Wir suchen den ersten Knoten mit dem geringsten Gini-Index (minimale Verunreinigung)

OUTLOOK ist ein nominales Merkmal und kann drei Werte annehmen: Sunny/overcast/rainfall.

OUTLOOK	Yes	No	#
Sunny	2	3	5
Overcast	4	0	4
Rainfall	3	2	5
			14

$$Gini(Outlook|Sunny) = 1 - \sum p_i^2 = 1 - \left(\frac{2}{5}\right)^2 - \left(\frac{3}{5}\right)^2 = 0{,}48$$

$$Gini(Outlook|Overc) = 1 - \left(\frac{4}{4}\right)^2 = 0$$

$$Gini(Outlook|Rainfall) = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2 = 0{,}48$$

Die gewichtete Summe von Outlook:

$$Gini(Outlook) = \frac{5}{14} \cdot 0{,}48 + \frac{4}{14} \cdot 0 + \frac{5}{14} \cdot 0{,}48 = 0{,}342$$

TEMPERATURE	Yes	No	#
HOT	2	2	4
MILD	4	2	6
COOL	3	1	4
			14

$$Gini(Temp\ Hot) = 1 - \sum p_i^2 = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0'5$$

$$Gini(Temp\ Mild) = 1 - \sum p_i^2 = 1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2 = 0'445$$

$$Gini(Temp\ Cool) = 1 - \sum p_i^2 = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 0'375$$

$$Gini(Temp) = \frac{4}{14} \cdot 0'5 + \frac{6}{14} \cdot 0'445 + \frac{4}{14} \cdot 0'375 = 0'43$$

HUMIDITY	Yes	No	#
HIGH	3	4	7
NORMAL	6	1	7

$$Gini(Humidity\ High) = 1 - \sum p_i^2 = 1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2 = 0'49$$

$$Gini(Humidity\ Normal) = 1 - \sum p_i^2 = 1 - \left(\frac{6}{7}\right)^2 - \left(\frac{1}{7}\right)^2 = 0'24$$

$$Gini(Humidity) = \frac{7}{14} \cdot 0'49 + \frac{7}{14} \cdot 0'24 = 0'367$$

WIND	Ja	Nein	#
WEAK	6	2	8
STRONG	3	3	6

$$Gini(Wind\ Weak) = 1 - \left(\frac{6}{8}\right)^2 - \left(\frac{2}{8}\right)^2 = 0'375$$

