Machine Learning Exercise Sheet 2

Lab Group L33

E1 (a)

$h(x) = \langle \perp \rangle$

$\Big\downarrow$ example 1

$h(x) = \langle$ Monday, no, easygoing, evening $\rangle$

$\Big\downarrow$ example 2

$h(x) = \langle$ Monday, no, easygoing, evening $\rangle$

$\Big\downarrow$ example 3

$h(x) = \langle$ Monday, no, easygoing, evening $\rangle$

$\Big\downarrow$ example 4

$h(x) = \langle$ Monday, no, easygoing, ? $\rangle$ = return value

(b) $S_0 = \{\langle \bot, \bot, \bot, \bot \rangle\}$

$\downarrow x_1$

$S_1 = \{\langle \text{Monday}, \text{no}, \text{easygoing}, \text{evening} \rangle\}$

$\downarrow x_4$

$\boxed{S_2 = \{\langle \text{Monday}, \text{no}, \text{easygoing}, ? \rangle\}}$

$\uparrow \qquad \uparrow$

$\boxed{G_3 = \{\langle \text{Mon.}, ?, \text{easyg.}, ? \rangle, \langle ?, \text{no}, \text{easygoing}, ? \rangle\}}$

$\uparrow x_4$

$G_2 = \{\langle \text{Monday}, ? \text{ easygoing}, ? \rangle, \langle ?, \text{no}, \text{easygoing}, ? \rangle, \overbrace{\langle ?, ?, \text{easyg.}, \text{evening} \rangle}^{\text{not consistent with } x_4 \to \text{remove}}\}$

$\uparrow x_3$

$G_1 = \{\langle ?, ?, \text{easygoing}, ? \rangle\}$

$\uparrow x_2$

$G_0 = \{\langle ?, ?, ?, ? \rangle\}$

$\Longrightarrow$ boundary Sets: $S = S_2$, $G = G_3$

(c) $V_{H,D} = S \cup G$

E2 (a) Yes, if $\exists h \in H: \exists s \in S\ \exists g \in G: g >_g h >_g s$

in other words, if there is a hypothesis h that is strictly more general than any hypothesis in S and strictly more specific than any hypothesis in G.

(b) ☑ $(S_2 \geq_g S_1) \vee (S_1 \geq_g S_2)$

☑ $(S_2 \geq_g S_1) \wedge (S_1 \geq_g S_2)$

☐ $(S_2 \ngeq_g S_1) \vee (S_1 \ngeq_g S_2)$

☐ $(S_2 \ngeq_g S_1) \wedge (S_1 \ngeq_g S_2)$

Counterexample

~~$S_1 \geq_g S_2 \Rightarrow S_2 \ngeq_g S_1$~~

$S_1 \ngeq_g S_2 \Rightarrow$ false

(c) Inductive BIAS: "the policy by which the algorithm generalizes from observed training examples to classify unseen instances is its inductive bias... "
The Find-S Algorithm has stronger inductive bias, because it makes the strong assumption of a consistent set D, therefore it ignores the negative examples. The other algorithm checks for consistency.
Apart from that the Algorithms work similar.

E3

(a) $\langle 1, 1, 10, 10 \rangle = g_0$

(b) $h := \langle x_1, y_1, x_2, y_2 \rangle \geq_g \langle x_1', y_1', x_2', y_2' \rangle$
iff $h((x_1', y_1')) = h((x_2', y_2')) = 1$

- [ ] $\langle 1, 2, 3, 4 \rangle \geq_g \langle 1, 1, 4, 4 \rangle$
- [x] $\langle 2, 3, 6, 7 \rangle \geq_g \langle 3, 4, 5, 7 \rangle$
- [ ] $\langle 1, 1, 2, 8 \rangle \geq_g \langle 1, 1, 3, 3 \rangle$
- [ ] $\langle 3, 3, 9, 9 \rangle \geq_g \langle 1, 1, 1, 1 \rangle$

(c) $h_1 = \langle 2, 3, 5, 6 \rangle$
$h_2 = \langle 3, 3, 5, 7 \rangle$

(d) See next 2 Pages.

E3 (d)

$S_0 = \{\langle \perp \rangle\}$

$\downarrow x_1 = (5,3)$

$S_1 = \{\langle 5,3,6,4\rangle, \langle 4,2,5,3\rangle, \langle 5,2,6,3\rangle, \langle 4,3,5,4\rangle\}$

$\downarrow x_5 = (4,4)$

$S_2 = \{\langle 4,3,5,4\rangle\}$

$\downarrow x_7 = (6,5)$

$\boxed{S_3 = \{\langle 4,3,6,5\rangle\}}$

$\Rightarrow G = G_5$; $S = S_3$. Chart on next page.

