4. Exercise Sheet
Introduction to Machine Learning (WiSe 2020/21)

Anton Lammert
Matrikelnummer: ??????

Henrik Leisdon
Matrikelnummer: ??????

Tony Zöppig
Matrikelnummer: ??????

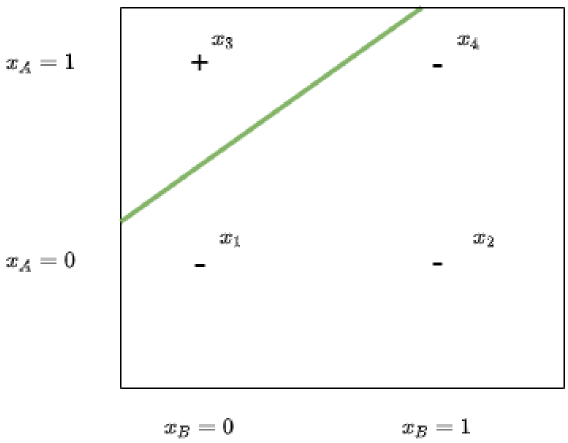
Bauhaus-Universität Weimar

January 28, 2021

Exercise 1

a)

	x_A	x_B	A∧¬B	c(x)
x_1	0	0	0	-
x_2	0	1	0	-
x_3	1	0	1	+
x_4	1	1	0	-

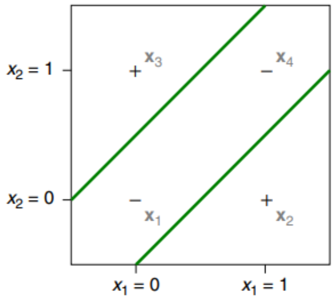


To determine the perceptron weights we have to look at the truth table. In x_3 is the only example that satisfies the formula. Therefore the combination of x_A=1 and x_B=0 has to be the only combination for which the threshold function $\phi(x)$ outputs 1.
So x_A=1 needs to have a positive value to increase the weight $\mathbf{w}$, x_B=1 needs to have a negative value to decrease the weight.

Therefore is w_0=0, w_1=1 and w_2=-1, so that only x_3 is true ($\mathbf{w}$=w_0+x_A*w_1+x_B*w_2= 0+1*1+0*(-1) = 1).
Every other example is below the threshold.

b)

	x_1	x_2	XOR	$c(\mathbf{x})$
$\mathbf{x}_1$	0	0	0	−
$\mathbf{x}_2$	1	0	1	+
$\mathbf{x}_3$	0	1	1	+
$\mathbf{x}_4$	1	1	0	−

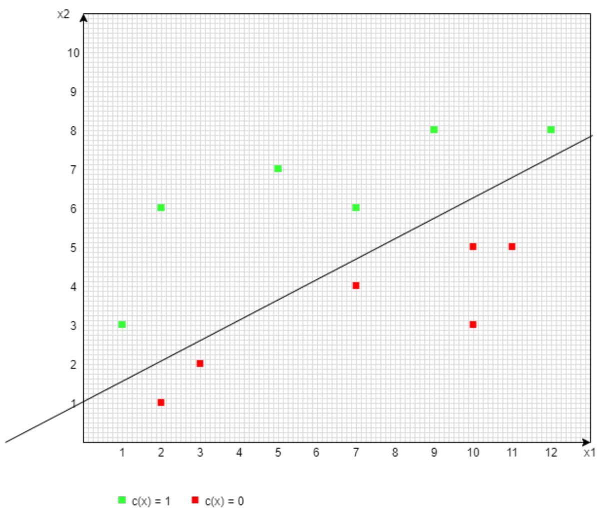


As mentioned in the lecture the **XOR** operation can not be realized in a single perceptron. If we look to the truth table, we can see that the combinations 0,0 and 1,1 are both 0. Therefore we need two linear decision boundaries. But this is only possible through a multi-perceptron.

Exercise 2

a,b,c (Tony)

a)



b)

$0 = w_0 \cdot 1 + w_1 \cdot x_1 + w_2 \cdot x_2$
$0 = 1 \cdot 1 + 1 \cdot x_1 - 2 \cdot x_2$

c)

If the result is negative we know that $c(x) = 1$ else we know that $c(x) = 0$.

