① $Gini(D) = 1 - \left(\frac{7}{10}\right)^2 - \left(\frac{3}{10}\right)^2 = 0.42$

3 values (High, Average, Low)

$2^3 = 8$ {Single, Married, Divorced}

$= \frac{8}{10}\left[1 - \left(\frac{6}{8}\right)^2 - \left(\frac{2}{8}\right)^2\right] + \frac{2}{10}\left[1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right]$

$= 0.30 + 0.1 = 0.4$

$Gini_{status \in \{Married, Divorced\}}(D)$

$= \frac{5}{10} Gini(D) + \frac{5}{10} Gini(D)$

$= \frac{5}{10}\left[1 - \left(\frac{4}{5}\right)^2 - \left(\frac{3}{5}\right)^2\right] + \frac{5}{10}\left[1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2\right]$

$= 0.16 + 0.24 = 0.4$

$Gini_{status \in \{Divorced, Single\}}(D)$

$= \frac{7}{10} Gini(D) + \frac{3}{10} Gini(D)$

$= \frac{7}{10}\left[1 - \left(\frac{4}{7}\right)^2 - \left(\frac{3}{7}\right)^2\right] + \frac{3}{10}\left[1 - \left(\frac{3}{3}\right)^2 - \left(\frac{0}{3}\right)^2\right]$

$= \frac{7}{10}(0.75) + \frac{3}{10}(0.00) = 0.3428 + 0.000 = 0.3428$

$Gini_{martial\ status = Single}(D)$

$= \frac{5}{10}\left[1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2\right] + \frac{5}{10}\left[1 - \left(\frac{4}{5}\right)^2 - \left(\frac{1}{5}\right)^2\right]$

$= 0.16 + 0.24 = 0.4$

$Gini_{martial\ status = Divorced}(D)$

$= \frac{2}{10}\left[1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right] + \frac{8}{10}\left[1 - \left[\frac{6}{8}\right]^2 - \left[\frac{2}{8}\right]^2\right]$

$= 0.3 + 0.1 = 0.4$

Continue →

For marital status reduction in impurity

$= Gini(D) - Gini_A(D)$

$= 0.42 - 0.3628 \quad , \quad 0.42 - 0.4$

$= 0.0772 \quad , \quad 0.02$

$\boxed{Min = 0.02}$

$Gini\ D_{(Low)} = Gini\ D \{High, \cancel{Av}\ Average\}$

$= \frac{3}{10} Gini\ D_1 + \frac{7}{10} Gini\ D_2$

$= \frac{3}{10}[1-(\frac{3}{3})^2 - (\frac{0}{3})^2] + \frac{7}{10}[1-(\frac{4}{7})^2 - (\frac{3}{7})^2]$

$= 0.3428$

$Gini\ D_{(average)} = Gini\ \{Low, high\}$

$= \frac{3}{10}[1-(\frac{0}{3})^2 - (\frac{6}{3})^2] + \frac{7}{10}[1-(\frac{7}{7})^2 - (\frac{0}{7})^2]$

$= 0 + 0 = 0$

$Gini\ D_{(high)} = Gini\ D\ \{Low, Average\}$

$= \frac{4}{10}[1-(\frac{4}{4})^2 - (\frac{0}{4})^2] + \frac{6}{10}[1-(\frac{3}{6})^2 - (\frac{3}{6})^2]$

$= 0 + \frac{6}{10}[1 - 0.5]$

$= 0 + 0.3 = 0.3$

Continue

1

Reducing Impurity $= 0.42 - 0, \quad 0.420 - 0.3428$

$= 0.42, \quad 0.072$

$0.42 - 0.320$ ½

$\boxed{\min = 0.072}$

For Employed = only 2 values (yes, no)