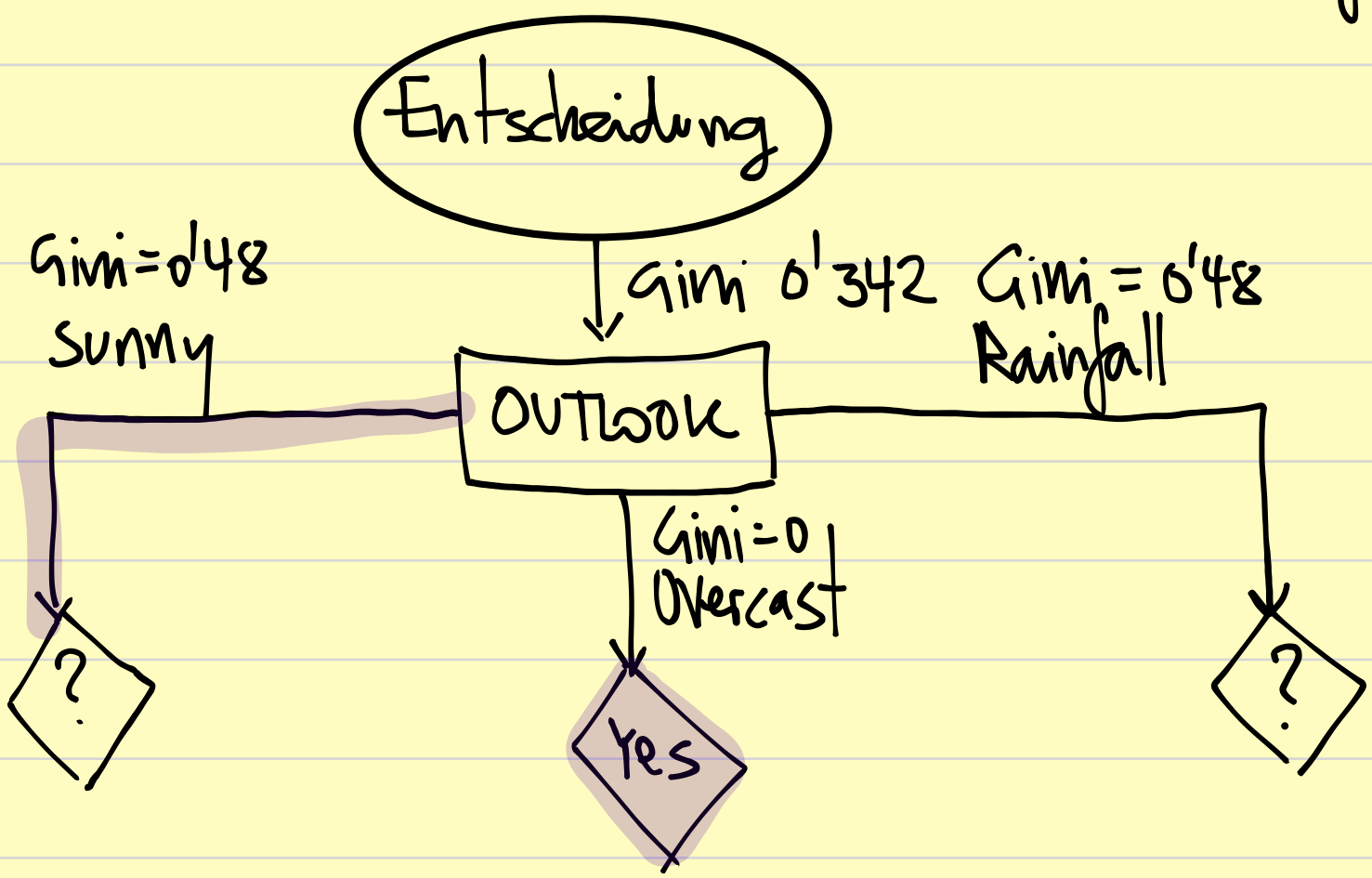
$$Gini(Wind\ Strong) = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = 0'5$$

$$Gini(Wind) = \frac{8}{14} \cdot 0'375 + \frac{6}{14} \cdot 0'5 = 0'428$$

Entscheidung:

Variablen	Gini
Outlook	0'342
Temp.	0'439
Humidity	0'367
Wind	0'428

→ Outlook hat die geringste Verunreinigung und wird als Knote gewählt.



Outlook SUNNY = { Temp. / Humi. / Wind.

SUNNY+Temp.	Ja	Nein	#
HOT	0	2	2
MILD	1	1	2
COOL	1	0	1
			5

$$Gini(Sunny+Temp.Hot) = 1-\left(\frac{0}{2}\right)^2-\left(\frac{2}{2}\right)^2 = 0$$