$\boxed{G_5 = \{\langle 2,2,8,5\rangle, \langle 3,2,8,7\rangle\}}$

$\uparrow x_8 = (2,6)$

$G_4 = \{\langle 2,2,8,7\rangle\}$

$\uparrow x_6 = (5,1)$

$G_3 = \{\langle 2,1,8,7\rangle\}$

$\uparrow x_4 = (5,8)$

$G_2 = \{\langle 2,1,8,10\rangle\}$

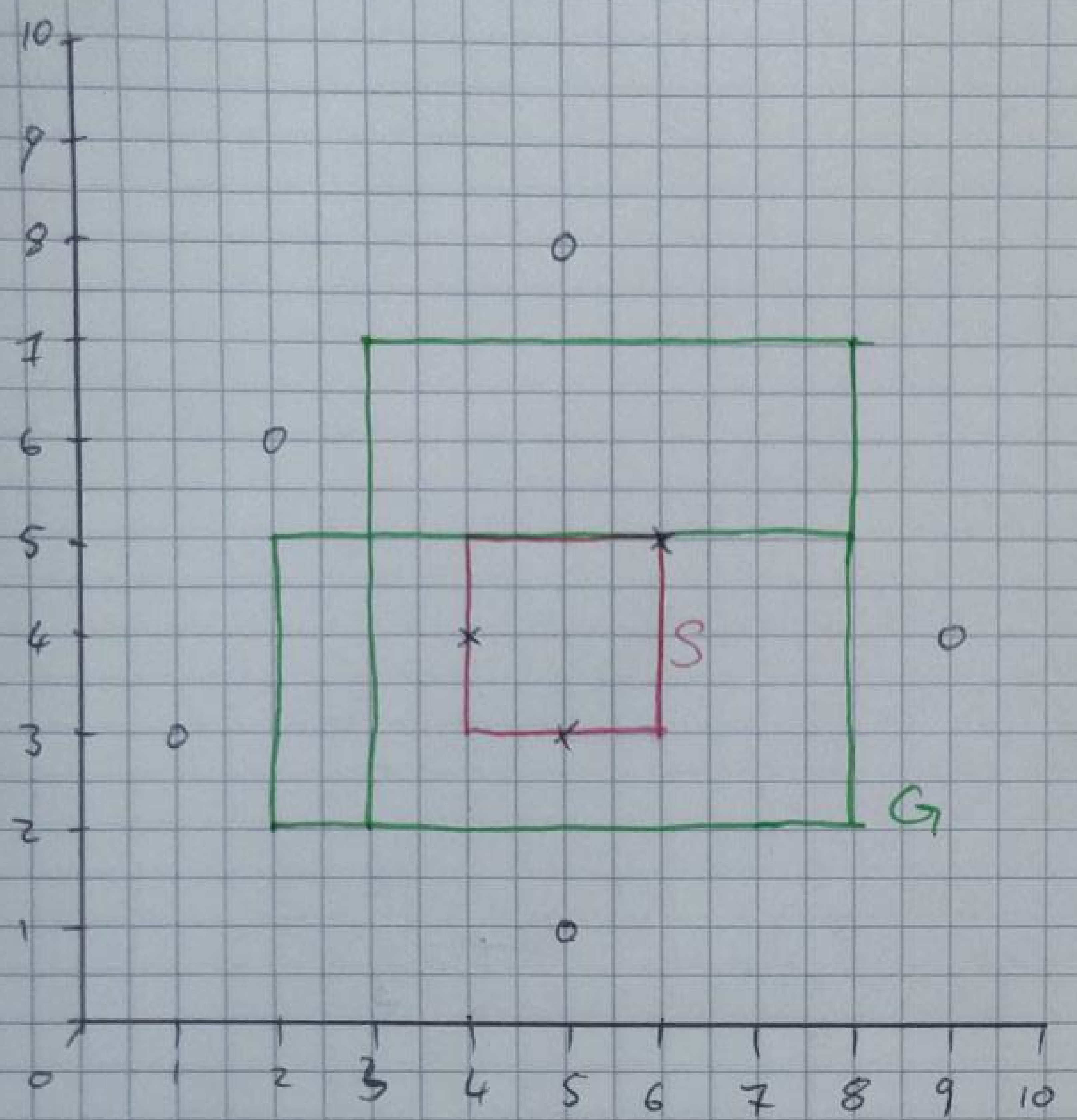
$\uparrow x_3 = (1,3)$

$G_1 = \{\langle 1,1,8,10\rangle\}$

$\uparrow x_2 = (9,4)$

$G_0 = \{\langle 1,1,10,10\rangle\}$

*d)



e) G and S and therefore the entire version space would be empty.

