05506005 ชื่อ ..... รหัสนักศึกษา .....

ใบงาน Machine Learning

วัตถุประสงค์ หัดแทนค่าเพื่อสร้าง C4.5 decision tree

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

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

โมเดล C4.5 decision tree เพื่อสร้างโมเดลพยากรณ์ เขียนได้ว่า กล่าวคือ Entropy, E, ของของข้อมูล S เมื่อใช้ Candidate, C_k, มีค่าเป็นผลรวมของ $-p_{j \text{ and } c=k}\log_2(p_{j \text{ and } c=k})$ โดย j คือ label ที่สนใจ (เช่น เล่น หรือ ไม่เล่น) และ k คือ ค่าของ candidate นั้นๆ และ $p_{j \text{ and } c=k}$ คือความน่าจะเป็นที่เหตุการณ์ค่า c=k มีค่าเป็น j

เมื่อนำ E_{CK} 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), $E_{C=สภาพอากาศ}(S)$, $E_{C=อุณหภูมิ}(S)$, $E_{C=สภาพลม}(S)$ ในตาราง
2. candidate ใด ให้ Ec(S) ต่ำที่สุด
3. สารสนเทศที่ได้จากข้อ 2 สร้างกฎอะไรได้บ้าง

E(S) = $\sum_{j=1}^{v} -p_j log_2(p_j)$=						
		P_j	P(j=y|k)	P(j=n|k)	$p_j c_k \log2(p_j c_k)$ /*ติด log2 ไว้ได้*/	Remark (P_j)
สภาพอากาศ	k=s	3/8	0/3	3/3		$P\{_{j=y\|อากาศ=s}\}$ = {} $P\{_{j=n\|อากาศ=s}\}$ = {#1,#2,#8}
	k=o	2/8	2/2	0/2		
	k=r	3/8	2/3	1/3		
$E_{C=สภาพอากาศ}(S)$ =						
อุณหภูมิ	k=h	3/8				
	k=m	2/8				
	k=c	3/8				
$E_{C=อุณหภูมิ}(S)$ =						
สภาพลม	k=T					
	k=F					
$E_{C=สภาพลม}(S)$ =						

กำหนดส่ง TBA