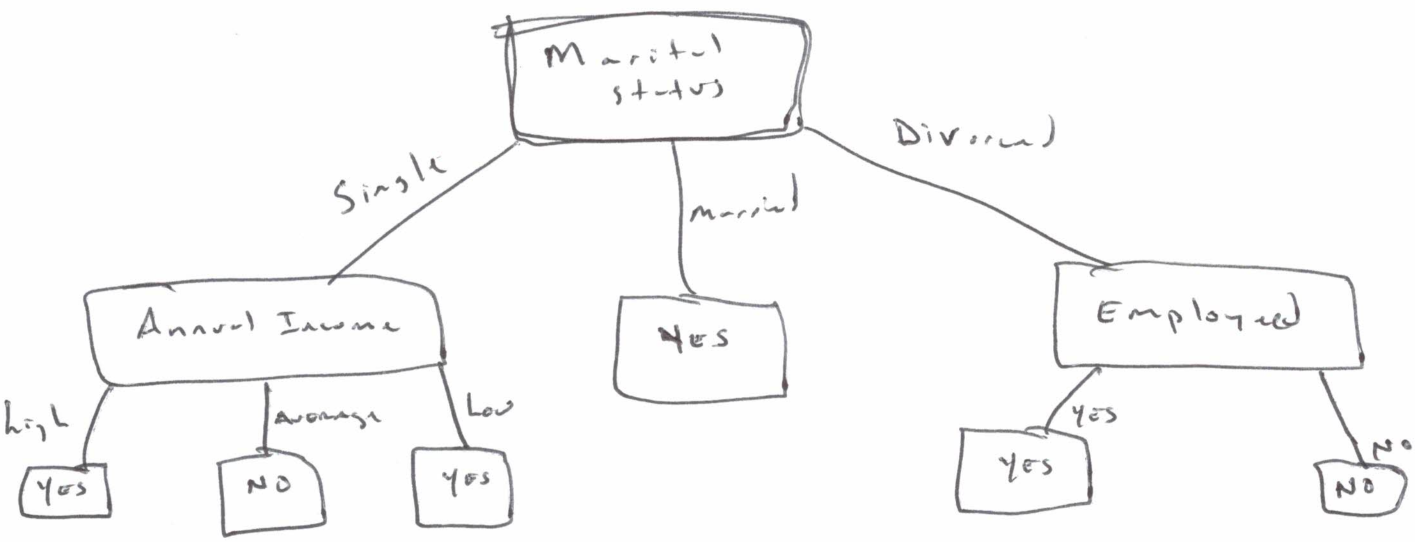
$2^2 = 4 \quad \{\phi, \{yes\}, \{no\}, \{yes, no\}\}$

$$\text{Gini}_{(yes)}(D) = \frac{3}{10}\left[1-\left(\frac{3}{3}\right)^2-\left(\frac{0}{3}\right)^2\right] + \frac{7}{10}\left[1-\left(\frac{4}{7}\right)^2-\left(\frac{3}{7}\right)^2\right]$$

$$= 0 + 0.3428 = 0.3428$$

Reducing Impurity $= 0.42 - 0.3428 = 0.072$

Minimum gini is given by attribute marital status $= 0.02$
Root of decision tree is marital status



For married we can directly assign to "Yes" because there is no class called "No" for an attribute. For Left and right minimum gini is the same. We can take annual income to left and Employed to right. For marital status, annual income high we have only "Yes" attribute and also for Average and for Low we have only the Yes class.

(2) Outlook = Sunny (2 yes, 3 NO)

$$Gini(2,3) = 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2\right] = 1 - \left[\frac{4}{25} + \frac{9}{25}\right] = 1 - \frac{13}{25} = \frac{25-13}{25} = \frac{12}{25}$$

Outlook = Overcast (4 yes, 0 no)

$$Gini(4,0) = 1 - \left[\left(\frac{4}{4}\right)^2\right] = (1-1) = 0$$

Outlook = Raining (3 yes, 2 no)

$$Gini(3,2) = 1 - \left[\left(\frac{3}{5}\right)^2 + \left(\frac{2}{5}\right)^2\right] = \frac{12}{25}$$

$$Gini(play, outlook) = \frac{5}{14}\left(\frac{12}{25}\right) + \frac{4}{14}(0) + \frac{5}{14}\left(\frac{12}{25}\right)$$

$$= \cancel{2} \times \frac{\cancel{5}}{\cancel{14}_7} \times \frac{12}{\cancel{25}_5} = \frac{12}{35} = 0.3428$$

Temperature = hot (2 yes, 2 no)

$$Gini(2,2) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right] = 1 - \left[\frac{4}{16} + \frac{4}{16}\right] = 1 - \frac{8}{16} = \frac{8}{16}$$

