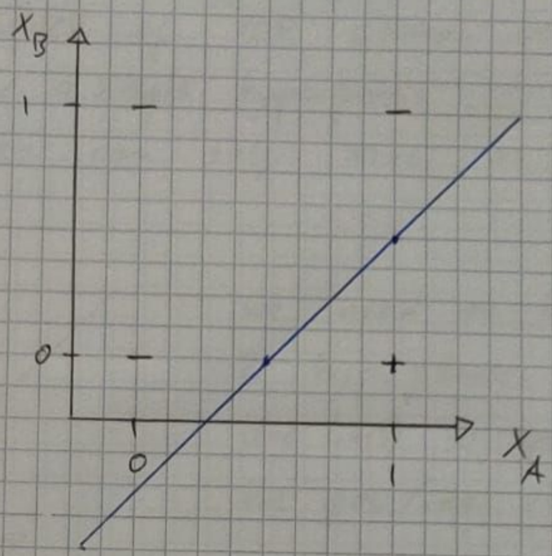
Ex. Sheet 4

E1

(a)

X_A	X_B	$c(x)$
0	0	0
0	1	0
1	0	1
1	1	0



$w_0 + X_A w_1 + X_B w_2 = 0$

Let $w_1 := 1$ Plug in $X_A = 0.5$, $X_B = 0$

$w_0 + 0.5 = 0$

$\Rightarrow w_0 = -0.5$

$w_1 > 0$ because $w_1 X_A$ should be increasing if X_A is increasing

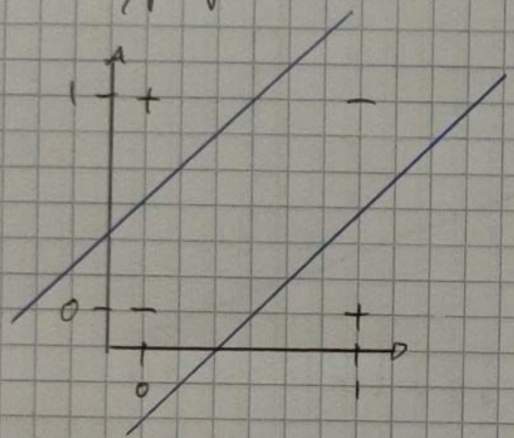
Plug in $X_A = 1$, $X_B = 0.5$

$-0.5 + 1 + 0.5 w_2 = 0$

$\Rightarrow w_2 = -1$

$\Rightarrow \underline{\underline{w = (-0.5, 1, -1)}}$

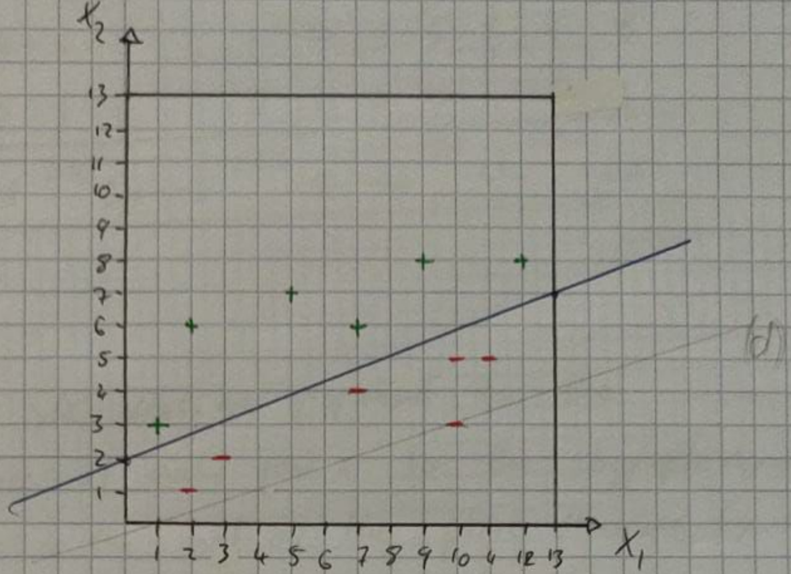
(b) Because in the feature space of $[0,1]^2$, positive and negative examples of XOR cannot be linearly separated by a single hyperplane.
A single perceptron can only serve as one hyperplane.



we need at least 2 hyperplanes (here blue lines) to separate the pos./neg. examples.