Number	1	2	3	4	5	6	7	8	9	10	11	12	13
x1	1	7	7	2	3	2	5	10	11	12	10	9	-2
x2	3	4	6	6	2	1	7	3	5	8	5	8	1
c(x)	1	0	1	1	0	0	1	0	0	1	0	1	0

d)

The equation for the weights $w_0 = 1$,$w_1 = -1$, $w_2 = 3$ has the following form: 1-1x_1+3x_2=0. It is used for classifying the first point of the training data.

e)

Manual computation of the first 3 weight vectors:

1.

x_0=1 $\quad w_0 = 1$
x_1=1 $\quad w_1 = -1$
x_2=3 $\quad w_2 = 3$
c(x) = 1

$$y(x) = (1\ -1\ 3) * \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = 9 \rightarrow y(x)=1$$

$$\delta = c(x)-y(x) = 1-1 = 0$$

$$\Delta\ w = 1*0* \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

new w = (1 -1 3)

2.

x_0=1 $\quad w_0 = 1$
x_1=7 $\quad w_1 = -1$
x_2=4 $\quad w_2 = 3$
c(x) = 0

$$y(x) = (1\ -1\ 3) * \begin{pmatrix} 1 \\ 7 \\ 4 \end{pmatrix} = 6 \rightarrow y(x)=1$$

$$\delta = c(x)-y(x) = 0-1 = -1$$

$$\Delta\ w = 1*-1* \begin{pmatrix} 1 \\ 7 \\ 4 \end{pmatrix} = \begin{pmatrix} -1 \\ -7 \\ 4 \end{pmatrix}$$

new w = (0 -8 -1)

3.

x_0=1 $\quad w_0 = 0$
x_1=7 $\quad w_1 = -8$
x_2=6 $\quad w_2 = -1$
c(x) = 1

$$y(x) = (0\ -8\ -1) * \begin{pmatrix} 1 \\ 7 \\ 6 \end{pmatrix} = -62 \rightarrow y(x)=0$$

$$\delta = c(x)-y(x) = 1-0 = 1$$

$$\Delta\ w = 1*1* \begin{pmatrix} 1 \\ 7 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 7 \\ 6 \end{pmatrix}$$

new w = (1 -1 5)

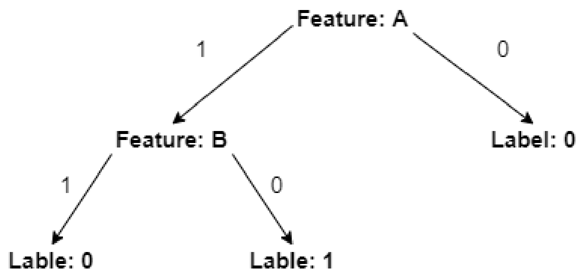
Remaining iterations:

w_0	w_1	w_2	y(x)	δ
1.	-1.	3.	9.	0.
0.	-8.	-1.	6.	-1.
2.	6.	11.	-62.	2.
2.	6.	11.	80.	0.
1.	3.	9.	42.	-1.
0.	1.	8.	16.	-1.
0.	1.	8.	61.	0.
-1.	-9.	5.	34.	-1.
0.	2.	10.	-75.	1.
0.	2.	10.	104.	0.
-1.	-8.	5.	70.	-1.
1.	10.	21.	-33.	2.
1.	10.	21.	74.	0.
0.	3.	17.	155.	-1.
0.	3.	17.	123.	0.
0.	3.	17.	108.	0.
-1.	0.	15.	43.	-1.
-2.	-2.	14.	14.	-1.
-2.	-2.	14.	86.	0.
-3.	-12.	11.	20.	-1.
-2.	-1.	16.	-80.	1.
-2.	-1.	16.	114.	0.
-3.	-11.	11.	68.	-1.
-1.	7.	27.	-14.	2.
-1.	7.	27.	87.	0.
-2.	0.	23.	156.	-1.
-2.	0.	23.	136.	0.
-2.	0.	23.	136.	0.
-3.	-3.	21.	44.	-1.
-4.	-5.	20.	12.	-1.
-4.	-5.	20.	111.	0.
-5.	-15.	17.	6.	-1.
-4.	-4.	22.	-85.	1.
-4.	-4.	22.	124.	0.
-5.	-14.	17.	66.	-1.
-5.	-14.	17.	5.	0.

