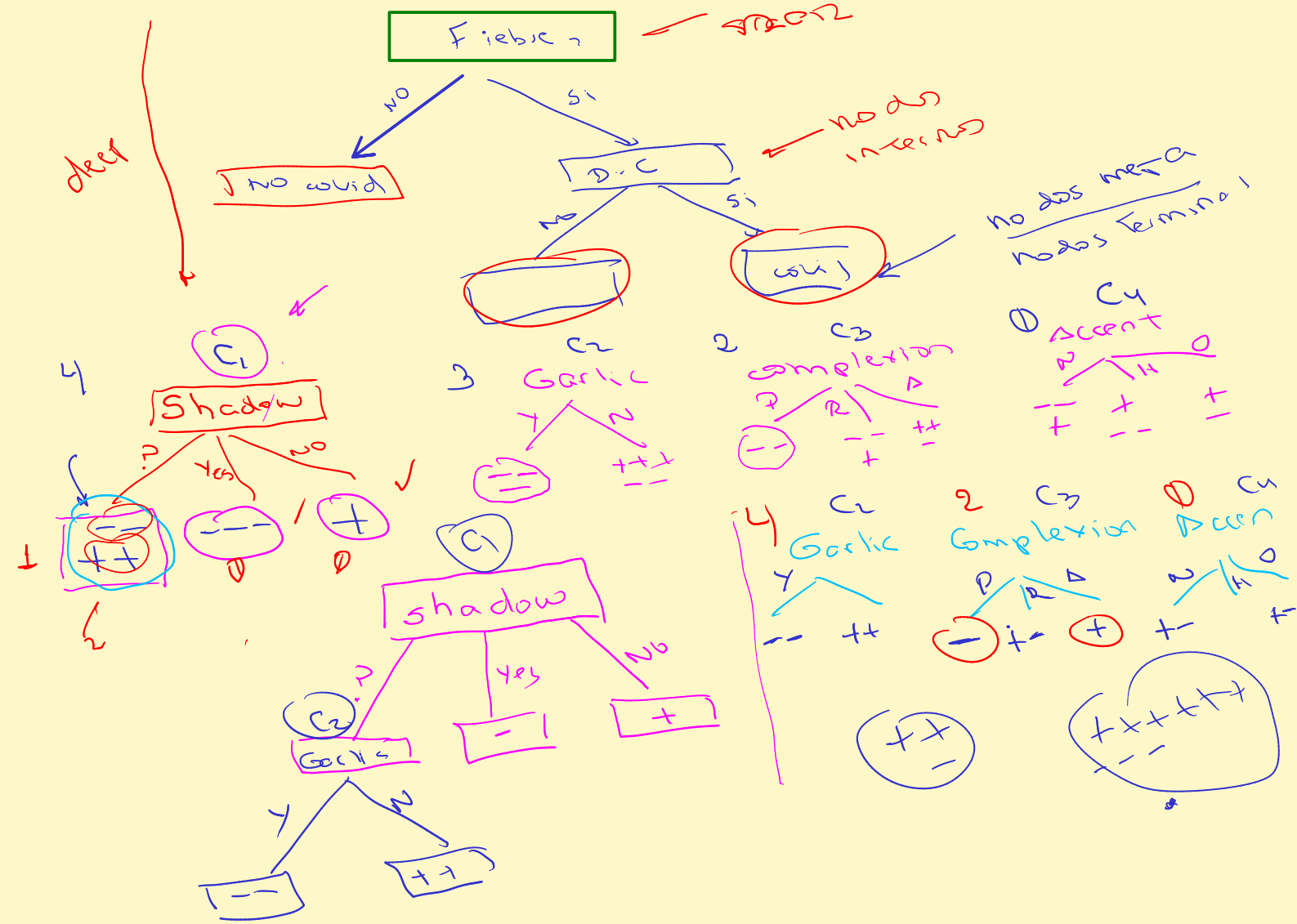
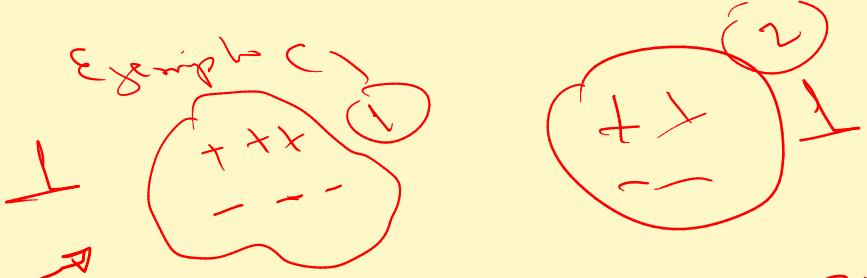
Romanian Data Base

