

ESERCIZIO ALBERO CLASSIFICAZIONE

$$\log_2(1) = 0$$

$$IG = 1 - \left[\frac{\text{right class}}{\text{right tot}} H\left(\frac{T}{T+F}, \frac{F}{T+F}\right) \right] \rightarrow \text{quanti True o False di quella classe}$$

$$\frac{\log N}{\log 2}$$

$$= 1 - \left[\frac{\text{right class}}{\text{right tot}} \cdot \left(-\frac{T_{\text{right}}}{T+F} \log_2 \left(\frac{T_{\text{right}}}{T+F} \right) - \frac{F_{\text{right}}}{T+F} \log_2 \left(\frac{F_{\text{right}}}{T+F} \right) \right]$$

TIPS: Per capire quale attributo è rappresentato meglio, vedere se la sua classe corrisponde a TRUE o FALSE (sempre).
Scegliere quel R' attributo che permette perfetta distinzione in classi.
E che permette in proporzione alle classi presentare migliore ^{CLASSI DISTINTE} risultato.

1° STEP

$$H\left(\frac{T}{T+F}, \frac{F}{T+F}\right) = H\left(\frac{6}{12}, \frac{6}{12}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{2} \log_2 \left(\frac{1}{2}\right) = +0.5 + 0.5 = 1$$

$$IG(\text{Patraw}) = 1 - \left[\frac{\text{wave}}{12} H\left(0, \frac{2}{2}\right) + \frac{\text{row}}{12} H\left(\frac{4}{4}, 0\right) + \frac{1}{2} \frac{\text{pale}}{12} H\left(\frac{2}{6}, \frac{4}{6}\right) \right]$$

$$= 1 - \left[0 + 0 + \frac{1}{2} H\left(\frac{1}{3}, \frac{2}{3}\right) \right]$$

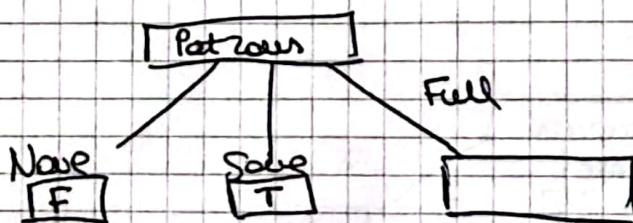
$$= 1 - \left[\frac{1}{2} \left(-\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right) \right) \right]$$

$$= 1 - \left[\frac{1}{2} \left(-\frac{1}{3} (-1.58) - \frac{2}{3} (-0.58) \right) \right] =$$

$$= 1 - \left[\frac{1}{2} (0.53 + 0.38) \right] = 0.54$$

~~No calcolo~~
Basta cancellare wave, row e pale.

(se non lo capisci da qui, allora calcola poi IG)



2° STEP:

Considera solo le regole riguarde e reeffettua i calcoli di H_{doget} e di $\text{gain information gain}$.

$$\underline{H_{\text{doget}}}\left(\frac{2}{6}, \frac{4}{6}\right) = H\left(\frac{1}{3}, \frac{2}{3}\right) = -\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) \\ = 0,53 + 0,39 = \underline{0,92}$$

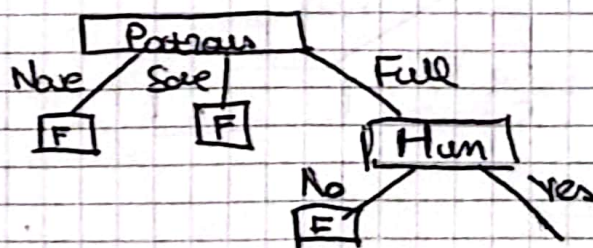
$$\begin{aligned} IG_{\text{HUN}} &= 0,92 - \left[\frac{4}{6} H_{\text{True}}\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{2}{6} H_{\text{False}}(0,1) \right] = \\ &= 0,92 - \left[\frac{2}{3} H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{3} H(0,1) \right] = \\ &= 0,92 - \left[\frac{2}{3} \left[-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] \right] \\ &= 0,92 - \left[\frac{2}{3} (0,5 + 0,5) \right] = \underline{0,25} \end{aligned}$$

$$\begin{aligned} IG_{\text{PRICE}} &= 0,92 - \left[\frac{4}{6} H\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{2}{6} H(0,1) \right] \\ &= 0,25 \end{aligned}$$

$$IG_{\text{RES}} = 0,92 - \left[\frac{2}{6} H(0,1) + \frac{4}{6} H\left(\frac{2}{4}, \frac{2}{4}\right) \right] = 0,25$$

Se erano uguali gli (IG), allora scegliere mo a caso tra quelli più alti per costruire l'albero.

Scegli HUN



Cancellare la regola che non correttamente rappresenta.

