SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Machine Learning HW5

Date No.

1 Using DFS, we know that there is 2 paths

① $A \rightarrow E^{\checkmark} \rightarrow H^{\times} \leftarrow F$

② $A \rightarrow E^{\checkmark} \rightarrow G^{\checkmark} \rightarrow I^{\times} \leftarrow H^{\checkmark} \leftarrow F$

Both paths are closed

① $H^*$ is not given, explaining away

② $I^*$ is not given, explaining away

Hence A & F are <u>independent</u>.

However, Given C & I, path ② becomes

② $A \rightarrow E^{\checkmark} \rightarrow G^{\checkmark} \rightarrow I^{\checkmark} \leftarrow H^{\checkmark} \leftarrow F$

~~Hen~~ With the new info for node I, path ② is now open. Hence, A & F is now <u>dependent</u>.

2. As all nodes takes $\{1, 2\}$, we may treat them as boolean.

Join probability $P(A \dots I)$ =

| | | |
|---|---|---|
| $P(A) \cdot P(B) \cdot P(C) \cdot P(E \mid A, B, C)$ | : | $1+1+1+8$ |
| $\cdot\, P(D) \cdot P(G \mid E, D)$ | : | $1 + 4$ |
| $\cdot\, P(F) \cdot P(H \mid E, F)$ | : | $1 + 4$ |
| $\cdot\, P(I \mid G, H)$ | : | $4$ + |
| | | 25 parameters ⁂ |

When D & F takes 5 values,

| | | |
|---|---|---|
| $P(A) \cdot P(B) \cdot P(C) \cdot P(E \mid A, B, C)$ | : | $1+1+1+8$ |
| $\cdot\, P(D) \cdot P(G \mid E, D)$ | : | $4 + 10$ |
| $\cdot\, P(F) \cdot P(H \mid E, F)$ | : | $4 + 10$ |
| $\cdot\, P(I \mid G, H)$ | : | $4$ + |
| | | 43 parameters ⁂ |

Date No.

3 $P(E=2 \mid C=1)$

$P(E \mid A,B)$ given $C=1$ :

| A | B | E | |
|---|---|---|---|
| 1 | 1 | 0.1 | 0.9 |
| 1 | 2 | 0.5 | 0.5 |
| 2 | 1 | 0.9 | 0.1 |
| 2 | 2 | 0.5 | 0.5 |

A

| | |
|---|---|
| 0.2 | 0.8 |

B

| | |
|---|---|
| 0.7 | 0.3 |

$\Downarrow$

| A = a | B = b | E = 2 | | |
|---|---|---|---|---|
| 0.2 | 0.7 | 0.9 | = | 0.126 |
| 0.2 | 0.3 | 0.5 | = | 0.03 |
| 0.8 | 0.7 | 0.1 | = | 0.056 |
| 0.8 | 0.3 | 0.5 | = | 0.12 |
| | | | | 0.332 # |

$\therefore P(E=2 \mid C=1) = 0.332$

4 $P(A=1) = \frac{C(A=1)}{12} = \frac{7}{12} = 0.583$

| A | |
|---|---|
| 1 | 2 |
| 0.583 | 0.416 |

#

| | E | F | H 1 | 2 |
|---|---|---|---|---|
| ① | 1 | 1 | 1 | 0 |
| ② | 1 | 2 | 0.66 | 0.33 |
| ③ | 2 | 1 | 0.66 | 0.33 |
| ④ | 2 | 2 | 0.5 | 0.5 |

#

obs

| | E | F | H | |
|---|---|---|---|---|
| ③ | 2 | 1 | 1 | |
| ③ | 2 | 1 | 1 | |
| ④ | 2 | 2 | 2 | |
| ③ | 2 | 1 | 2 | |
| | 1 | 1 | 1 | ① |
| | 1 | 2 | 1 | ② |
| | 1 | 2 | 2 | ② |
| ③ | 2 | 1 | 2 | |
| ④ | 2 | 2 | 1 | |
| ③ | 2 | 1 | 1 | |
| ③ | 2 | 1 | 1 | |
| | 1 | 2 | 1 | ② |

5 For the original structure, the BIC penalty term is

$$-\frac{\dim(G)}{2}\log(m) = -\frac{25}{2}\log(12)$$
$$= -13.4897$$

whereas the new structure:

$$P(A)\cdot P(B)\cdot P(C)\cdot P(E\mid A,B,C) : 1+1+1+8$$
$$P(D)\cdot P(G\mid E,D) : 1+4$$
$$P(F)\cdot P(H\mid E,F) : 1+4$$
$$P(I\mid G) : 2 \quad +$$

23 parameters

$$\therefore \text{ BIC penalty} : -\frac{23}{2}\log 12 = -12.4108$$

For the log-likelihood, the only difference is parameter for I

| original | new |
|---|---|
| $P(I\mid G,H)$ | $P(I, G)$ |
| $\sum_{t=1}^{m}\log\theta_I(x_I^{(t)}\mid x_{pa_I}^{(t)})$ | $\sum_{t=1}^{m}\log\theta_I(x_I^{(t)}\mid x_{pa_I}^{(t)})$ |
| $:\ \sum_{t=1}^{M}\log\theta_I(x_I^{(t)}\mid x_G^{(t)}, x_H^{(t)})$ | $\sum_{t=1}^{M}\log\theta_I(x_I^{(t)}\mid x_G^{(t)})$ |

| G | H | I = 1 | I = 2 |
|---|---|---|---|
| 1 | 1 | 0.5 | 0.5 |
| 1 | 2 | 0.5 | 0.5 |
| 2 | 1 | 0.5 | 0.5 |
| 2 | 2 | 0.5 | 0.5 |

| G | I = 1 | I = 2 |
|---|---|---|
| 1 | 0.5 | 0.5 |
| 2 | 0.5 | 0.5 |

Since I ~~takes~~ 0.5 regardless of G & H,

$$\sum_{t=1}^{m}\log\theta_I(x_I^{(t)}\mid x_G^{(t)}, x_H^{(t)}) = \sum_{t=1}^{m}\log\theta_I(x_I^{(t)}\mid x_G^{(t)})$$

log-likelihood is the same for both models.

Since both have the same log-likelihood, but the new model has less penalty, the new model has higher BIC score and hence the new model is the better model.