Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
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

TABLE 3.2
Training examples for the target concept *PlayTennis*.

Entropy(D) / Info(D)

$$Info(D) = -\sum_{i=1}^{2} P_i \log_2 P_i$$

$$P_1 = \frac{9}{14} \qquad P_2 = \frac{5}{14}$$

$$Entropy(D) = -\left[\frac{9}{14}\log_2\left(\frac{9}{14}\right) + \frac{5}{14}\log_2\left(\frac{5}{14}\right)\right]$$

$$= -\left[0.642 \times (-0.637) + 0.357\,(-1.485)\right]$$

$$= -\left[-0.408 - 0.530\right]$$

$$\cancel{P_2} = \simeq \underline{\underline{0.940}} \text{ bits.}$$

$$Info_{outlook}(D) = \cancel{Info} \sum_{j=1}^{v} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \sum_{j=1}^{3} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \frac{|D_{sunny}|}{|D|} Info(D_{sunny}) + \frac{|D_{overcast}|}{|D|} Info(D_{overcast}) + \frac{|D_{Rain}|}{|D|} Info(D_{Rain})$$

$$\mathrm{Info}_{Outlook}(D) = \frac{5}{14}\,\mathrm{Info}(D_{Sunny}) + \frac{4}{14}\,\mathrm{Info}(D_{Overcast}) + \frac{5}{14}\,\mathrm{Info}(D_{Rain})$$

$$= \frac{5}{14}\left[-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right]$$

$$+ \frac{4}{14}\left[\cancelto{0}{-\frac{4}{4}\log_2\left(\frac{4}{4}\right)} - \cancelto{0}{\frac{0}{4}\log_2\left(\frac{0}{4}\right)}\right]$$

$$+ \frac{5}{14}\left[-\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right]$$

$$= 0.346 + 0 + 0.346$$

$$\simeq \underline{\underline{0.693}}$$

$$\mathrm{Gain}(Outlook) = \mathrm{Info}(D) - \mathrm{Info}_{Outlook}(D)$$

$$= 0.940 - 0.693$$

$$\simeq \underline{\underline{0.246}}$$

$$\underset{Temperature}{Info(D)} = \sum_{i=1}^{v} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \sum_{i=1}^{3} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \frac{|D_{Hot}|}{|D|} Info(D_{Hot}) + \frac{|D_{Mild}|}{|D|} Info(D_{Mild}) + \frac{|D_{cool}|}{|D|} Info(D_{Cool})$$

$$= \frac{4}{14} Info(D_{Hot}) + \frac{6}{14} Info(D_{Mild}) + \frac{4}{14} Info(D_{cool})$$

$$= \frac{4}{14}\left[-\left(\frac{2}{4}\right)\log_2\left(\frac{2}{4}\right) - \left(\frac{2}{4}\right)\log_2\left(\frac{2}{4}\right)\right] + \frac{6}{14}\left[-\left(\frac{4}{6}\right)\log_2\left(\frac{4}{6}\right) - \left(\frac{2}{6}\right)\log_2\left(\frac{2}{6}\right)\right] + \frac{4}{14}\left[-\left(\frac{3}{4}\right)\log_2\left(\frac{3}{4}\right) - \left(\frac{1}{4}\right)\log_2\left(\frac{1}{4}\right)\right]$$

$$= 0.285(1) + 0.393 + 0.231$$

$$= \underline{0.908} \text{ bits}$$

Gain (Temperature)

$$= Info(D) - Info_{Temperature}(D)$$

$$= 0.940 - 0.908$$

$$= \underline{0.031}$$

$$Info_{Humidity}(D) = \sum_{i=1}^{2} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \frac{|D_{high}|}{|D|} Info(D_{high})$$

$$+ \frac{|D_{Normal}|}{|D|} Info(D_{Normal})$$

$$= \frac{7}{14} \times Info(D_{high}) + \frac{7}{14} \times Info(D_{Normal})$$

$$= \frac{7}{14} \cdot \left[-\frac{3}{7} \times \log_2\left(\frac{3}{7}\right) - \left(\frac{4}{7}\right) \times \log_2\left(\frac{4}{7}\right)\right]$$

