Swarali Paygude
PA22

Panel 1

PAGE NO.
DATE: / /

# Decision Tree

- buys_comp = 'yes' : Y
  buys_comp = 'no' : N

- Entropy for class labels :

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

$$= 0.940$$

- Info of each attribute :

| Age | $y_i$ | $n_i$ | $I(y_i, n_i)$ |
|---|---|---|---|
| <= 30 | 2 | 3 | 0.971 |
| 30..40 | 4 | 0 | 0 |
| >40 | 3 | 2 | 0.971 |

$$\text{Info}_{age}(D) = \frac{5}{14} \times I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2)$$

$$= 0.694$$

| Income | $y_i$ | $n_i$ | $I(y_i, n_i)$ |
|---|---|---|---|
| high | 2 | 2 | 1 |
| medium | 4 | 2 | 0.918 |
| low | 3 | 1 | 0.811 |

$$\text{Info}_{income}(D) = \frac{4}{14} I(2,2) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1)$$

$$= 0.911$$

| Student | $y_i$ | $n_i$ | $I(y_i, n_i)$ |
|---|---|---|---|
| yes | 6 | 1 | 0.592 |
| no | 3 | 4 | 0.985 |

$$Info_{student}(D) = \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4)$$

$$= 0.789$$

| Credit rating | $y_i$ | $n_i$ | $I(y_i, n_i)$ |
|---|---|---|---|
| excellent | 3 | 3 | 1 |
| Fair | 6 | 2 | 0.811 |

$$Info_{cr}(D) = \frac{6}{14} I(3,3) + \frac{8}{14} I(6,2)$$

$$= 0.892$$

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

$$= -0.694 + 0.940 = 0.246$$

$$Gain(income) = Info(D) - Info_{income}(D)$$

$$= 0.940 - 0.911 = 0.029$$

$$Gain(student) = Info(D) - Info_{student}(D)$$

$$= 0.940 - 0.789 = 0.151$$

$$Gain(credit\ rating) = Info(D) - Info_{cr}(D)$$

$$= 0.940 - 0.892 = 0.041$$

For age >40

$$\text{Info}(D) = I(3,2)$$

$$= -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)$$

$$= 0.971$$

Information of each attr.:

| Income | $y_i$ | $n_i$ | $I(y_i, n_i)$ |
|---|---|---|---|
| high | 0 | 0 | 0 |
| medium | 1 | 1 | 1 |
| low | 2 | 1 | 0.918 |

$$\text{Info}_{\text{income}}(D) = 0 + \frac{2}{5} I(1,1) + \frac{3}{5} I(2,1)$$

$$= 0.951$$

| Student | $y_i$ | $n_i$ | $I(y_i, n_i)$ |
|---|---|---|---|
| yes | 2 | 1 | 0.918 |
| no | 1 | 1 | 1 |

$$\text{Info}_{\text{student}}(D) = \frac{3}{5} I(2,1) + \frac{2}{5} I(1,1)$$

$$= 0.951$$

| Cred. rating | $y_i$ | $n_i$ | $I(y_i, n_i)$ |
|---|---|---|---|
| Fair | 3 | 0 | 0 |
| excellent | 0 | 2 | 0 |

$$Info_{cr}(D) = \frac{3}{5} I(3,0) + \frac{2}{5} I(0,2)$$

$$= 0$$

$$Gain(income) = Info(D) - Info_{income}(D)$$

$$= 0.971 - 0.951$$

$$= 0.020$$

$$Gain(student) = Info(D) - Info_{stud}(D)$$

$$= 0.971 - 0.951$$

$$= 0.020$$

$$Gain(cred.\ rating) = Info(D) - Info_{cr}(D)$$

$$= 0.971 - 0$$

$$= 0.971$$

For age $\leq 30$

$$Info(D) = I(2,3) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)$$

$$= 0.971$$

| Income | $y_i$ | $n_i$ | $I(y_i, n_i)$ |
|---|---|---|---|
| h. | 0 | 2 | 0 |
| m. | 1 | 1 | 1 |
| l. | 1 | 0 | 0 |

$$Info_{income}(D) = \frac{2}{5} I(1,1)$$

$$= 0.40$$

| student | $y_i$ | $n_i$ | $I(y_i, n_i)$ |
|---|---|---|---|
| yes | 2 | 0 | 0 |
| no | 0 | 3 | 0 |

$Info_{student}(D) = 0$

| cred. rating | $y_i$ | $n_i$ | $I(y_i, n_i)$ |
|---|---|---|---|
| excellent | 1 | 1 | 1 |
| fair | 1 | 2 | 0.918 |

$Info_{cr}(D) = \frac{2}{5} I(1,1) + \frac{3}{5} I(1,2)$

$= 0.951$

Gain (income) = ~~0.048~~ $0.971 - 0.4 = 0.571$

Gain (student) = $0.971 - 0 = 0.971$

Gain (cred. rating) = $0.971 - 0.951 = 0.020$

Decision Tree: