PER ALTRI APPUNTI CONSULTARE IL SITO:

https://luigi-v.github.io/Appunti_Universita/

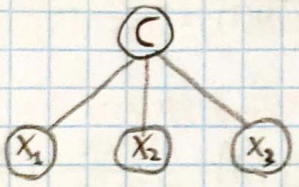
Per altri appunti consultare il sito: *https://luigi-v.github.io/Appunti_Universita/*

By *Vollono Luigi*

ESERCIZIO MONETE (RETE BAYESIANA E CPT)

1. MONETA ESTRATTA CASUALMENTE →

3 RISULTATI POSSIBILI →



CPT PER C

C	P(C)
a	1/3
b	1/3
c	1/3

CPT DI OTTENERE TESTA

C	X_1	P(C)
a	TESTA	0.2
b	TESTA	0.6
c	TESTA	0.8

2. APPLICARE LA REGOLA DI BAYES → $P(\text{CAUSE} \mid \text{EFFECT}) = P(\text{EFFECT} \mid \text{CAUSE})\, P(\text{CAUSE}) / P(\text{EFFECT})$

$$P(\underset{\text{CAUSA}}{C} \mid \underset{\text{EFFETTI}}{2\ \text{TESTA}\ 1\ \text{CROCE}}) = P(2\ \text{TESTA}\ 1\ \text{CROCE} \mid C)\, P(C) / P(2\ \text{TESTA}\ 1\ \text{CROCE})$$

DENOMINATORE DIVENTA α E' FATT. DI NORMALIZZAZIONE

$$= \alpha\, P(2\ \text{TESTA}\ 1\ \text{CROCE} \mid C)\, P(C)$$

$$= \alpha\, P(2\ \text{TESTA}\ 1\ \text{CROCE} \mid C)$$

PROB. CHE SIA LA MONETA a:

$$P(X_1 = \text{CROCE}, X_2 = \text{TESTA}, X_3 = \text{TESTA} \mid C = a) = P(X_1 = \text{CROCE} \mid C = a)\, P(X_2 = \text{TESTA} \mid C = a)\, P(X_3 = \text{TESTA} \mid C = a) =$$

$$= 0.8 \cdot 0.2 \cdot 0.2 = 0.032 \xrightarrow{\times 3\ \text{monete}} 0.032 \cdot 3 = 0.096$$

PROB. CHE SIA LA MONETA b:

$$P(X_1 = \text{croce}, X_2 = \text{testa}, X_3 = \text{testa} \mid C = b) = P(X_1 = \text{croce} \mid C = b)\, P(X_2 = \text{testa} \mid C = b)\, P(X_3 = \text{testa} \mid C = b) =$$

$$= 0.4 \cdot 0.6 \cdot 0.6 = 0.144 \rightarrow 0.144 \cdot 3 = 0.432$$

PROBAB. CHE SIA LA MONETA c:

$$P(X_1 = \text{croce}, X_2 = \text{testa}, X_3 = \text{testa} \mid C = c) =$$

$$= 0.2 \cdot 0.8 \cdot 0.8 = 0.128 \times 3 = 0.384$$

DTL TENNIS:

DAY	OUTLOOK	TEMP	HUMIDITY	WIND	PLAY
D1	SUNNY	HOT	HIGH	WEAK	NO
D2	SUNNY	HOT	HIGH	STRONG	NO
D3	OVERCAST	HOT	HIGH	WEAK	YES
D4	RAIN	MILD	HIGH	WEAK	YES
D5	RAIN	COOL	NORMAL	WEAK	YES
D6	RAIN	COOL	NORMAL	STRONG	NO
D7	OVERCAST	COOL	NORMAL	STRONG	YES
D8	SUNNY	MILD	HIGH	WEAK	NO
D9	SUNNY	COOL	NORMAL	WEAK	YES
D10	RAIN	MILD	NORMAL	WEAK	YES
D11	SUNNY	MILD	NORMAL	STRONG	YES
D12	OVERCAST	MILD	HIGH	STRONG	YES
D13	OVERCAST	HOT	NORMAL	WEAK	YES
D14	RAIN	MILD	HIGH	STRONG	NO

(2) CALCOLO ENTROPIA DELL'ATTRIBUTO TARGET (PLAY):

$$I(\text{TARGET}) = \frac{N^\circ SI}{N^\circ TOT} \cdot \log_2 \frac{N^\circ TOT}{N^\circ SI} + \frac{N^\circ NO}{N^\circ TOT} \cdot \log_2 \frac{N^\circ TOT}{N^\circ NO}$$

(2) CALCOLO REMINDER DEGLI ALTRI ATTRIBUTI PER TROVARE LA ROOT:

$$\text{REMINDER}(A) = \frac{N^\circ 1^\circ VAL}{N^\circ TOT\ VAL} \cdot \left(\frac{N^\circ SI}{N^\circ 1^\circ VAL} \cdot \log_2 \frac{N^\circ 1^\circ VAL}{N^\circ SI} + \frac{N^\circ NO}{N^\circ 1^\circ VAL} \log_2 \frac{N^\circ 1^\circ VAL}{N^\circ NO} \right) + \frac{N^\circ 2^\circ VAL}{N^\circ TOT\ VAL} \cdot \left(\frac{N^\circ SI}{N^\circ 2^\circ VAL} \cdot \log_2 \frac{N^\circ 2^\circ VAL}{N^\circ SI} + \frac{N^\circ NO}{N^\circ 2^\circ VAL} \cdot \log_2 \frac{N^\circ 2^\circ VAL}{N^\circ NO} \right) + \dots$$

(2.1) CALCOLO GUADAGNO DEGLI ATTRIBUTI

$$\text{GAIN}(A) = I(\text{TARGET}) - \text{REMINDER}(A)$$

(3) CALCOLO ENTROPIA FOGLIA, REMINDER CONDIZIONATO PER RAMIFICARE

$$I(A=VAL1) = \frac{N^\circ SI \text{ DI } A=VAL1}{N.TOT \text{ DI } A=VAL1} \cdot \log_2 \frac{N^\circ TOT \text{ DI } A=VAL1}{N^\circ SI \text{ DI } A=VAL1} + \text{CON "NO" DI } A=VAL1$$

$$R(B|A=VAL1) = \frac{N^\circ B=VAL1 \text{ DI } A}{N^\circ TOT\ A=VAL1} \cdot \left(\frac{N^\circ SI \text{ "} B=VAL1}{N^\circ B=VAL1 \text{ DI } A} \log_2 \frac{N^\circ B=VAL1 \text{ DI } A}{N^\circ SI \text{ "} B=VAL1} + \text{CON "NO"} \right)$$

1. ENTROPIA DI PLAY

$$I(PLAY) = \frac{9}{14} \cdot \log_2 \frac{14}{9} + \frac{5}{14} \cdot \log_2 \frac{14}{5} = 0.64 \cdot 0,64 + 0.36 \cdot 1.48 = 0.95$$