Temp = mild (4 yes, 2 no)

$$Gini(4,2) = 1 - \left[\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right] = 1 - \left[\frac{16}{36} + \frac{4}{36}\right] = 1 - \frac{20}{36} = \frac{16}{36}$$

Temp = cool (3 yes, 1 no)

$$Gini(3,1) = 1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right] = 1 - \left[\frac{9}{16} + \frac{1}{16}\right] = 1 - \frac{10}{16} = \frac{6}{16}$$

$$Gini(play, temp) = \frac{4}{14}\left(\frac{8}{16}\right) + \frac{6}{14}\left(\frac{16}{36}\right) + \frac{4}{14}\left(\frac{6}{16}\right)$$

$$= \frac{1}{14}\left(\cancel{4} \times \frac{8}{\cancel{16}_4} + \cancel{6} \times \frac{16}{\cancel{36}_6} + \cancel{4} \times \frac{6}{\cancel{16}_4}\right)$$

$$= \frac{1}{14}\left(\frac{8}{4} + \frac{16}{6} + \frac{6}{4}\right)$$

$$= \frac{1}{14}\left(2 + \frac{8}{3} + \frac{3}{2}\right)$$

$$= \frac{1}{14}\left(\frac{12 + 16 + 9}{6}\right)$$

$$= \frac{37}{84} = 0.4404$$

Continue →

(2) Humidity = High (3 yes, 4 No)

$$Gini(3,4) = 1 - \left[\left(\tfrac{3}{7}\right)^2 + \left(\tfrac{4}{7}\right)^2\right] = 1 - \left[\tfrac{9}{49} + \tfrac{16}{49}\right] = 1 - \tfrac{25}{49} = \tfrac{24}{49}$$

Humidity = Normal (6 yes, 1 No)

$$Gini(6,1) = 1 - \left[\left(\tfrac{6}{7}\right)^2 + \left(\tfrac{1}{7}\right)^2\right] = 1 - \left[\tfrac{36}{49} + \tfrac{1}{49}\right] = 1 - \tfrac{37}{49} = \tfrac{12}{49}$$

$$Gini(play, humidity) = \tfrac{7}{14}\left(\tfrac{24}{49}\right) + \tfrac{7}{14}\left(\tfrac{12}{49}\right)$$
$$= \tfrac{1}{2}\left(\tfrac{24}{49}\right) + \tfrac{1}{2}\left(\tfrac{12}{49}\right)$$
$$= \tfrac{12}{49} + \tfrac{6}{49} = \tfrac{18}{49} = 0.3673$$

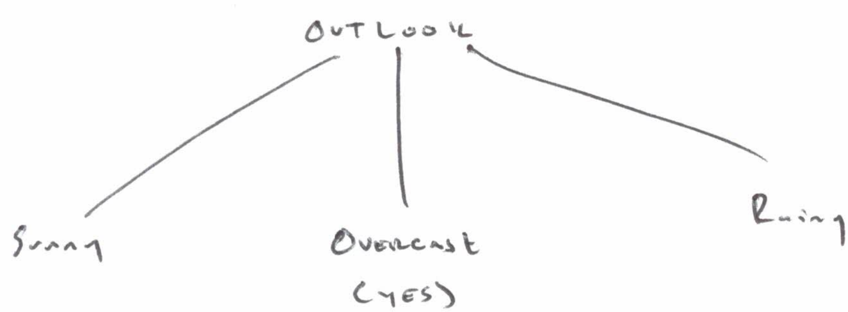
Windy = True (3 No, 3 yes)

$$Gini(3,3) = 1 - \left[\left(\tfrac{3}{6}\right)^2 + \left(\tfrac{3}{6}\right)^2\right] = 1 - \left[2 \times \tfrac{9}{36}\right] = 1 - \tfrac{18}{36} = \tfrac{18}{36}$$

