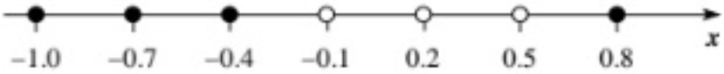
Pg: 597

4.2.2 Algoritmul AdaBoost

59. **(Algoritmul AdaBoost: întrebări în legătură cu aplicarea algoritmului pe un set de date din $\mathbb{R}$)**

• ○ ***CMU, 2011 fall, T. Mitchell, A. Singh, HW6, pr. 3.2-8***

Considerăm setul de date de antrenament din figura următoare. Simbolurile • şi ○ desemnează etichete pozitive şi respectiv etichete negative. Vom folosi algoritmul AdaBoost cu compaşi de decizie în rolul de ipoteze „slabe“.



a. Determinaţi separatorul decizional corespunzător primei ipoteze „slabe“, h_1. Desenaţi-l pe figura de mai sus şi indicaţi [eventual printr-o mică săgeată perpendiculară pe acest separator] care este zona clasificată cu $+$.

b. Calculaţi ε_1 şi α_1. Cât este acurateţea obţinută de AdaBoost la antrenare dacă oprim acum algoritmul?

c. Cât va fi valoarea noilor probabilităţi / ponderi $D_2(i)$ pentru fiecare dintre cele şapte exemple de antrenament?

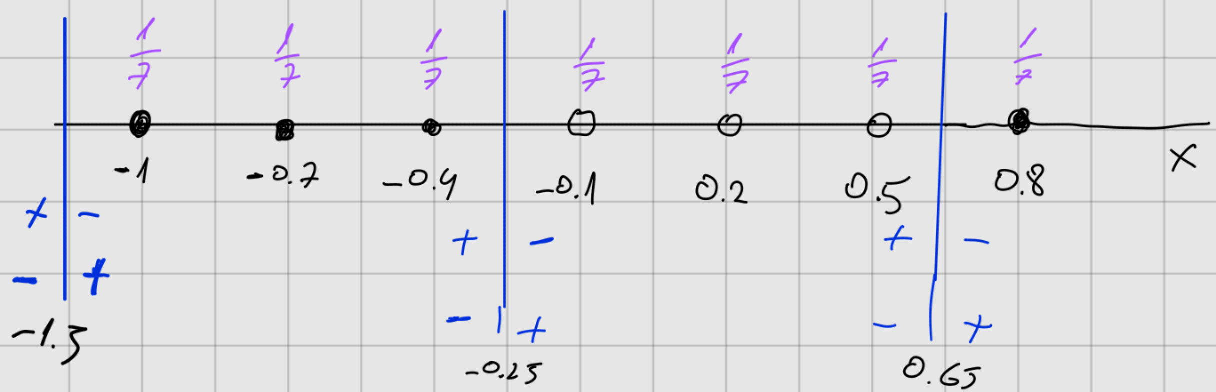
d. Identificaţi în mod riguros separatorul decizional corespunzător celei de-a doua ipoteze „slabe“, h_2 şi apoi includeţi-l în desen. Indicaţi iarăşi zona de decizie corespunzătoare clasei $+$.

e. Care sunt exemplele de antrenament cărora le va fi asignată cea mai mică pondere / probabilitate după ce algoritmul AdaBoost va fi terminat cea de-a doua sa iteraţie?

f. Se îmbunătăţeşte oare acurateţea la antrenare obţinută de AdaBoost la a doua iteraţie în raport cu cea obţinută la prima iteraţie?

a, b, c)

Setăm ponderile:



Încercăm separatorii:

n	-0.25	0.65	-1.5
$err_{D_1}(X<n)$	$1/7$	$3\cdot\frac{1}{7}+\frac{1}{7}=\frac{4}{7}$	$\frac{4}{7}$
$err_{D_1}(X\geq n)$	$\frac{6}{7}$	$\frac{3}{7}$	$\frac{3}{7}$

alegem minimul: $\varepsilon_1 = \frac{1}{7}$

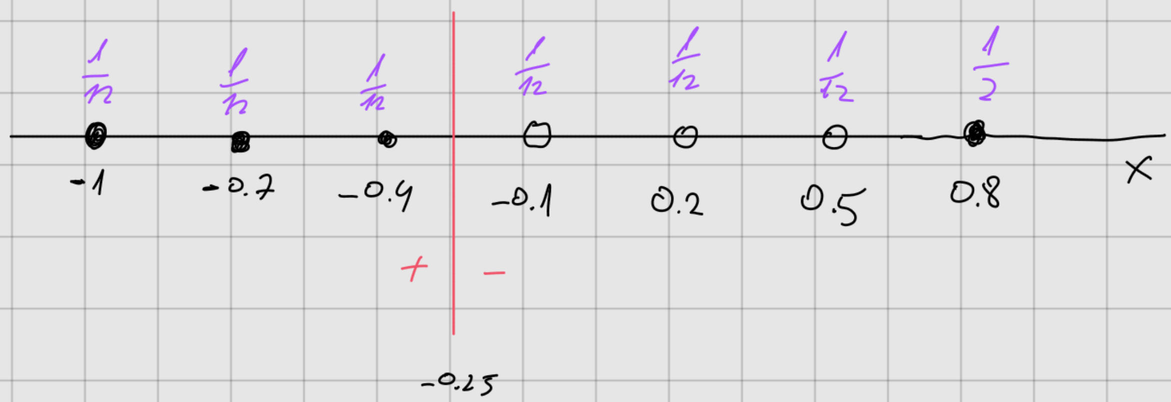
$h_1 = (X < -0.25) = sign(-0.25 - X)$

$$\gamma_1 = \frac{1}{2} - \varepsilon_1 = \frac{1}{2} - \frac{1}{7} = \frac{5}{7}$$

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

Actualizăm ponderile:

$$\begin{cases} \dfrac{D_t(i)}{2 \cdot \varepsilon_t}, & i \in \{0.8\} \quad (\text{greșit}) \\ \dfrac{D_t(i)}{2 \cdot (1-\varepsilon_T)}, & i \notin \{0.8\} \quad (\text{corect}) \end{cases}$$



$-1, -0.7, -0.4$: $\dfrac{\frac{1}{\cancel{7}}}{2\frac{6}{\cancel{7}}} = \dfrac{1}{12}$

$-0.1, 0.2, 0.5$: $\dfrac{\frac{1}{\cancel{7}}}{2 \cdot \frac{6}{\cancel{7}}} = \dfrac{1}{12}$

0.8: $\dfrac{\frac{1}{\cancel{7}}}{2 \cdot \frac{1}{\cancel{7}}} = \dfrac{1}{2}$

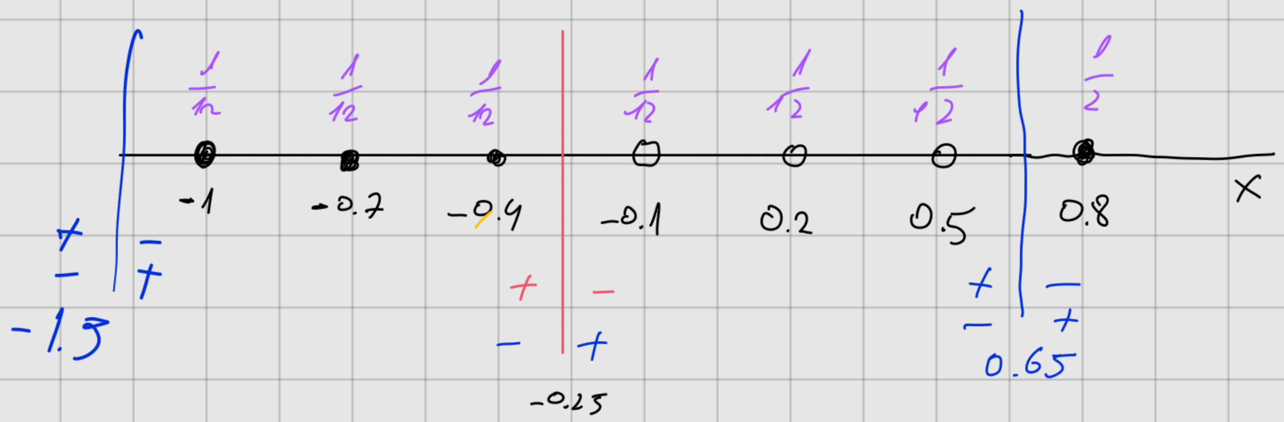
Dacă am opri alg., acuratețea ar fi:

$$H_1(X) = sign(\alpha_1 \cdot h_1(X)) = sign\left(\frac{1}{7} \cdot sign(-0.25 - X)\right)$$

	target	predictio $H_3(x)$
$x_1 = -1$	$+1$	$+$
$x_2 = -0.7$	$+1$	$+$
$x_3 = -0,4$	-1	$+$
$x_4 = -0.1$	-1	$-$
$x_5 = 0.2$	-1	$-$
$x_6 = 0.5$	-1	$-$
$x_7 = 0.8$	$+1$	$-$

err medio $= \frac{1}{7}$

d)



