z|X,\theta)) = -\sum_z q(z) \log\left(\frac{P(z|X,\theta)}{q(z)}\right)$;

$\sum_z q(z) \log \frac{P(z|X,\theta)}{q(z)} = -KL(q(z) \| P(z|X,\theta))$ (6)

• q este o distribuție de probabilitate $\Rightarrow$

$\sum_z q(z) = 1$ (suma probabilităților dintr-o distribuție de probabilitate este 1) (7);

• (5), (6), (7) $\Leftrightarrow F(q(z), \theta) = -KL(q(z) \| P(z|X,\theta)) + \log P(X|\theta) \cdot 1$;

$F(q(z), \theta) = \log P(X|\theta) - KL(q(z) \| P(z|X,\theta))$;

$\log P(X|\theta) = F(q(z), \theta) + KL(q(z) \| P(z|X,\theta))$;

↳ adevărat;

Ex. 3

demonstrație maximul lui F în raport cu $q(z)$ este atins pentru distribuția $P(z|X, \theta^{(t)})$ și valoarea maximului este $\max_{q(z)} F(q(z), \theta^{(t)}) = H(p(z|x,\theta^{(t)})) +$

$+ E_{p(z|x,\theta^{(t)})}[\log p(x,z|\theta^{(t)})]$;

- din Ex. 1 $\Rightarrow F(q(z),\theta) \le \log p(x|\theta)$;

  ↳ $F(q(z),\theta^{(t)}) \le \log p(x|\theta^{(t)})$ (1);

- din Ex. 2 $\Rightarrow \log p(x|\theta) = F(q(z),\theta) + KL(q(z) \| p(z|x,\theta))$

  ↳ $\log p(x|\theta^{(t)}) = F(q(z),\theta^{(t)}) + KL(q(z) \| p(z|x,\theta^{(t)}))$ (2);

- $\exists\, KL(p \| p) = 0$ $(KL(p\|q) = 0 \Leftrightarrow p = q)$ (3);

- (2),(3), $q(z) = p(z|x,\theta^{(t)}) \Rightarrow \log p(x|\theta^{(t)}) = F(p(z|x,\theta^{(t)}),\theta^{(t)}) + KL(p(z|x,\theta^{(t)}) \| p(z|x,\theta^{(t)})) =$

  $= F(p(z|x,\theta^{(t)}),\theta^{(t)}) + 0$;

  $\log p(x|\theta^{(t)}) = F(p(z|x,\theta^{(t)}),\theta^{(t)})$ (4);

  Din (1) și (4) $\Rightarrow$ <u>maximul lui F se obține pentru distribuția $q(z) = p(z|x,\theta^{(t)})$</u> (5).

- $\log p(x|\theta^{(t)}) = \max_{q(z)} F(q(z),\theta^{(t)}) =$

  $\overset{(5)}{=} F(p(z|x,\theta^{(t)}),\theta^{(t)}) = \sum_z p(z|x,\theta^{(t)}) \log\left(\frac{p(x,z|\theta^{(t)})}{p(z|x,\theta^{(t)})}\right) =$

  $= E_{p(z|x,\theta^{(t)})}\left[\log \frac{p(x,z|\theta^{(t)})}{p(z|x,\theta^{(t)})}\right]$ (6);

- $\log \frac{a}{b} = \log a - \log b$ (7);

- (6),(7) $\Rightarrow \max_{q(z)} F(q(z),\theta^{(t)}) =$

  $= E_{p(z|x,\theta^{(t)})}[\log p(x,z|\theta^{(t)}) - \log p(z|x,\theta^{(t)})]$ (8);

- $E[X-Y] = E[X] - E[Y]$ (9);
- (8), (9) $\Rightarrow \max_{q(z)} F(q(z), \theta^{(t)}) =$ $E_{p(z|x,\theta^{(t)})}[\log p(x,z|\theta^{(t)})] - E_{p(z|x,\theta^{(t)})}[\log p(z|x,\theta^{(t)})]$

$= E_{p(z|x,\theta^{(t)})}[\log p(x,z|\theta^{(t)})] - H[p(z|x,\theta^{(t)})]$;

$= E_{p(z|x,\theta^{(t)})}[\log p(x,z|\theta^{(t)})]$

↳ adevărat;

Ex. 4

- demonstrație $\log p(x|\theta) - \log p(x|\theta^{(t)}) \geq Q(\theta|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t)})$, $\forall t$ fixat și $\forall \theta$ astfel încât $Q(\theta|\theta^{(t)}) \geq Q(\theta^{(t)}|\theta^{(t)})$; ~~log a·b = log a + log b~~

- $p(B \cap C|A) = p(B|A) \cdot p(C|A,B)$ (1);
- $p(x,z|\theta) \overset{(1)}{=} p(x|\theta) \cdot p(z|x,\theta)$;
- $\log p(x,z|\theta) = \log(p(x|\theta) \cdot p(z|x,\theta))$;

$\log p(x,z|\theta) \overset{(2)}{=} \log p(x|\theta) + \log p(z|x,\theta)$;

$\log p(x|\theta) = \log p(x,z|\theta) - \log p(z|x,\theta)$;

$\log p(x|\theta) = \sum_z p(z|x,\theta^{(t)}) \cdot \log p(x|\theta) = \sum_z (\log p(x,z|\theta) - \log p(z|x,\theta)) \cdot p(z|x,\theta^{(t)})$ (3);