Windy = False (6 yes, 2 No)

$$Gini(6,2) = 1 - \left[\left(\tfrac{6}{8}\right)^2 + \left(\tfrac{2}{8}\right)^2\right] = 1 - \left[\tfrac{36}{64} + \tfrac{4}{64}\right] = 1 - \tfrac{40}{64} = \tfrac{24}{64}$$

$$Gini(play.windy) = \tfrac{\cancel{6}}{14} \times \tfrac{\cancel{18}\,3}{\cancel{36}\,\cancel{6}} + \tfrac{\cancel{8}}{14} \times \tfrac{24\,3}{\cancel{64}\,8} = \tfrac{3}{14} + \tfrac{3}{14} = \tfrac{6}{14} = 0.4288$$



Outlook = Sunny

Temp	Humidity	Windy	Play
Hot	high	False	No
Hot	high	True	No
mild	high	false	No
~~Cold~~ cool	Normal	false	yes
cool	Normal	False	yes

Continue →

(2)

(1) Temperature = Hot (0 yes, 2 no)

$Gini(0,2) = 1 - (\frac{2}{2})^2 = 1 - 1 = 0$

Temp = mild (0 yes, 1 no)

$Gini(0,1) = 1 - (1)^2 = 1 - 1 = 0$

Temp = cool (2 yes, 0 no)

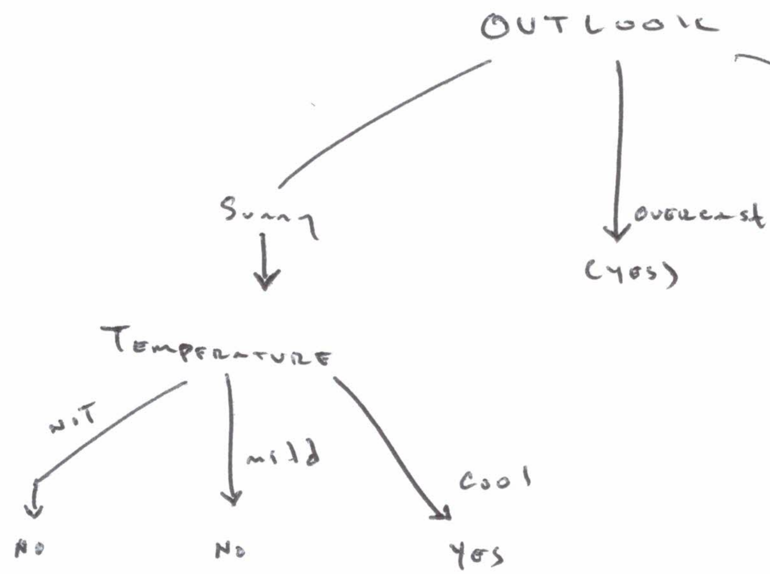
$Gini(2,0) = 1 - (\frac{2}{2})^2 = 1 - 1 = 0$

Gini (play, temp)

$= \frac{2}{5}(0) + \frac{1}{5}(0) + \frac{2}{5}(0)$

$= 0$ (min possible)

Root node



Outlook = Rainy

Temp	Humidity	Windy	Play
mild	high	False	Yes
cool	Normal	False	yes
cool	Normal	True	NO
mild	normal	False	Yes
mild	high	true	NO

Continue →

(2)

(1) Temp = mild (2 yes, 1 no)

$Gini(2,1) = 1 - [(\frac{2}{3})^2 + (\frac{1}{3})^2] = 1 - [\frac{4}{9} + \frac{1}{9}] = 1 - \frac{5}{9} = \frac{4}{9}$

Temp = cool (1 yes, 1 no)

$Gini(1,1) = 1 - [(\frac{1}{2})^2 + (\frac{1}{2})^2] = 1 - [\frac{1}{4} + \frac{1}{4}] = 1 - \frac{2}{4} = \frac{2}{4} = \frac{1}{2}$

$Gini(play, temp) = \frac{3}{5}(\frac{4}{9}) + \frac{2}{5}(\frac{1}{2}) = \frac{12}{45} + \frac{2}{10} = \frac{24+18}{90} = \frac{42}{90} = \frac{21}{45}$

(2) Humidity = High (1 yes, 1 no)

$Gini(1,1) = 1/2$

Humidity = Normal (2 yes, 1 no)

$Gini(2,1) = 4/9$

$Gini(play, Humidity) = \frac{2}{5}(\frac{1}{2}) + \frac{3}{5}(\frac{4}{9}) = \frac{21}{45}$