η	-0.25	0.65	-1.5
err D_2 $(x<\eta)$	$\frac{1}{2}$	$\frac{1}{2} + 3 \cdot \frac{1}{12} = \frac{3}{4}$	$\frac{3}{12} + \frac{1}{2} = \frac{9}{12} = \frac{3}{4}$
err D_2 $(x \geq \eta)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

$$\varepsilon_2 = \frac{1}{4}$$

$$h_2 = (x \geq 0.65) = (x - 0.65) \geq 0$$

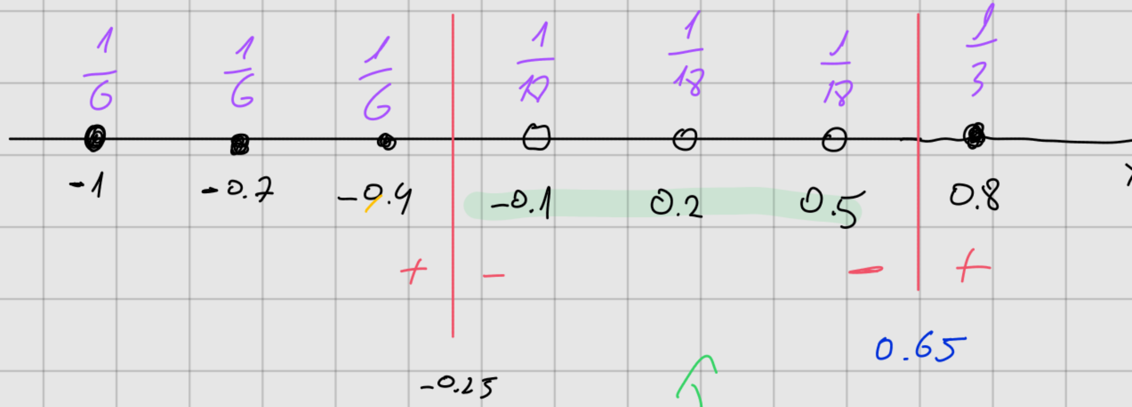
$$\alpha_2 = \frac{1}{2} \ln \frac{\frac{3}{\cancel{4}}}{\frac{1}{\cancel{4}}} = \ln \sqrt{3}$$

$$\begin{cases} \dfrac{D_2(i)}{2 \cdot \varepsilon_2}, & i \in \{-1, -0.7, -0.4\} \text{ (greșite)} \\ \dfrac{D_2(i)}{2 \cdot (1-\varepsilon_2)} & i \notin \{ \dots \} \text{ (corecte)} \end{cases}$$

i	-1	-0.7	-0.4	-0.1	0.2	0.5	0.8
$D_3(i)$							

$$\frac{\frac{1}{\cancel{12}\,3}}{2 \cdot \frac{1}{\cancel{4}}} = \frac{1}{6} \qquad \frac{\frac{1}{\cancel{12}\,3}}{2 \cdot \frac{3}{\cancel{4}}} = \frac{1}{18} \qquad \frac{\frac{1}{2}}{2 \cdot \frac{3}{\cancel{4}\,2}} = \frac{1}{3}$$

greșite



e) exemplele vor fi cele asignate corect, cum ar fi: $i = -0.1,\ 0.2,\ 0.5$ cu probabilitatea $\frac{1}{18}$

$$\underset{0.89}{\ln\sqrt{6}} > \underset{0.5..}{\ln\sqrt{3}}$$

f) ~~Dec, pe îmbunătățește~~: $H_2(x_i) = \text{sign}\left(\alpha_1 \,\text{sign}(-0.25 - x) + \alpha_2 \,\text{sign}(x - 0.65)\right)$

	t	α_t	D_1^i	D_2^i	D_3^i	D_4^i	D_5^i	D_6^i	D_7^i
$h_1 \to$	1	$\ln\sqrt{6}$	+	+	+	−	−	−	−
$h_2 \to$	2	$\ln\sqrt{3}$	−	−	−	−	−	−	+
		$H_T(x_i)$	+	+	+	−	−	−	−

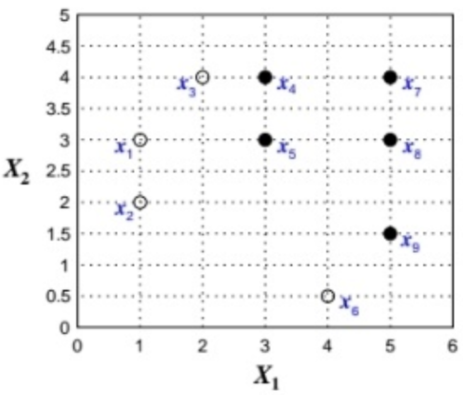
Aparent acuratețea tot nu s-a îmbunătățit este la fel.

60. **(AdaBoost: aplicare pe un set de date din $\mathbb{R}^2$)**

adaptare facută de Liviu Ciortuz, după
• ◦ *MIT, 2004 fall, Tommi Jaakkola, final, pr. 1*

Fie setul de exemple de antrenament (instanţe etichetate) din figura de mai jos. Simbolurile • şi ◦ desemnează etichete pozitive şi respectiv etichete negative. Pentru a rezolva această problemă de clasificare, vă cerem să aplicaţi algoritmul AdaBoost folosind drept ipoteze „slabe“ compaşi de decizie (engl., decision stump). La fiecare iteraţie de boosting veţi selecta acel compas de decizie care minimizează eroarea ponderată la antrenare (engl., weighted training error). În cazul în care există mai mulţi compaşi de decizie care au [o

aceeaşi!] cea mai bună eroare ponderată la antrenare, veţi putea alege unul dintre ei în mod arbitrar.



a. Pe figura dată, desenaţi primul compas de decizie şi etichetaţi-l cu h_1, indicând de asemenea cu $+/-$ cele două zone de decizie pe care le determină acest compas.

b. Pe aceeaşi figură, scrieţi în apropierea fiecărei instanţe probabilitatea asignată ei după prima iteraţie executată de algoritmul AdaBoost. De asemenea, încercuiţi instanţele care au probabilitatea cea mai mare. Justificaţi răspunsul în mod riguros.

c. Cât este eroarea ponderată la antrenare produsă de primul compas de decizie după prima iteraţie, adică după ce probabilităţile asociate instanţelor au fost recalculate? Justificaţi.

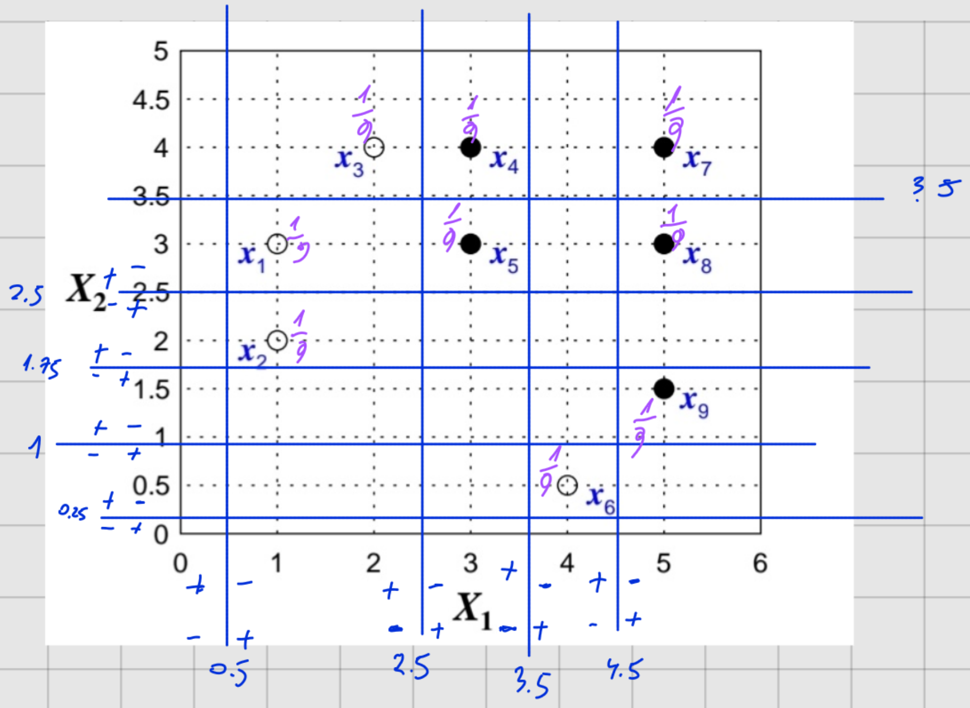
d. Desenaţi pe figura dată cel de-al doilea compas de decizie şi etichetaţi-l cu h_2, indicând de asemenea cu $+/-$ cele două zone de decizie pe care le determină acest compas.

e. Există oare instanţe de antrenament care sunt clasificate eronat de către H_2, *ipoteza combinată* produsă de AdaBoost după două iteraţii? Justificaţi răspunsul în mod riguros.

f. Dacă răspunsul pe care l-aţi dat la punctul e este pozitiv, ce se poate spune dacă se execută încă o iteraţie (adică, a treia)? Desenaţi pe figura dată cel de-al treilea compas de decizie şi etichetaţi-l cu h_3, indicând de asemenea cu $+/-$ cele două zone de decizie pe care le determină acest compas. Dacă eroarea la antrenare este acum 0, trasaţi separatorul decizional determinat de algoritmul AdaBoost.

