

Sandu Mihai Alexandru

matricola: 0321308

Domanda n° 2

1) Costruire l'albero di decisione (DT), per farlo dobbiamo prima scegliere un attributo che divida l'esempio in due sottoinsiemi, idealmente uno "tutti positivi" e uno "tutti negativi".

L'attributo che andiamo a prendere per creare un possibile Decision Tree è:

= 1

= 10

is Heavy → Edible : A, B (2)
 ↘ Not Edible: E, F, G (3)

C (1)
 D, H (2)

is Smelly → Edible : C (1)
 ↘ Not Edible: E, H (2)

A, B (2)
 D, F, G (3)

is Spotted → Edible : B (1)
 ↘ Not Edible: E, F (2)

A, C (2)
 D, G, H (3)

is Smooth → Edible : C (1)
 ↘ Not Edible: D, F, G (3)

A, B (2)
 E, H (2)

Prendiamo l'attributo ~~is Heavy~~ per costruire il DT:
 is Smelly

2) Calcolare l'Information Gain per ogni nodo del DT.

Per farlo prima ci dobbiamo calcolare l'entropia:

$$I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

per e il reminder (A)

Però $IG(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - \text{reminder}(A)$ dove A è l'attributo

~~Per l'attributo isHeavy~~

Per isHeavy abbiamo:

$$I_{\text{isHeavy } 1} = -\left(\frac{2}{5} \log_2 \frac{2}{5} + \frac{3}{5} \log_2 \frac{3}{5}\right) = 0,92 \quad \text{isHeavy} = 1$$

$$I_{\text{isHeavy } 0} = -\left(\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3}\right) = 0,91 \quad \text{isHeavy} = 0$$

Per isSmelly abbiamo:

$$I_{\text{isSmelly } 1} = -\left(\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3}\right) = 0,91 \quad \text{isSmelly} = 1$$

$$I_{\text{isSmelly } 0} = -\left(\frac{2}{5} \log_2 \frac{2}{5} + \frac{3}{5} \log_2 \frac{3}{5}\right) = 0,92 \quad \text{isSmelly} = 0$$

Per isSpotted abbiamo:

$$I_{\text{isSpotted } 1} = 0,91 \quad \text{isSpotted} = 1$$

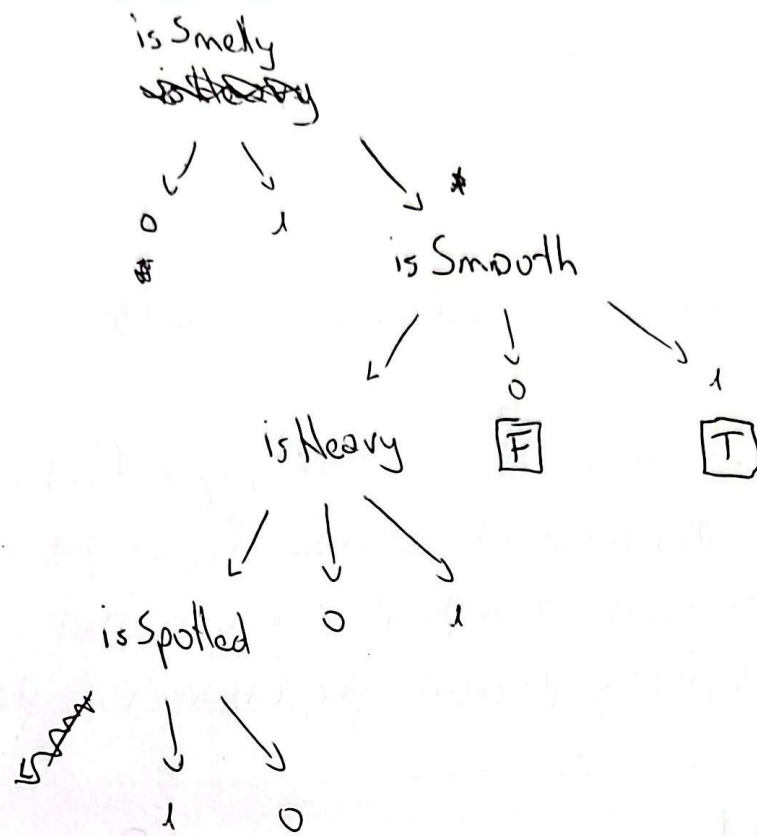
$$I_{\text{isSpotted } 0} = 0,92 \quad \text{isSpotted} = 0$$

Per isSmooth con lo stesso procedimento: isSmooth = 1

$$I_{\text{isSmooth } 1} = 0,90 \quad \text{isSmooth} = 1$$

$$I_{\text{isSmooth } 0} = 1$$

pk



Orz ci dobbiamo calcolare il reminder $(A) = \sum_{i=0}^n \frac{p_i + n_i}{p+n} I(A)$

$$\text{Reminder}(\text{isSmooth}) = \left(\frac{4}{8} \cdot 1\right) + \left(\frac{4}{8} \cdot 0,9\right) = 0,90$$

$$I_g(\text{isSmooth}) = \text{Entropia}(\text{isSmooth}) - \text{Reminder}(\text{isSmooth}) = 0,1$$

Applichiamo ora lo stesso procedimento per ~~tutte~~ tutti ~~gli~~ gli altri nodi del DT

3) Per calcolare ~~la~~ ~~precisione~~ ~~la~~ ~~precisione~~ usiamo le formule:

$$\text{Precision}(P) = \frac{TP}{TP + FP}$$

$$\text{Recall}(R) = \frac{TP}{TP + FN}$$

$$\text{F-measure}(F_1) = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Dove per TP abbiamo le istanze positive che ~~sono~~ devono essere positive, per FP abbiamo i falsi positivi e per FN i falsi negativi