2. REMINDER DEGLI ALTRI ATTRIBUTI

$$R(OUTLOOK) = \frac{5}{14} \cdot \left(\frac{2}{5} \log_2 \frac{5}{2} + \frac{3}{5} \cdot \log_2 \frac{5}{3} \right) + \frac{4}{14} \left(\frac{4}{4} \log_2 \frac{4}{4} + 0 \right) + \frac{5}{14} \left(\frac{3}{5} \log_2 \frac{5}{3} + \frac{2}{5} \log_2 \frac{5}{2} \right) =$$

$$= 0.36 \left((0.4 + 1,32) + (0.6 \cdot 0.74) \right) + 0.36 \left((0.6 \cdot 0.74) + (0.4 \cdot 1,32) \right) = \underline{0,70} \text{ MIGLIORE}$$

$$R(TEMP) = \frac{4}{14} \cdot \left(\frac{2}{4} \log_2 \frac{4}{2} + \frac{2}{4} \log_2 \frac{4}{2} \right) + \frac{6}{14} \left(\frac{4}{6} \log_2 \frac{6}{4} + \frac{2}{6} \log_2 \frac{6}{2} \right) + \frac{4}{14} \left(\frac{3}{4} \log_2 \frac{4}{3} + \frac{1}{4} \log_2 \frac{4}{1} \right) =$$

$$= 0,36 \left((0.5 \cdot 1) + (0.5 \cdot 1) \right) + 0.43 \left((0.67 \cdot 0.58) + (0.34 \cdot 1.58) \right) + 0.36 \left((0.75 \cdot 0.42) + (0.25 \cdot 2 \right)$$

$$\approx 0.36 + 0.43 + 0,29 \approx 0,92$$

$$R(HUMIDITY) = \frac{7}{14} \left(\left(\frac{3}{7} \log_2 \frac{7}{3} \right) + \left(\frac{4}{7} \log_2 \frac{7}{4} \right) \right) + \frac{7}{14} \left(\left(\frac{6}{7} \log_2 \frac{7}{6} \right) + \left(\frac{1}{7} \log_2 7 \right) \right) = 0,92$$

$$R(WIND) = \frac{8}{14} \left(\left(\frac{6}{8} \log_2 \frac{8}{6} \right) + \left(\frac{2}{8} \log_2 \frac{8}{2} \right) \right) + \frac{6}{14} \left(\left(\frac{3}{6} \log_2 \frac{6}{3} \right) + \left(\frac{3}{6} \log_2 \frac{6}{3} \right) \right) = 0.92$$

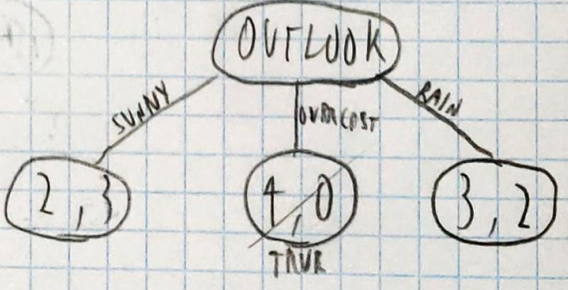
2.2 GUADAGNO DEGLI ATTRIBUTI

$$GAIN(OUTLOOK) = 0.95 - 0,70 = \underline{0.25} \text{ SI PRENDE OUTLOOK}$$

$$GAIN(TEMP) = 0.95 - 0.92 = 0.03$$

$$GAIN(HUMIDITY) = 0.95 - 0.92 = 0.03$$

$$GAIN(WING) = \text{//} \approx 0.03$$



3. CALCOLO ENTROPIA PONDERATA

$I(\text{OUTLOOK}=\text{SUNNY}) = \left(\frac{2}{5}\log_2\frac{2}{2}\right) + \left(\frac{3}{5}\log_2\frac{3}{3}\right) = (0.4 \cdot 1.32) + (0.6 \cdot 0.74) = 0,97$

3.2 CALCOLO REMINDER CONDIZIONATO

$R(\text{TEMP} \mid \text{OUTLOOK}=\text{SUNNY}) = \frac{2}{5}\left(\frac{2}{2}\log_2\frac{2}{2} + 0\right) + \frac{2}{5}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) + \frac{1}{5}\left(1\log_2 1 + 0\right) =$

$= 0.4\,((0.5 \cdot 1) + (0.5 \cdot 1)) = 0.4$

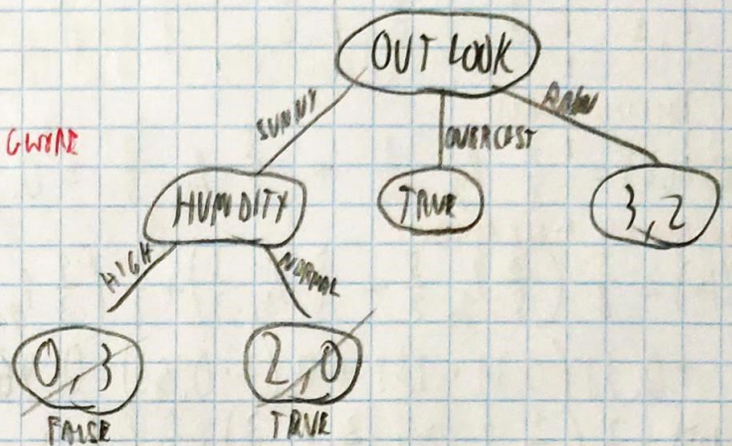
$R(H \mid O=\text{SUNNY}) = \frac{3}{5}\left(\frac{0}{3}\log\frac{3}{0} + \frac{3}{3}\log\frac{3}{3}\right) + \frac{2}{5}\left(\frac{0}{2}\log_2\frac{2}{0} + \frac{0}{2}\log_2\frac{2}{0}\right) = 0$

$R(W \mid O=\text{SUNNY}) = \frac{3}{5}\left(\frac{1}{3}\log\frac{3}{1} + \frac{2}{3}\log\frac{3}{2}\right) + \frac{2}{5}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) = (>0)$

$\text{GAIN}(T \mid O=\text{SUNNY}) = 0.97 - 0.4 < 0.97$

$\text{GAIN}(H \mid O=\text{SUNNY}) = 0.97 - 0 = \underline{0.97}$ MIGLIORE

$\text{GAIN}(W \mid O=\text{SUNNY}) < 0.97$



$R(T \mid O=\text{RAIN}) = \frac{3}{5}\left(\frac{2}{3}\log_2\frac{3}{2} + \frac{1}{3}\log_2 3\right) + \frac{2}{5}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) = 0.6\,(0.67 \cdot 0.58 +$

$+ 0.34 \cdot 1.58) + 0.4(1) = 0.95$

$R(H \mid O=\text{RAIN}) = \frac{2}{5}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) + \frac{3}{5}\left(\frac{2}{3}\log_2\frac{3}{2} + \frac{1}{3}\log_2 3\right) = 0.4 + 0.6\,(0.39 + 0.53) =$

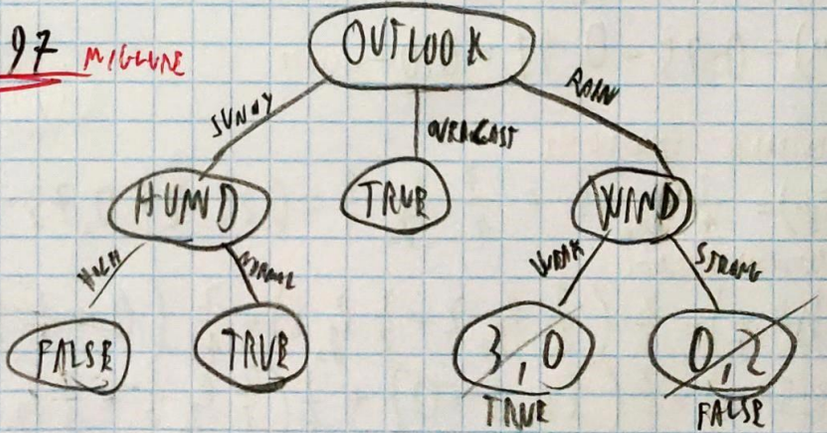
0.95

$R(W \mid O=\text{RAIN}) = \frac{3}{5}\left(\frac{3}{3}\log_2\frac{3}{3} + 0\right) + \frac{2}{5}\left(\frac{0}{2}\log_2\frac{2}{0} + \frac{2}{2}\log_2\frac{2}{2}\right) = 0$