E2

(a)



(b) $w \cdot \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = 0$, $w \cdot \begin{bmatrix} 1 \\ 13 \\ 7 \end{bmatrix} = 0$

$$w_0 + 0 \cdot w_1 + 2 \cdot w_2 = 0$$

let $w_1 := -1$

$$w_0 - 13 + 7 w_2 = 0$$

$$\Rightarrow w_0 = 13 - 7 w_2$$

$$13 - 7w_2 + 2 w_2 = 0 = 13 - 5 w_2$$

$$w_2 = \frac{13}{5}$$

$$\Rightarrow w_0 = 13 - 7 \cdot \frac{13}{5} = -\frac{26}{5}$$

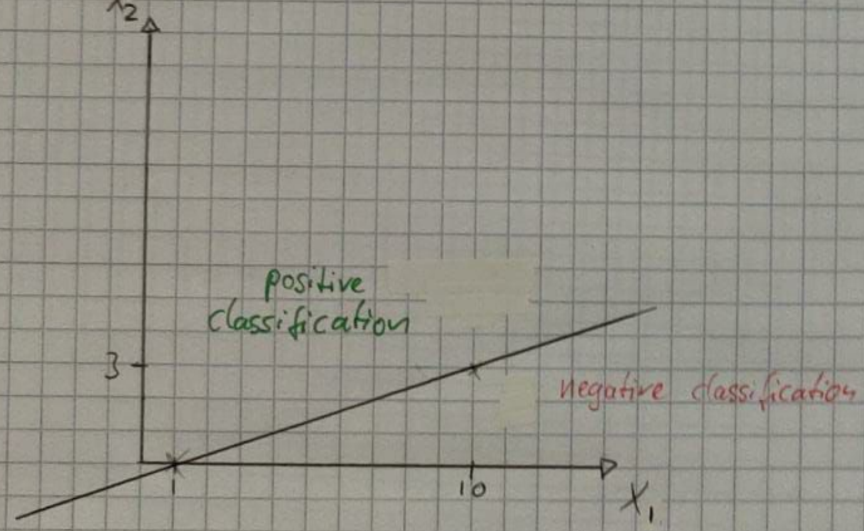
$$\Rightarrow \underline{\underline{w = \left[-\frac{26}{5}, -1, \frac{13}{5} \right]}}$$

(c)

No	1	2	3	4	5	6	7	8	9	10	11	12
x_1	1	7	7	2	3	2	5	10	11	12	10	9
x_2	3	4	6	6	2	1	7	3	5	8	5	8
$c(x)$	1	0	1	1	1	0	1	0	0	1	0	1
$y(x) = w_0 + w_1 x_1 + w_2 x_0$	1.6	-1.8	3.4	8.4	-3	-4.6	8	-7.4	-3.2	3.6	-2.2	6.6

pattern: the $y(x)$ values are proportional to the distance of each point x to the hyperplane, which is the root of the plane described by the w vector. Thus, the sign of an example's $y(x)$ corresponds to the classification of the example x.

(d)



let $x_1 = 1 \Rightarrow [1, -1, 3] \cdot \begin{bmatrix} 1 \\ 1 \\ x_2 \end{bmatrix} = 0$