Care ar fi rezultatul dacă AdaBoost ar alege ca a treia ipoteză „slabă“ un alt (cel mai bun!) compas de decizie (h_3') în locul lui h_3?

a)



x_2:	θ	0.25	1	1.75	2.5	3.5
err $D_1(x_2<\theta)$		$\frac{5}{9}$	6/9	5/9	6/9	5/9
err $D_1(x_2\geq\theta)$		$\frac{4}{9}$	3/9	4/9	3/9	4/9

x_1:	θ	0.5	2.5	3.5	4.5
err $D_1(x_1<\theta)$		5/9	8/9	6/9	7/9
err $D_1(x_1\geq\theta)$		4/9	1/9	3/9	2/9

$$\varepsilon_1 = \frac{1}{9} \qquad h_1 = (x_1 \geq 2.5) = (x_1 - 2.5 \geq 0)$$

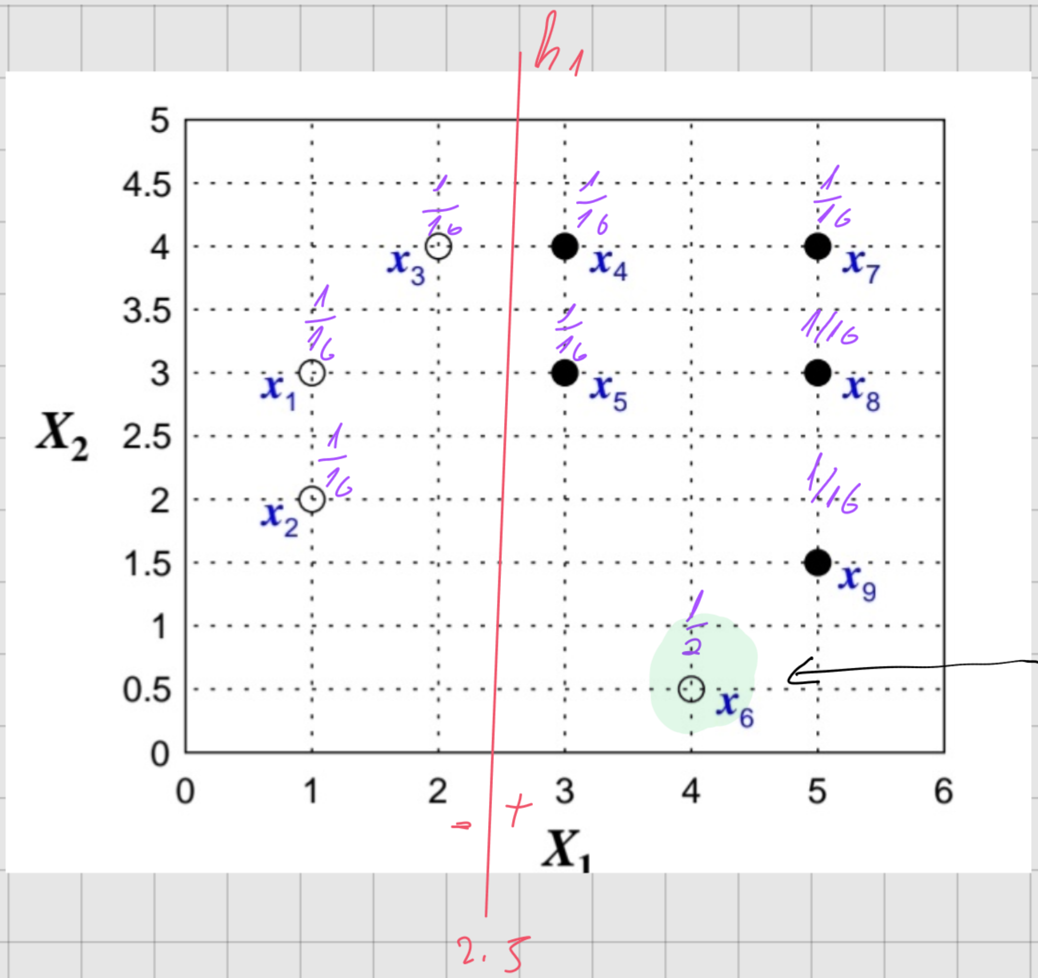
$$\alpha_1 = \frac{1}{2} \ln\left(\frac{\frac{8}{\cancel{9}}}{\frac{1}{\cancel{9}}}\right) = \ln\sqrt{8}$$

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
$D_2(x_i)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{2}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

$$\begin{cases} \dfrac{D_1(i)}{2\cdot\varepsilon_1}, & i \in \{6\} \\[2ex] \dfrac{D_1(i)}{2\cdot(1-\varepsilon_1)}, & i \notin \{6\} \end{cases}$$

$$x_6: \frac{\frac{1}{9}}{2\cdot\frac{1}{9}} = \frac{1}{2}$$

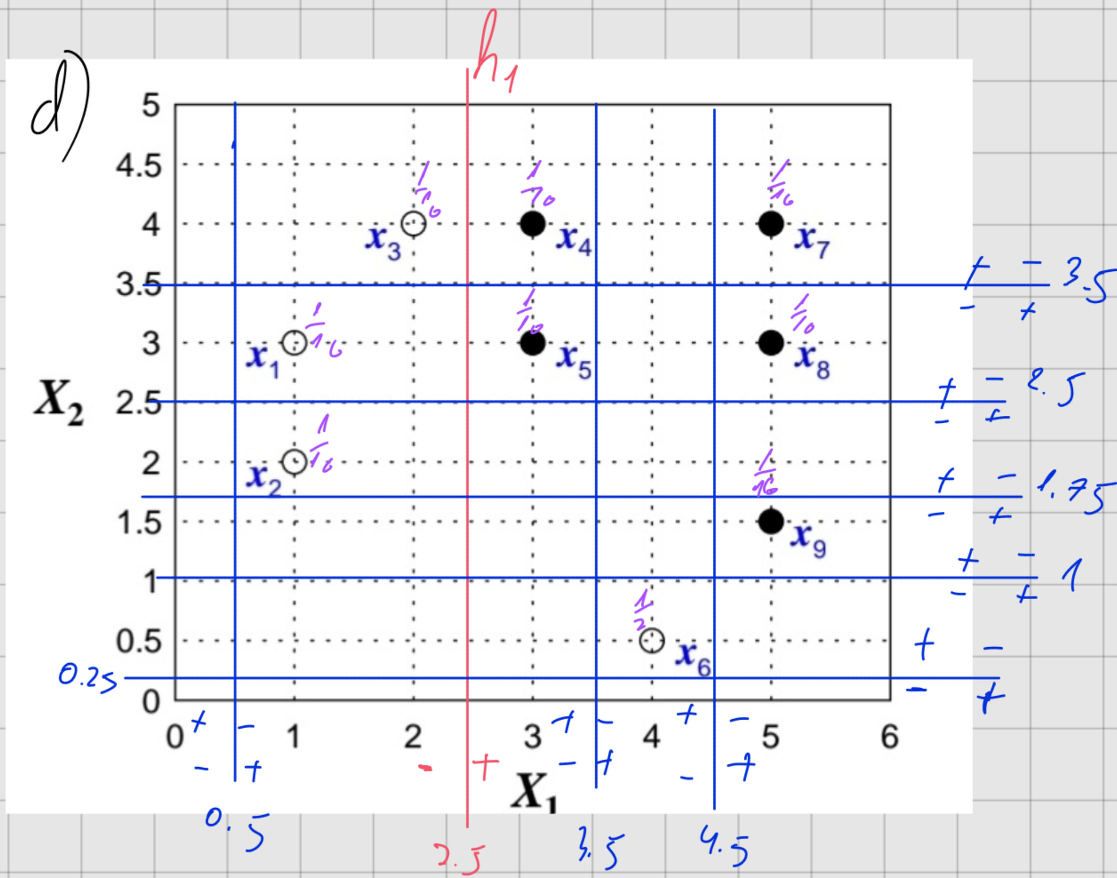
$$x_{\{1,2,\dots 9\}-\{6\}}: \frac{\frac{1}{\cancel{9}}}{2\cdot\frac{8}{\cancel{9}}} = \frac{1}{16}$$



inst. clas. greşit.

c) err. pond. la antrenare este de $\frac{1}{9}$, deoarece doar x_6 a fost clasificat greşit.

d)