$\text{GAIN}(T \mid O=\text{RAIN}) = 0.97 - 0.95 = 0.02$

$\text{GAIN}(H \mid O=\text{RAIN}) = 0.97 - 0.95 = 0.02$

$\text{GAIN}(W \mid O=\text{RAIN}) = 0.97 - 0 = \underline{0.97}$ MIGLIORE



ESERCIZIO SLIDE FUNGHI

RXPL	NOT HEAVY	SMELLY	SPOTTED	SMOOTH	EDIBLE
A	1	0	0	0	1
B	1	0	1	0	1
C	0	1	0	1	1
D	0	0	0	1	0
E	1	1	1	0	0
F	1	0	1	1	0
G	1	0	0	1	0
H	0	1	0	0	0
U	0	1	1	1	? → 1
V	1	1	0	1	? → 1
W	1	1	0	0	? → 0

CALCOLO ENTROPIA TARGET

$$I(\text{EDIBLE}) = \frac{3}{8} \log_2 \frac{8}{3} + \frac{5}{8} \log_2 \frac{8}{5} = 0.37 \cdot 1.41 + 0.63 \cdot 0.68 = 0.95$$

CALCOLO RENYOR DEGLI ATTRIBUTI

$$R(NH) = \frac{5}{8}\left(\frac{2}{5}\log_2 \frac{5}{2} + \frac{3}{5}\log_2 \frac{5}{3}\right) + \frac{3}{8}\left(\frac{1}{3}\log_2 3 + \frac{2}{3}\log_2 \frac{3}{2}\right) = 0.63\left(0.4 \cdot 1.32 + 0.6 \cdot 0.74\right) +$$

$$+ 0.38\left(0.34 \cdot 1.58 + 0.67 \cdot 0.58\right) = 0.95$$

$$R(SME) = \frac{3}{8}\left(\frac{1}{3}\log_2 3 + \frac{2}{3}\log_2 \frac{3}{2}\right) + \frac{5}{8}\left(\frac{2}{5}\log_2 \frac{5}{2} + \frac{3}{5}\log_2 \frac{5}{3}\right) = 0.37\left(0.99\right) + 0.63\left(0.97\right) = 0.95$$

$$R(SP) = \frac{3}{8}\left(\frac{1}{3}\log_2 3 + \frac{2}{3}\log_2 \frac{3}{2}\right) + \frac{5}{8}\left(\frac{2}{5}\log_2 \frac{5}{2} + \frac{3}{5}\log_2 \frac{5}{3}\right) = 0.37\left(0.53 + 0.39\right) + 0.63\left(0.53 + 0.44\right) = 0.95$$

$$R(SM) = \frac{4}{8}\left(\frac{1}{4}\log_2 4 + \frac{3}{4}\log_2 \frac{4}{3}\right) + \frac{4}{8}\left(\frac{2}{4}\log_2 2 + \frac{2}{4}\log_2 2\right) = 0,91$$

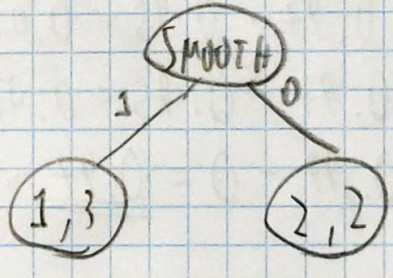
CALCOLO GUADAGNO DEGLI ATTRIBUTI

$$GAIN(NH) = 0.95 - 0.95 = 0$$

$$GAIN(SME) = 0.95 - 0.95 = 0$$

$$GAIN(SP) = 0.95 - 0.95 = 0$$

$$GAIN(SM) = 0.95 - 0.91 = \underline{0.04}$$ MIGLIORE



CALCOLO ENTROPIA SMOOTH=1

$$I(SM_{=1}) = \frac{1}{4}\log_2 4 + \frac{3}{4}\log_2 \frac{4}{3} = 0.25 \cdot 2 + 0.75 \cdot 0.42 = 0,81$$

CALCOLO RENYOR ATTRIBUTI

$$R(NH \mid SM_{=1}) = \frac{2}{4}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) + \frac{2}{4}\left(0 + \frac{2}{2}\log_2 1\right) = 0.5$$

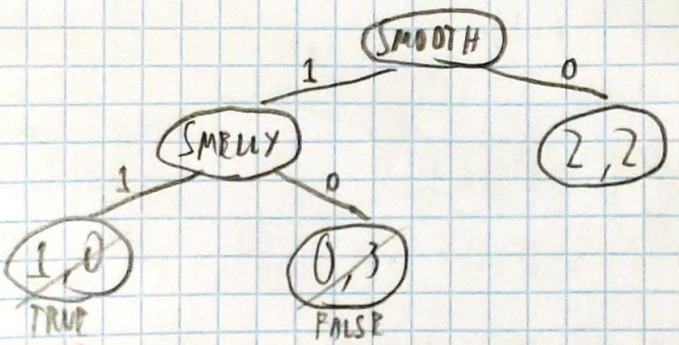
$$R(SME \mid SM_{=1}) = \frac{1}{4}\left(1 \log_2 1 + 0\right) + \frac{3}{4}\left(\frac{0}{3}\log_2 \frac{3}{0} + \frac{3}{3}\log_2 1\right) = 0$$

$$R(SP \mid SM_{=1}) = \frac{1}{4}\left(\frac{0}{1}\log_2 \frac{1}{0} + \frac{1}{1}\log_2 1\right) + \frac{3}{4}\left(\frac{1}{3}\log_2 3 + \frac{2}{3}\log_2 \frac{3}{2}\right) = 0,75\left(0.54 + 0,39\right) = 0.70$$

$$GAIN(NH \mid SM_{=1}) = 0.82 - 0.5 < 0.88$$

$$GAIN(SME \mid SM_{=1}) = 0.82 - 0 = \underline{0.82}$$ MIGLIORE

$$GAIN(SP \mid SM_{=1}) = 0.82 - 0.70 < 0.82$$



ENTROPIA SMOOTH = 0

$$I(SMOOTH=0) = \frac{2}{4} \log_2 \frac{1}{2} + \frac{2}{4} \log_2 2 = 1$$

RESIDUO ALTRI ATTRIBUTI

$$R(NH \mid SM=0) = \frac{3}{4}\left(\frac{2}{3}\log_2 \frac{3}{2} + \frac{1}{3}\log_2 3\right) + \frac{1}{4}\left(\frac{1}{1}\log_2 1 + 0\right) = 0{,}75(0{,}39 + 0{,}54) = 0.70$$

$$R(SME \mid SM=0) = \frac{2}{4}\left(0 + \frac{2}{2}\log_2 1\right) + \frac{2}{4}\left(\frac{2}{2}\log_2 1 + 0\right) = 0$$

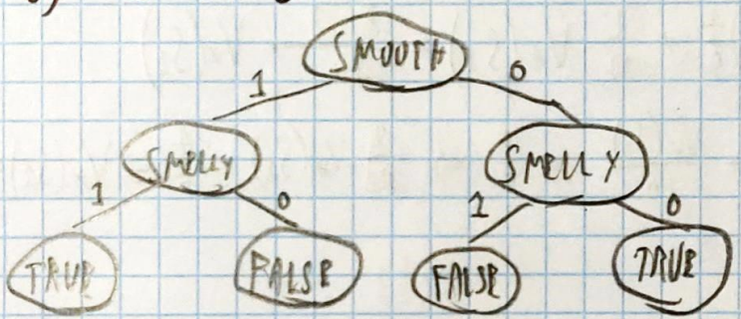
$$R(SP \mid SM=0) = \frac{2}{4}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) + \frac{2}{4}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) = 1$$