$$+ \frac{7}{14} \left[-\frac{6}{7} \times \log_2\left(\frac{6}{7}\right) - \left(\frac{1}{7}\right) \times \log_2\left(\frac{1}{7}\right)\right]$$

$$= 0.492 + 0.295$$

$$= \underline{0.786}$$

$$Gain(Humidity) = Info(D) - Info_{Humidity}(D)$$

$$= 0.940 - 0.786$$

$$\boxed{Gain(Humidity) = 0.153}$$

$$Info_{Wind}(D) = \sum_{i=1}^{2} \frac{|D_j|}{|D|} \times Info(D_j)$$

$$= \frac{|D_{weak}|}{|D|} \times Info(D_{weak}) + \frac{|D_{strong}|}{|D|} \times Info(D_{strong})$$

$$= \frac{8}{14} \times Info(D_{weak}) + \frac{6}{14} \times Info(D_{strong})$$

$$= \left(\frac{8}{14}\right) \times \left[-\left(\frac{6}{8}\right) \times \log_2\left(\frac{6}{8}\right) - \left(\frac{2}{8}\right) \times \log_2\left(\frac{2}{8}\right)\right]$$

$$+ \left(\frac{6}{14}\right) \times \left[-\left(\frac{3}{6}\right) \log_2\left(\frac{3}{6}\right) - \left(\frac{3}{6}\right) \log_2\left(\frac{3}{6}\right)\right]$$

$$= 0.463 + 0.428 = 0.891$$

$$Gain(Wind) = Info(D) - Info_{Wind}(D)$$

$$= 0.940 - 0.891$$

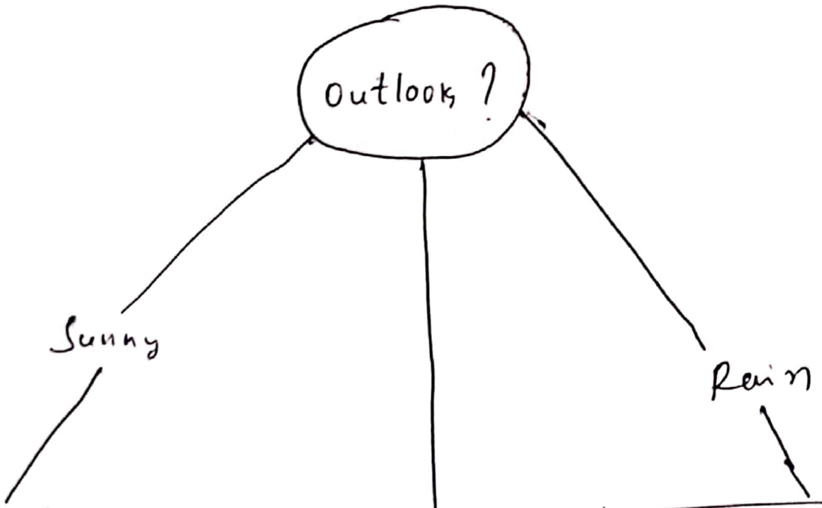
$$= \underline{0.048}$$

Gain (outlook) = 0.246 ✓

Gain (Temperature) = 0.031

Gain (Humidity) = 0.153

Gain (Wind) = 0.048



Sunny

Temp	Humd	Wind	play Tennis
~~Hot~~	high	weak	no
Hot	high	strong	no
Mild	high	weak	no
Cool	Normal	Weak	yes
Mild	Normal	strong	yes

Rain

Temp	Humd	Wind	play Tennis
Mild	High	weak	yes
Cool	Normal	weak	yes
Cool	Normal	Strong	No
Mild	Normal	weak	yes
Mild	High	Strong	No

Overcast

Temp	Humd	Wind	Play Tennis
Hot	high	weak	yes
Cool	Normal	Strong	yes
Mild	high	Strong	yes
Hot	Normal	weak	yes