- $\sum_i a - b = \sum_i a - \sum_i b$ (4);
- $\log a \cdot b = \log a + \log b$ (2);
- $a = b - c \Rightarrow xa = x(b-c) \Rightarrow \sum_i xa = \sum_i x(b-c) \Rightarrow$ (3);
- (3), (4) $\Rightarrow \sum_z p(z|x,\theta^{(t)}) \log p(x|\theta) =$

$= \sum_z \left(P(z|x,\theta^{(t)}) \log P(x,z|\theta) - P(z|x,\theta^{(t)}) \cdot \log P(z|x,\theta) \right)$ și ;

$\sum_z P(z|x,\theta^{(t)}) \log P(x|\theta) = \sum_z P(z|x,\theta^{(t)}) \log P(x,z|\theta) - \sum_z P(z|x,\theta^{(t)}) \log P(z|x,\theta)$ (5);

- $\sum_z P(z|x,\theta^{(t)}) \log P(x|\theta) = \log P(x|\theta) \sum_z P(z|x,\theta^{(t)}) = \log P(x|\theta) \cdot 1 = \log P(x|\theta)$ (6);

  ↳ $q(z) = P(z|x,\theta^{(t)})$ este o distribuție de probabilitate $\Rightarrow \sum_z q(z) = 1$;

- $Q(\theta|\theta^{(t)}) = E_{P(z|x,\theta^{(t)})} [\log P(x,z|\theta)] = \sum_z P(z|x,\theta^{(t)}) \log P(x,z|\theta)$ (7);

- $CH(\theta,\theta^{(t)}) = -\sum_z P(z|x,\theta^{(t)}) \log P(z|x,\theta) \Leftrightarrow$ cross-entropie (8);

- (5), (6), (7), (8) $\Rightarrow \log P(x|\theta) = Q(\theta|\theta^{(t)}) + CH(\theta|\theta^{(t)})$, $\forall \theta$ (9);

- (9), $\theta = \theta^{(t)} \Rightarrow \log P(x|\theta^{(t)}) = Q(\theta^{(t)}|\theta^{(t)}) + CH(\theta^{(t)}|\theta^{(t)})$ (10);

- inegalitatea lui Gibbs $\Leftrightarrow CH(\theta|\theta^{(t)}) \geq CH(\theta^{(t)}|\theta^{(t)})$ (11);

- (9) - (10) $\Rightarrow \log P(x|\theta) - \log P(x|\theta^{(t)}) = Q(\theta|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t)}) + CH(\theta|\theta^{(t)}) - CH(\theta^{(t)}|\theta^{(t)})$ (12);

- (11), (12) $\Rightarrow \log P(x|\theta) - \log P(x|\theta^{(t)}) \geq Q(\theta|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t)})$;

  ↳ adevărat;

Ex. 5

2 monede; heads cu probabilitate p (prima);
heads cu probabilitate q (a doua);
n aruncări;
pentru fiecare aruncare, alegem prima monedă
cu probabilitate π;
rezultatul fiecărei aruncări i este $x_i \in \{0,1\}$;
$\theta = \{p, q, \pi\}$;
$z_i \in \{0,1\}$ este moneda folosită pentru aruncarea i;

a) $E[z_i | x_i, \theta] = P(z_i = 1 | x_i, \theta)$;

- $E[z_i | x_i, \theta] = \sum_{z_i \in \{0,1\}} z_i \cdot P(z_i | x_i, \theta) =$

$= \underbrace{0 \cdot P(z_i = 0 | x_i, \theta)}_{0} + 1 \cdot P(z_i = 1 | x_i, \theta)$;

$\hookrightarrow \underline{E[z_i | x_i, \theta] = P(z_i = 1 | x_i, \theta)}$;

b) $P(z_i = 1 | x_i, \theta^{(t)})$ cu Bayes;

- teorema lui Bayes $\Leftrightarrow P(B|A) = \dfrac{P(A|B)P(B)}{P(A|B)P(B) + P(\neg A|B)P(B)}$ (1)

- $\cancel{P(z_i = 1 | x_i, \theta^{(t)}) = \dfrac{P(x_i, \theta^{(t)} | z_i = 1) P(z_i = 1)}{P(x_i, \theta^{(t)} | z_i = 1) P(z_i = 1) + P(x_i, \theta^{(t)} | z_i = 0) P(z_i = 0)}}$;

- teorema lui Bayes pentru 3 variabile $\Leftrightarrow$

$P(R | H, S) = \dfrac{P(H | R, S) P(R | S)}{P(H | R, S) P(R | S) + P(H | \neg R, S) P(\neg R | S)}$ (2);