GUADAGNO

$$GAIN(NH \mid SM=0) = 1 - 0.70$$

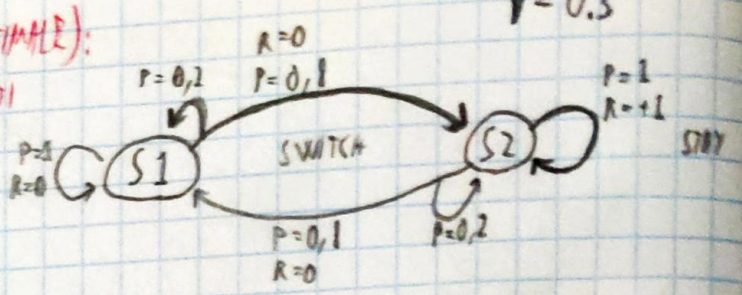
$$GAIN(SME \mid SM=0) = 1 - 0 = 1 \text{ MIGLIORE}$$

$$GAIN(SP \mid SM=0) = 1 - 1 = 0$$



ESERCIZIO 1 (POLICY OTTIMALE & VALUE FUNCTION OTTIMALE):

$\gamma = 0.5$



① POLICY OTTIMALE → PER OTTENERE IL MAX GUADAGNO DI TUTTI GLI STATI
LA SI VEDE AD OCCHIO

$\pi_{S2} \Rightarrow$ SWITH

$\pi_{S2} \rightarrow$ STAY

② CALCOLO VALUE FUNCTION OTTIMALE COL SISTEMA DI EQUAZIONE CORRISPONDENTE ALLA POLICY OTTIMALE

TEOREMA PRINCIPIO DI OTTIMALITÀ: $V_*(s) = \max_{a \in A} \left\{ R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V_*(s') \right\}$

$$V_*(S_2) = \max_{STAY} \left\{ 1 + \frac{1}{2}\left[1 \cdot V_*(S_2) \right] \right\} = 1 + \frac{V_*(S_2)}{2}$$

($R_{S_2}^{STAY}$, $P_{S_2S_2}^{STAY}$, γ)

$$V_*(S_2) = \frac{2 + V_*(S_2)}{2} \Rightarrow V_*(S_2) = 2$$

$$V_*(S_1) = 0 + \frac{1}{2}\left[\frac{8}{10} \cdot 2 + \frac{2}{10} \cdot V_*(S_1) \right]$$

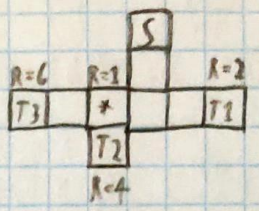
($R_{S_1}^{SWITCH}$, $P_{S_1S_2}^{SWITCH}$, $V_*(S_2)$, $P_{S_2S_1}^{SWITCH}$)

$$= \frac{1}{2}\left(\frac{8}{5} + \frac{2}{10} \cdot V_*(S_1) \right)$$

$$= \frac{2}{10} + \frac{2}{20} \cdot V_*(S_1) = \frac{4}{5} + \frac{1}{10} \cdot V_*(S_1)$$

$$V_*(S_1) = \frac{4}{5} + \frac{V_*(S_1)}{10} \Rightarrow V_*(S_1) - \frac{V_*(S_1)}{10} = \frac{4}{5} \Rightarrow \frac{9}{10} V_*(S_1) = \frac{4}{5} \Rightarrow V_*(S_1) = \frac{10}{9} \cdot \frac{4}{5} = \frac{8}{9}$$

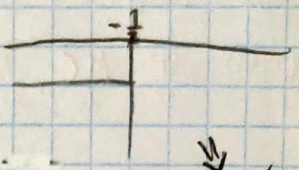
ESERCIZIO 2 (CALCOLO γ GAMMA) $\quad G_1 = R_2 + \gamma R_3 + \_ + \gamma^{T-2} R_T$



① VALORI DI γ CHE PONTA L'AGENTE IN T1:

$$\begin{cases} G_{T_1} > G_{T_2} \\ G_{T_1} > G_{T_3} \end{cases} = \begin{cases} 0 + \gamma\cdot 0 + \gamma^2\cdot 0 + \gamma^3\cdot 2 > 0 + \gamma\cdot 0 + \gamma^2\cdot 1 + \gamma^3\cdot 4 \\ 0 + \gamma\cdot 0 + \gamma^2\cdot 0 + \gamma^3\cdot 2 > 0 + \gamma\cdot 0 + \gamma^2\cdot 1 + \gamma^3\cdot 0 + \gamma^4\cdot 6 \end{cases}$$

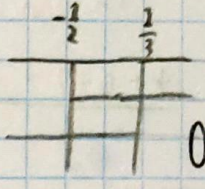
$$= \begin{cases} 2\gamma^3 > \gamma^2 + 4\gamma^3 \\ 2\gamma^3 > \gamma^2 + 6\gamma^4 \end{cases} = \begin{cases} 2\gamma > 1 + 4\gamma \\ 2\gamma > 1 + 6\gamma^2 \end{cases} = \begin{cases} \gamma < -\frac{1}{2} \\ 6\gamma^2 - 2\gamma + 1 < 0 \end{cases} \quad \Delta = 2 - 24 = -20 \Rightarrow \emptyset$$



NON HA SOLUZIONI SICCOME $0 \le \gamma \le 1$

② // IN T2:

$$\begin{cases} G_{T_2} > G_{T_1} \\ G_{T_2} > G_{T_3} \end{cases} = \begin{cases} 1\gamma^2 + 4\gamma^3 > 2\gamma^3 \\ \gamma^2 + 4\gamma^3 > \gamma^2 + 6\gamma^4 \end{cases} = \begin{cases} 4\gamma - 2\gamma > -1 \\ 6\gamma^2 - 4\gamma < 0 \end{cases} = \begin{cases} \gamma > -\frac{1}{2} \\ \gamma < \frac{2}{3} \end{cases}$$



$0 < \gamma < \frac{2}{3}$

③ // IN T3:

$\Delta = 4 - 24 = -20$

$$\begin{cases} G_{T_3} > G_{T_1} \\ G_{T_3} > G_{T_2} \end{cases} = \begin{cases} \gamma^2 + 6\gamma^4 > 2\gamma^3 \\ \gamma^2 + 6\gamma^4 > \gamma^2 + 4\gamma^3 \end{cases} = \begin{cases} 6\gamma^2 - 2\gamma + 1 > 0 \\ 6\gamma - 4 > 0 \end{cases} = \begin{cases} \emptyset \\ \gamma > \frac{2}{3} \end{cases}$$

$\frac{2}{3} < \gamma < 1$

ESERCIZIO 3 (MC-TD)

9 EPISODI GENERATI DA MDP (S-A-R):

1) A,0, B,4
2) A,2
3) A,1, B,0, A,2
4) B,0, A,1, B,2
5) B,0, A,2
6) B,2
7) B,4
8) B,2
9) B,4

1. FIRST VISIT MC, PRENDO LA PRIMA OCCORRENZA DELLO STATO TARGET E SOMMO LE RICOMPENSE FINO ALLA FINE, TUTTO DIVISO IL NUMERO DI PRIME OCCORRENZE DELLO STATO IN QUESTIONE.

$$V(A) = \frac{\overbrace{0+4}^{1E} + \overbrace{2}^{2E} + \overbrace{1+0+2}^{3E} + \overbrace{1+2+2}^{4E+5E}}{5} = \frac{14}{5}$$

NUM. EPISODI CHE CONTANO A

$$V(B) = \frac{4+0+2+0+1+2+0+2+2+4+2+4}{8} = \frac{23}{8}$$

2. TD(0), PER OGNI EPISODIO VEDO QUAL'È IL PRIMO STATO E APPLICO LA FORMULA METTENDO AI $V(S)$ e $V(S')$ INIZIALMENTE I VALORI DI DEFAULT E SUCCESSIVAMENTE MISTO I VMAN CALCOLATI IMMEDIATAMENTE PRIMA. SE LO STATO SUCCESSIVO È TERMINALE IL $V(S')=0$.

$$V(S_t) = V(S_t) + \alpha\left[\underbrace{R_{T+1}}_{\text{RICOMPENSA SUCC.}} + \underbrace{\gamma V(S_{T+1})}_{\text{SCONTO}} - \underbrace{V(S_t)}_{\text{TIME STEP}}\right] \xrightarrow{\text{SE } \alpha=\gamma=1} = R_{T+1} + V(S_{T+1})$$

Ponendo $\alpha = \gamma = 1$, $V(A) = V(B) = 0$

1E: $V(A) = 0 + 1[0 + 1\cdot 0 - 0] = 0$

$V(B) = 0 + 1[4 + 1\cdot \underbrace{0}_{V(T)=0} - 0] = 4$

2E: $V(A) = 0 + 1[2 + 1\cdot 0 - 0] = 2$

3E: $V(A) = 2 + 1 + 4 - 2 = 5$

$V(B) = 4 + 0 + 5 - 4 = 5$

$V(A) = 5 + 2 + 0 - 5 = 2$

4E: $V(B) = 0 + 2 = 2$

$V(A) = 1 + 2 = 3$

$V(B) = 2 + 0 = 2$

5E: $V(B) = 0 + 3 = 3$

$V(A) = 2 + 0 = 2$

6E: $V(B) = 2 + 0 = 2$

7E: $V(B) = 4 + 0 = 4$

8E: $V(B) = 2 + 0 = 2$

9E: $V(B) = 4 + 0 = 4$

ESERCIZIO 3 (Q-LEARNING) TRACCIA/ESEMPIO

2	0	0	2
1	-1	-10	0
0	0	-1	0
	0	1	2

F

1E: $(0,0) \xrightarrow{\rightarrow} (1,0) \xrightarrow{\uparrow} (1,1) \xrightarrow{\rightarrow} (2,1) \xrightarrow{\uparrow} (2,2)$

2E: $(2,1) \xrightarrow{\downarrow} (2,0) \xrightarrow{\uparrow} (2,1)$

3E: $(1,1) \xrightarrow{\downarrow} (1,0) \xrightarrow{\rightarrow} (2,0)$

4E: $(0,0) \xrightarrow{\rightarrow} (1,0) \xrightarrow{\uparrow} (1,1)$

FATTORE DI SCONTO $\gamma = 0,9$

TASSO DI APPRENDIMENTO $\alpha = 1$

1Q-TABLE INIZIALIZZATA CON LA RICOMPENSA IMMEDIATA:

	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)	(2,0)	(2,1)	(2,2)
↑	-1	0	X	-10	0	X	0	2	0
↓	X	0	-1	X	-1	-10	X	0	0
←	X	X	X	0	-1	0	-1	-10	0
→	-1	-10	0	0	0	2	X	X	0

2. COSTRUZIONE Q-TABLE MEDIANTE FORMULA →

$$Q(S,A) = Q(S,A) + \alpha\left(R_s^A + \gamma \max_{a'} Q(S',A') - Q(S,A)\right)$$

(max delle azioni successive)

1E:

$Q((0,0),\rightarrow) = -1 + 1\cdot(-1 + 0.9 \max\{0,0,-10\} + 1) = -1 - 1 + 0 + 1 = -1$

$Q((1,0),\uparrow) = -10 + 1(-10 + 0.9 \max\{0,-1,-1,0\} + 10) = -10$

$Q((1,1),\rightarrow) = 0 + 0 + 0.9 \cdot \max\{-10,0,2\} - 0 = 1.8$

$Q((2,1),\uparrow) = 2 + 2 + 0.9 \max\{0,0\} - 2 = 2$

2E:

$Q((2,1),\downarrow) = 0 + 0 + 0.9 \max\{0,-1\} - 0 = 0$

$Q((2,0),\uparrow) = 0 + 0 + 0.9 \max\{-10,2,0\} - 0 = 1.8$

3E:

$Q((1,1),\downarrow) = -1 + (-1) + 0.9 \max\{0,0,-10\} + 1 = -1$

$Q((1,0),\rightarrow) = 0 - 0 + 0.9 \max\{1.8,-1\} - 0 = 1.62$

4E:

$Q((0,0),\rightarrow) = -1 - 1 + 0.9 \max\{-10,1.62\} + 1 = -1 + 1.46 = 0,46$

$Q((1,0),\uparrow) = -10 - 10 + 0.9 \max\{0,-1,-1,1.8\} + 10 = -10 + 1.62 = -8,38$

	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)	(2,0)	(2,1)	(2,2)
↑	-1	0	X	-8,38	0	X	1.8	2	0
↓	X	0	-1	X	-1	-10	X	0	0
←	X	X	X	0	-1	0	-1	-10	0
→	0,46	-10	0	1.62	1.8	2	X	X	0

POLICY GREEDY DOPO AGGIORNAMENTI Q-TABLE:

$\pi_{(0,0)} = \rightarrow$

$\pi_{(0,1)} = \uparrow / \downarrow$

$\pi_{(0,2)} = \rightarrow$

$\pi_{(1,1)} = \rightarrow$

$\pi_{(1,2)} = \rightarrow$

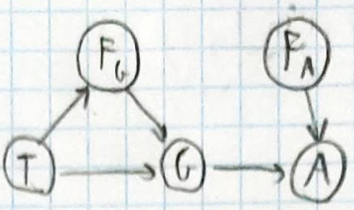
$\pi_{(2,0)} = \uparrow$

$\pi_{(2,1)} = \uparrow$

$\pi_{(2,2)} = X$

ESERCIZIO CENTRALE NUCLEARE (NODI BAYESIANI)

- A - suona allarme
- F_A - allarme è difettoso
- F_G - indicatore difettoso
- G - lettura indicatore
- T - temperatura del nucleo



		CAUSE			
		$T=m$		$T=h$	
		F_G	$\bar{F}_G$	F_G	$\bar{F}_G$
EFFETTO	$G=m$	y	x	$1-y$	$1-x$
	$G=h$	$1-y$	$1-x$	y	x

	$T=m$		$T=h$	
	F_A	$\bar{F}_A$	F_A	$\bar{F}_A$
A	0	0	0	1
$\bar{A}$	1	1	1	0

INFERENZA PER ENUMERAZIONE CON:

$P(T)=\tau$, $P(F_G|T)=q$ e $P(F_G|\bar{T})=h$

$$P(T=h \mid \bar{F}_A, \bar{F}_G, A) = \frac{\sum_G P(T=h, \bar{F}_A, \bar{F}_G, A, G)}{\alpha = \sum_{T,G} P(\bar{F}_A, \bar{F}_G, A, T, G)} =$$

$$\overset{NUM}{=} \sum_G P(T=h)\, P(\bar{F}_A)\, P(\bar{F}_G|T)\, P(A|\bar{F}_A, G)\, P(G|T, \bar{F}_G) =$$

$$= P(T=h)\, P(\bar{F}_A)\, P(\bar{F}_G|T) \sum_G P(A|\bar{F}_A, G)\, P(G|T, \bar{F}_G) =$$

$$= \tau \cdot P(\bar{F}_A)(1-q)\left[P(A|\bar{F}_A, G)\, P(G|T, \bar{F}_G) + P(A|\bar{F}_A, \bar{G})\, P(\bar{G}|T, \bar{F}_G)\right]$$