X_1:

s	0.5	2.5	3.5	4.5
$err_{D_2}(X_1 < s)$	5/16	8/16	6/16	$\frac{6}{16}+\frac{1}{2}=\frac{14}{16}$
$err_{D_2}(X_1 \geq s)$	11/16	8/16	10/16	2/16

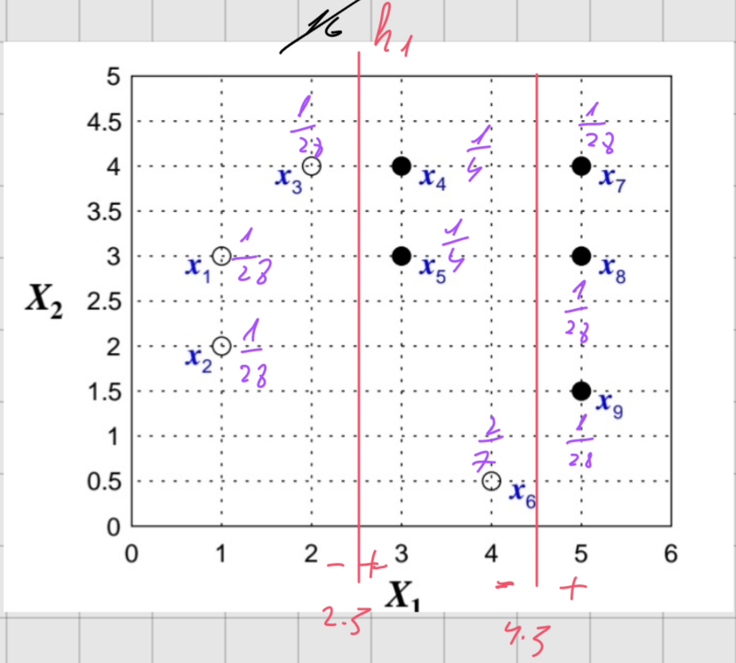
X_2:

s	0.25	1	1.75	2.5	3.5
$err_{D_2}(X_2 < s)$	5/16	$\frac{5}{16}+\frac{1}{2}=\frac{13}{16}$	$\frac{12}{16}$	13/16	12/16
$err_{D_2}(X_2 \geq s)$	11/16	$\frac{3}{16}$	4/16	3/16	4/16

$\varepsilon_2 = \frac{2}{16}$ $\qquad h_2 = (X_1 \geq 4.5) = (X_1 - 4.5 \geq 0)$ $\qquad \alpha_2 = \frac{1}{2}\ln\left(\frac{\frac{14}{\cancel{16}}}{2\cdot\frac{2}{\cancel{16}}}\right) = \ln\sqrt{\frac{7}{2}}$

$$D_3: \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 \\ \frac{1}{28} & \frac{1}{28} & \frac{1}{28} & \frac{1}{4} & \frac{1}{4} & \frac{2}{7} & \frac{1}{28} & \frac{1}{28} & \frac{1}{28} \end{pmatrix}$$

X_4, X_5: $\frac{\frac{1}{\cancel{16}}}{2\cdot\frac{2}{\cancel{16}}} = \frac{1}{4}$ $\qquad X_6$: $\frac{\frac{1}{\cancel{2}}}{\frac{\cancel{2}\cdot\cancel{14}\,7}{\cancel{16}\,\cancel{8}\,\cancel{4}\,2}} = \frac{2}{7}$ $\qquad$ restul: $\frac{\frac{1}{\cancel{16}}}{2\cdot\frac{14}{\cancel{16}}} = \frac{1}{28}$



$$H_T(x) = \text{sign}\left(\alpha_1 \cdot h_1(x) + \alpha_2 \cdot h_2(x)\right)$$

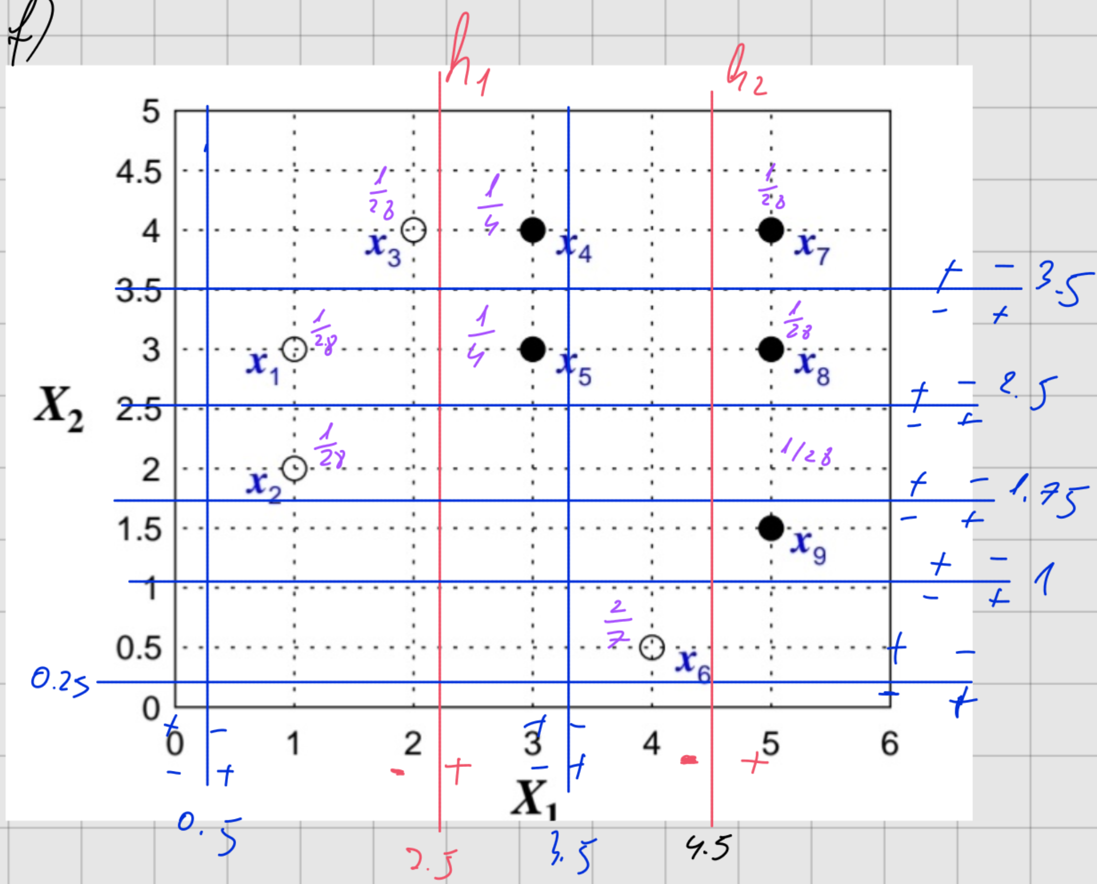
$$H_2(X) = \text{sign}\Big(\underbrace{\ln\sqrt{8}}_{1.039} \cdot \text{sign}(X_1 - 2.5) + \underbrace{\ln\sqrt{\tfrac{7}{2}}}_{0.626}\, \text{sign}(X_1 - 4.5)\Big)$$

$1.039 > 0.626$

	t	α_t	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
$h_1 \to$	1	$\ln\sqrt{8} = 1.039$	−	−	−	+	+	+	+	+	+
$h_2 \to$	2	$\ln\sqrt{\frac{7}{2}} = 0.626$	−	−	−	−	−	−	+	+	+
		$H_T(x_i)$	−	−	−	+	+	+	+	+	+

Obs. că $H_2(X)$ clasifică greșit instanța x_6.

f)



X_1:

θ	0.5	2.5	3.5	4.5
$err_{D_3}(X_1<\theta)$	$\frac{1}{2}+\frac{3}{28}=\frac{17}{28}$	$\frac{6}{28}+\frac{1}{2}=\frac{20}{28}$	$\frac{6}{28}$	$\frac{1}{2}$
$err_{D_3}(X_1\geq\theta)$	$\frac{11}{28}$	$\frac{8}{28}$	$\frac{22}{28}$	$\frac{1}{2}$

X_2:

θ	0.25	1	1.75	2.5	3.5
$err_{D_3}(X_2<\theta)$	$\frac{17}{28}$	$25/28$	$\frac{24}{28}$	$23/28$	$\frac{1}{4}+\frac{1}{28}+\frac{2}{28}+\frac{2}{7}=\frac{3}{28}+\frac{7}{28}+\frac{8}{28}=\frac{18}{28}$
$err_{D_3}(X_2\geq\theta)$	$\frac{11}{28}$	$\frac{3}{28}$	$\frac{4}{28}$	$3/28$	$10/28$

$$\varepsilon_3=\frac{3}{28} \qquad h_3=(X_2\geq 1)=(X_2-1)$$

$$\alpha_3=\frac{1}{2}\ln\left(\frac{\frac{25}{\cancel{28}}}{\frac{3}{\cancel{28}}}\right)=\ln\sqrt{\frac{25}{3}}$$