• $p(z_i=1 \mid x_i, \theta^{(t)}) \overset{(2)}{=}$

$$= \frac{p(x_i \mid z_i=1, \theta^{(t)})\, p(z_i=1 \mid \theta^{(t)})}{p(x_i \mid z_i=1, \theta^{(t)})\, p(z_i=1 \mid \theta^{(t)}) + p(x_i \mid z_i=0, \theta^{(t)})\, p(z_i=0 \mid \theta^{(t)})}$$

$$= \frac{p^{x_i} \cdot (1-p)^{x_i} \cdot \pi}{p^{x_i} \cdot (1-p)^{x_i} \cdot \pi + q^{x_i} \cdot (1-q)^{x_i} \cdot (1-\pi)};$$

$$\hookrightarrow \underline{p(z_i=1 \mid x_i, \theta^{(t)}) = \frac{p^{x_i}(1-p)^{x_i}\pi}{p^{x_i}(1-p)^{x_i}\pi + q^{x_i}(1-q)^{x_i}(1-\pi)}};$$

c) $\log p(x, z \mid \theta);$

• $\log p(x, z \mid \theta) = \log \prod_{i=1}^{n} p(x_i, z_i \mid \theta) =$

$= \log \prod_{i=1}^{n} p(x_i \mid \theta) \cdot p(z_i \mid x_i, \theta) =$

$\hookrightarrow p(B \cap C \mid A) = p(B \mid A) \cdot p(C \mid A, B);$

$= \log \prod_{i=1}^{n} p(z_i \mid \theta) \cdot p(x_i \mid z_i, \theta) =$

$= \log \prod_{i=1}^{n} (p^{x_i} \cdot (1-p)^{x_i} \cdot \pi)^{z_i} \cdot (q^{x_i} \cdot (1-q)^{1-x_i} \cdot (1-\pi))^{1-z_i} =$

$= \sum_{i=1}^{n} \log ((p^{x_i} \cdot (1-p)^{x_i} \cdot \pi)^{z_i} + \log (q^{x_i} \cdot (1-q)^{1-x_i} \cdot (1-\pi))^{1-z_i};$

$\hookrightarrow \log a \cdot b = \log a + \log b;$

$= \sum_{i=1}^{n} (\log (p^{x_i} \cdot (1-p)^{1-x_i} \cdot \pi)^{z_i} +$

• $\log a^x = x \log a;$

$$\hookrightarrow \underline{\log p(x, z \mid \theta) = \sum_{i=1}^{n} \left(z_i \log p^{x_i}(1-p)^{1-x_i}\pi + (1-z_i) \log q^{x_i}(1-q)^{1-x_i}(1-\pi) \right)};$$

d) $Q(\theta|\theta^{(t)}) = E_{p(z|x,\theta^{(t)})}[\log p(x,z|\theta)] =$

$= \sum_{i=1}^{n} E[z_i|x_i,\theta^{(t)}](\log \pi + x_i \log \eta + (1-x_i)\log(1-\eta) +$

$+ (1 - E[z_i|x_i,\theta^{(t)}])(\log(1-\pi) + x_i \log q +$

$+ (1-x_i)\log(1-q))$;

- $Q(\theta|\theta^{(t)}) = E_{p(z|x,\theta^{(t)})}[\log p(x,z|\theta)] =$

$\overset{c)}{=} E_{p(z|x,\theta^{(t)})}\left[\sum_{i=1}^{n}\left(z_i \log \eta^{x_i}(1-\eta)^{1-x_i}\pi + (1-z_i)\cdot \log q^{x_i}(1-q)^{1-x_i}(1-\pi)\right)\right] =$

$= \sum_{i=1}^{n}\left(E[z_i|x_i,\theta^{(t)}] \log \eta^{x_i}(1-\eta)^{1-x_i}\pi + (1-E[z_i|x_i,\theta^{(t)}]) \log q^{x_i}(1-q)^{1-x_i}(1-\pi)\right) =$

$= \sum_{i=1}^{n}\left(E[z_i|x_i,\theta^{(t)}] \log \eta^{x_i}(1-\eta)^{1-x_i}\pi - E[z_i|x_i,\theta^{(t)}] \log q^{x_i}(1-q)^{1-x_i}(1-\pi) + \log q^{x_i}(1-q)^{1-x_i}(1-\pi)\right) =$

$= \sum_{i=1}^{n}\left(E[z_i|x_i,\theta^{(t)}]\left(\log \eta^{x_i}(1-\eta)^{1-x_i} + \log \pi - \log q^{x_i}(1-q)^{1-x_i}(1-\pi)\right) + \log q^{x_i}(1-q)^{1-x_i}(1-\pi)\right) =$

$= \sum_{i=1}^{n}\left(E[z_i|x_i,\theta^{(t)}]\left(\log \eta^{x_i} + \log(1-\eta)^{1-x_i} + \log(1-\pi) + \log q^{x_i} + \log(1-q)^{1-x_i} + \log(1-\pi)\right)\right)$;

$+ \log q^{x_i} + \log(1-q)^{1-x_i} + \log(1-\pi)$;

$+ \log q^{x_i} + \log b$;

$\hookrightarrow \log a \cdot b = \log a + \log b$;

- $\log a^n = n \log a \quad (1)$;
- $Q(\theta,\theta^{(t)}) \overset{(1)}{=} \sum_{i=1}^{n}\left(E[z_i|x_i,\theta^{(t)}](x_i \log \eta + (1-x_i)\log(1-\eta) + \log \pi + x_i \log q + (1-x_i)\log(1-q) + \log(1-\pi)\right) +$

$+ x_i \log q + (1-x_i)\log(1-q) + \log(1-\pi)$;

$\hookrightarrow Q(\theta, \theta^{(t)}) = \sum_{i=1}^{n} \Big(E[z_i | x_i, \theta^{(t)}]\big(x_i \log p + (1-x_i)\log(1-p) + \log \pi\big) + (1 - E[z_i | x_i, \theta^{(t)}]) \cdot \big(\log q \cdot x_i + (1-x_i)\log(1-q) + \log(1-\pi)\big)\Big)$;