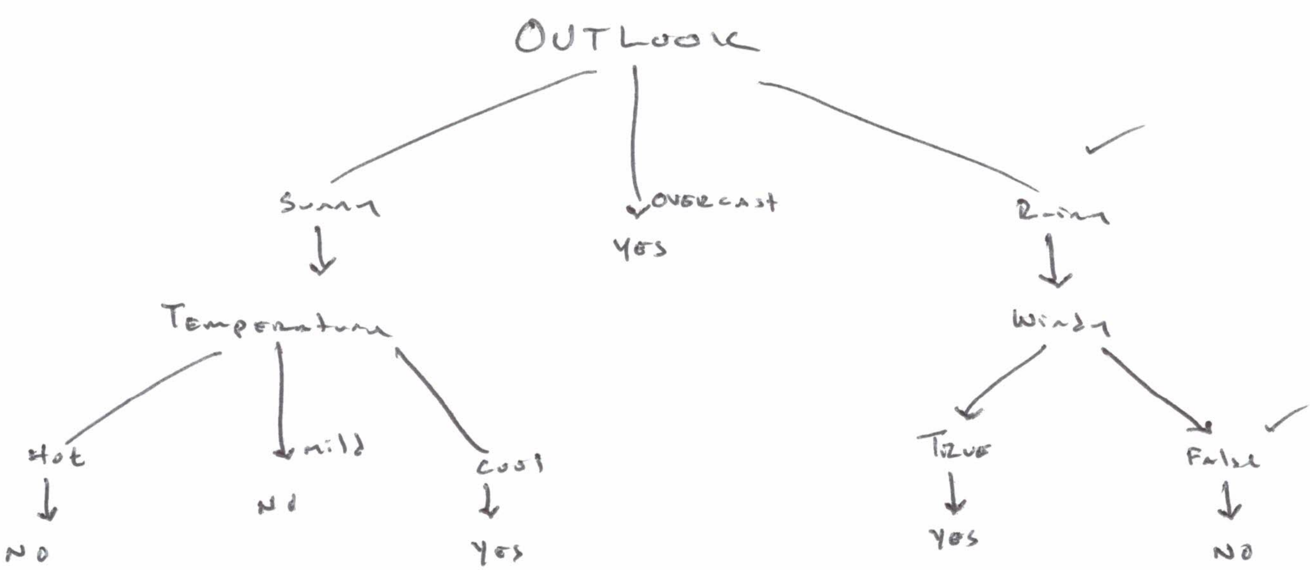
(3) Windy = False (3 yes, 0 no)

$Gini(3,0) = 0$

Windy = True (0 yes, 2 no)

$Gini(0,2) = 0$

$Gini(play, windy) = \frac{3}{5}(0) + \frac{2}{5}(0) = 0$ (min) => Root



X = (Outlook = Rainy, Temp = Hot, Humidity = High, Windy = False) = NO

(4) Yes = 3 , No = 7

$X = (HM = NO, MS = \text{Divorced}, AI = 120000) = ?$

$P(x|yes) = \pi^{n}\ p(x_i|yes) = P(y) \times (P_1)(P_2) \ldots (P_n)$

$p(Ho = No|yes) = 3/3$

$P(ms = \text{divorced}|yes) = 1/3$

$p(AI > 91000|yes) = 1/3$

$P(x|yes) = 3/10 \times (3/3 \times 1/3 \times 1/3) = \frac{1}{10 \times 3} = 0.0333$

$p(x|no) = \pi^{n}_{i=1}\ P(x_i|No) = p(No) \times (P_1)(P_2) \ldots (P_n)$

$P(Ho = no|no) = 4/7$

$P(ms = \text{divorced}|No) = 1/3$

$P(AI > 91000|No) = 4/7$

$P(x|No) = (7/10) \times (4/7 \times 1/3 \times 4/7) = \frac{16}{10 \times 7 \times 3} = 0.0533$

Since p(x|yes) < p(x|No), so X will be labeled with "NO" class

P(x|yes) has less probability than p(x|no), so p(x|no) will dominate p(x|yes) and is why X instance will be labeled as "NO"