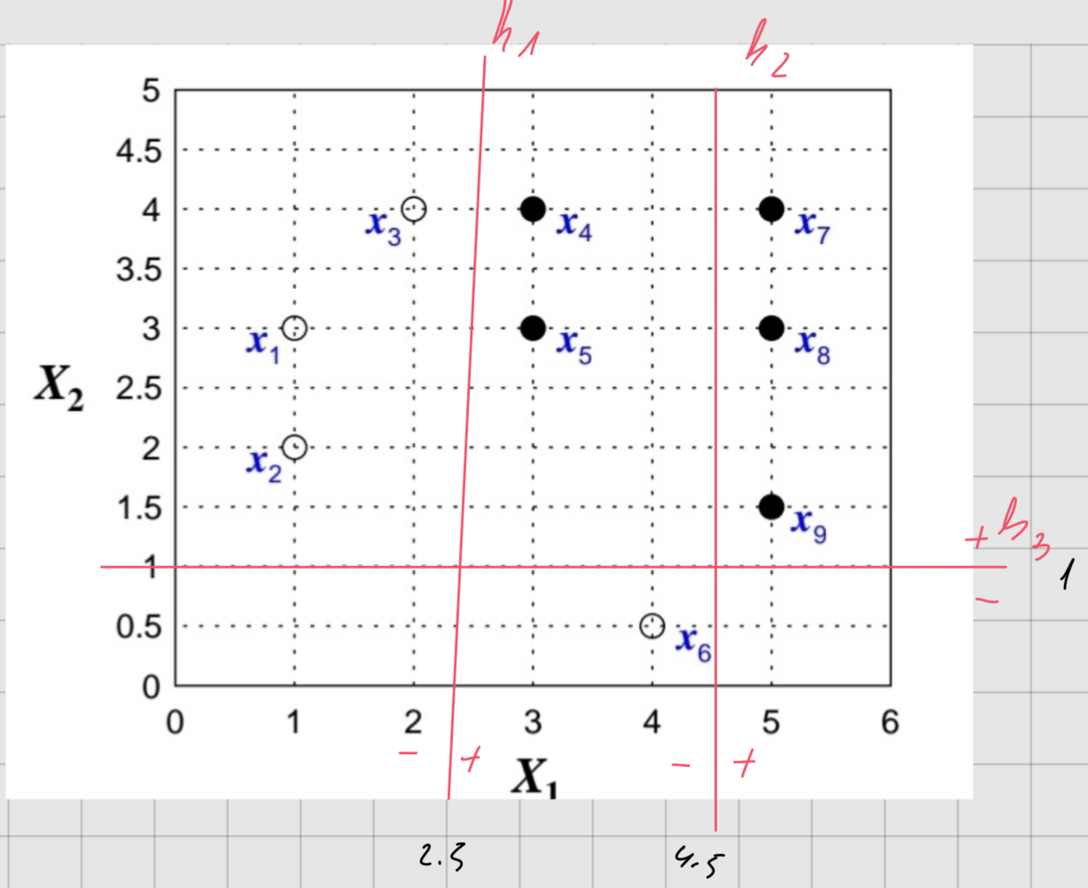
$$D_4(i): \begin{cases} \dfrac{D_3(i)}{2\cdot\varepsilon_3} : & i\in\{x_1,x_2,x_3\} \\[2ex] \dfrac{D_3(i)}{2\cdot(1-\varepsilon_3)} : & i\notin\{x_1,x_2,x_3\} \end{cases}$$

x_1,x_2,x_3: $\frac{1}{6}$

x_4,x_5: $\dfrac{\frac{1}{4}}{\cancel{2}\cdot\frac{25}{\cancel{28}\,14}}=\frac{7}{50}$

x_7,x_8,x_9: $\frac{1}{50}$

x_6: $\dfrac{\frac{2}{7}}{\cancel{2}\cdot\frac{25}{\cancel{28}\,4}}=\frac{4}{25}$



t	α_t	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
$h_1 \to$ 1	$\ln\sqrt{8} = 1{,}039$	−	−	−	+	+	+	+	+	+
$h_2 \to$ 2	$\ln\sqrt{\frac{7}{2}} = 0{,}626$	−	−	−	−	−	−	+	+	+
$h_3 \to$ 3	$\ln\sqrt{\frac{25}{3}} = 1.060$	+	+	+	+	+	−	+	+	+
	$H_T(x_i)$	−	−	−	+	+	−	+	+	+

err. la antr. este acum 0.

Dacă ada boost ar fi ales celălalt compas: $h_2 = (X_2 - 2.5)$

x_9, x_1, x_3 ar fi fost greşite, însă $H_3(x)$-final ar fi fost la fel. Deoarece ponderile majoritare pt h_1, h_2 nt x_1, x_3, x_9 clasifică corect semnul.

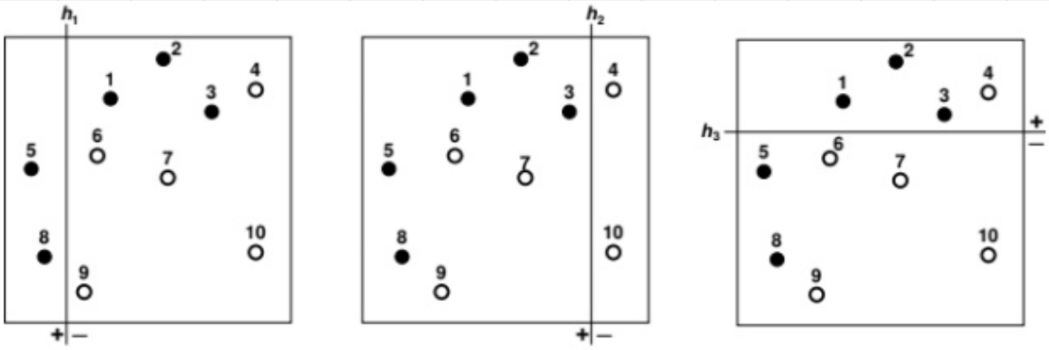
61. (Algoritmul AdaBoost: aplicare pe un set de date din $\mathbb{R}^2$)

Liviu Ciortuz, 2021, pornind de la un slide al lui Virgil Pavlu, Northeastern University, USA

Se consideră că aplicăm algoritmul AdaBoost pe dataset-ul din figura alăturată. (Pentru uşurinţa exprimării la calcule, am notat pe figură indicii instanţelor de antrenament, în imediata apropiere a acestora.)

Folosim convenţia noastră obişnuită de notare: simbolul • desemnează instanţe pozitive, iar simbolul ∘ instanţe negative.

La primele trei iteraţii ale algoritmului au fost selectaţi compaşii de decizie h_1, h_2 şi h_3 (în această ordine), aşa cum se indică în figurile de mai jos.



Obiectivul **acestei probleme este să determinăm dacă la sfârşitul celor trei iteraţii algoritmul AdaBoost reuşeşte să clasifice perfect toate instanţele de antrenament.**

a. Calculaţi

– distribuţiile probabiliste corespunzătoare celor 3 iteraţii, adică $D_1(x_i)$, $D_2(x_i)$ şi $D_3(x_i)$, pentru $i = 1, \ldots, 10$;

– pentru fiecare dintre cele 3 iteraţii ($t = 1, 2, 3$): eroarea ponderată la antrenare (ε_t) produsă de compasul de decizie h_t, precum şi ponderea (α_t) asociată ipotezei / compasului de decizie h_t.

Veţi completa tabelele următoare şi veţi indica succint(!) modul în care aţi procedat pentru a ajunge la rezultatele respective.

i	1	2	3	4	5	6	7	8	9	10
$D_1(x_i)$	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10
$D_2(x_i)$	1/6	1/6	1/6	1/14	1/14	1/14	1/14	1/14	1/14	1/14
$D_3(x_i)$	7/66	7/66	7/66	1/22	1/22	1/6	1/6	1/22	1/6	1/22

$$\varepsilon_1 = \frac{3}{10} \qquad \alpha_1 = \frac{1}{2} \ln \frac{\frac{7}{\cancel{10}}}{\frac{3}{\cancel{10}}} = \ln \sqrt{\frac{7}{3}}$$

$$D_2(x_i): \begin{cases} \dfrac{D_1\, i}{2 \cdot \varepsilon_1}, & i \in \{1,2,3\} \\[2ex] \dfrac{D_1\, i}{2 \cdot (1-\varepsilon_1)}, & i \notin \{1,2,3\} \end{cases} \qquad 1,2,3: \ \frac{\frac{1}{10}}{2 \cdot \frac{3}{10}} = \frac{1}{6} \qquad 4,5,6,\ldots 10: \ \frac{\frac{1}{\cancel{10}}}{2 \cdot \frac{7}{\cancel{10}}} = \frac{1}{14}$$

$$\varepsilon_2 = \frac{3}{14} \qquad \alpha_2 = \frac{1}{2}\ln\frac{\frac{11}{\cancel{14}}}{\frac{3}{\cancel{14}}} = \ln\sqrt{\frac{11}{3}}$$