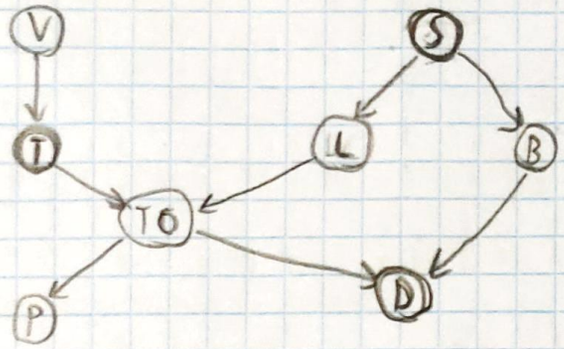
$$\overset{DEN}{\propto} \sum_{T,G} P(\bar{F}_A)\, P(\bar{F}_G|T)\, P(A|G, \bar{F}_A)\, P(T)\, P(G|\bar{F}_G, T) =$$

$$= P(\bar{F}_A) \sum_T P(\bar{F}_G|T)\, P(T) \sum_G P(A|G, \bar{F}_A)\, P(G|\bar{F}_G, T) =$$

$$= P(\bar{F}_A)\Big\{P(\bar{F}_G|T)\, P(T)\left[P(A|G, \bar{F}_A)\, P(G|\bar{F}_G, T) + P(A|\bar{G}, \bar{F}_A)\, P(\bar{G}|\bar{F}_G, T)\right] +$$
$$+ P(\bar{F}_G|\bar{T})\, P(\bar{T})\left[P(A|G, \bar{F}_A)\, P(G|\bar{F}_G, \bar{T}) + P(A|\bar{G}, \bar{F}_A)\, P(\bar{G}|\bar{F}_G, \bar{T})\right]\Big\}$$

$$= P(\bar{F}_A)\Big\{(1-q)\tau\left[P(A|G, \bar{F}_A)\, P(G|\bar{F}_G, T) + \;//\;\right]\Big\}$$

TRACCIA A - ESEMPIO (RETI BAYESIANE):



$$P(D|S,\bar{T}) = \frac{\sum_{TO}\sum_{B}\sum_{V}\sum_{L} P(D,S,\bar{T},TO,B,V,L)}{\sum_{V} P(S,\bar{T},V)} =$$

$$= \frac{\sum_{TO}\sum_{B}\sum_{V}\sum_{L} P(D|TO,B)P(S)P(\bar{T}|V)P(TO|\bar{T},L)P(B|S)P(V)P(L|S)}{\sum_{V} P(S)P(\bar{T}|V)P(V)} =$$

$$= \frac{P(S)\sum_{L}P(L|S)\sum_{TO}P(TO|\bar{T},L)\sum_{B}P(B|S)P(D|TO,B)\sum_{V}P(V)P(\bar{T}|V)}{P(S)\sum_{V}P(\bar{T}|V)P(V)}$$

$$= P(L|S)\{P(TO|\bar{T},L)[P(B|S)P(D|TO,B)+P(\bar{B}|S)P(D|TO,\bar{B})]+$$
$$+P(\bar{TO}|\bar{T},L)[P(B|S)P(D|\bar{TO},B)+P(\bar{B}|S)P(D|\bar{TO},\bar{B})]\}+$$
$$+P(\bar{L},S)\{P(TO|\bar{T},\bar{L})[P(B|S)P(D|\bar{TO},B)+P(\bar{B}|S)P(D|TO,\bar{B})]+$$
$$+P(\bar{TO}|\bar{T},\bar{L})[P(B|S)P(D|\bar{TO},B)+P(\bar{B}|S)P(D|\bar{TO},\bar{B})]\} =$$

$$= 0.1\{1[0.2\cdot 0.9+0.8\cdot 0.7]+0[\sim]\}+0.9\{0[\sim]+1[0.2\cdot 0.6+0.8\cdot 0.05]\}$$

$$= 0.1\{0.74\}+0.9\{0.16\} = 0.218$$

TRACCIA - ESEMPIO (DT1)

	A	I	S	C	
RID	AGE	INCOME	STUDENT	CREDIT	PC
1	≤ 30	HIGH	NO	FAIR	NO
2	≤ 30	HIGH	NO	EX	NO
3	31…40	HIGH	NO	FAIR	YES
4	> 40	MED	NO	FAIR	YES
5	> 40	LOW	YES	FAIR	YES
6	> 40	LOW	YES	EX	NO
7	31…40	LOW	YES	EX	YES
8	≤ 30	MED	NO	FAIR	NO
9	≤ 30	LOW	YES	FAIR	YES
10	> 40	MED	YES	FAIR	YES
11	≤ 30	MED	YES	EX	YES
12	31…40	MED	NO	EX	YES
13	31…40	HIGH	YES	FAIR	YES
14	> 40	MED	NO	EX	NO
15	≤ 30	MED	YES	FAIR	? → YES

CALCOLO ENTROPIA DI TARGET = PC:

$$I(PC) = \frac{9}{14}\log_2\frac{14}{9} + \frac{5}{14}\log_2\frac{14}{5} = 0.409 + 0.53 = 0.939$$

CALCOLO REMAINDER

$$R(A) = \frac{5}{14}\left(\frac{2}{5}\log_2\frac{5}{2} + \frac{3}{5}\log_2\frac{5}{3}\right) + \frac{4}{14}\left(\frac{4}{4}\log_2 1 + 0\right) + \frac{5}{14}\left(\frac{3}{5}\log_2\frac{5}{3} + \frac{2}{5}\log_2\frac{5}{2}\right) =$$

$$= 0.346 + 0.346 = 0.692$$

0,39 0,53

$$R(I) = \frac{4}{14}\left(\frac{2}{4}\log_2\frac{4}{2} + \frac{2}{4}\log_2\frac{4}{2}\right) + \frac{6}{14}\left(\frac{4}{6}\log_2\frac{6}{4} + \frac{2}{6}\log_2\frac{6}{2}\right) = 0.286 + 0.42 = 0.706$$

$$R(S) = \frac{7}{14}\left(\frac{6}{7}\log_2\frac{7}{6} + \frac{1}{7}\log_2 7\right) + \frac{7}{14}\left(\frac{3}{7}\log_2\frac{7}{3} + \frac{4}{7}\log_2\frac{7}{4}\right) = 0.296 + 0.493 = 0.789$$

$$R(C) = \frac{6}{14}\left(\frac{3}{6}\log_2\frac{6}{3} + \frac{3}{6}\log_2\frac{6}{3}\right) + \frac{8}{14}\left(\frac{6}{8}\log_2\frac{8}{6} + \frac{2}{8}\log_2\frac{8}{2}\right) = 0.429 + 0.464 = 0.893$$

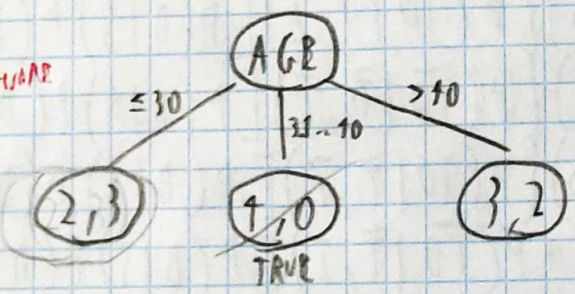
CALCOLO GUADAGNO PER SCEGLIERE LA ROOT

$$GAIN(A) = 0.939 - 0.692 = 0.247$$ MIGLIORE

$$GAIN(I) = 0.939 - 0.706 = 0.233$$

$$GAIN(S) = 0.939 - 0.789 = 0.15$$

$$GAIN(C) = 0.939 - 0.893 = 0.046$$



CALCOLO ENTROPIA DI AGE = "≤ 30" E RIMANDO DEI RESTANTI ATTRIBUTI

$$I(A_{\leq 30}) = \frac{2}{5}\log_2\frac{5}{2} + \frac{3}{5}\log_2\frac{5}{3} = 0.97$$

$$R(I \mid A_{\leq 30}) = \frac{2}{5}\left(0 + \frac{2}{2}\log_2 1\right) + \frac{2}{5}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) + \frac{1}{5}(0) = 0.4$$

$$R(S \mid A_{\leq 30}) = \frac{2}{5}\left(\frac{2}{2}\log_2 1 + 0\right) + \frac{3}{5}\left(0 + \frac{3}{3}\log_2 1\right) = 0$$