Exercise 3

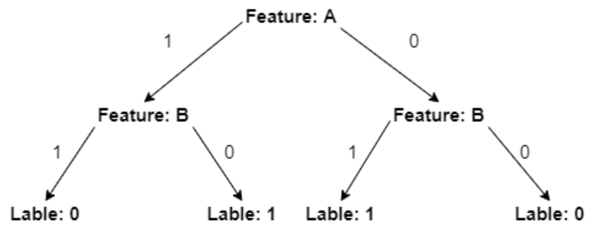
a)

A	B	c(x)
1	1	0
1	0	1
0	1	0
0	0	0



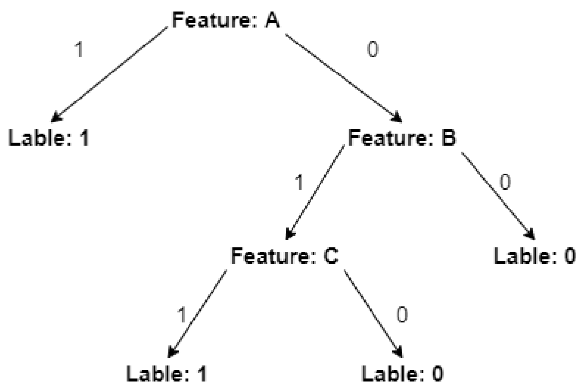
b)

A	B	c(x)
1	1	0
1	0	1
0	1	1
0	0	0



c)

A	B	C1	B$\wedge C$	c(x)
1	1	1	1	1
0	1	1	1	1
	0	1	0	0
	0	0	0	0
	1	0	0	0



Exercise 4

(a) $iGain(D, A) \equiv H(D) - \sum_{a \in A} \frac{|D_a|}{|D|} \cdot H(D_a)$ with $H(D) = -p_{\oplus} \log_2(p_{\oplus}) - p_{\ominus} \log_2(p_{\ominus})$

Color	Fur	Size	Class
brown	ragged	small	well-behaved
black	ragged	big	dangerous
black	smooth	big	dangerous
black	curly	small	well-behaved
white	curly	small	well-behaved
white	smooth	small	dangerous
red	ragged	big	well-behaved

We want to know which attribute should be used first for splitting. For this we compute $iGain(D, Color)$, $iGain(D, Fur), iGain(D, Size)$ and chose the attribute with the maximum average information gain.

$H(D) = -\frac{4}{7} \cdot \log_2(\frac{4}{7}) - \frac{3}{7} \cdot \log_2(\frac{3}{7}) = 0{,}985.$
$iGain(D, Color) = H(D) - \Big(\frac{|D_{brown}|}{|D|} \cdot H(D_{brown}) + \frac{|D_{black}|}{|D|} \cdot H(D_{black}) + \frac{|D_{white}|}{|D|} \cdot H(D_{white}) + \frac{|D_{red}|}{|D|} \cdot H(D_{red}) \Big)$
$\frac{|D_{brown}|}{|D|} = \frac{1}{7}, \quad \frac{|D_{black}|}{|D|} = \frac{3}{7}, \quad \frac{|D_{white}|}{|D|} = \frac{2}{7}, \quad \frac{|D_{red}|}{|D|} = \frac{1}{7}$

D_{brown} would look as follows:

Color	Fur	Size	Class
brown	ragged	small	well-behaved

Which means that $H(D_{brown}) = -1 \log_2(1) - 0 \log_2(0) = 0$

D_{black} would look as follows:

Color	Fur	Size	Class
black	ragged	big	dangerous
black	smooth	big	dangerous
black	curly	small	well-behaved

Which means that $H(D_{black}) = -\frac{1}{3} \log_2(\frac{1}{3}) - \frac{2}{3} \log_2(\frac{2}{3}) = 0,918$

D_{white} would look as follows:

Color	Fur	Size	Class
white	curly	small	well-behaved
white	smooth	small	dangerous

Which means that $H(D_{black}) = -\frac{1}{2} \log_2(\frac{1}{2}) - \frac{1}{2} \log_2(\frac{1}{2}) = 1$

D_{red} would look as follows:

Color	Fur	Size	Class
red	ragged	big	well-behaved

Which means that $H(D_{red}) = -1 \log_2(1) - 0 \log_2(0) = 0$

Entering those values into the computation of $iGain(D, Color)$ we get:
$iGain(D, Color) = 0,985 - \big(\frac{1}{7} \cdot 0 + \frac{3}{7} \cdot 0,918 + \frac{2}{7} \cdot 1 + \frac{1}{7} \cdot 0 \big) = 0,305$