$$D_3(x_i): \begin{cases} \dfrac{D_2\, i}{2\cdot\varepsilon_2} & : i \in \{6,7,9\} \\[2ex] \dfrac{D_2\, i}{2\cdot(1-\varepsilon_2)} & : i \notin \{6,7,9\} \end{cases}$$

$$6,7,9: \quad \frac{\frac{1}{\cancel{14}}}{2\cdot\frac{3}{\cancel{14}}} = \frac{1}{6} \qquad 4,5,8,10: \quad \frac{\frac{1}{\cancel{14}}}{2\cdot\frac{11}{\cancel{14}}} = \frac{1}{22}$$

$$1,2,3: \quad \frac{\frac{1}{\cancel{6}\,3}}{2\cdot\frac{11}{\cancel{14}\,7}} = \frac{7}{66}$$

$$\varepsilon_3: \frac{3}{22} \qquad \alpha_3 = \frac{1}{2}\ln\frac{\frac{19}{\cancel{22}}}{\frac{3}{\cancel{22}}} = \ln\sqrt{\frac{19}{3}}$$

t	1	2	3
ε_t	$3/10$	$3/14$	$3/22$
α_t	$\ln\sqrt{\frac{7}{3}}$	$\ln\sqrt{\frac{11}{3}}$	$\ln\sqrt{\frac{19}{3}}$

***Atenţie*! Pentru a vă uşura munca, am completat noi câteva dintre elementele tabelului precedent. Bazaţi-vă pe valorile indicate de noi, ca să nu faceţi calcule laborioase, păstrând însă rigurozitatea / corectitudinea raţionamentelor.**

b. Folosind ipoteza combinată obţinută de algoritmul AdaBoost la finalul celei de-a treia iteraţii, stabiliţi:

***i*. eroarea la antrenare produsă (pentru calcularea ei, puteţi folosi tabelul de mai jos),**

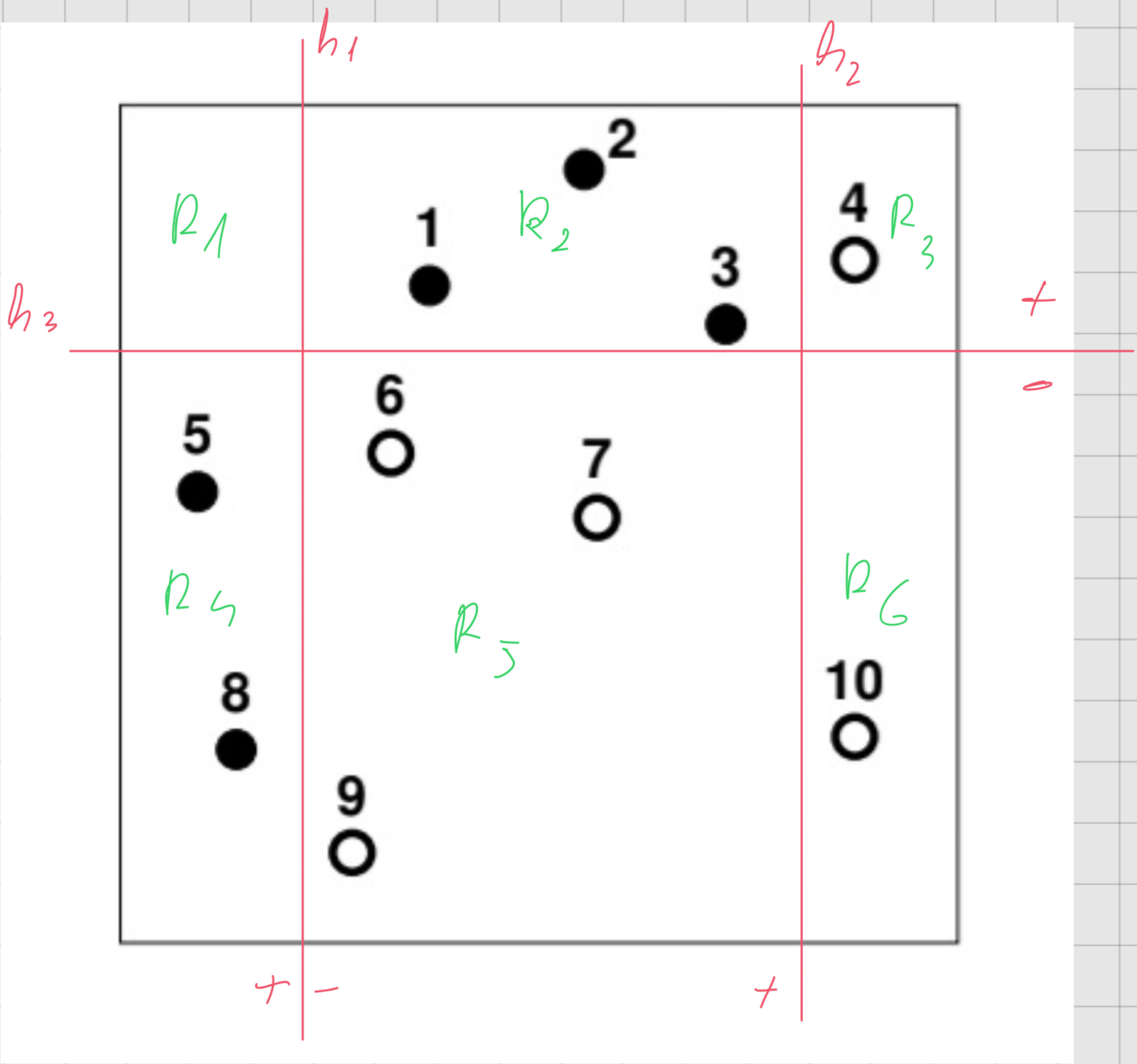
***ii*. zonele de decizie corespunzătoare acestui clasificator (veţi justifica modul în care aţi procedat!). Veţi indica aceste zone de decizie, precum şi graniţele de decizie, pe primul desen din enunţ.**

b

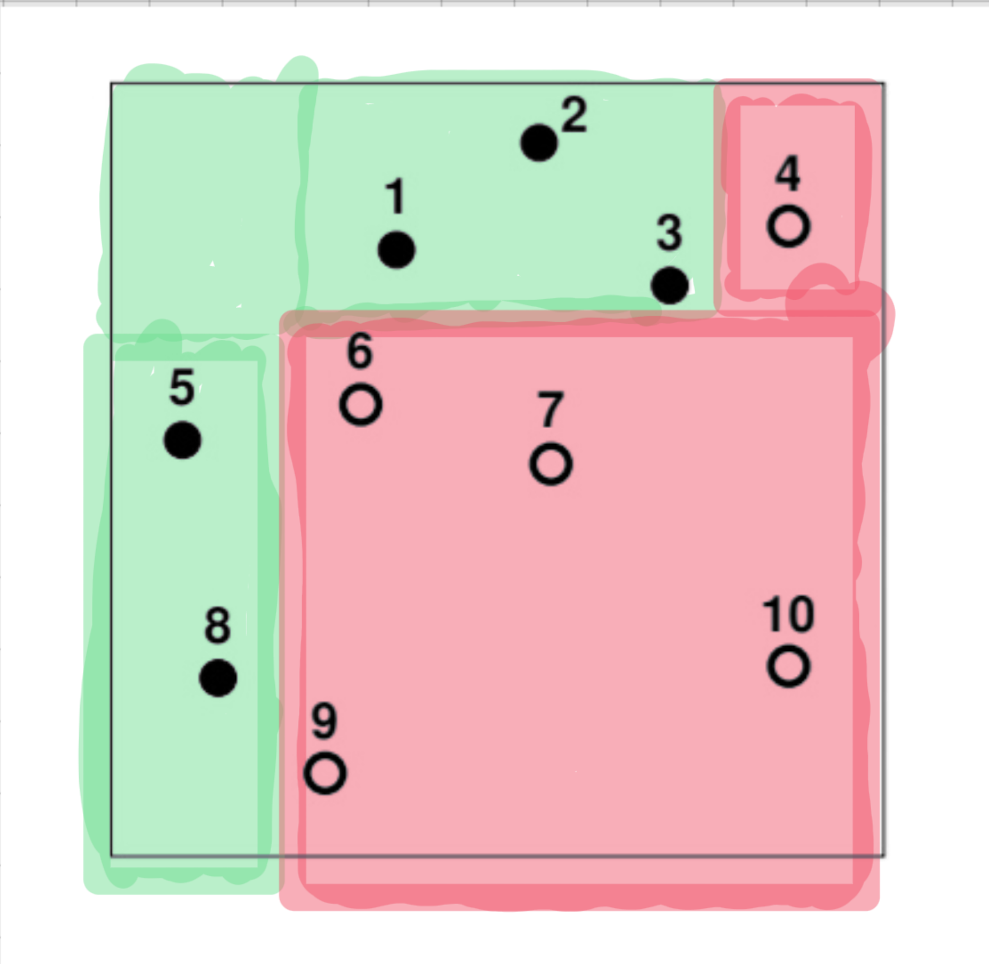
t	α_t	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	$\alpha_1 = 0.423$	−	−	−	−	+	−	−	+	−	−
2	$\alpha_2 = 0.649$	+	+	+	−	+	+	+	+	+	−
3	$\alpha_3 = 0.922$	+	+	+	+	−	−	−	−	−	−
	$H_3(x_i)$	+	+	+	−	+	−	−	+	−	−