$$1 - 1 + 3x_2 = 0$$

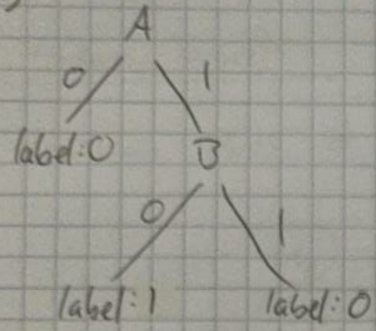
$$x_2 = 0 \rightarrow p(1, 0)$$

let $x_1 = 10 \Rightarrow 1 - 10 + 3x_2 = 0$

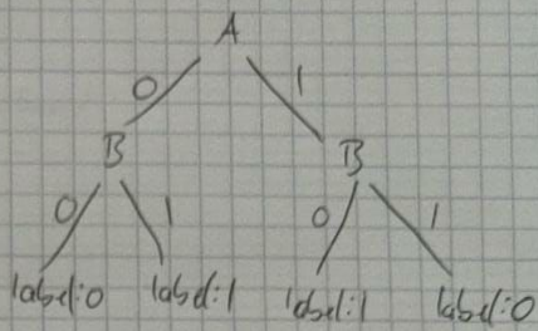
$$x_2 = 3 \rightarrow p(10, 3)$$

E3

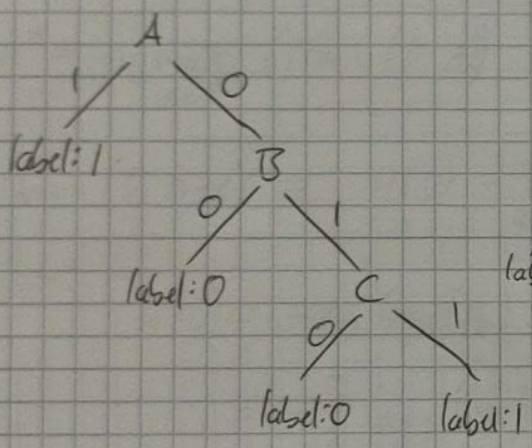
(a) $A \wedge \neg B$



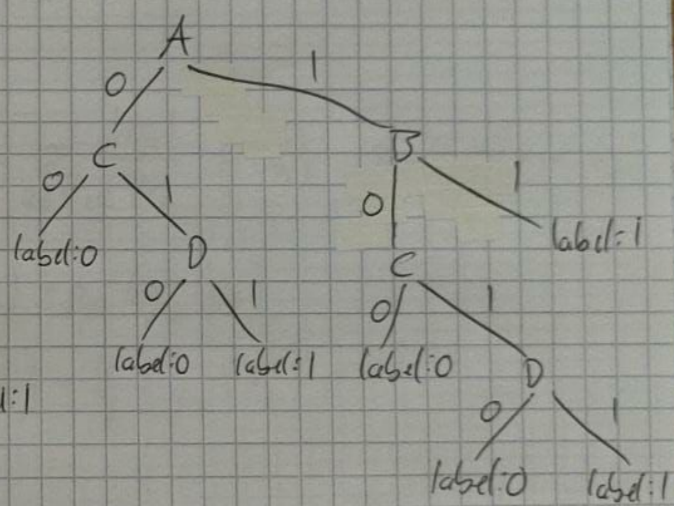
(b) A XOR B



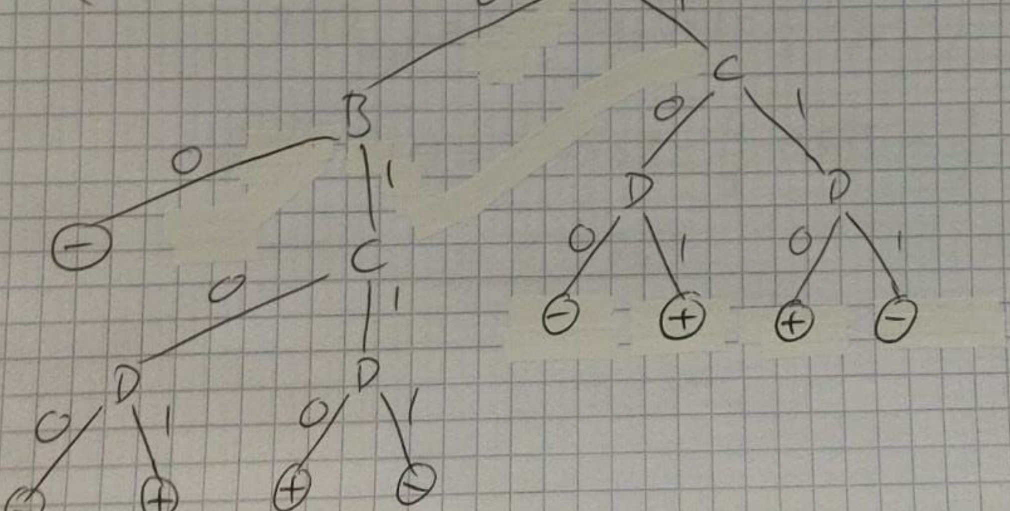
(c) $A \vee (B \wedge C)$



(d) $(A \wedge B) \vee (C \wedge D)$



(e) $(A \vee B) \wedge (C \text{ XOR } D)$



(E4)

