EXERCISE DECISION TREES

$$I_m^* = -\sum_{j=1}^{n} \frac{N_{m,j}}{N_m} \left(\sum_{i=1}^{K} p_{m,j}^i \log_2 p_{m,j}^i \right)$$

Looking for the most discriminant attribute:

- **Attribute CONDITION**

condition	TOTAL	play	no_play
sunny	5	2	3
cloudy	4	4	0
rainy	5	3	2
SUM	14		

$$I_0^* = -\left(\frac{5}{14} \cdot \left(\frac{2}{5} \cdot \log_2 \frac{2}{5} + \frac{3}{5} \cdot \log_2 \frac{3}{5} \right) + \frac{4}{14} \cdot \left(\frac{4}{4} \cdot \log_2 \frac{4}{4} + \frac{0}{4} \cdot \log_2 \frac{0}{4} \right) + \frac{5}{14} \cdot \left(\frac{3}{5} \cdot \log_2 \frac{3}{5} + \frac{2}{5} \cdot \log_2 \frac{2}{5} \right) \right) = 0.693536$$

- **Attribute TEMPERATURE**

tempperature	TOTAL	play	no_play
<72	6	4	2
>=72	8	5	3
SUM	14		

$$I_0^* = -\left(\frac{6}{14} \cdot \left(\frac{4}{6} \cdot \log_2 \frac{4}{6} + \frac{2}{6} \cdot \log_2 \frac{2}{6} \right) + \frac{8}{14} \cdot \left(\frac{5}{8} \cdot \log_2 \frac{5}{8} + \frac{3}{8} \cdot \log_2 \frac{3}{8} \right) \right) = 0.938946$$

- **Attribute HUMIDITY**

humidity	TOTAL	play	no_play
<80	6	5	1
>=80	8	4	4
SUM	14		

$$I_0^* = -\left(\frac{6}{14}\cdot\left(\frac{5}{6}\cdot\log_2\frac{5}{6}+\frac{1}{6}\cdot\log_2\frac{1}{6}\right)+\frac{8}{14}\cdot\left(\frac{4}{8}\cdot\log_2\frac{4}{8}+\frac{4}{8}\cdot\log_2\frac{4}{8}\right)\right)$$
$$= 0.85001$$

- **Attribute WIND**

wind	TOTAL	play	no_play
yes	6	3	3
no	8	6	2
SUM	14		

$$I_0^* = -\left(\frac{6}{14}\cdot\left(\frac{3}{6}\cdot\log_2\frac{3}{6}+\frac{3}{6}\cdot\log_2\frac{3}{6}\right)+\frac{8}{14}\cdot\left(\frac{6}{8}\cdot\log_2\frac{6}{8}+\frac{2}{8}\cdot\log_2\frac{2}{8}\right)\right)$$
$$= 0.892159$$

Therefore, first split will be with attribute "condition" because its entropy is the lowest. This will be the attribute in the root node.