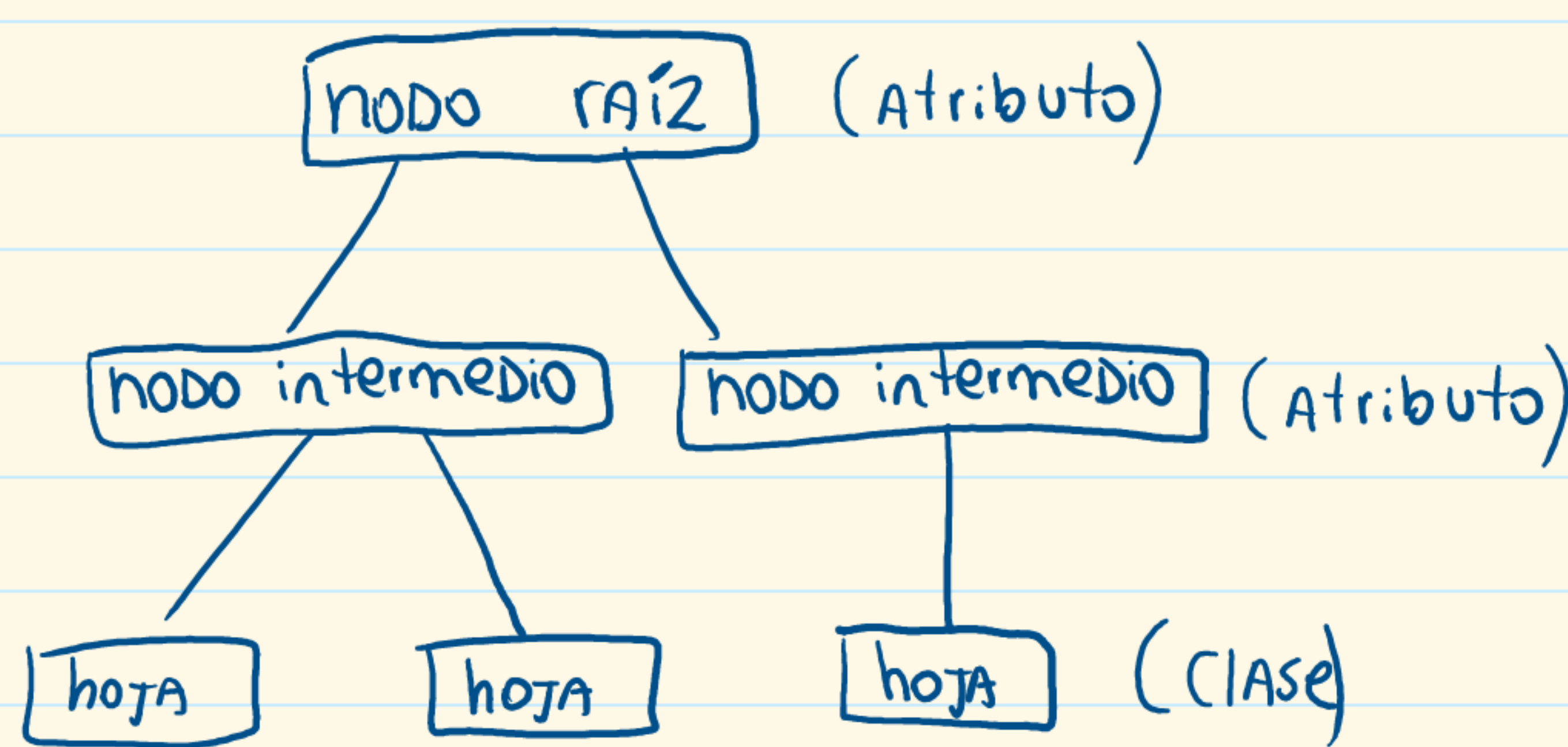


<https://drive.google.com/file/d/1VpgNysD5th2RrW1sXgWXmXHMqg6IOSH5/view?usp=drivesdk>

Control 4 Árboles de Clasificación

fases

input (Data) → **Clasificador** (Clasificador) → output (Clasificación)
Se asigna a una clase