Since D_{sunny} is impure, Split D_{sunny}

~~Let~~ Now our D is D_{sunny}

$$\therefore \; Info(D) = -\sum_{i=1}^{2} p_i \log_2 p_i$$

$$\Rightarrow \; P_1 = \frac{2}{5} \qquad P_2 = \frac{3}{5}$$

$$= -\left[\frac{2}{5}\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right]$$

$$= \underline{\underline{0.970}} \text{ bits}$$

~~Info (D)~~ ~~Temp~~ To split D_{sunny}, we have to select best among $\{$Temp, Humd, Wind$\}$

$$\therefore \; Info_{Temp}(D) = \sum_{j=1}^{v=3} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \frac{|D_{hot}|}{|D|} \times Info(D_{hot}) + \frac{|D_{mild}|}{|D|} \times Info(D_{mild})$$

$$+ \frac{|D_{cool}|}{|D|} \times Info(D_{cool})$$

$$\cancel{= \frac{2}{5}\, Info}$$

$$= \frac{2}{5}\left[-\frac{0}{2} \times \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \times \log_2\left(\frac{2}{2}\right)\right]$$

$$+ \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right]$$

$$+ \left(\frac{1}{5}\right)\left[-\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \left(\frac{0}{1}\right)\log_2\left(\frac{0}{1}\right)\right]$$

$$\text{Info}_{Temp}(D) = 0.4$$

$$\boxed{\text{Gain}(Temp) = 0.970 - 0.4 = 0.57}$$

~~Info~~

$$\text{Info}_{Humd}(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \text{Info}(D_j)$$

$$= \frac{|D_{high}|}{|D|} \text{Info}(D_{high}) + \frac{|D_{Normal}|}{|D|} \times \text{Info}(D_{Normal})$$

$$= \frac{3}{5} \times \left[\underbrace{-\frac{0}{3} \times \log_2\left(\frac{0}{3}\right) - \frac{3}{3} \times \log_2\left(\frac{3}{3}\right)}_{0} \right]$$

$$+ \frac{2}{5} \times \left[\underbrace{-\frac{2}{2} \times \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \log_2\left(\frac{0}{2}\right)}_{0} \right]$$

$$= 0$$

$$\therefore \boxed{\text{Gain}(Humidity) = \text{Info}(D) - \text{Info}_{Hund}(D) = 0.970 - 0 = 0.970}$$

(11)

$$Info_{Wind}(D) = \sum_{j=1}^{v=2} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \frac{|D_{weak}|}{|D|} Info(D_{weak})$$

$$+ \frac{|D_{strong}|}{|D|} \times Info(D_{strong})$$

$$= \frac{3}{5}\left[-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right]$$

