ex 17

Sonntag, 12. Juni 2022 15:49

a) root of the tree:

Y = soccer 2

no. of games: 14

$P(\text{soccer} = \text{yes}) = \frac{9}{14}$

$P(\text{soccer} = \text{no}) = \frac{5}{14}$

$$H(Y) = -\sum_z P(Y=z) \cdot \log_2(P(Y=z))$$

$$= -P(\text{soccer}=\text{yes}) \cdot \log_2(P(\text{soccer}=\text{yes})) - P(\text{soccer}=\text{no}) \cdot \log_2(P(\text{soccer}=\text{no}))$$

$$= -\frac{9}{14} \cdot \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \cdot \log_2\left(\frac{5}{14}\right)$$

$$\approx \underline{0,94}$$

b) X = wind

$P(\text{wind} = \text{true}) = \frac{6}{14}$

$P(\text{wind} = \text{false}) = \frac{8}{14}$

information gain when cutting on X:

$$IG(X, Y) = H(Y) - H(Y|X)$$

$$H(Y|X) = -\sum_m P(X=m) \sum_z P(Y=z \mid X=m) \log\left(P(Y=z \mid X=m)\right)$$

NR:

$P(\text{soccer}=\text{yes} \mid \text{wind}=\text{true}) = \frac{3}{6}$ $\quad$ $P(\text{soccer}=\text{yes} \mid \text{wind}=\text{false}) = \frac{6}{8}$

$P(\text{soccer}=\text{no} \mid \text{wind}=\text{true}) = \frac{3}{6}$ $\quad$ $P(\text{soccer}=\text{no} \mid \text{wind}=\text{false}) = \frac{2}{8}$

$$= 0,94 + \frac{6}{14} \cdot \left[\frac{3}{6} \cdot \log_2\left(\frac{3}{6}\right) + \frac{3}{6} \cdot \log_2\left(\frac{3}{6}\right)\right] + \frac{8}{14} \cdot \left[\frac{6}{8} \cdot \log_2\left(\frac{6}{8}\right) + \frac{2}{8} \cdot \log_2\left(\frac{2}{8}\right)\right]$$

$$\approx \underline{0,048}$$