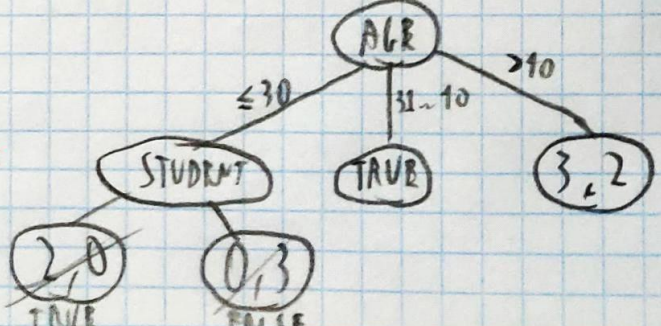
$$R(C \mid A_{\leq 30}) = \frac{2}{5}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) + \frac{3}{5}\left(\frac{1}{3}\log_2 3 + \frac{2}{3}\log_2\frac{3}{2}\right) = 0.4 + 0.6 = 1$$

CALCOLO GUADAGNO PER SCEGLIERE LA PROLE

$$GAIN(I \mid A_{\leq 30}) = 0.97 - 0.4 = 0.93$$

$$GAIN(S \mid A_{\leq 30}) = 0.97 - 0 = 0.97$$ MIGLIORE

$$GAIN(C \mid A_{\leq 30}) = 0.97 - 1 = -0.03$$



CALCOLO ENTROPIA DI AGE > 40

$$I(AGE>40) = \frac{3}{5} \log_2 \frac{5}{3} + \frac{2}{5} \log_2 \frac{5}{2} = 0.97$$

CALCOLO ENTROPIA DEGLI ATTRIBUTI RIMANENTI

$$R(1 \mid AGE>40) = \frac{3}{5}\left(\frac{2}{3} \log_2 \frac{3}{2} + \frac{1}{2} \log_2 2\right) + \frac{2}{5}\left(\frac{1}{2} \log_2 2 + \frac{1}{2} \log_2 2\right) = 0.53 + 0.4 = 0.93$$

$$R(S \mid A>40) = \frac{3}{5}\left(\frac{2}{3} \log_2 \frac{3}{2} + \frac{1}{3} \log_2 3\right) + \frac{2}{5}\left(\frac{1}{2} \log_2 2 + \frac{1}{2} \log_2 2\right) = 0.55 + 0.4 = 0.95$$

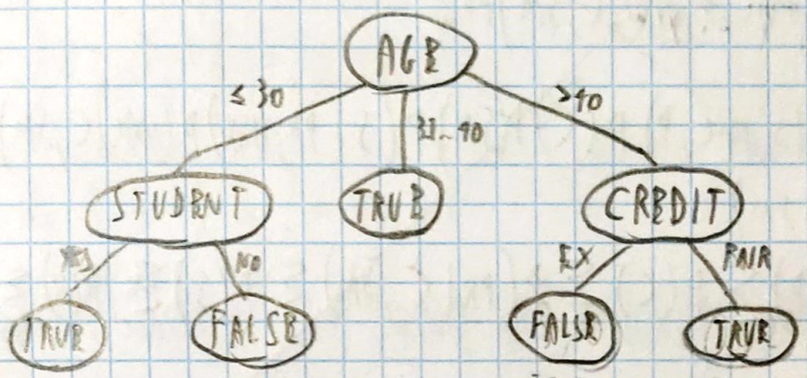
$$R(C \mid A>40) = \frac{2}{5}\left(0 + \frac{2}{2} \log_2 1\right) + \frac{3}{5}\left(\frac{3}{3} \log_2 1 + 0\right) = 0$$

CALCOLO GUADAGNO PER SCEGLIERE LA FOGLIA

$$GAIN(1 \mid A>40) = 0.97 - 0.93 = 0.04$$

$$GAIN(S \mid A>40) = 0.97 - 0.95 = 0.02$$

$$GAIN(C \mid A>40) = 0.97 - 0 = 0.97$$ MIGLIORE



TRACCIA 12-01-2013 - ESERCIZIO 2 (DTL)

MEDIE	STUDIATO	SUPERATO
BASSA	F	F ←
BASSA	V	V
MEDIA	F	F ←
MEDIA	V	V
ALTA	F	V ←
ALTA	V	V

CALCOLO ENTROPIA DI "SUPERATO"

$$I(SUP) = \frac{4}{6} \log_2 \frac{6}{4} + \frac{2}{6} \log_2 \frac{6}{2} = 0.92$$

RIMANDO DEGLI ATTRIBUTI PER SCEGLIERE ROOT

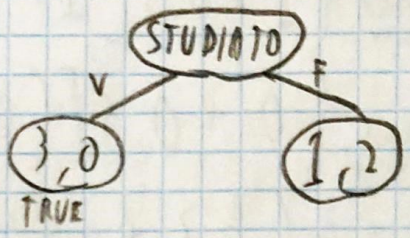
$$R(M) = \frac{2}{6}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) + \frac{2}{6}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) + \frac{2}{6}\left(\frac{2}{2}\log_2 1 + 0\right) = \frac{1}{3} + \frac{1}{3} = 0.67$$

$$R(S) = \frac{3}{6}\left(\frac{3}{3}\log_2 1 + 0\right) + \frac{3}{6}\left(\frac{1}{3}\log_2 3 + \frac{2}{3}\log_2 \frac{3}{2}\right) = 0.46$$

CALCOLO GUADAGNO PER SCEGLIERE ROOT

$$GAIN(M) = 0.92 - 0.67 = 0.25$$

$$GAIN(S) = 0.92 - 0.46 = \underline{0.46}$$ MIGLIORE



CALCOLO ENTROPIA "STUDIATO = F"

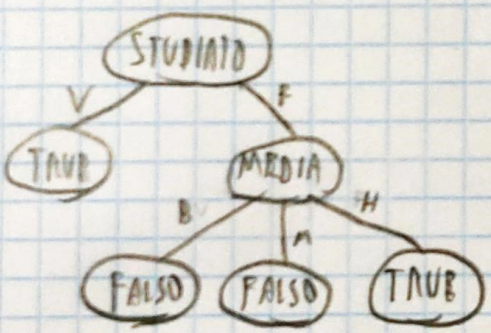
$$I(S=F) = \frac{1}{3}\log_2 3 + \frac{2}{3}\log_2 \frac{3}{2} = 0.92$$

CALCOLO RIMANDO PER "MEDIA"

$$R(M|S=F) = \frac{1}{3}(0 + 1\log_2 1) + \frac{1}{3}(0 + 1\log_2 1) + \frac{1}{3}(1\log_2 1 + 0) = 0$$

GUADAGNO DI "MEDIA"

$$G(M|S=F) = 0.92 - 0 = 0.92$$



Вариант 4 (№ L-1)

$\gamma = 0,5$ $R_1 = -1$, $R_2 = 2$, $R_3 = 6$, $R_4 = 3$, $R_5 = 2$ и $T = 5$

$G_0 = -1 + \frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 6 + \frac{1}{8} \cdot 3 + \frac{1}{16} \cdot 2 = -1 + 1 + \frac{3}{2} + \frac{3}{8} + \frac{1}{8} = 2$

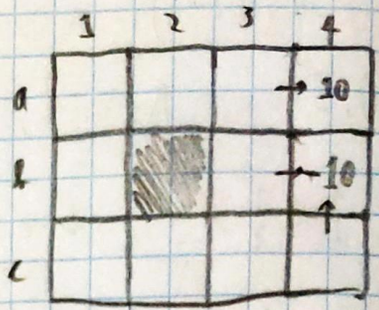
$G_1 = 2 + \frac{1}{2} \cdot 6 + \frac{1}{4} \cdot 3 + \frac{1}{8} \cdot 2 = 2 + 3 + \frac{3}{4} + \frac{1}{4} = 6$