Ex. 6

x_i	X_1	X_2	y_i
x_1	1	2	+1
x_2	2	3	+1
x_3	3	4	−1
x_4	3	2	−1
x_5	3	1	−1
x_6	4	4	−1
x_7	5	4	−1
x_8	5	2	+1
x_9	5	1	+1

$T=3$ iterații AdaBoost;

- $T=1 \Rightarrow D_1(i) = \frac{1}{9}, \forall i = \overline{1,9}$;
- pragurile exterioare $\Leftrightarrow \frac{1}{2}$ pentru X_1 și X_2;
- pragurile interioare $\Leftrightarrow \frac{5}{2}$ și $\frac{9}{2}$ pentru X_1 și

$\frac{3}{2}, \frac{5}{2}$ și $\frac{7}{2}$ pentru X_2;

• eroarea ponderată minimă la antrenare este $\varepsilon_3 = \frac{1}{8}$ și se obține pentru $x_1 \geq \frac{9}{2}$;

$\hookrightarrow h_3 = \text{sign}\left(x_1 - \frac{9}{2}\right)$;

• $\alpha_3 = \frac{1}{2} \ln \frac{1-\varepsilon_3}{\varepsilon_3} = \frac{1}{2} \ln \frac{1-\frac{1}{8}}{\frac{1}{8}} = \frac{1}{2} \ln \frac{\frac{7}{8}}{\frac{1}{8}} =$

$= \frac{1}{2} \ln \frac{7}{8} \cdot \frac{8}{1} = \frac{1}{2} \ln 7 = \ln \sqrt{7}$;

t	ε_t	α_t	$D_t(1)$	$D_t(2)$	$D_t(3)$	$D_t(4)$	$D_t(5)$	$D_t(6)$	$D_t(7)$
1	$\frac{2}{9}$	$\ln\sqrt{\frac{7}{2}}$	1/9	1/9	1/9	1/9	1/9	1/9	1/9
2	$\frac{2}{14}$	$\ln\sqrt{6}$	1/14	1/14	1/14	1/4	1/4	1/14	1/14
3	$\frac{1}{8}$	$\ln\sqrt{7}$	1/24	1/24	1/24	7/42	7/42	1/24	1/24

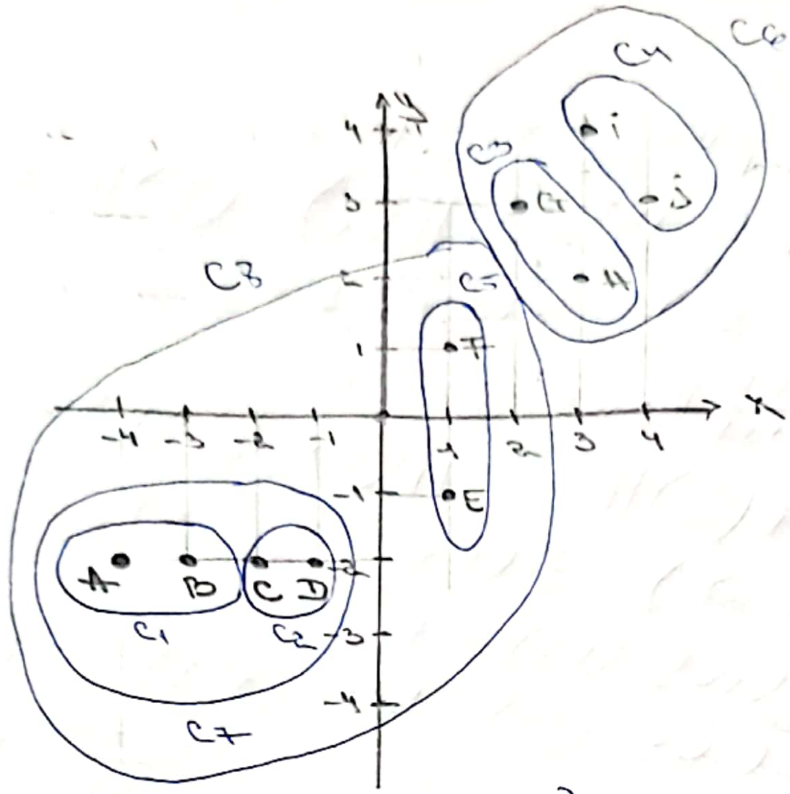
t	ε_t	α_t	$D_t(8)$	$D_t(9)$	eroare(H_t)
1	$\frac{2}{9}$	$\ln\sqrt{\frac{7}{2}}$	1/9	1/9	$\frac{2}{9}$
2	$\frac{2}{14}$	$\ln\sqrt{6}$	1/14	1/14	$\frac{2}{9}$
3	$\frac{1}{8}$	$\ln\sqrt{7}$	1/4	1/4	0

•

t	α_t	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	$\ln\sqrt{\frac{7}{2}}$	+1	+1	−1	+1	+1	−1	−1	+1	+1
2	$\ln\sqrt{6}$	+1	+1	−1	−1	−1	−1	−1	−1	−1
3	$\ln\sqrt{7}$	−1	−1	−1	−1	−1	−1	+1	+1	+1
	$H_3(x_i)$	+1	+1	−1	−1	−1	−1	−1	+1	+1