$\ominus$ = well-behaved, $\oplus$ = dangerous

<u>ID3(D, Features, Target):</u>

most common class = well-behaved

t_0

label = $\ominus$

impure $\to$ yes, and Features $\neq \emptyset$

$\hookrightarrow$ we split.

$$A^* = \underset{A \in F}{\operatorname{argmax}} \left(iGain(D, A) \right)$$

$$iGain(D, size) \equiv H(D) - \sum_{a \in Size} \frac{|D_a|}{|D|} \cdot H(D_a) \quad , Dom(size) = \{small, big\}$$

$$H(D) = -p_{\oplus} \log_2(p_{\oplus}) - p_{\ominus} \log_2(p_{\ominus})$$

$$= -\frac{3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right)$$

$$H(D) \approx 0.9852$$

$$H(D_{small}) = -\frac{1}{4} \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right) \approx 0.8113$$

$$H(D_{big}) = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) \approx 0.9183$$

$$iGain(D, Size) = 0.9852 - \frac{4}{7} \cdot 0.8113 - \frac{3}{7} \cdot 0.9183$$

$$\underline{\underline{\approx 0.1280}}$$

iGain(D, Fur): Dom(Fur) = {raggad, smooth, curly}

$$iGain(D, Fur) = H(D) - \frac{|D_{ragged}|}{|D|} \cdot H(D_{ragged}) - \frac{|D_{smooth}|}{|D|} \cdot H(D_{smooth})$$

$$- \frac{|D_{curly}|}{|D|} \cdot H(D_{curly})$$

$$H(D_{ragged}) = -\frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right) \approx 0.9183$$

$$H(D_{smooth}) = -\frac{2}{2} \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) = 0$$

$$H(D_{curly}) = 0$$

$$iGain(D, Fur) = 0.9852 - \frac{3}{7} \cdot 0.9183 = \underline{\underline{0.5916}}$$

iGain(Color): Dom(Color) = {brown, black, white, red}

$H(D_{brown}) = 0$ because just one example...

$$H(D_{black}) = -\frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) = 0.9183$$

$$H(D_{white}) = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

$$H(D_{red}) = 0$$

$$iGain(Color) = 0.9852 - \frac{3}{7} \cdot 0.9183 - \frac{2}{7}$$

$$\underline{\underline{= 0.3059}}$$

$$\Rightarrow \underset{A \in F}{\operatorname{argmax}} \left(iGain(A) \right) = \underline{\underline{Fur}}$$

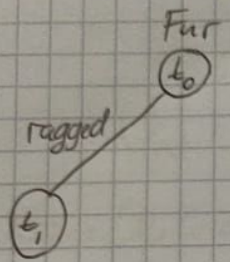
The Attribute "Fur" is the feature that best classifies the examples in the Dataset. $A^* = Fur$.

Now iterate values in A^* and do recursive calls.

$$D_{ragged} = \{(X, c(X)) \in D : X|_{Fur} = ragged\}$$

$$|D_{ragged}| = 3 \Rightarrow D_{ragged} \neq \emptyset$$

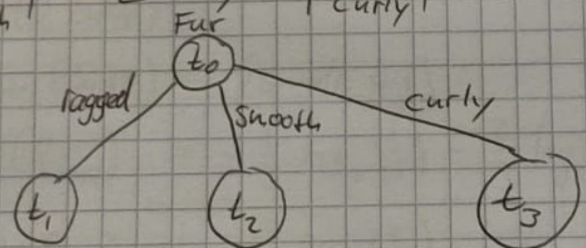
$\Rightarrow$ createEdge$(t_0$, ragged, ID3$(D_{ragged}, F \setminus \{Fur\}, Target))$



likewise for smooth, curly:

$|D_{smooth}| = 2$, $|D_{curly}| = 2$

↳



remaining calls before return:

ID3$(D_{ragged}, F \setminus \{Fur\}, Target)$,
ID3$(D_{smooth}, F \setminus \{Fur\}, Target)$,
ID3$(D_{curly}, F \setminus \{Fur\}, Target)$

ID3$(D_{ragged}, F \setminus \{Fur\}, Target)$:

label(t_1) = mostCommonClass$(D_{ragged}, Target)$ = well-behaved

$D := D_{ragged}$, $F := F \setminus \{Fur\}$

impure → yes, $F \neq \emptyset$
↳ Split.

$$A^* = \underset{A \in F}{\operatorname{argmax}} \left(iGain(D, A) \right)$$

D:

Color	Size	Class
brown	small	⊖
black	big	⊕
red	big	⊖

$F = \{Color, Size\}$

$iGain(D, Color)$: $\quad Dom(Color) = \{brown, black, white, red\}$

$$H(D) = -\frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right) = 0.9183$$

$$H(D_{brown}) = 0, \quad |D_{brown}| = 1$$

$$H(D_{black}) = 0 \quad |D_{black}| = 1$$

? $D_{white} = \emptyset$ $\quad H(D_{white}) = 0 \quad |D_{white}| = 0$

$$H(D_{red}) = 0 \quad |D_{red}| = 1$$

$$iGain(D, Color) = H(D) - \frac{1}{3} \cdot 0 - \frac{1}{3} \cdot 0 - 0 \cdot 0 - \frac{1}{3} \cdot 0$$

$$= \underline{\underline{0.9183}}$$

$iGain(D, Size)$:

$Dom(Size) = \{small, big\}$

$$H(D_{small}) = -\frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right) = 0.9183$$

$$H(D_{big}) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.9183$$

$$iGain(D, Size) = 0.9183 - \frac{1}{3} 0.9183 - \frac{2}{3} \cdot 0.9183$$

$$= \underline{\underline{0}}$$

$$\Rightarrow \underset{A \in F}{\operatorname{argmax}} \left(iGain(D, A)\right) = Color$$

The feature "color" is the most discriminative feature, in the remaining dataset.

$A^* = Color.$

Iterate $a \in A^*$, $A^* = \{brown, black, white, red\}$

$|D_{brown}| = 1$

$\Rightarrow createEdge(t_1, brown, ID3(D_{brown}, F \setminus \{Color\}, Target))$

$|D_{black}| = 1$

$\Rightarrow create\ Edge(t_1, black, ID3(D_{black}, F \setminus \{Color\}, Target))$

$|D_{white}| = 0$

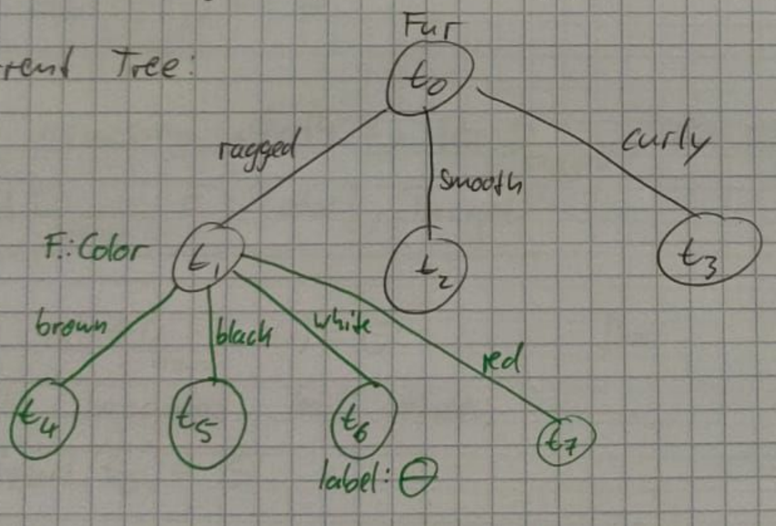
$\Rightarrow t_1'$ is leaf Node, $label(t_1') =$ most Com. class in D $= \ominus$

create Edge $(t_1, white, t_1')$

$|D_{red}| = 1$

$\Rightarrow createEdge(t_1, red, ID3(D_{red}, F \setminus \{Color\}, Target))$

Current Tree:



recursion depth	remaining function calls
1	$ID3(D_{smooth}, F \setminus \{Fur\}, Target)$
	$ID3(D_{curly}, F \setminus \{Fur\}, Target)$
2	$ID3(D_{brown}, F \setminus \{Color\}, Target)$
	$ID3(D_{black}, F \setminus \{Color\}, Target)$
	$ID3(D_{red}, F \setminus \{Color\}, Target)$