$iGain(D, Fur) = H(D) - \left(\frac{|D_{ragged}|}{|D|} \cdot H(D_{ragged}) + \frac{|D_{curly}|}{|D|} \cdot H(D_{curly}) + \frac{|D_{smooth}|}{|D|} \cdot H(D_{smooth})\right)$
$\frac{|D_{ragged}|}{|D|} = \frac{3}{7}, \quad \frac{|D_{curly}|}{|D|} = \frac{2}{7}, \quad \frac{|D_{smooth}|}{|D|} = \frac{2}{7}$

D_{ragged} would look as follows:

Color	Fur	Size	Class
brown	ragged	small	well-behaved
black	ragged	big	dangerous
red	ragged	big	well-behaved

Which means that $H(D_{ragged}) = -\frac{2}{3}\log_2(\frac{2}{3}) - \frac{1}{3}\log_2(\frac{1}{3}) = 0,918$

D_{curly} would look as follows:

Color	Fur	Size	Class
black	curly	small	well-behaved
white	curly	small	well-behaved

Which means that $H(D_{curly}) = -1\log_2(1) - 0\log_2(0) = 0$

D_{smooth} would look as follows:

Color	Fur	Size	Class
black	smooth	big	dangerous
white	smooth	small	dangerous

Which means that $H(D_{smooth}) = -0\log_2(0) - 1\log_2(1) = 0$

Entering those values into the computation of $iGain(D, Fur)$ we get:
$iGain(D, Fur) = 0,985 - (\frac{3}{7} \cdot 0,918 + \frac{2}{7} \cdot 0 + \frac{2}{7} \cdot 0) = 0,591$

$iGain(D, Size) = H(D) - (\frac{|D_{small}|}{|D|} \cdot H(D_{small}) + \frac{|D_{big}|}{|D|} \cdot H(D_{big}))$
$\frac{|D_{small}|}{|D|} = \frac{4}{7}, \quad \frac{|D_{big}|}{|D|} = \frac{3}{7}$

D_{small} would look as follows:

Color	Fur	Size	Class
brown	ragged	small	well-behaved
black	curly	small	well-behaved
white	curly	small	well-behaved
white	smooth	small	dangerous

Which means that $H(D_{smooth}) = -\frac{3}{4}\log_2(\frac{3}{4}) - \frac{1}{4}\log_2(\frac{1}{4}) = 0,811$

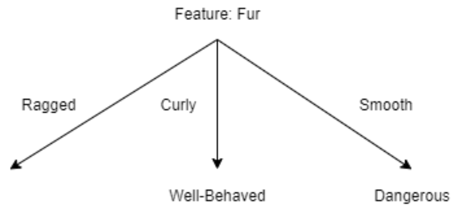
D_{big} would look as follows:

Color	Fur	Size	Class
black	ragged	big	dangerous
black	smooth	big	dangerous
red	ragged	big	well-behaved

Which means that $H(D_{big}) = -\frac{1}{3}\log_2(\frac{1}{3}) - \frac{2}{3}\log_2(\frac{2}{3}) = 0,918$

Entering those values into the computation of $iGain(D, Size)$ we get:
$iGain(D, Size) = 0,985 - (\frac{4}{7} \cdot 0,811 + \frac{3}{7} \cdot 0,918) = 0,128$

This means that we have the largest average information gain for $iGain(D, Fur)$ so we split D into D_{ragged}, D_{smooth} and D_{curly}. Our decision tree looks now as follows:



Since our decision tree is not complete yet we have to decide if it is better to split D_{ragged} for the attribute $Size$ or the attribute $Color$. This means we want to determine $iGain(D_{ragged}, Color)$ and $iGain(D_{ragged}, Size)$.