Ex. 7

$A(-4,-2)$, $B(-3,-2)$, $C(-2,-2)$, $D(-1,-2)$, $E(1,-1)$;
$F(1,1)$, $G(2,3)$, $H(3,2)$, $I(3,4)$, $J(4,3)$;
metrica lui Ward;
distanța dintre 2 perechi de clustere este egală $\Leftrightarrow$
prioritate alfabetică;



• $\Delta(\{A\},\{B\}) = \frac{1\cdot 1}{1+1}\|\mu_{\{A\}}-\mu_{\{B\}}\|^2$;

• $\mu_{\{A\}} = \frac{1}{1}\cdot \begin{pmatrix}-4\\-2\end{pmatrix} = \begin{pmatrix}-4\\-2\end{pmatrix}$;

• $\mu_{\{B\}} = \frac{1}{1}\cdot \begin{pmatrix}-3\\-2\end{pmatrix} = \begin{pmatrix}-3\\-2\end{pmatrix}$,

• $\|\mu_{\{A\}}-\mu_{\{B\}}\|^2 = \left(\sqrt{(-4+3)^2+(-2+2)^2}\right)^2 = 1^2+0^2 = 1$;

$\hookrightarrow \Delta(\{A\},\{B\}) = \frac{1}{2}\cdot 1 = \frac{1}{2}$;

• $\Delta(\{D\},\{E\}) = \frac{1\cdot 1}{1+1}\|\mu_{\{D\}}-\mu_{\{E\}}\|^2 = \frac{1}{2}\left\|\begin{pmatrix}-1\\-2\end{pmatrix}-\begin{pmatrix}1\\-1\end{pmatrix}\right\|^2 =$

$= \frac{1}{2}\left\|\frac{1}{1}\cdot\begin{pmatrix}-1\\-2\end{pmatrix} - \frac{1}{1}\cdot\begin{pmatrix}1\\-1\end{pmatrix}\right\|^2 = \frac{1}{2}(2^2+1) = \frac{5}{2}$;

$= \left((-1-1)^2+(-2+1)^2\right)\cdot\frac{1}{2} = \frac{1}{2}$

• $\Delta(\{E\},\{F\}) = \frac{1\cdot 1}{1+1}\|\mu_{\{E\}}-\mu_{\{F\}}\|^2 = \frac{1}{2}\left\|\begin{pmatrix}1\\-1\end{pmatrix}-\begin{pmatrix}1\\1\end{pmatrix}\right\|^2 =$

$= \frac{1}{2}\left\|\frac{1}{1}\cdot\begin{pmatrix}1\\-1\end{pmatrix} - \frac{1}{1}\cdot\begin{pmatrix}1\\1\end{pmatrix}\right\|^2 = \frac{1}{2}(0^2+2^2) = \frac{4}{2} = 2$;

$= \frac{1}{2}\left((1-1)^2+(-1-1)^2\right) =$

• $\Delta(\{B\},\{C\}) = \Delta(\{C\},\{D\}) = \Delta(\{A\},\{B\}) = \frac{1}{2}$;

• $\Delta(\{F\},\{G\}) = \frac{1\cdot 1}{1+1}\,\|\mu_{\{F\}} - \mu_{\{G\}}\|^2 =$

$= \frac{1}{2}\left\| \frac{1}{1}\cdot\begin{pmatrix}1\\1\end{pmatrix} - \frac{1}{1}\cdot\begin{pmatrix}2\\3\end{pmatrix}\right\|^2 = \frac{1}{2}\left\|\begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}2\\3\end{pmatrix}\right\|^2 =$

$= \frac{1}{2}\left((1-2)^2 + (1-3)^2\right) = \frac{1}{2}(1^2+4^2) = \frac{5}{2};$

• $\Delta(\{G\},\{H\}) = \frac{1\cdot 1}{1+1}\,\|\mu_{\{G\}} - \mu_{\{H\}}\|^2 =$

$= \frac{1}{2}\left\| \frac{1}{1}\cdot\begin{pmatrix}2\\3\end{pmatrix} - \frac{1}{1}\cdot\begin{pmatrix}3\\2\end{pmatrix}\right\|^2 = \frac{1}{2}\left\|\begin{pmatrix}2\\3\end{pmatrix} - \begin{pmatrix}3\\2\end{pmatrix}\right\|^2 =$

$= \frac{1}{2}\left((2-3)^2 + (3-2)^2\right) = \frac{1}{2}(1^2+1^2) = \frac{2}{2} = 1;$

• $\Delta(\{I\},\{J\}) = \Delta(\{G\},\{H\}) = 1;$

• $\Delta(\{I\},\{H\}) = \Delta(\{G\},\{J\}) = \Delta(\{E\},\{F\}) = 2;$

↳ distanța minimă este $\frac{1}{2} \Rightarrow \underline{C_1 = \{A, B\}};$

• $\mu_{C_1} = \frac{1}{2}\left(\begin{pmatrix}-4\\-2\end{pmatrix} + \begin{pmatrix}-3\\-2\end{pmatrix}\right) = \frac{1}{2}\cdot\begin{pmatrix}-7\\-4\end{pmatrix} = \begin{pmatrix}-3{,}5\\-2\end{pmatrix};$

• $\Delta(C_1, C) = \frac{2\cdot 1}{2+1}\,\|\mu_{C_1} - \mu_{\{C\}}\|^2 =$

$= \frac{2}{3}\left\|\begin{pmatrix}-3{,}5\\-2\end{pmatrix} - \begin{pmatrix}-2\\-2\end{pmatrix}\right\|^2 = \frac{2}{3}\left\|\begin{pmatrix}-1{,}5\\0\end{pmatrix}\right\|^2 =$