$\Leftrightarrow G = S = V_{H,D} = \emptyset$

f) We could add to our possible hypotheses H all disjunctions of two rectangles.

$H' = H \cup (\vee(H \times H))$

Then a hypothesis $h \in H'$ would assign a point (x, y) to 1 if ① (h is a rectangle $\langle x_1, y_1, x_2, y_2 \rangle$ and $x_1 \leq x \leq x_2$ and $y_1 \leq y \leq y_2$)

or ② (h is a disjunction $h_1 \vee h_2$ of two rectangles and $h_1((x, y)) = 1$ or $h_2((x, y)) = 1$.)

6

E4

(a) $\pi = 1$

$y_1(x) = w \cdot x_1$

Learn(D, 1):

$\mathcal{L}_+ = 0$	$\mathcal{L}_- = 0$	Example No.	x_1	$c(x)$	$I(x_1, c(x))$	$I(-x_1, c(x))$
0	1	0	1	1	0	1
0	2	1	-1	-1	0	1
1	2	2	1	-1	1	0
1	3	3	1	1	0	1
2	3	4	1	-1	1	0
3	3	5	-1	1	1	0
4	3	6	1	-1	1	0
4	4	7	-1	-1	0	1
4	5	8	1	1	0	1
4	6	9	1	1	0	1

$\mathcal{L}_+ \leq \mathcal{L}_- \Rightarrow \underline{\underline{w = 1}}$

$\Rightarrow \underline{\underline{y_1(x) = x_1}}$

$$Err_{tr}(y_1) = \frac{|\{(x, c(x)) \in D_{tr} : y_1(x) \neq c(x)\}|}{|D_{tr}|}, \quad D_{tr} = D$$

$$= \frac{|\{(x, c(x)) \in D_{tr} : x_1 \neq c(x)\}|}{|D_{tr}|}$$

$$= \frac{|\{(1,-1), (1,-1), (-1,1), (1,-1)\}|}{|D_{tr}|} = \frac{4}{10} = \underline{\underline{\frac{2}{5}}}$$

(b) $D_{tr} = D \setminus D_{test}$

Example No	x_1	$c(x)$	$\mathcal{L}_+ = 0$	$\mathcal{L}_- = 0$	$I(x_1, c(x))$	$I(-x_1, c(x))$
0	1	1	0	1	0	1
1	-1	-1	0	2	0	1
2	1	-1	1	2	1	0
3	1	1	1	3	0	1
4	1	-1	2	3	1	0
5	-1	1	3	3	1	0
6	1	-1	4	3	1	0

$\mathcal{L}_+ < \mathcal{L}_- \Rightarrow w = -1$

$\Rightarrow y_1'(x) = -x_1$

holdout error:

$$Err(y, D_{test}) = \frac{|\{(x, c(x)) \in D_{test} : y_1'(x) \neq c(x)\}|}{|D_{test}|}$$

$$= \frac{|\{(x, c(x)) \in D_{test} : -x_1 \neq c(x)\}|}{|D_{test}|}$$

$$= \frac{3}{3} = \underline{\underline{1}}$$

(c) $k=2$

$D_{val_1} \rightsquigarrow 0,1,2,3$

$D_{val_2} \rightsquigarrow 4,5,6$

$D_{test} \rightsquigarrow 7,8,9$

π^* ?

$$\pi^* = \underset{\pi_j, j=1,2}{\operatorname{argmin}} \left(\sum_{i=1}^{k} Err(y'_{i\pi_j}, D_{val_i}) \right)$$

$$= \underset{\pi_j, j=1,2}{\operatorname{argmin}} \left(Err(y'_{1\pi_j}, D_{val_1}) + Err(y'_{2\pi_j}, D_{val_2}) \right)$$

$y'_{1\pi_1}(x) = w_{1\pi_1} \cdot x_1$ | trained on $D_{tr} = D \setminus (D_{test} \cup D_{val1})$

$j=1$
$k=1$ (I) On $D \setminus (D_{test} \cup D_{val1})$: $(\rightsquigarrow$ Examples $\overbrace{4,5,6}^{D_{tr}})$

$\mathcal{L}_+ = 3 \geq \mathcal{L}_- = 0$

$\Rightarrow w_{1\pi_1} = -1$

$$Err(y'_{1\pi_1}, D_{val1}) = \frac{|\{(x, c(x)) \in D_{val1} : x_1 \neq c(x)\}|}{|D_{val1}|}$$

$$= \frac{3}{4} = \frac{3}{4}$$

$j=1$
$k=2$ (II) On $D \setminus (D_{test} \cup D_{val2})$: $(\rightsquigarrow$ Examples $\overbrace{0,1,2,3}^{D_{tr}})$

