

ใบงาน Machine Learning

วัตถุประสงค์ หัดแทนค่าเพื่อสร้าง C4.5 decision tree

$$E_{C_k}(S) = - \sum_{j=1}^v p_{j \text{ and } c=k} \log_2(p_{j \text{ and } c=k})$$

$$E_C(S) = \sum_{k=1}^u p_k E_{C_k}(S)$$

โมเดล C4.5 decision tree เพื่อสร้างโมเดลพยากรณ์ เขียนได้ว่า
กล่าวคือ Entropy, E, ของของข้อมูล S เมื่อใช้ Candidate, C_k , มีค่าเป็น
ผลรวมของ $-p_{j \text{ and } c=k} \log_2(p_{j \text{ and } c=k})$ โดย j คือ label ที่สนใจ (เช่น เล่น
หรือไม่เล่น) และ k คือ ค่าของ candidate นั้นๆ และ $p_{j \text{ and } c=k}$ คือความ
น่าจะเป็นที่เหตุการณ์ค่า $c=k$ มีค่าเป็น j
เมื่อนำ E_{C_k} มาผลรวมก็เพียงถ่วงน้ำหนักด้วยสัดส่วนของแต่ละ $c=k$
ก็จะได้ Entropy เมื่อใช้ C เป็น candidate

...จากข้อมูลต่อไปนี้

Weather	Temp	Wind	label
s	h	F	n
s	h	T	n
o	h	F	y
r	m	F	y
r	c	F	y
r	c	T	n
o	c	T	y
s	m	F	n

x	$\log_2(x)$
0	0
1/2	-1
1	0
1/3	-1.585
2/3	-0.585
1/4	-2
3/4	-0.415
1/5	-2.3219
2/5	-1.3219
3/5	-0.737
4/5	-0.3219
1/6	-2.585
5/6	-0.263
1/7	-2.8074
2/7	-1.8074
3/7	-1.2224
4/7	-0.8074
5/7	-0.4854
6/7	-0.2224
1/8	-3
3/8	-1.415
5/8	-0.6781
7/8	-0.1926

คำสั่ง

1. คำนวณ $E(S)$, $E_{C=\text{สภาพอากาศ}}(S)$, $E_{C=\text{อุณหภูมิ}}(S)$, $E_{C=\text{สภาพลม}}(S)$ ในตาราง

2. candidate ไດ ให้ $E_C(S)$ ต่ำที่สุด **Ans** $E_C = \text{สภาพอากาศ}$

3. สารสนเทศที่ได้จากข้อ 2 สร้างกฎอะไรได้บ้าง

Ans สามารถสร้าง tree จากสภาพอากาศ
ถ้าอากาศเป็น s จะไม่ออกไปเล่นเสมอ
ถ้าอากาศเป็น o จะออกไปเล่นเสมอ
ถ้าอากาศ r wind = F จะเล่น
r wind = T ไม่เล่น

$E(S) = \sum_{j=1}^v -p_j \log_2(p_j) =$						
		P_j	$P(j=y k)$	$P(j=n k)$	$p_{j=C_k} \log_2(p_{j=C_k})$ /*ติด \log_2 ไว้ได้*/	Remark (P_j)
สภาพอากาศ	$k=s$	3/8	0/3	3/3	$-\frac{0}{3} \log_2(\frac{0}{3}) + -(\frac{3}{3}) \log_2(\frac{3}{3}) = 0$	$P\{j=y \text{อากาศ}=s\} = \{\}$ $P\{j=n \text{อากาศ}=s\} = \{\#1, \#2, \#8\}$
	$k=o$	2/8	2/2	0/2	$-\frac{2}{2} \log_2(\frac{2}{2}) + -(\frac{0}{2}) \log_2(\frac{0}{2}) = 0$	$P\{j=y \text{อากาศ}=o\} = \{\#3, \#7\}$ $P\{j=n \text{อากาศ}=o\} = \{\}$
	$k=r$	3/8	2/3	1/3	$-\frac{2}{3} \log_2(\frac{2}{3}) + -(\frac{1}{3}) \log_2(\frac{1}{3}) = 0.91829$	$P\{j=y \text{อากาศ}=r\} = \{\#4, \#5\}$ $P\{j=n \text{อากาศ}=r\} = \{\#6\}$
$E_{C=\text{สภาพอากาศ}}(S) = \frac{3}{8} I(0, \frac{3}{8}) + \frac{2}{8} I(\frac{2}{2}, 0) + \frac{3}{8} I(\frac{2}{3}, \frac{1}{3}) = \frac{3}{8} (0) + \frac{2}{8} (0) + \frac{3}{8} (0.91829) = 0.34436$						
อุณหภูมิ	$k=h$	3/8	1/3	2/3	$-\frac{1}{3} \log_2(\frac{1}{3}) + -(\frac{2}{3}) \log_2(\frac{2}{3}) = 0.91829$	$P\{j=y \text{อุณหภูมิ}=h\} = \{\#5\}$ $P\{j=n \text{อุณหภูมิ}=h\} = \{\#1, \#2\}$
	$k=m$	2/8	1/2	1/2	$-\frac{1}{2} \log_2(\frac{1}{2}) + -(\frac{1}{2}) \log_2(\frac{1}{2}) = 1$	$P\{j=y \text{อุณหภูมิ}=m\} = \{\#4\}$ $P\{j=n \text{อุณหภูมิ}=m\} = \{\#8\}$
	$k=c$	3/8	2/3	1/3	$-\frac{2}{3} \log_2(\frac{2}{3}) + -(\frac{1}{3}) \log_2(\frac{1}{3}) = 0.91829$	$P\{j=y \text{อุณหภูมิ}=c\} = \{\#5, \#7\}$ $P\{j=n \text{อุณหภูมิ}=c\} = \{\#6\}$
$E_{C=\text{อุณหภูมิ}}(S) = \frac{3}{8} (0.91829) + \frac{2}{8} (1) + \frac{3}{8} (0.91829) = 0.95872$						
สภาพลม	$k=T$	3/8	1/3	2/3	$-\frac{1}{3} \log_2(\frac{1}{3}) + -(\frac{2}{3}) \log_2(\frac{2}{3}) = 0.91829$	$P\{j=y \text{สภาพลม}=T\} = \{\#2, \#6\}$ $P\{j=n \text{สภาพลม}=T\} = \{\}$
	$k=F$	5/8	3/5	2/5	$-\frac{3}{5} \log_2(\frac{3}{5}) + -(\frac{2}{5}) \log_2(\frac{2}{5}) = 0.97095$	$P\{j=y \text{สภาพลม}=F\} = \{\#3, \#4, \#5\}$ $P\{j=n \text{สภาพลม}=F\} = \{\#1, \#7\}$
$E_{C=\text{สภาพลม}}(S) = \frac{3}{8} (0.91829) + \frac{5}{8} (0.97095) = 0.95120$						