$= \frac{2}{3}(1{,}5^2 + 0^2) = \frac{2}{3}\cdot 2{,}25 = \frac{4{,}5}{3};$

↳ $\Delta(C_1,\{C\}) > \Delta(\{C\},\{D\})$ $\left(\frac{4{,}5}{3} = 1{,}5 > \frac{1}{2} = 0{,}5\right) \Rightarrow$

distanța minimă este $\frac{1}{2} \Rightarrow \underline{C_2 = \{C, D\}};$

• $\mu_{C_2} = \frac{1}{2}\left(\begin{pmatrix}-2\\-2\end{pmatrix} + \begin{pmatrix}-1\\-2\end{pmatrix}\right) = \frac{1}{2}\cdot\begin{pmatrix}-3\\-4\end{pmatrix} = \begin{pmatrix}-1{,}5\\-2\end{pmatrix};$

• $\Delta(C_1, C_2) = \frac{2\cdot 2}{2+2}\,\|\mu_{C_1} - \mu_{C_2}\|^2 = \frac{4}{4}\left\|\begin{pmatrix}-3{,}5\\-2\end{pmatrix} - \begin{pmatrix}-1{,}5\\-2\end{pmatrix}\right\|^2 =$

$= 1\cdot\left\|\begin{pmatrix}-2\\0\end{pmatrix}\right\|^2 = 2^2 + 0^2 = 4;$

• $\Delta(C_2, \{E\}) = \frac{2\cdot 1}{2+1}\,\|\mu_{C_2} - \mu_E\|^2 = \frac{2}{3}\left\|\begin{pmatrix}-1{,}5\\-2\end{pmatrix} - \begin{pmatrix}1\\-1\end{pmatrix}\right\|^2 =$

$= \frac{2}{3}\left\|\begin{pmatrix}-2{,}5\\-1\end{pmatrix}\right\|^2 = \frac{2}{3}(2{,}5^2 + 1^2) = \frac{2}{3}(6{,}25+1) = \frac{2}{3}\cdot 7{,}25 = 4{,}8(3);$

$\hookrightarrow 1 < 2 < \frac{5}{2} < 4 < 4,8(3) \Rightarrow$ distanța minimă este $1 \Rightarrow \underline{C_3 = \{G, H\}}$;

- $\mu_{C_3} = \frac{1}{2}\left(\begin{pmatrix}2\\3\end{pmatrix} + \begin{pmatrix}3\\2\end{pmatrix}\right) = \frac{1}{2}\cdot\begin{pmatrix}5\\5\end{pmatrix} = \begin{pmatrix}2,5\\2,5\end{pmatrix}$;
- $\Delta(C_3, \{I\}) = \frac{2\cdot 1}{2+1} \| \mu_{C_3} - \mu_{\{I\}} \|^2 = \frac{2}{3} \left\| \begin{pmatrix}2,5\\2,5\end{pmatrix} - \begin{pmatrix}3\\4\end{pmatrix} \right\|^2 =$
$= \frac{2}{3} \left\| \begin{pmatrix}-0,5\\-1,5\end{pmatrix} \right\|^2 = \frac{2}{3}(0,5^2 + 1,5^2) = \frac{2}{3}\cdot 2,5 = 1,(6)$;

$\hookrightarrow 1,(6) > 1 \Rightarrow$ distanța minimă este $1 \Rightarrow$ $\underline{C_4 = \{I, J\}}$;

- $\mu_{C_4} = \frac{1}{2}\left(\begin{pmatrix}3\\4\end{pmatrix} + \begin{pmatrix}4\\3\end{pmatrix}\right) = \frac{1}{2}\cdot\begin{pmatrix}7\\7\end{pmatrix} = \begin{pmatrix}3,5\\3,5\end{pmatrix}$;
- $\Delta(C_3, C_4) = \frac{2\cdot 2}{2+2} \| \mu_{C_3} - \mu_{C_4} \|^2 = \cancel{\frac{4}{4}}_1 \left\| \begin{pmatrix}2,5\\2,5\end{pmatrix} - \begin{pmatrix}3,5\\3,5\end{pmatrix} \right\|^2 =$
$= \left\| \begin{pmatrix}-1\\-1\end{pmatrix} \right\|^2 = 1^2 + 1^2 = 2$;

$\hookrightarrow$ distanța minimă este $2 \Rightarrow C_5 = \{E, F\}$ $(\Delta(C_3, C_4) = \Delta(\{E\}, \{F\})$, ordine alfabetică);

- $\mu_{C_5} = \frac{1}{2}\left(\begin{pmatrix}1\\-1\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix}\right) = \frac{1}{2}\cdot\begin{pmatrix}2\\0\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix}$;
- $\Delta(C_5, C_3) = \frac{2\cdot 2}{2+2} \| \mu_{C_5} - \mu_{C_3} \|^2 = \cancel{\frac{4}{4}}_1 \left\| \begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}2,5\\2,5\end{pmatrix} \right\|^2 =$
$= \left\| \begin{pmatrix}-1,5\\-2,5\end{pmatrix} \right\|^2 = 1,5^2 + 2,5^2 = 8,5$;