$H_3(x) = \text{sign}(\alpha_1 \cdot h_1(x) + \alpha_2 h_2(x) + \alpha_3 h_3(x))$

err. di ant. $= 0$.



t	α_1	R_1	R_2	R_3	R_4	R_5	R_6
1	0.423	+	−	−	+	−	−
2	0.649	+	+	−	+	+	−
3	0.922	+	+	+	−	−	−
	$H_3(R_i)$	+	+	−	+	−	−

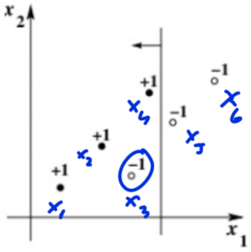


62. (AdaBoost: întrebări în legătură cu aplicarea algoritmului pe un set de date din $\mathbb{R}^2$)

• ∘ *CMU, 200X spring, midterm, pr. 3*
MIT, 2006 fall, Tommi Jaakkola, final, pr. 2

Folosind algoritmul AdaBoost, vrem să obţinem un ansamblu de compaşi de decizie (engl., decision stumps) h_t, de forma $H(x) = sign\left(\sum_{i=1}^{T} \alpha_t h_t(x)\right)$.

În figura alăturată sunt desenate câteva puncte (instanţe) etichetate în planul bidimensional, precum şi primul compas de decizie care a fost ales de către algoritmul AdaBoost. Un compas de decizie oarecare produce valori binare ± 1, ţinând cont doar de un anumit *prag* (engl., the split point). Săgeata mică din figură, care este perpendiculară pe dreapta care reprezintă compasul de decizie indică *zona de decizie* pentru care compasul de decizie va produce valoarea $+1$.



a. Încercuiţi toate acele instanţe din figură pentru care ponderea / probabilitatea [atribuită de către AdaBoost] va creşte ca urmare a incorporării [în ipoteza combinată H a] primului compas de decizie. Justificaţi răspunsul în mod riguros.

b. Desenaţi pe aceeaşi figură un compas de decizie care va putea fi selectat la următoarea iteraţie a algoritmului AdaBoost. Veţi trasa atât dreapta care reprezintă compasul de decizie cât şi [o săgeată care să indice] zona sa de decizie pozitivă. Justificaţi în mod riguros.

c. Va fi oare coeficientul / votul α_2, care este asociat celui de-al doilea compas de decizie mai mare decât α_1, coeficientul [din ansamblul H] pentru primul compas de decizie? Cu alte cuvinte, vom avea oare $\alpha_2 > \alpha_1$? Justificaţi în mod riguros.

Recapitulare:

sign (X≥n)

$H_3(x) = sign(\alpha_1 \cdot h_1(x) + \alpha_2 \cdot h_2(x) \ldots)$

a) deoarece inst. sunt clas. greșit => pt urm. it. acele inst. clasif. greșit vor avea o prob. mai mare.

Inițial: $D_1(i) = \frac{1}{6}$, $i \in \{1 \ldots 6\}$

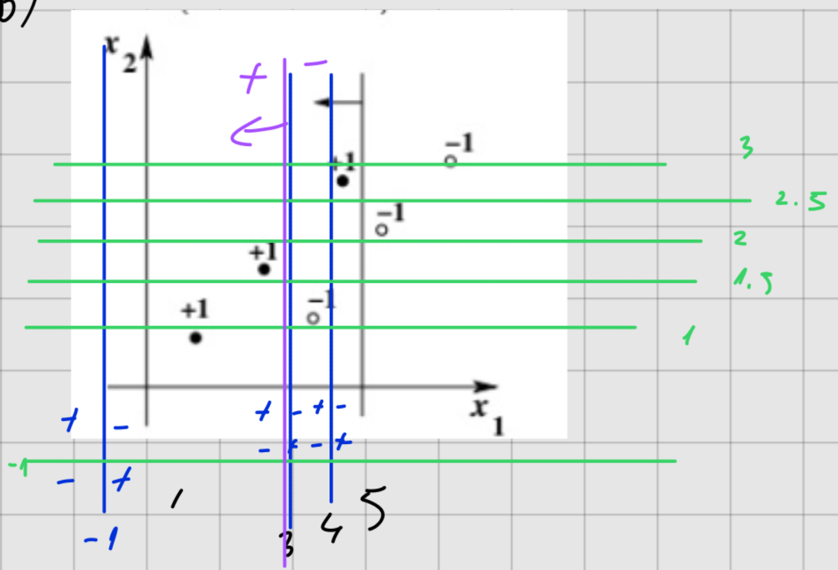
$\varepsilon_1 = \frac{1}{6}$

$$\begin{cases} \dfrac{D_1(i)}{2 \cdot \varepsilon_1}, & i = 3 \Rightarrow \dfrac{\frac{1}{6}}{2 \cdot \frac{1}{6}} = \frac{1}{2} \\[2ex] \dfrac{D_1(i)}{2(1-\varepsilon_1)}, & i \in \{restul\} \Rightarrow D_1(i) = \frac{1}{6}, \; i \in \{1 \ldots 6\} \Rightarrow \end{cases}$$

$$\Rightarrow \frac{\frac{1}{6}}{2 \cdot \frac{5}{6}} = \frac{1}{10}, \quad i \in \{1 \ldots 6\} / \{3\}$$

$$D_2(X): \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{10} & \frac{1}{10} & \frac{1}{2} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \end{pmatrix}$$

↳ se obs. inst. 3 (cea încercuită) a crescut, b)



n	-1	3	4	5
$err(X_1 < n)$	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{1}{2}$
$err(X_1 \geq n)$	$\frac{7}{10}$	$\frac{9}{10}$	$\frac{4}{10}$	$\frac{1}{2}$

n	-1	1	1.5	2	2.5	3
$err(X_2 < n)$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{7}{10}$	$\frac{6}{10}$	$\frac{7}{10}$	$\frac{6}{10}$
$err(X_2 \geq n)$	$\frac{7}{10}$	$\frac{8}{10}$	$\frac{3}{10}$	$\frac{4}{10}$	$\frac{3}{10}$	$\frac{4}{10}$

$$\varepsilon_2 = \frac{1}{10} \qquad \gamma_2 = \frac{1}{2} - \varepsilon_2 = \frac{4}{10}, \quad \alpha_2 = \frac{1}{2} \ln \frac{(1-\varepsilon_2)}{\varepsilon_2} =$$

$$= \frac{1}{2} \cdot \ln \frac{\frac{9}{\cancel{10}}}{\frac{1}{\cancel{10}}} = \frac{1}{2} \cdot \ln 9$$

$h_1 = (X_1 < 3) = \text{sign}(3 - X_1)$

$$\begin{cases} \dfrac{D_2(i)}{2 \cdot \varepsilon_2}, & i \in \{X_4\}, \quad \dfrac{\cancel{\frac{1}{10}}}{2 \cdot \cancel{\frac{1}{10}}} = \dfrac{1}{2} \quad \text{Greşite} \\[2ex] \dfrac{D_2(i)}{2 \cdot (1 - \varepsilon_2)}, & i \in \{\text{restul}\}, \quad \dfrac{\frac{1}{\cancel{10}}}{2 \cdot \frac{9}{\cancel{10}}} = \dfrac{1}{18} \quad \text{corecte} \end{cases}$$

$$D_3(x): \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{18} & \frac{1}{18} & \frac{1}{18} & \frac{1}{2} & \frac{1}{18} & \frac{1}{18} \end{pmatrix}$$

$H_2(x) = sign(\alpha_1 \cdot h_1(x))$

c) $\alpha_1 = \frac{1}{2} \cdot \ln \frac{(1-\varepsilon_1)}{\varepsilon_1} = \frac{1}{2} \ln \frac{\frac{5}{\cancel{6}}}{\frac{1}{\cancel{6}}} = \frac{1}{2} \cdot \ln 5$

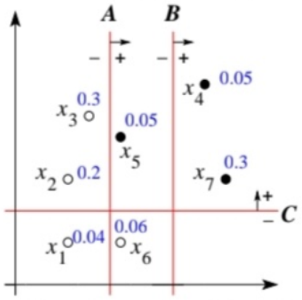
$\alpha_2 = \frac{1}{2} \ln 9$

$\rightarrow \alpha_2 > \alpha_1$ Da vom avea.