3° STEP: Calcolare leigh separate e ricomporre (IG)

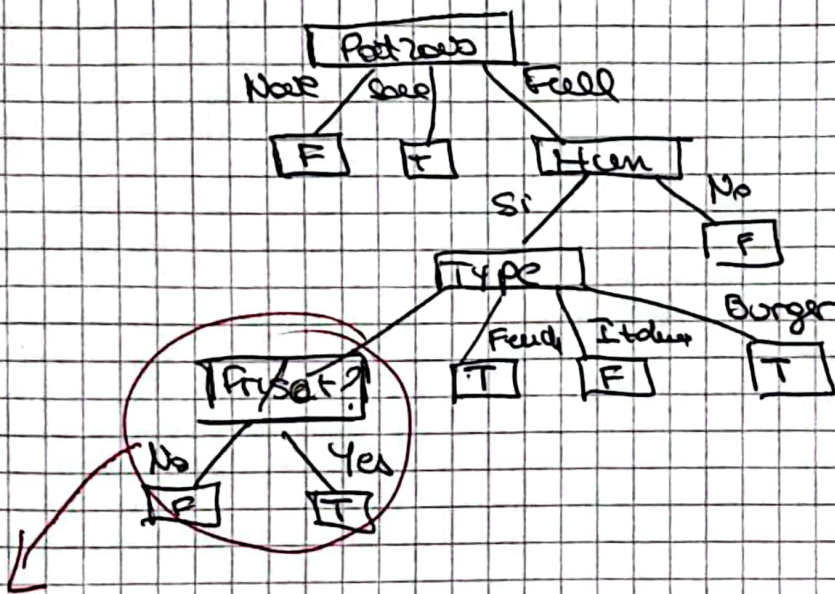
$$H\left(\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

$$IG_{type} = \frac{1}{4} H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{4} H(0, 1) + \frac{1}{4} H(1, 0) = 1 - \frac{1}{2} = 0.5$$

N.B

Per capire quale correttamente rappresentato in quale essere ^{5°} classe che corrisponda ad una diversa T o F.
Se \exists solo $Thi = T \rightarrow$ allora correttamente rappresent

Caratteristiche Italiane Burger, correttamente rappresentate



4° STEP:

Considero solo le due leigh.
E leigh effettuate calcolo vedo quale attributo mi permette di ~~bettere~~ distinguere due classi.

Se ce ne sono di più, con lo stesso IG, allora scegliere uno dei due convenienti.

N.B

Per scegliere attributo, oltre al calcolo di IG, devi considerare un trucco:

SCEGLIERE ATTRIBUTO LE QUALI CLASSI SONO SEPARATE T o F RISPETTO ALLA VARIABILE TARGET.

\Rightarrow CHE PERMETTE DI CLASSIFICARE PIÙ CLASSI.

CASO PARTICOLARE \rightarrow 3 CLASSI: 1 IMPRES CORIST.

\rightarrow A PARTIRE DA \rightarrow considero leigh di una classe \rightarrow calcolo IG degli attributi
 \rightarrow A PARTIRE DA \rightarrow considero leigh dell'altra classe \rightarrow calcolo IG degli altri attributi

P.S. Se rimane una classe \rightarrow con T o F \rightarrow quella è classificata correttamente.

~~da capire come si calcola~~

12 = n° delle tabella

Building a Classification Tree

$$H\left(\frac{t}{t+f}, \frac{f}{t+f}\right) = -\frac{6}{12} \cdot \log_2 \frac{6}{12} - \frac{6}{12} \cdot \log_2 \frac{6}{12} = 1$$

$$IG(Patrons) = 1 - \left[\frac{2}{12} \cdot H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{12} \cdot H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{6}{12} \cdot H\left(\frac{2}{6}, \frac{4}{6}\right) \right] = 1 - \left[\frac{2}{12} \cdot (0 \cdot \log_2 0 - 1 \cdot \log_2 1) \dots \right]$$

$$IG(Type) = 1 - \left[\frac{2}{12} \cdot H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{12} \cdot H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{12} \cdot H\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{4}{12} \cdot H\left(\frac{2}{4}, \frac{2}{4}\right) \right] = 0$$

Select the attribute with the highest gain

PATRONS

Selctare e' attributo con il massimo MAGGIORE (16)

ULTIMA PAGINA

Example	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	Target
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	F
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	F
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T

CALCOLO E SPIEGAZIONE

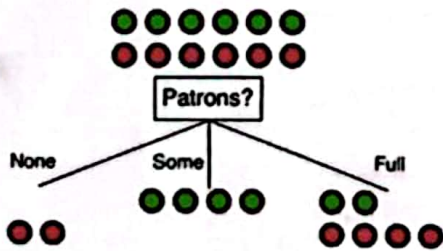
PRECEDENTE: H(y) = 1

Building a Classification Tree

$$IG(x_{\text{new feature}}) = 1 - [\dots]$$

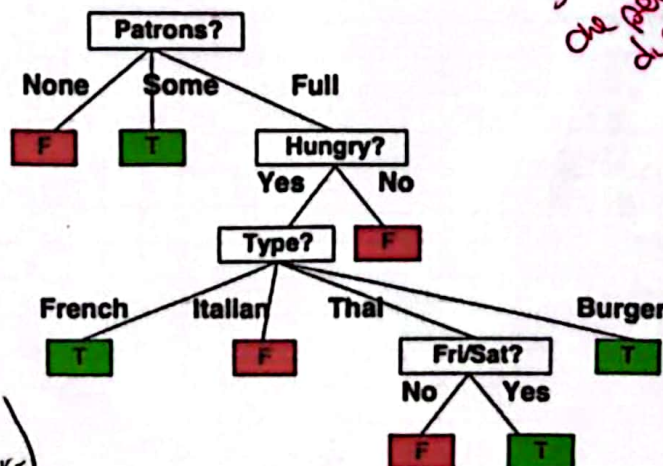
scegli (6) + 120 → calcola con i coefficienti rappresentati e ridotti

$$\frac{\log(6)}{\log(12)}$$



Example	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	Target
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	F
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	F
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T

Building a Classification Tree



Scegliere la FEATURE + INFORMATION (scegliere non importa perché cause OVERFITTING)

non è rappresentata correttamente sceglilo con la che permette di classificarla