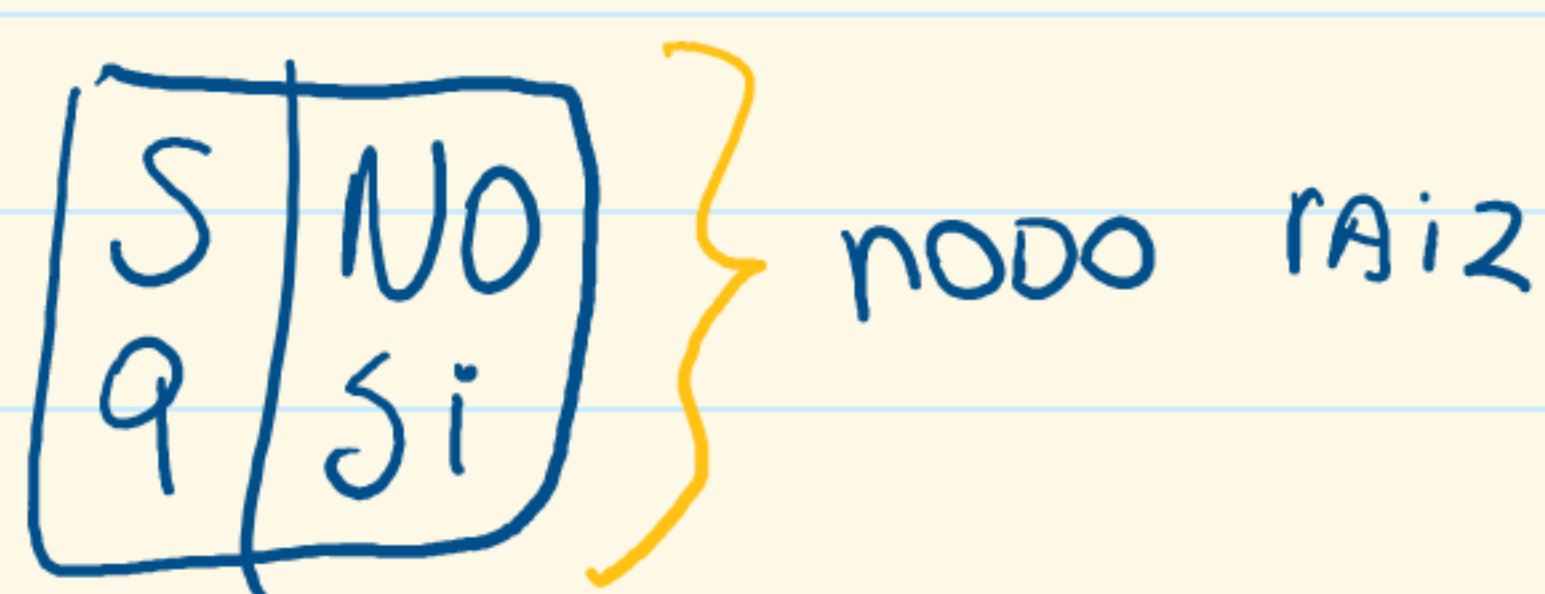
muestra de aprendizaje

le indico a la $f(x)$ cuales son las clasificaciones de cada elemento (iA)

Muestra de evaluación

no se conoce la clase y la $f(x)$ las tiene que clasificar. (Ahora él tiene que reconocer)

ejm: Jugar tenis (si/no)