$$+ \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$$

$$= 0.550 + 0.4$$

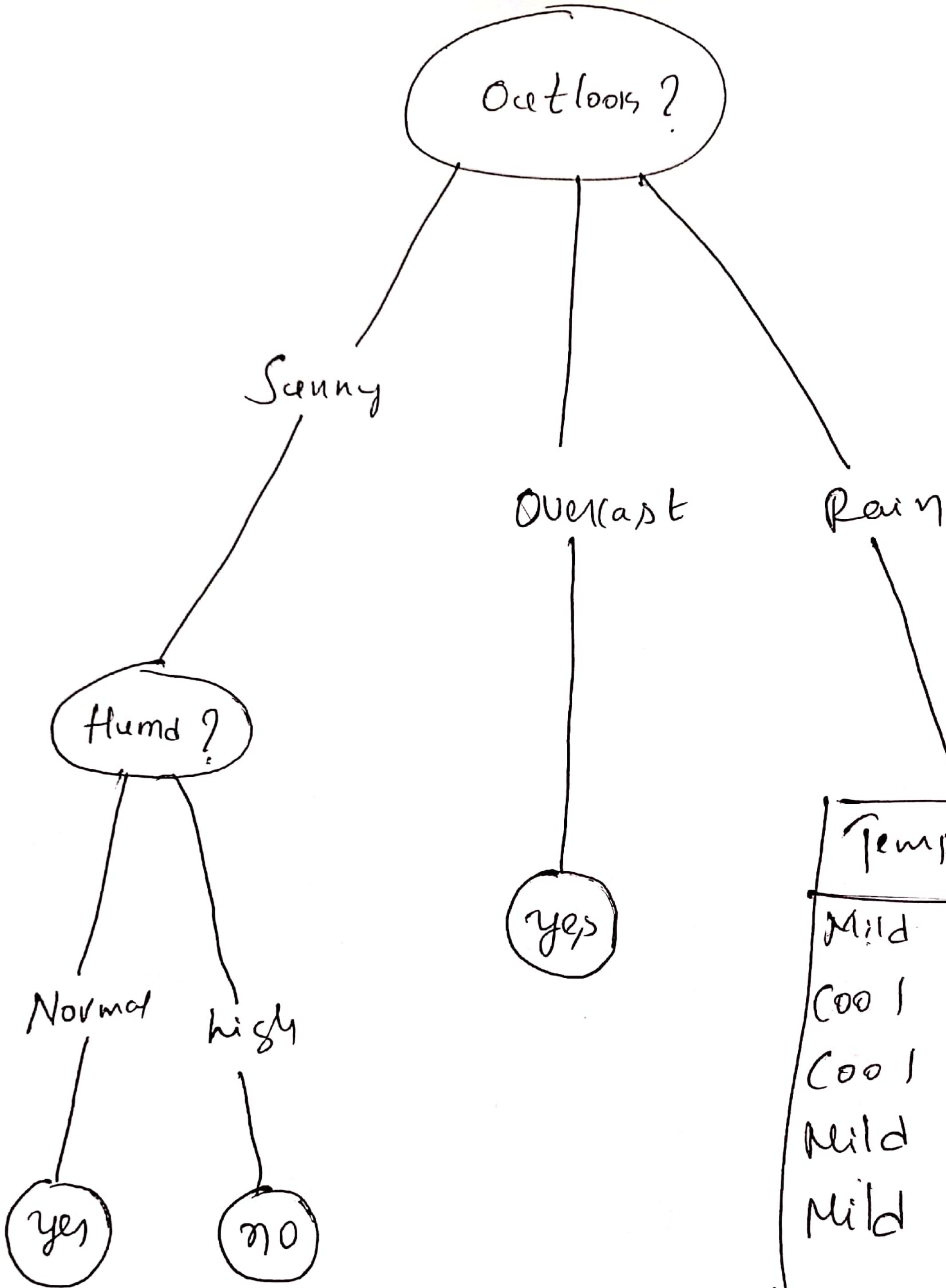
$$= \underline{0.950}$$

$$Gain(Wind) = Info(D) - Info_{Wind}(D)$$

$$= 0.970 - 0.950$$

$$= \underline{0.020}$$

$$\boxed{\begin{array}{l} Gain(Temp) = 0.57 \\ Gain(Humid) = 0.970 \\ Gain(Wind) = 0.020 \end{array}}$$



Temp	Humd	Wind	play Tennis
Mild	high	weak	yes
Cool	Normal	weak	yes
Cool	Normal	Strong	No
Mild	Normal	weak	yes
Mild	high	strong	No

Since D_{Rain} is impure,

Split D_{Rain} Using best Spliting attribute from the attribute list {Temp, Wind}

$$Info(D) = -\sum_{i=1}^{2} P_i \log_2 P_i$$

$$P_1 = \frac{3}{5} \qquad P_2 = \frac{2}{5}$$

$$= -\left[\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right]$$

$$= 0.970 \text{ bits}$$

$$Info_{Temp}(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \sum_{j=1}^{2} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \frac{|D_{Mild}|}{|D|} Info(D_{Mild}) + \frac{|D_{cool}|}{|D|} Info(D_{cool})$$

$$= \left(\frac{3}{5}\right)\left[-\left(\frac{2}{3}\right)\log_2\left(\frac{2}{3}\right) - \left(\frac{1}{3}\right)\log_2\left(\frac{1}{3}\right)\right] + \left(\frac{2}{5}\right)\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$$