$\hookrightarrow 2 < 8,5 \Rightarrow$ ~~C6~~ distanța minimă este $2 \Rightarrow$ $\underline{C_6 = \{G, H, I, J\}}$;

- $\Delta(C_5, C_2) = \frac{2\cdot 2}{2+2} \| \mu_{C_5} - \mu_{C_2} \|^2 = \cancel{\frac{4}{4}}_1 \left\| \begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}-1,5\\-2\end{pmatrix} \right\|^2 =$
$= \left\| \begin{pmatrix}2,5\\2\end{pmatrix} \right\|^2 = 2,5^2 + 2^2 = 10,25$;

• $\mu_{C_6} = \frac{1}{2}\left(\begin{pmatrix}3,5\\3,5\end{pmatrix}+\begin{pmatrix}2,5\\2,5\end{pmatrix}\right) = \frac{1}{2}\begin{pmatrix}6\\6\end{pmatrix} = \begin{pmatrix}3\\3\end{pmatrix}$;

• $\Delta(C_5,C_6) = \frac{2\cdot 4}{2+4}\|\mu_{C_5}-\mu_{C_6}\|^2 = \frac{8}{6}\left\|\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}3\\3\end{pmatrix}\right\|^2 =$

$= \frac{4}{3}\left\|\begin{pmatrix}-2\\-3\end{pmatrix}\right\|^2 = \frac{4}{3}(2^2+3^2) = \frac{4}{3}\cdot 13 = 17,(3)$;

↳ $4 < 17,(3) \Rightarrow$ distanța minimă este $4 \Rightarrow$

$\underline{C_7 = \{A,B,C,D\}}$;

• $\mu_{C_7} = \frac{1}{2}\left(\begin{pmatrix}-3,5\\-2\end{pmatrix}+\begin{pmatrix}-1,5\\-2\end{pmatrix}\right) = \frac{1}{2}\begin{pmatrix}-5\\-4\end{pmatrix} = \begin{pmatrix}-2,5\\-2\end{pmatrix}$;

• $\Delta(C_7,C_5) = \frac{2\cdot 4}{2+4}\|\mu_{C_7}-\mu_{C_5}\|^2 = \frac{8}{6}\left\|\begin{pmatrix}-2,5\\-2\end{pmatrix}-\begin{pmatrix}1\\0\end{pmatrix}\right\|^2 =$

$= \frac{4}{3}\left\|\begin{pmatrix}-3,5\\-2\end{pmatrix}\right\|^2 = \frac{4}{3}(3,5^2+2^2) = \frac{4}{3}\cdot 16,25 = 21,(6)$;

↳ $17,(3) < 21,(6) \Rightarrow$ distanța minimă este $17,(3) \Rightarrow \underline{C_8 = \{E,F,G,H,I,J\}}$;

↳ $\underline{C_9 = \{A,B,C,D,E,F,G,H,I,J\}}$;

Ex. 8

$P_1(1.90, 0.97)$, $P_2(1.76, 0.84)$, $P_3(2.32, 1.63)$, $P_4(2.31, 2.09)$;
$P_5(1.14, 2.11)$, $P_6(5.02, 3.02)$, $P_7(5.74, 3.84)$, $P_8(2.25, 3.47)$;
$P_9(4.71, 3.60)$, $P_{10}(3.17, 4.96)$;
K-means; 2 clustere; $\mu_1^{(0)} = P_1$, $\mu_2^{(0)} = P_{10}$;

a) prima iterație;

• $d(P_2,P_1) = \sqrt{(1,76-1,90)^2+(0,84-0,97)^2} = \sqrt{0,14^2+0,13^2} = \sqrt{0,0365} = 0,19$;

• $d(P_2,P_{10}) = \sqrt{(1,76-3,17)^2+(0,84-4,96)^2} = \sqrt{1,41^2+4,12^2} = \sqrt{18,96} = 4,35$;

↳ $d(P_2,P_1) < d(P_2,P_{10}) \Rightarrow C_1 = \{P_1, P_2\}$;

- $d(P_3, P_{10}) = \sqrt{(2,32-3,17)^2 + (1,63-4,96)^2} =$
$= \sqrt{0,85^2 + 3,33^2} = \sqrt{11,81} = 3,44;$
- $d(P_3, P_1) = \sqrt{(2,32-1,90)^2 + (1,63-0,97)^2} =$
$= \sqrt{0,42^2 + 0,66^2} = \sqrt{0,612} = 0,78;$
$\hookrightarrow d(P_3, P_1) < d(P_3, P_{10}) \Rightarrow C_1 = \{P_1, P_2, P_3\};$
- $d(P_4, P_1) = \sqrt{(2,31-1,90)^2 + (2,09-0,97)^2} =$
$= \sqrt{0,41^2 + 1,12^2} = \sqrt{1,4225} = 1,19;$
- $d(P_4, P_{10}) = \sqrt{(2,31-3,17)^2 + (2,09-4,96)^2} =$
$= \sqrt{0,86^2 + 2,87^2} = \sqrt{8,9765} = 2,99;$
$\hookrightarrow d(P_4, P_1) < d(P_4, P_{10}) \Rightarrow C_1 = \{P_1, P_2, P_3, P_4\};$
- $d(P_5, P_1) = \sqrt{(1,14-1,90)^2 + (2,11-0,97)^2} =$
$= \sqrt{0,76^2 + 1,14^2} = \sqrt{1,8772} = 1,37;$
- $d(P_5, P_{10}) = \sqrt{(1,14-3,17)^2 + (2,11-4,96)^2} =$
$= \sqrt{2,03^2 + 2,85^2} = \sqrt{12,2434} = 3,49;$
$\hookrightarrow d(P_5, P_1) < d(P_5, P_{10}) \Rightarrow C_1 = C_1 \cup \{P_5\};$
- $d(P_6, P_1) = \sqrt{(5,02-1,90)^2 + (3,02-0,97)^2} =$
$= \sqrt{3,12^2 + 2,05^2} = \sqrt{13,9369} = 3,73;$
- $d(P_6, P_{10}) = \sqrt{(5,02-3,17)^2 + (3,02-4,96)^2} =$
$= \sqrt{1,85^2 + 1,94^2} = 2,68;$
$\hookrightarrow d(P_6, P_{10}) < d(P_6, P_1) \Rightarrow C_2 = C_2 \cup \{P_6\};$
- $d(P_7, P_1) = \sqrt{(5,74-1,90)^2 + (3,84-0,97)^2} =$
$= \sqrt{3,84^2 + 2,87^2} = 4,79;$
- $d(P_7, P_{10}) = \sqrt{(5,74-3,17)^2 + (3,84-4,96)^2} =$
$= \sqrt{2,57^2 + 1,12^2} = 2,80;$