D_{ragged} would look as follows:

Color	Fur	Size	Class
brown	ragged	small	well-behaved
black	ragged	big	dangerous
red	ragged	big	well-behaved

$H(D_{ragged}) = 0,918$ (as computed before)

$iGain(D_{ragged}, Color) = H(D_{ragged}) -$
$\left(\frac{|D_{ragged,brown}|}{|D_{ragged}|} \cdot H(D_{ragged,brown}) + \frac{|D_{ragged,black}|}{|D_{ragged}|} \cdot H(D_{ragged,black}) + \frac{|D_{ragged,red}|}{|D_{ragged}|} \cdot H(D_{ragged,red})\right)$

$\frac{|D_{ragged,brown}|}{|D_{ragged}|} = \frac{1}{3}, \quad \frac{|D_{ragged,black}|}{|D_{ragged}|} = \frac{1}{3}, \quad \frac{|D_{ragged,red}|}{|D_{ragged}|} = \frac{1}{3}$

$D_{ragged,brown}$ would look as follows:

Color	Fur	Size	Class
brown	ragged	small	well-behaved

Which means that $H(D_{ragged,brown}) = -1\log_2(1) - 0\log_2(0) = 0$

$D_{ragged,black}$ would look as follows:

Color	Fur	Size	Class
black	ragged	big	dangerous

Which means that $H(D_{ragged,black}) = -0\log_2(0) - 1\log_2(1) = 0$

$D_{ragged,red}$ would look as follows:

Color	Fur	Size	Class
red	ragged	big	well-behaved

Which means that $H(D_{ragged,black}) = -1\log_2(1) - 0\log_2(0) = 0$

Entering those values into the computation of $iGain(D_{ragged}, Color)$ we get:
$iGain(D_{ragged}, Color) = 0,918 - (\frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 0) = 0,918$.

$iGain(D_{ragged}, Size) = H(D_{ragged}) - \left(\frac{|D_{ragged,small}|}{|D|} \cdot H(D_{ragged,small}) + \frac{|D_{ragged,big}|}{|D|} \cdot H(D_{ragged,big})\right)$
$\frac{|D_{ragged,small}|}{|D|} = \frac{1}{3}, \quad \frac{|D_{ragged,big}|}{|D|} = \frac{2}{3}$.

$D_{ragged,small}$ would look as follows:

Color	Fur	Size	Class
brown	ragged	small	well-behaved

Which means that $H(D_{ragged,small}) = -1\log_2(1) - 0\log_2(0) = 0$

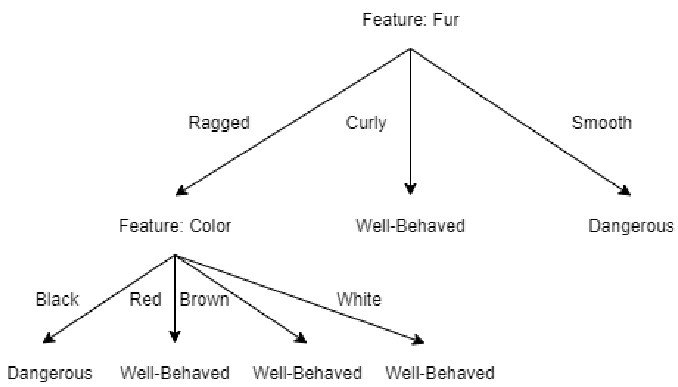
$D_{ragged,big}$ would look as follows:

Color	Fur	Size	Class
black	ragged	big	dangerous
red	ragged	big	well-behaved

Which means that $H(D_{ragged,big}) = -\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2}) = 1$

Entering those values into the computation of $iGain(D_{ragged}, Size)$ we get:
$iGain(D_{ragged}, Size) = 0,918 - (\frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1) = 0,251$.

This means that the attribute $Color$ is chosen for the new split since the average information gain is largest for the $Color$. Our decision tree now looks as follows:



(b) If we classify the new example we first check for the fur. Since the fur is ragged we next check for the color. Since the color is black we classify the new example as dangerous.

Exercise 5

c)

Approach 1 - Cross validation: In this approach, the data set is divided into several small training sets. The algorithm is then trained with a part of the training sets. Afterwards, the algorithm can be tested with the initial sets that are still unknown to it.

Approach 2 - Removing Features: In this approach, irrelevant features are removed from the dataset. This enhances the algorithm's ability to generalize.

Approach 3 - Early stopping: The algorithm is trained iteratively with an increasing number of runs and then the error is determined. Up to a certain point, the error usually decreases before it increases again and generalizability decreases. At this point, the training should be stopped.

Approach 4 - Using more examples for training. In case of images as input new examples can be generated by adding a bit of noise, flipping, rotating and slightly modifying the original images.