

Example: Attribute Selection with Information Gain

□ Class P: buys_computer = "yes"

□ Class N: buys_computer = "no"

$$Info(D) = I(9,5) = -\frac{9}{14} \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \log_2\left(\frac{5}{14}\right) = 0.940$$

age	p _i	n _i	I(p _i , n _i)
<=30	2	3	0.971
31...40	4	0	0
>40	3	2	0.971

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

$$Info_{age}(D) = \frac{5}{14} I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2) = 0.694$$

$\frac{5}{14} I(2,3)$ means "age <=30" has 5 out of 14 samples, with 2 yes'es and 3 no's.

Hence

$$Gain(age) = Info(D) - Info_{age}(D) = 0.246$$

Similarly, we can get

$$Gain(income) = 0.029$$

$$Gain(student) = 0.151$$

$$Gain(credit_rating) = 0.048$$

12

$$Info(D) = I(8,4) = -(8/12) \log_2(8/12) - (4/12) \log_2(4/12) = 0.918$$

age	p _i	n _i	I(p _i , n _i)	
<=30	2	2	1.000	
31...40	3	0	0	
>40	3	2	0.971	

$$Info_{age}(D) = 4/12 I(2,2) + 3/12 I(3,0) + 5/12 I(3,2) = 0.738$$

$$Info_{income}(D) = 3/12 I(2,1) + 5/12 I(4,1) + 4/12 I(2,2) =$$

$$Gain(age) = Info(D) - Info_{age}(D) = 0.918 - 0.738$$

$$= 0.18$$

$$Gain(income) =$$

$$Gain(Student) =$$

$$Gain(credit_rating) =$$