R05631027_楊皓文_機器學習_作業#8

Problem 1 (Decision tree)

R05631027 楊皓文 ML HW#8

1. $Entropy(K) = -\left(\frac{9}{14}\log_2\frac{9}{14} + \frac{5}{14}\log_2\frac{5}{14}\right) = 0.9403$ (Yes / No)

$Gain(K, Temp) = Entropy(K) - \frac{6}{14}Entropy(K_1) - \frac{4}{14}Entropy(K_2) - \frac{4}{14}Entropy(K_3)$ (<15° / 15-25° / >25°)

(其中 $Entropy(K_1) = -\left(\frac{2}{6}\log_2\frac{2}{6} + \frac{4}{6}\log_2\frac{4}{6}\right) = 0.9183$

$Entropy(K_2) = -\left(\frac{4}{4}\log_2\frac{4}{4}\right) = 0$

$Entropy(K_3) = -\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right) = 0.8113$)

$= 0.9403 - \frac{6}{14}\cdot 0.9183 - 0 - \frac{4}{14}\cdot 0.8113 = 0.3149$ ✓

$Gain(K, Midterm) = Entropy(K) - \frac{4}{14}Entropy(K_1) - \frac{6}{14}Entropy(K_2) - \frac{4}{14}Entropy(K_3)$ (low / medium / high)

(其中 $Entropy(K_1) = -\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right) = 0.8113$

$Entropy(K_2) = -\left(\frac{4}{6}\log_2\frac{4}{6} + \frac{2}{6}\log_2\frac{2}{6}\right) = 0.9183$

$Entropy(K_3) = -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right) = 1$)

$= 0.9403 - \frac{4}{14}\cdot 0.8113 - \frac{6}{14}\cdot 0.9183 - \frac{4}{14}\cdot 1 = 0.029$

$Gain(K, HW\_done) = Entropy(K) - \frac{7}{14}Entropy(K_1) - \frac{7}{14}Entropy(K_2)$ (Yes / No)

(其中 $Entropy(K_1) = -\left(\frac{6}{7}\log_2\frac{6}{7} + \frac{1}{7}\log_2\frac{1}{7}\right) = 0.5917$

$Entropy(K_2) = -\left(\frac{3}{7}\log_2\frac{3}{7} + \frac{4}{7}\log_2\frac{4}{7}\right) = 0.9852$)

$= 0.9403 - \frac{7}{14}\cdot 0.5917 - \frac{7}{14}\cdot 0.9852 = 0.1519$

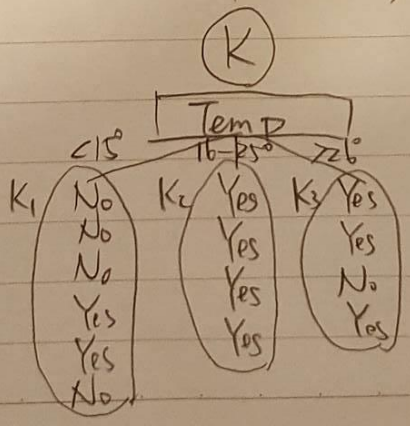
$Gain(K, Quiz) = Entropy(K) - \frac{6}{14}Entropy(K_1) - \frac{8}{14}Entropy(K_2)$ (Good / Bad)

(其中 $Entropy(K_1) = -\left(\frac{3}{6}\log_2\frac{3}{6} + \frac{3}{6}\log_2\frac{3}{6}\right) = 1$

$Entropy(K_2) = -\left(\frac{6}{8}\log_2\frac{6}{8} + \frac{2}{8}\log_2\frac{2}{8}\right) = 0.8113$)

$= 0.9403 - \frac{6}{14}\cdot 1 - \frac{8}{14}\cdot 0.8113 = 0.0481$

∴ $Gain(K, Temp)$ 最大，因此選"Temp"做為 root node



No.
Date : :

接下來搭配觀察法，可看出 <15°C 再選擇 HW_done attribute，>26°C 再選擇 Quiz attribute，即可完全分類。

<15°：$En(K) = -\left(\frac{2}{6}\log_2\left(\frac{2}{6}\right) + \frac{4}{6}\log_2\left(\frac{4}{6}\right)\right) = 0.9183$

$Gain(K, HW\_done) = En(K) - \frac{2}{6}En(K_1) - \frac{4}{6}En(K_2)$ (Yes, No)

（其中 $En(K_1) = -\left(\frac{2}{2}\log_2\left(\frac{2}{2}\right)\right) = 0$

$En(K_2) = -\left(\frac{4}{4}\log_2\left(\frac{4}{4}\right)\right) = 0$）

$= 0.9183 - 0 - 0 = 0.9183$ ✓

>26°：$En(K) = -\left(\frac{3}{4}\log_2\left(\frac{3}{4}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right) = 0.8113$

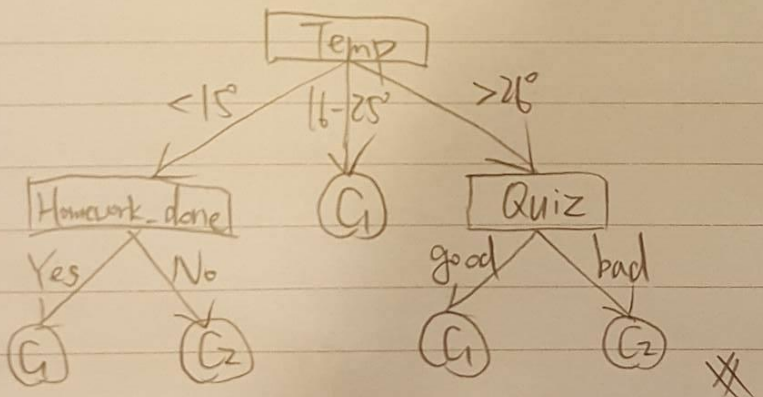
$Gain(K, Quiz) = En(K) - \frac{1}{4}En(K_1) - \frac{3}{4}En(K_2)$ (good, bad)

（其中 $En(K_1) = -\left(\frac{1}{1}\log_2\left(\frac{1}{1}\right)\right) = 0$

$En(K_2) = -\left(\frac{3}{3}\log_2\left(\frac{3}{3}\right)\right) = 0$）

$= 0.8113 - 0 - 0 = 0.8113$ ✓

所以 model 為：



※