$$= 0.550 + 0.4$$

$$= \underline{0.950}$$

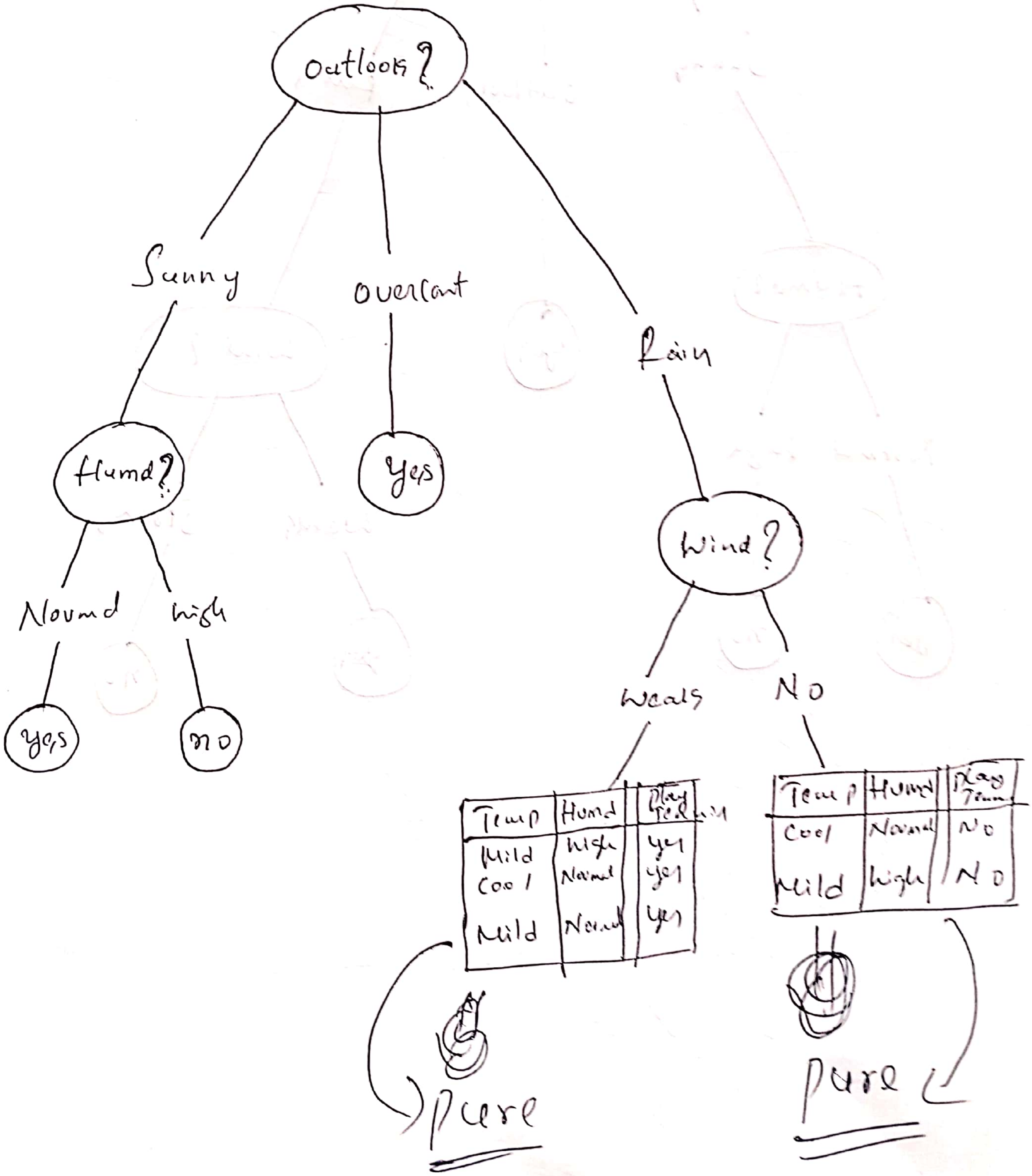
Scanned with CamScanner

Gain (Temp) = 0.020

Gain (Wind) = 0.970

∴

Therefore ~~our~~ Decision Tree is



∴ Final
Decision Tree