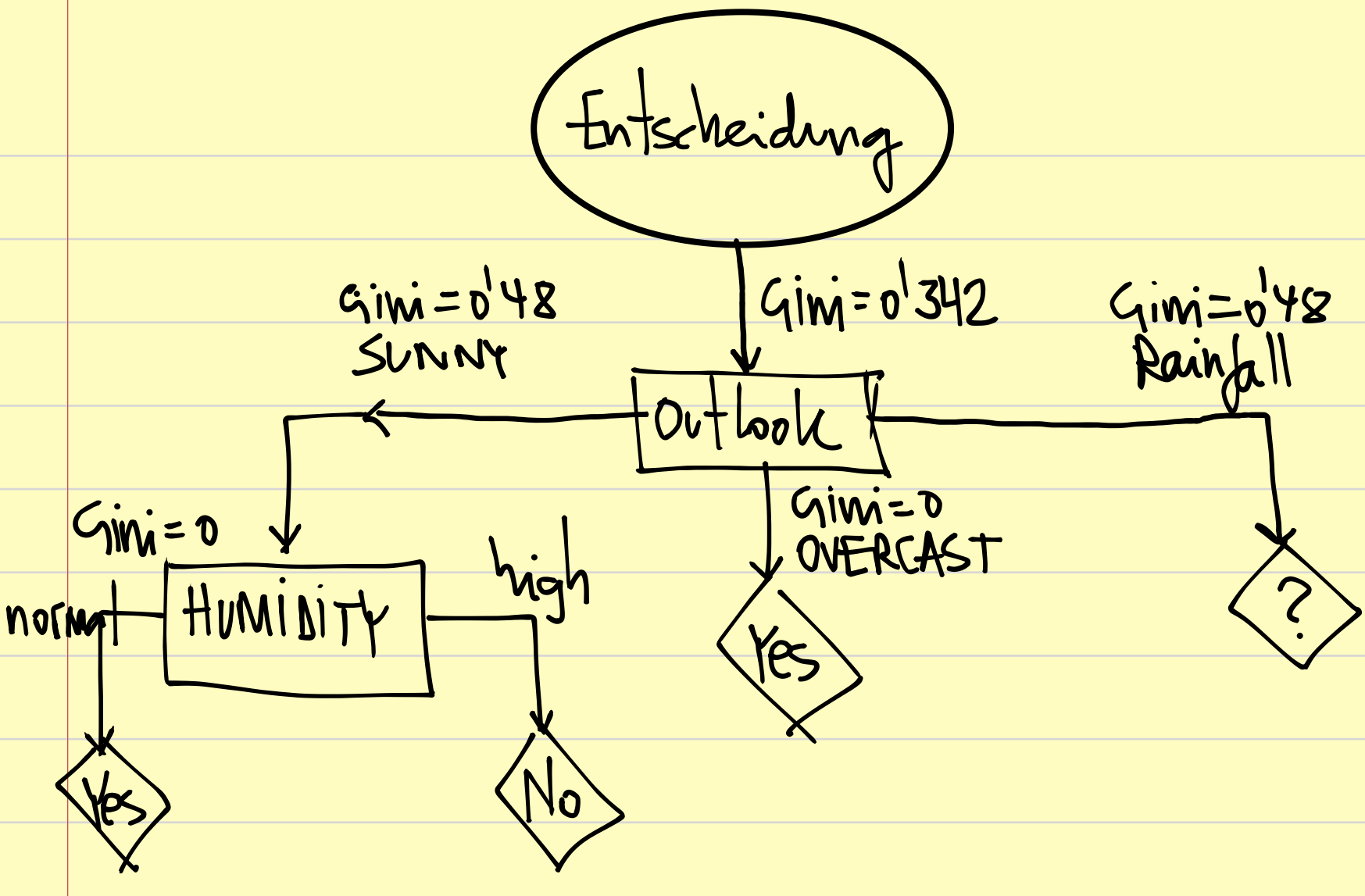
$$Gini(Sunny+TempMild) = 1-\left(\frac{1}{2}\right)^2-\left(\frac{1}{2}\right)^2 = 0{,}5$$

$$Gini(Sunny+TempCool) = 1-\left(\frac{1}{1}\right)^2-\left(\frac{0}{1}\right)^2 = 0$$

$$Gini(Sunny+Temp) = 0\cdot\frac{2}{5}+0{,}5\cdot\frac{2}{5}+0\cdot\frac{1}{5} = 0{,}2$$

Sunny+Humidity	Ja	Nein	#
High	0	3	3
NORMAL	2	0	2

$$Gini(Sunny+Humidity) = 0$$



Outlook Rainfall { Temperature / Wind

Rainfall + Temperature	Ja	Nein	#
HOT	1	1	2
MILD	2	1	3
COOL	0	0	0
			5

$$Gini(Rainfall + Temp\,Hot) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0'5$$