	Vampire?	Shadow?	Garlic?	Complexion?	Accent?
P_1	No	?	Yes	Pale	None
P_2	No	Yes	Yes	Ruddy	None
	Yes	?	No	Ruddy	None
	Yes	No	No	Average	Heavy
	Yes	?	No	Average	Odd
	No	Yes	No	Pale	Heavy
	No	Yes	No	Average	Heavy
P_n	No	?	Yes	Ruddy	Odd



Exemple C)



$$D(Set) = -(P_+ \log P_+ + P_- \log P_- + P_\Delta \log P_\Delta)$$

$$= -\sum_{i \in C} P_i \log P_i$$

$$D(Set) = -P_+ \log P_+$$

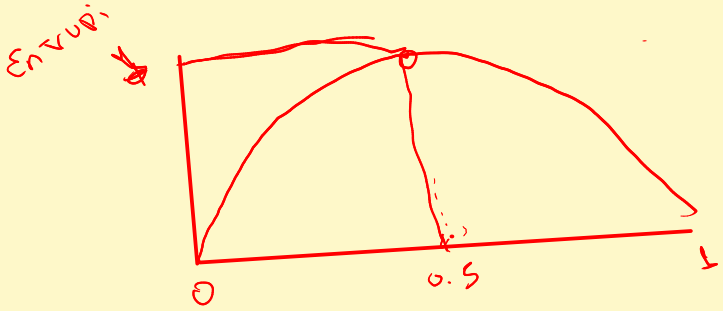
$$= -1 \log 1 = 0$$

$$D(Set) = -(P_+ \log P_+ + P_- \log P_-)$$

$$= -\left(\frac{1}{2}\log_2 2^{-1} + \frac{1}{2}\log_2 \frac{1}{2^2}\right)$$

$$= -\left[\frac{1}{2}\log_2 2^{-1} + \frac{1}{2}\log_2 2^{-1}\right]$$

$$= -\left[-\frac{1}{2} - \frac{1}{2}\right] = 1$$



Romanian Data Base

Vampire?	Shadow?	Garlic?	Complexion?	Accent?
No	?	Yes	Pale	None
No	Yes	Yes	Ruddy	None
Yes	?	No	Ruddy	None
Yes	No	No	Average	Heavy
Yes	?	No	Average	Odd
No	Yes	No	Pale	Heavy
No	Yes	No	Average	Heavy
No	?	Yes	Ruddy	Odd

$$D_f = \frac{4}{8} \cdot 1 + \frac{3}{8} \cdot 0 + \frac{1}{8} \cdot 0$$

$$D_{f_1} = \frac{1}{2} = 0.5$$

$$D_{f_2} = \frac{5}{8} \cdot 0.92 = 0.6$$

$$D_{f_3} = 0.92 \cdot (3/8) + 0.92(3/8)$$

$$D_{f_3} = 0.92\left(\frac{3}{4}\right) =$$

$$D_s = -\sum_{i \in C} P_c \log P_c$$

$$D_f = \frac{\#\text{ Elementos en grupo}}{\#\text{ Elementos Total}} \cdot D_s$$

$$D_f = \frac{|S|}{T} \cdot D_s$$

Tree (X, Y):

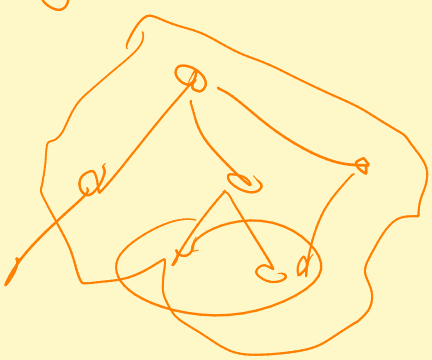
For fi in Features (X)

$D_s = 0$

For C in fi

$D_s = D_s - P_c \log P_c$

overfitting



$D_f = D_s \cdot \frac{|c|}{size(X)}$

If ($D_f = 0$)

return Nodo ( )

Else

Tree (,)

Gini — Entropia