ex 17

a) root of the tree:

Y = soccer ?

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:

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

$H(Y|X) = -\sum_{m} P(X=m) \sum_{z} P(Y=z \mid X=m) \log (P(Y=z \mid X=m))$

NR:

$P(\text{soccer}=\text{yes} \mid \text{wind}=\text{true}) = \frac{2}{6}$ $P(\text{soccer}=\text{yes} \mid \text{wind}=\text{false}) = \frac{6}{8}$

$P(\text{soccer}=\text{no} \mid \text{wind}=\text{true}) = \frac{4}{6}$ $P(\text{soccer}=\text{no} \mid \text{wind}=\text{false}) = \frac{2}{8}$

$= 0,94 - \frac{6}{14} \cdot \left[\frac{2}{6} \cdot \log_2 \left(\frac{2}{6}\right) + \frac{4}{6} \cdot \log_2 \left(\frac{4}{6}\right)\right] - \frac{8}{14} \cdot \left[\frac{6}{8} \cdot \log_2 \left(\frac{6}{8}\right) + \frac{2}{8} \cdot \log_2 \left(\frac{2}{8}\right)\right]$

$\approx \underline{0,083}$