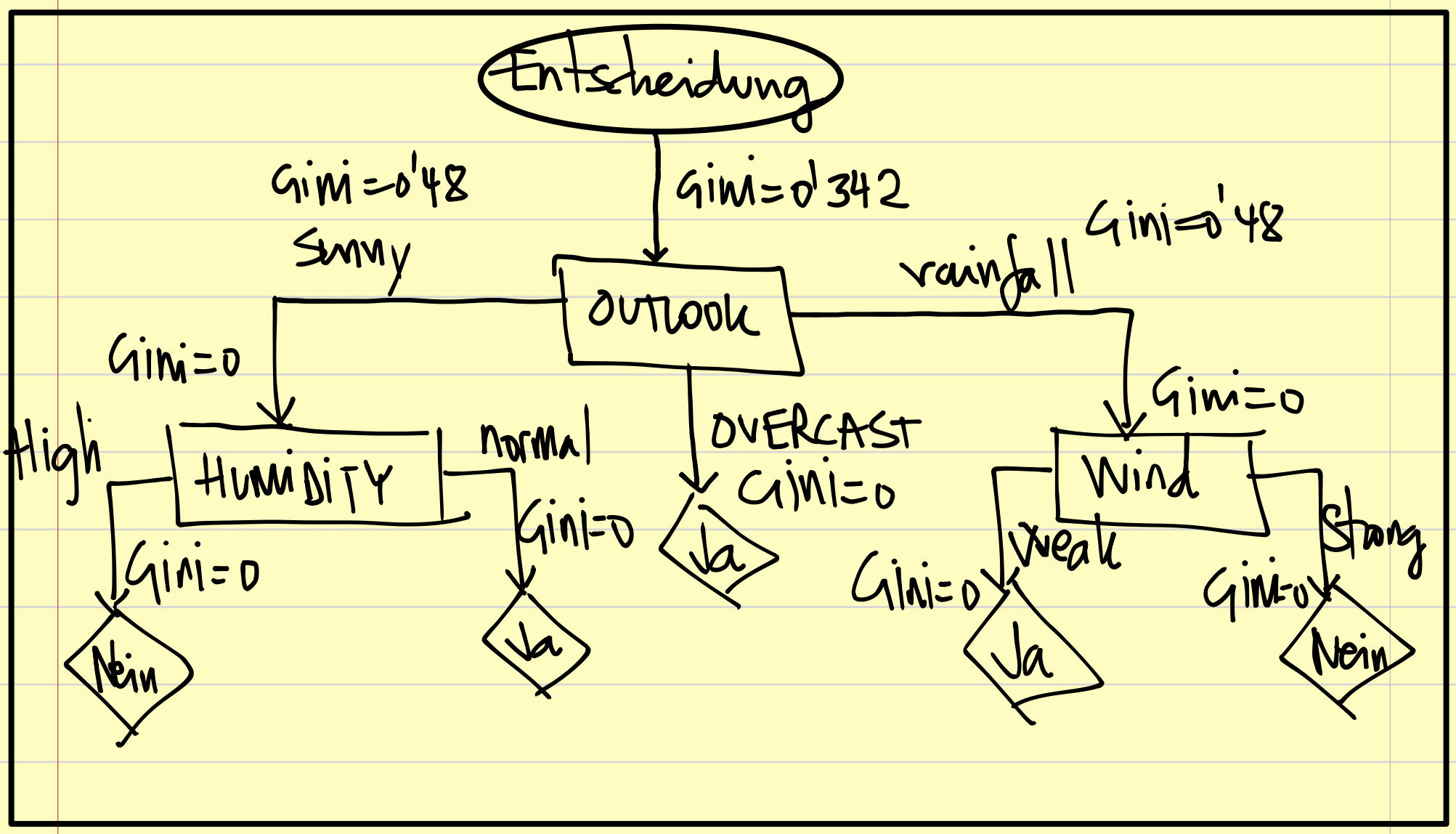
$$Gini(Rainfall + Temp\,Mild) = 1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2 = 0'44$$

$$Gini(Rainfall + Temp) = 0'5 \cdot \frac{2}{5} + 0'44 \cdot \frac{3}{5} = 0'467$$

Rainfall+Wind	Ja	Nein	#
WEAK	3	0	3
STRONG	0	2	2

$$Gini(Rainfall+Wind) = 0$$

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No



Beispiel. SEX Ja/Nein.

	Wohnungs. Verschmutzung	Sinnvolle Gespräche	fitness Niveau	Mond	Sex
1.	stark	oft	hoch	voll	Ja
2.	schwach	oft	gering	wachsend	Nein
3.	sauber	selten	hoch	voll	Ja
4.	stark	oft	mittel	abnehmend	Ja
5.	stark	selten	hoch	voll	Nein
6.	sauber	oft	hoch	wachsend	Ja
7.	schwach	oft	mittel	voll	Nein
8.	stark	oft	gering	voll	Ja
9.	schwach	selten	gering	neu	Ja
10.	sauber	oft	hoch	neu	Nein