Overfitting (S. 26 → ML3.pdf)

Tema 3, ex 4

H(Habitable)

$P(Hab) = \left[\frac{20+170+139+45}{800}, \frac{130+30+17+25}{800}\right] =$

$= \left[\frac{374}{800}, \frac{426}{800}\right]$

$H(\text{Habitable}) = \frac{374}{800}\log_2\frac{800}{374} + \frac{426}{800}\log_2\frac{800}{426} =$

$= 0,46\log_2 2,13 + 0,53\log_2 1,87 =$

$= 0,46 \cdot 1,09 + 0,53 \cdot 0,9 = 0,97$

H(Big)

$P(Hab \mid BIG = big) = \left[\frac{20+170}{350}, \frac{160}{350}\right] = \left[\frac{190}{350}, \frac{160}{350}\right]$

$= \left[\frac{19}{35}, \frac{16}{35}\right]$

$H(Hab \mid BIG = big) = \frac{19}{35}\log_2\frac{35}{19} + \frac{16}{35}\log_2\frac{35}{16}$

$= 0,54\log_2 1,84 + 0,45\log_2 2,18 =$

$= 0,54 \cdot 0,87 + 0,45 \cdot 1,12 =$

$= 0,46 + 0,5 = 0,96$

$P(Hab \mid BIG = small) = \left[\frac{184}{450}, \frac{266}{450}\right.$

$H(Hab|BIG=small) = 0,4 \cdot \log_2 2,44 + 0,59 \log_2 1,69 =$

$= 0,4 \cdot 1,28 + 0,59 \cdot 0,75 = 0,95$

$H(Big) = \frac{350}{800} \cdot 0,96 + \frac{450}{800} \cdot 0,95 = 0,43 \cdot 0,96 + 0,56 \cdot 0,95 =$

$= 0,94$

$IG(Hab : Big) = 0,97 - 0,94 = 0,03$

$H(Orbit)$

$P(ORbit = Near) = \left[\frac{20+139}{300}, \frac{147}{300}\right] = \left[\frac{159}{300}, \frac{147}{300}\right]$

Hab

$H(Or = Near) = 0,53 \log_2 1,88 + 0,47 \log_2 2,12 =$

$= 0,53 \cdot 0,91 + 0,47 \cdot 1,084 = 0,48 + 0,5 = 0,98$

$P(Or = Far) = \left(\frac{275}{500}, \frac{285}{500}\right)$

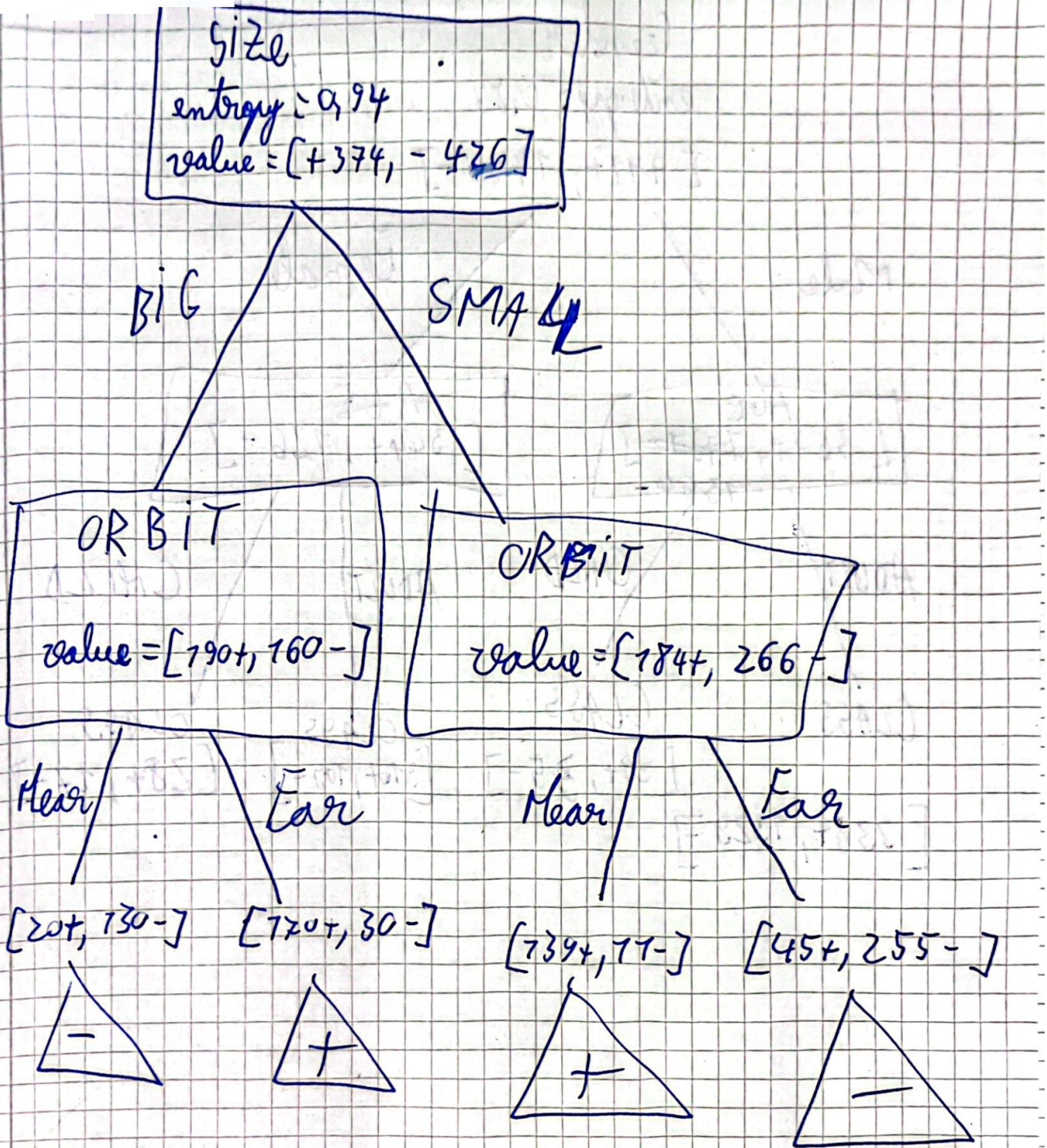
$H(Or = Far) = 0,43 \log_2 2,32 + 0,57 \log_2 1,75 =$

$= 0,43 \cdot 1,21 + 0,57 \cdot 0,8 = 0,52 + 0,45 = 0,97$

$H(Orbit) = \frac{300}{800} \cdot 0,98 + \frac{500}{800} \cdot 0,97 = 0,36 + 0,6 = 0,96$

$IG(Hab : Orbit = 0,97 - 0,96 = 0,01 < IG(Hab : Big)$

$\Rightarrow$ Big = root



training accuracy : $\frac{130 + 170 + 139 + 235}{800}$

$= \frac{674}{800} = 0,84 = 84\%$