$\left.\right\}$ where $F = F \setminus \{Fur\}$, $D = D_{ragged}$

$ID3(D_{smooth}, F\setminus\{Fur\}, Target)$

$D := D_{smooth}$, $F := F\setminus\{Fur\}$

$D =$

Color	Size	Class
black	big	dangerous
white	small	dangerous

$label(t_2) = \oplus$ = dangerous.

$\forall (x, c(x)) \in D : c(x) = \oplus \Rightarrow$ return t_2

$\Rightarrow t_2$ becomes leaf Node, $label = \oplus$

$ID3(D_{curly}, F\setminus\{Fur\}, Target)$

$D =$

Color	Size	Class
black	small	$\ominus$
white	small	$\ominus$

$\forall (x, c(x)) \in D_{cur} : c(x) = \ominus$

$\Rightarrow$ return t_3

$\Rightarrow t_3$ becomes leaf Node,

$label(t_3) = \ominus$

Now recursion depth 2, $D = D_{ragged}$, $F = F\setminus F\{Fur\}$

$ID3(D_{brown}, F\setminus\{Color\}, Target)$:

$D =$

Size	Class
small	$\ominus$

$\forall (x, c(x)) \in D_{brown} : c(x) = \ominus$

$\Rightarrow t_4$ becomes leaf Node,

$label(t_4) = \ominus$

$ID3(D_{black}, F\setminus\{Color\}, Target)$:

$D =$

Size	Class
big	$\oplus$

$\forall (x, c(x)) \in D_{black} : c(x) = \oplus$

$\Rightarrow t_5$ leaf Node, $label(t_5) = \oplus$

$ID3(D_{red}, F\setminus\{Color\}, Target)$:

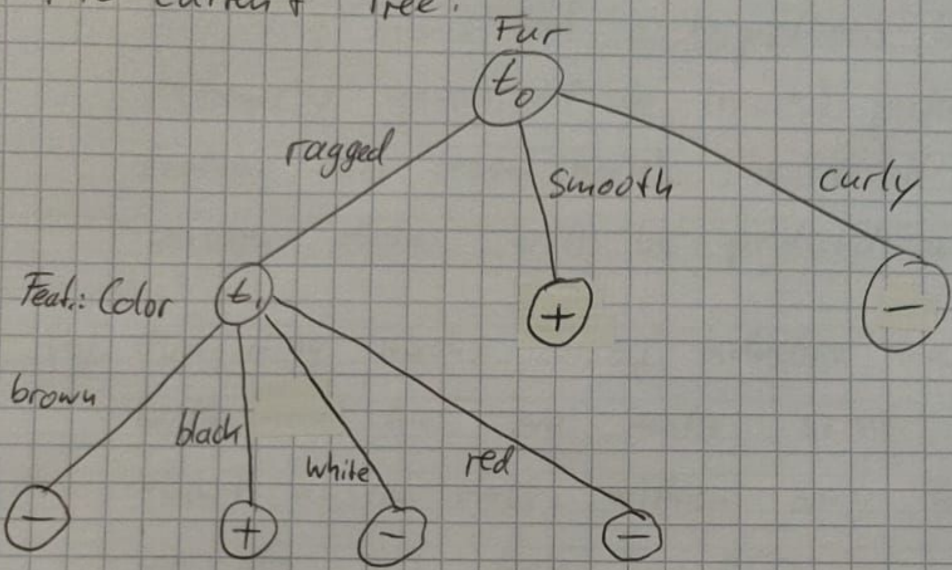
$D =$

Size	Class
big	$\ominus$

$\Rightarrow t_7$ leaf Node, $label(t_7) = \ominus$

All recursive calls have returned.

The current Tree:



Finally, the initial call of ID3 will return t_0.

(b) dangerous, by rule: IF $(Fur = ragged) \wedge (Color = black)$

THEN label $= \oplus$.