$\hookrightarrow d(P_7, P_{10}) < d(P_7, P_1) \Rightarrow C_2 = C_2 \cup \{P_7\};$

- $d(P_8, P_1) = \sqrt{(2,25-1,90)^2 + (3,47-0,97)^2} =$
$= \sqrt{0,35^2 + 2,5^2} = 2,52;$

- $d(P_8, P_{10}) = \sqrt{(2,25-3,17)^2 + (3,47-4,96)^2} =$
$= \sqrt{0,92^2 + 1,49^2} = 1,75;$

$\hookrightarrow d(P_8, P_{10}) < d(P_8, P_{10}) \Rightarrow C_2 = C_2 \cup \{P_8\};$

- $d(P_9, P_1) = \sqrt{(4,71-1,90)^2 + (3,60-0,97)^2} =$
$= \sqrt{2,81^2 + 2,63^2} = 3,85;$

- $d(P_9, P_{10}) = \sqrt{(4,71-3,17)^2 + (3,60-4,96)^2} =$
$= \sqrt{1,54^2 + 1,36^2} = 2,05;$

$\hookrightarrow d(P_9, P_{10}) < d(P_9, P_1) \Rightarrow C_2 = C_2 \cup \{P_9\};$

$\hookrightarrow \begin{cases} C_1^{(0)} = \{P_1, P_2, P_3, P_4, P_5\}; \\ C_2^{(0)} = \{P_6, P_7, P_8, P_9, P_{10}\}; \end{cases}$

compozitia clusterelor când K-means converge;

b)

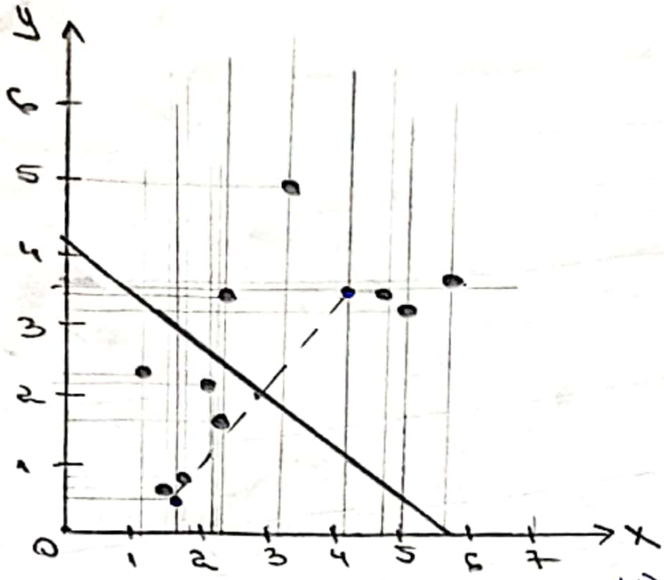
- $x_{\mu_1} = \dfrac{x_{P_1} + x_{P_2} + x_{P_3} + x_{P_4} + x_{P_5}}{5} =$
$= \dfrac{1,90 + 1,76 + 2,32 + 2,31 + 1,14}{5} = 1,886;$

- $y_{\mu_1} = \dfrac{y_{P_1} + y_{P_2} + y_{P_3} + y_{P_4} + y_{P_5}}{5} =$
$= \dfrac{0,97 + 0,84 + 1,63 + 2,09 + 2,11}{5} = 1,528;$

- $x_{\mu_2} = \dfrac{5,02 + 5,74 + 2,25 + 4,71 + 3,17}{5} = 4,178;$

- $y_{\mu_2} = \dfrac{3,02 + 3,84 + 3,47 + 3,60 + 4,96}{5} = 3,778;$

$\hookrightarrow \mu_1^{(1)}(1.886, 1.528); \mu_2^{(1)}(4.178, 3.778);$



↳ $C_1^{(1)} = \{P_1, P_2, P_3, P_4, P_5\}$ și $C_2^{(1)} = \{P_6, P_7, P_8, P_9, P_{10}\}$

ca la iterația anterioară $\Leftrightarrow$ algoritmul converge;