63. **(Algoritmul AdaBoost: exemplu de aplicare pe date din $\mathbb{R}^2$; întrebări calitative)**

prelucrare de Liviu Ciortuz, 2021, după
• ∘ MIT, 2007 fall, Tommi Jaakkola, final ex, pr. 1

Fie setul de date de antrenament din figura alăturată. Am marcat cu semnul ∘ exemplele / instanţele negative ($y_i = -1$) şi cu semnul • instanţele pozitive ($y_i = +1$). Figura conţine de asemenea ponderile normalizate (adică, probabilităţile) asociate exemplelor de antrenament, aşa cum au rezultat în urma executării unui anumit număr de iteraţii ale algoritmului AdaBoost. În figură sunt trasaţi şi trei compaşi de decizie, $h(x;\theta_A)$, $h(x;\theta_B)$ şi $h(x;\theta_C)$ sau, pe scurt, A, B şi C.



a. Care dintre aceşti trei compaşi de decizie consideraţi că a fost folosit la *precedenta iteraţie* a algoritmului AdaBoost, în aşa fel încât să rezulte ponderile [asociate exemplelor] prezentate în figură? Veţi răspunde indicând A, B sau C şi veţi justifica în mod riguros alegerea pe care aţi făcut-o.

b. Pe care dintre aceşti trei compaşi de decizie consideraţi că-l va selecta algoritmul AdaBoost la *iteraţia următoare*? Veţi răspunde indicând A, B sau C şi veţi justifica riguros, prin calcule, alegerea pe care aţi făcut-o.

c. În figura dată, încercuiţi instanţele de antrenament (este posibil să nu fie niciuna!) pe care ansamblul (i.e., combinaţia liniară de compaşi de decizie) $H_2(x) = sign(\alpha_A h(x;\theta_A) + \alpha_C h(x;\theta_C))$, cu $\alpha_A = 0.3$ şi $\alpha_C = 0.5$ nu le poate clasifica în mod corect.

a) C. → deoarece suma inst. clas. greșit $= \frac{1}{2}$, conf. prop. Ada boost.

$X_2 + X_3 = 0.2 + 0.3 = 0.5 = \frac{1}{2}$

b)

A: clas. greșit $x_6 \Rightarrow 0.06$

B: Clas. greșit x_5 => 0.05

C: a fost selectat ant => 0.5

Alg. va alege la urm. it pond. cea mai mică.

Urm. clas. ales va fi B.

C.

$H_2(x) = sign(\alpha_A h(x;\theta_A) + \alpha_C h(x;\theta_C))$, cu $\alpha_A = 0.3$
$\alpha_C = 0.5$

nu avem decât 2 in => $H_2(x)$ va clasifica după cel mai mare α => $\alpha_C = 0.5$ => dacă nu va clasifica corect x_2, x_3 vor fi clas. tot greșit.

70. (Concepte din $\mathbb{R}$ *reprezentabile* cu ajutorul combinaţiilor liniare de compaşi de decizie)

Liviu Ciortuz, 2021, pornind de la
• · CMU, 2011 spring, Roni Rosenfeld, HW10, pr. 4.a

Presupunem că avem o problemă de clasificare a unor instanţe pe axa reală: fiecare instanţă x_i este un număr real, iar etichetele pe care urmează să le preziceţi sunt binare, $y_i \in \{-1, +1\}$.

Pentru această problemă de clasificare, veţi folosi *ansambluri*, adică nişte combinaţii liniare de separatori / ipoteze „slabe". (Atenţie! NU trebuie să folosiţi algoritmul AdaBoost!) Vă readucem aminte că un astfel de *clasificator* are forma următoare:

$$\hat{y} = sign\Big(\sum_{t=1}^{T} \alpha_t h_t(x)\Big), \quad (269)$$

unde $\hat{y}$ este eticheta prezisă, $sign(x)$ este $+1$ dacă $x > 0$ şi respectiv -1 în cazul contrar, α_t este o *pondere* (număr real strict pozitiv), iar $h_t(x)$ este predicţia făcută de către ipoteza „slabă" h_t. Fiecare h_t ia una dintre următoarele forme:

$$h_t(x;s,+) = \begin{cases} -1 & \text{dacă } x < s \\ +1 & \text{dacă } x \geq s \end{cases} \qquad h_t(x;s,-) = \begin{cases} +1 & \text{dacă } x < s \\ -1 & \text{dacă } x \geq s, \end{cases}$$

pentru un anumit *prag de separare* (engl., split threshold) $s \in \mathbb{R}$.[588]

Considerăm următorul set de date, format din 5 instanţe situate pe axa reală:

0 1 2 3 4

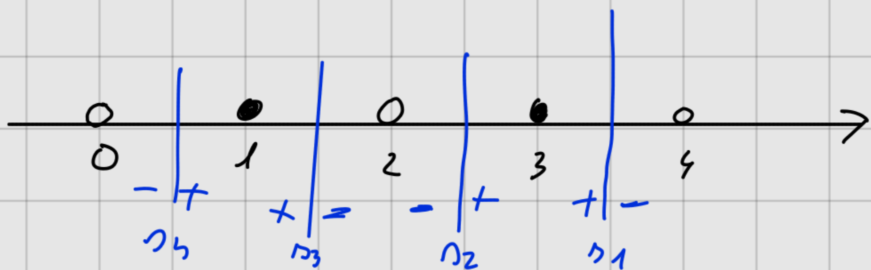
a. Arătaţi că acest set de date (LC: sau, acest *concept*) este *reprezentabil* cu ajutorul unei combinaţii liniare care este formată din 4 ipoteze „slabe". Aşadar, vă cerem să identificaţi în mod explicit 4 ipoteze „slabe" $h_1, \ldots, h_4$, precum şi ponderile lor $\alpha_1, \ldots, \alpha_4$, astfel încât ipoteza combinată $sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$ să fie *consistentă* cu datele de mai sus.
Puneţi în evidenţă cele 4 ipoteze „slabe" $h_1, \ldots, h_4$ pe desenul de mai sus, folosind praguri de separare $s_1, \ldots, s_4$ precum şi linii verticale, cărora le veţi asocia o etichetare adecvată. Concret, veţi desemna cu semnele + şi − la stânga şi la dreapta fiecărei linii verticale — asociată unui anumit prag s_j — zonele de decizie determinate de această ipoteză „slabă".

Observaţie: Ţinând cont de modul în care a fost definită funcţia $sign$ mai sus, *regula de predicţie* (269) va trata cazurile de „paritate" etichetând cu -1 instanţele pentru care suma ponderată (engl., weighted sum) a predicţiilor făcute de ipotezele „slabe" este 0. Acest mod de tratare a cazurilor de „paritate" este [foarte] util pentru rezolvarea punctului *a*.

[588] Remarcaţi faptul că definiţia dată aici pentru noţiunea de *ipoteză „slabă"* corespunde bine-cunoscutului *compas de decizie* (engl., decision stump), pentru cazul particular când datele sunt din $\mathbb{R}$.

606

obs. pt. cazul de egalitate.



$D_0: \left(\begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{array} \right)$

it$_1$ => $h_1(x; s_1, -) = \begin{cases} +1 & \text{dacă } x < s_1 \\ -1 & \text{dacă } x \geq s_1 \end{cases}$

it$_2$ => $h_2(x; s_2, +) = \begin{cases} -1 & \text{dacă } x < s_2 \\ +1 & \text{dacă } x \geq s_2 \end{cases}$

it$_3$ => $h_3(x; s_3, -)$

it$_4$ => $h_4(x; s_4, +)$

$\hat{y} = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$

$\hat{y} = sign\left(\alpha_1 h_1(x) + \alpha_2 h_2(x) + \alpha_3 h_3(x) + \alpha_4 h_4(x)\right)$

Indicaţie: Pentru a justifica uşor consistenţa *ansamblului* ales de dumneavoastră cu datele din figura de mai sus, vă cerem să completaţi un tabel similar cu cel pe care-l folosim la algoritmul AdaBoost atunci când facem calculul erorii la antrenare produse de ipoteza combinată $sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$.

α_t	h_t	0	1	2	3	4
$\alpha_1 =$	$h_1(x_i)$	+	−	+	+	−
$\alpha_2 =$	$h_2(x_i)$	−	−	−	+	+
$\alpha_3 =$	$h_3(x_i)$	+	+	−	−	−
$\alpha_4 =$	$h_4(x_i)$	−	+	+	+	+
$sign\left(\sum_{t=1}^{4} \alpha_t h_t(x_i)\right)$		−	+	−	+	−

b. Demonstraţi că setul de date de mai sus NU este reprezentabil cu mai puţin de 4 ipoteze „slabe".

c. *Generalizaţi* rezultatul obţinut la punctul a, referindu-vă la posibilitatea (sau, dimpotrivă, imposibilitatea) de a *reprezenta* — folosind combinaţii liniare de ipoteze „slabe" aşa cum au fost definite mai sus — dataset-uri arbitrare [formate din instanţe situate] pe axa reală.

b) . . .