$G_2 = 6 + \frac{1}{2} \cdot 3 + \frac{1}{4} \cdot 2 = 8$

$G_3 = 3 + \frac{1}{2} 2 = 4$

$G_4 = 2$ $G_5 = 0$

ESERCIZIO Q-LEARNING



1E: a_1, a_2, a_3, b_3, b_4

2E: $c_2, c_3, b_2, a_2, a_2, a_3, a_4$

3E: c_4, c_3, b_3, a_3, a_4

$\gamma = 0.9$

FORMULA Q-LEARNING:

$$Q(S,A) = Q(S,A) + \alpha\left[R_{t+1}^{a} + \gamma \max_{a' \in A} Q(S',A') - Q(S,A)\right] \xrightarrow{\alpha=1} R_{t+1}^{a} + \gamma \max_{a' \in A} Q(S',A')$$

1E:

b_4 è terminale

$$Q(b_3, dx) = -10 + 0.9 \max\{0\} = -10$$
$$Q(a_3, \text{down}) = 0 + 0.9 \max\{0, 0, -10\} = 0$$
$$Q(a_2, dx) = 0 + 0.9 \max\{0, 0, 0\} = 0$$
$$Q(a_1, dx) = 0 + 0.9 \max\{0, 0\} = 0$$

2E:

a_4 è terminale

$$Q(a_3, dx) = 10 + 0.9 \max\{0\} = 10$$
$$Q(a_2, dx) = 0 + 0.9 \max\{0, 0, 10\} = 9$$
$$Q(a_1, dx) = 0 + 0.9 \max\{0, 9\} = 8.1$$
$$Q(b_1, a_4) = 0 + 0.9 \max\{0, 8.1\} = 7.29$$
$$Q(c_1, a_4) = 0 + 0.9 \max\{7.29, 0\} = 6.56$$
$$Q(c_2, sx) = 0 + 0.9 \max\{6.56, 0\} = 5.9$$

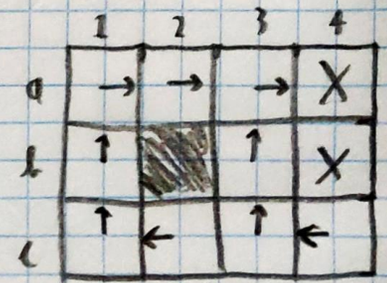
3E:

a_4 è terminale

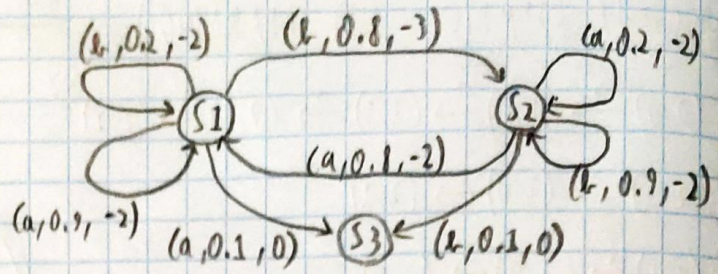
$$Q(a_3, dx) = 10 + 0.9 \max\{0\} = 10$$
$$Q(b_3, a_4) = 0 + 0.9 \max\{0, 10\} = 9$$
$$Q(b_3, a_4) = 0 + 0.9 \max\{9, 0, -10\} = 8.1$$
$$Q(c_4, sx) = 0 + 0.9 \max\{0, 8.1, 0\} = 7.29$$

POLICY OTTIMALE

$\pi_* \Rightarrow$



TRACCIA 12/01/2023 - ESERCIZIO 3



1. POLITICA OTTIMALE

La politica ottimale è andare nello stato S_3 siccome la ricompensa è 0, mentre andando negli altri stati la ricompensa è negativa, pertanto:

$\pi_{S_1} = (a, 0.1, 0)$, $\pi_{S_2} = (b, 0.1, 0)$ e $\pi_{S_3} = \times$ (rimango nello stato S_3)

2. VALUE ITERATION:

$$V(s) = \max_a \sum_{s',r} T(s', r \mid s, a)[r + \gamma V(s')]$$

1° ITERAZIONE:

azione a | azione b

$$V(S_1) = \max\{0.9[-2+0] + 0.1[0+0],\ 0.2[-2+0] + 0.8[-3+0]\} =$$
$$= \max\{-1.8, -2.8\} = -1.8 \quad \text{prendo l'azione a}$$

$$V(S_2) = \max\{0.2[-2+0] + 0.8[-2+(-1.8)],\ 0.9[-2+0] + 0.1[0+0]\} =$$
$$= \max\{-3.44, -1.8\} = -1.8 \quad \text{prendo l'azione b}$$

2° ITERAZIONE

$$V(S_1) = \max\{0.9[-2+(-1.8)] + 0.1[0+0],\ 0.2[-2-1.8] + 0.8[-3+(-1.8)]\} =$$
$$= \max\{-3.42, -4.6\} = -3.42 \quad \text{prendo l'azione a}$$

$$V(S_2) = \max\{0.2[-2+(-1.8)] + 0.8[-2+(-3.42)],\ 0.9[-2+(-1.8)] + 0.1[0+0]\} =$$
$$= \max\{-5.096, -3.42\} = -3.42 \quad \text{prendo l'azione b}$$

3° ITERAZIONE

$$V(S_1) = \max\{0.9[-2-3.42] + 0.1[0+0],\ 0.2[-2-3.42] + 0.8[-3-3.42]\} =$$
$$= \max\{-4.878, -6.22\} = -4.878 \quad \text{prendo l'azione a}$$

$$V(S_2) = \max\{0.2[-2-3.42] + 0.8[-2-4.878],\ 0.9[-2-3.42] + 0.1[0+0]\}$$
$$= \max\{-6.59, -4.88\} = -4.88 \quad \text{prendo l'azione b}$$

TRACCIA 12/01/2023 - ESERCIZIO 4

1E: $(S1, a, 0, S2)(S2, a, +1, T2)$

2E: $(S1, a, 0, S2)(S2, b, 0, S1)(S1, b, +1, T1)$

3E: $(S1, a, 1, S2)(S2, b, +1, T2)$

$\gamma = 1$, $\alpha = \frac{1}{2}$

S1, 0, S2, 1

S1, 0, S2, 0, S1, 1

S1, 1, S2, +1

2 MC - FIRST VISIT

$$V(S_1) = \frac{0+1+0+0+1+1+1}{3} = \frac{4}{3} \quad , \quad V(S_2) = \frac{1+0+1+1}{3} = 1$$

TD(0): $V(S_T) = V(S_T) + \alpha\left[R_{T+1} + \gamma V(S_{T+1}) - V(S_T)\right]$

1E:

$$V(S_1) = V(S_1)^0 + \frac{1}{2}\left[0 + 1 \cdot V(S_2)^0 - V(S_1)^0\right] = 0$$

$$V(S_2) = 0 + \frac{1}{2}\left[1 + 1 \cdot 0 - 0\right] = \frac{1}{2}$$

2E:

$$V(S_1) = 0 + \frac{1}{2}\left[0 + 1 \cdot \frac{1}{2} - 0\right] = \frac{1}{4}$$

$$V(S_2) = \frac{1}{2} + \frac{1}{2}\left[0 + 1 \cdot \frac{1}{4} - \frac{1}{2}\right] = \frac{3}{8}$$

$$V(S_1) = \frac{1}{4} + \frac{1}{2}\left[1 + 1 \cdot 0 - \frac{1}{4}\right] = \frac{5}{8}$$

3E:

$$V(S_1) = \frac{5}{8} + \frac{1}{2}\left[1 + 1 \cdot \frac{3}{8} - \frac{5}{8}\right] = 1$$

$$V(S_2) = \frac{3}{8} + \frac{1}{2}\left[1 + 1 \cdot 0 - \frac{3}{8}\right] = \frac{11}{16}$$