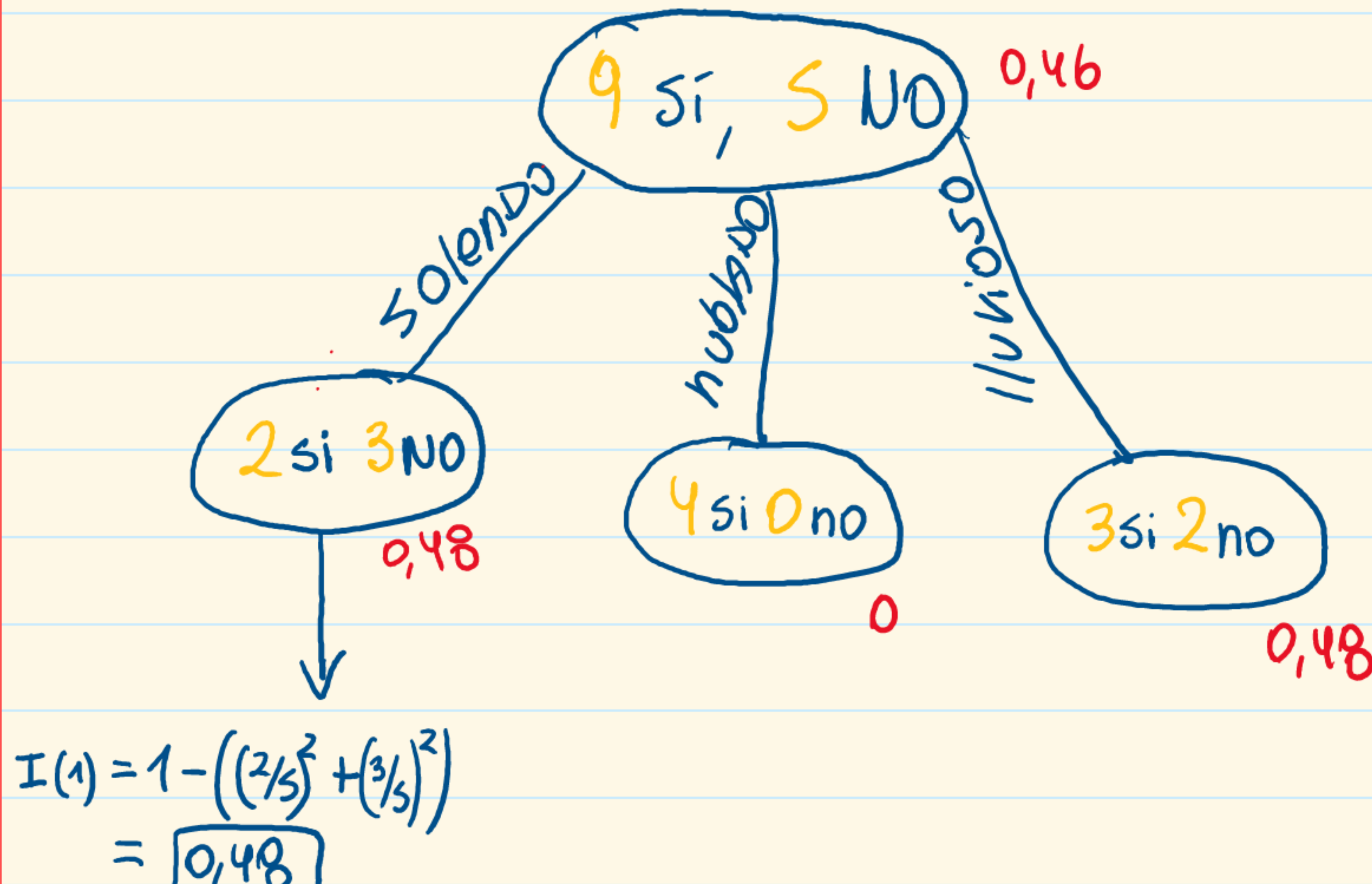
$N(K|t)$

$$N(Si|t) = 9 \quad \left. \begin{array}{l} p(Si|t) = 9/14 \\ p(NO|t) = 5/14 \end{array} \right\} p(Si|t) = 7/14 \quad \left. \begin{array}{l} p(Si|t) = 14/14 \\ p(NO|t) = 0/14 \end{array} \right\}$$

impureza nodo raíz (índice de gini)

$$I(RAIZ) = 1 - \left(\left(\frac{9}{14} \right)^2 + \left(\frac{5}{14} \right)^2 \right) = 0,46$$