$\mathcal{L}_+ = 1 \leq 3 = \mathcal{L}_- \Rightarrow w_{2\pi_1} = -1$

$y'_{2\pi_1}(x) = w_{2\pi_1} \cdot x_1$

$= x_1$

$$Err(Y'_{2_{JI_1}}, D_{val_2}) = \frac{|\{(x, c(x)) \in D_{val2} : -x_1 \neq c(x)\}|}{|D_{val2}|}$$

$$= \frac{3}{3} = 1$$

$j=2$
$k=1$ (III) $D_{tr} = D \setminus (D_{test} \cup D_{val1}) \rightsquigarrow$ Examples 4, 5, 6

$$Y'_{1_{JI_2}}(x) = w_{1_{JI_2}} \cdot x_2$$

Example No	x_2	$c(x)$	$\mathcal{L}_+ = 0$	$\mathcal{L}_- = 0$	$I(x_2, c(x))$	$I(-x_2, c(x))$
4	1	-1	1	0	1	0
5	1	1	1	1	0	1
6	1	-1	2	1	1	0

$$\mathcal{L}_+ > \mathcal{L}_- \Rightarrow w_{1_{JI_2}} = -1$$

$$\Rightarrow Y'_{1_{JI_2}}(x) = -x_2$$

$$Err(Y_{1_{JI_2}}, D_{val1}) = \frac{|\{(x, c(x)) \in D_{val1} : x_2 \neq c(x)\}|}{|D_{val1}|}$$

Examples in D_{val1}:

No	x_2	$c(x)$	$I(x_2, c(x))$
0	1	1	0
1	1	-1	1
2	1	-1	1
3	1	1	0

$$Err(Y_{1_{JI_2}}, D_{val_1}) = \frac{2}{4} = \frac{1}{2}.$$

$j=2$
$k=2$ (IV) $D_{tr} = D \setminus (D_{test} \cup D_{val2}) \rightsquigarrow$ Examples $0,1,2,3$

$$y'_{2_{\pi_2}}(x) = w_{2_{\pi_2}} \cdot x_2$$

on D_{tr}: $\mathcal{L}_+ = 2 = \mathcal{L}_- \rightarrow \mathcal{L}_+ \leq \mathcal{L}_- \Rightarrow w_{2_{\pi_2}} = 1$

$$y'_{2_{\pi_2}}(x) = x_2$$

$$Err(y'_{2_{\pi_2}}, D_{val2}) = \frac{|\{(x, c(x)) \in D_{val2} : x_2 \neq c(x)\}|}{|D_{val2}|}$$

$$= \frac{|\{(1,-1),(1,-1)\}|}{3} = \frac{2}{3}$$

$$\pi^* = \underset{\pi_j,\ j=1,2}{\operatorname{argmin}} \left(Err(y'_{1_{\pi_j}}, D_{val1}) + Err(y'_{2_{\pi_j}}, D_{val2}) \right)$$

$$\pi^* = \begin{cases} \pi_1 & \text{if } Err(y'_{1_{\pi_1}}, D_{val1}) + Err(y'_{2_{\pi_1}}, D_{val2}) < \\ & \quad Err(y'_{1_{\pi_2}}, D_{val1}) + Err(y'_{2_{\pi_2}}, D_{val2}) \\ \pi_2 & \text{else} \end{cases}$$

$$\cancel{\min\left(\frac{1}{4} + 0, \frac{1}{2} + \frac{2}{3}\right) = \min\left(\frac{1}{4}, \frac{7}{6}\right)}$$

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

$\pi^* = \underline{\underline{\pi_2}}$ since $\frac{3}{4} + 1 = \frac{7}{4} > \frac{1}{2} + \frac{2}{3} = \frac{7}{6}$

holdout error for y_{π^*}:

$\cancel{y_{\pi^*} = y_{\pi_2}(x) = w \cdot x_1 = y_1(x) = x_1}$ ~~(see 4(a))~~

$y'_{\pi^*} = y'_{\pi_2}(x) \rightsquigarrow$ trained on $D \setminus D_{test}$

$= -x_2$ ~~(see 4(b))~~

holdout error:

$$Err(y_{\pi^*}, D_{test}) = \frac{|\{(x, c(x)) \in D_{test} : -x_2 \neq c(x)\}|}{|D_{test}|}$$

$$= \frac{2}{3}$$