ganancia de información

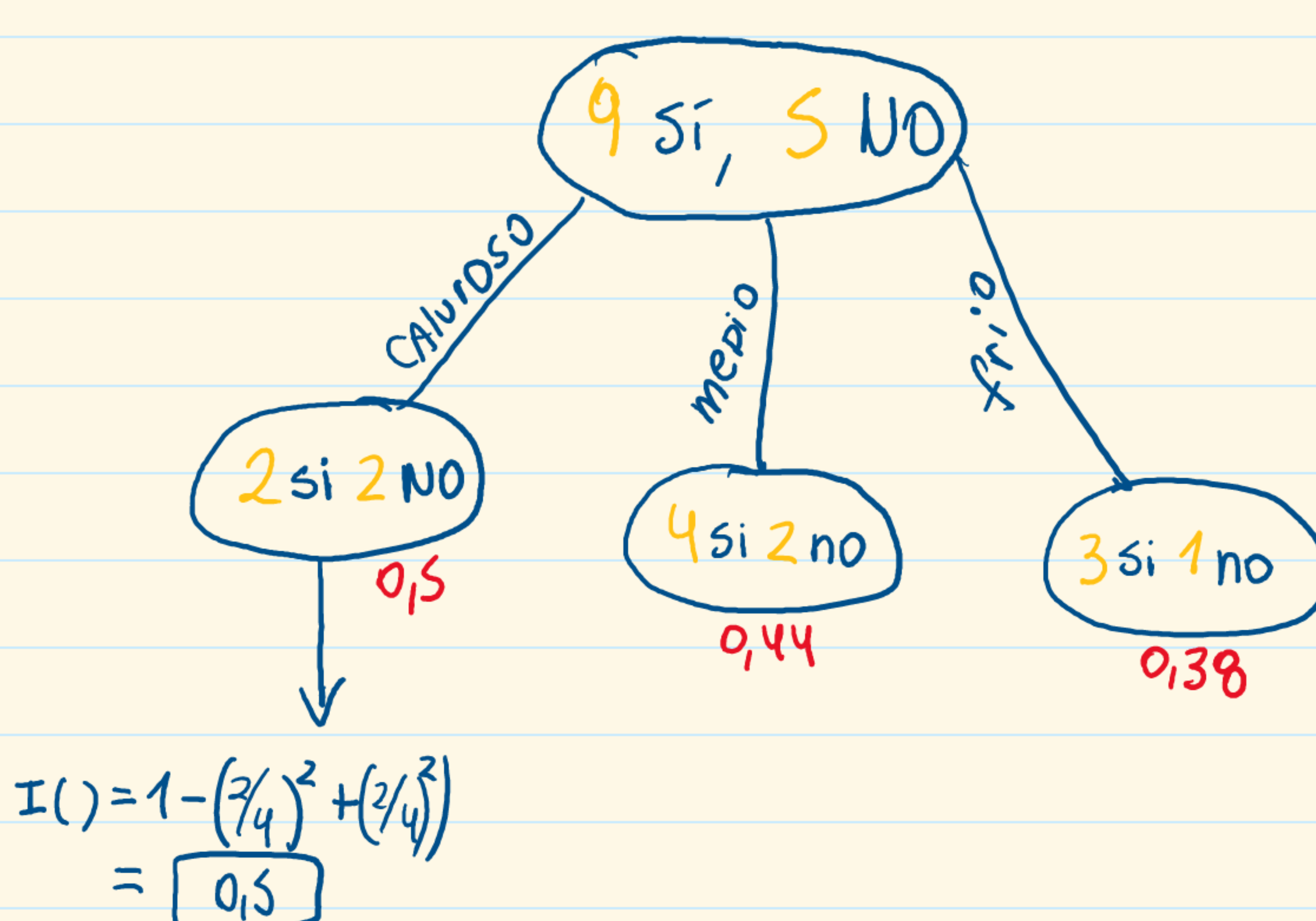


partición 1 (Clima)

$\Delta I(p_1) = I_{\text{padre}} - \text{pond hijo} \cdot \text{cant. datos hijo}$

$$0,46 - \frac{5}{14} \cdot (0,48) - \frac{4}{14} \cdot (0) - \frac{5}{14} (0,48)$$

$$= 0,12$$



partición 2 (Atributo +°)

$\Delta I(p_1) = I_{\text{padre}} - \text{pond hijo} \cdot \text{cant. datos hijo}$

$$0,46 - \frac{4}{14} \cdot (0,5) - \frac{6}{14} \cdot (0,44) - \frac{4}{14} (0,38)$$

$$= 0,02$$

