Exo apprentissage

Exercice 1 – Limites de la gourmandise

Soit un réseau bayésien G sur n variables aléatoires **binaires**.

Q 1.1 (1 point) On note i une variable aléatoire, r_i le nombre de ses modalités, p_i le nombre de ses parents, j une configuration possible de ses p_i nœuds parents.
Montrer que q_i le nombre de configurations des nœuds parents d'un nœud de G est forcément une puissance de 2. Quelle forme prend $\mathrm{Dim}(G)$?

Q 1.2 (2 points) Soit D une base de données complète sur ces n variables aléatoires **binaires**. On note N le nombre de cas total de la base, N_{ij} le nombre de cas où les parents de i ont la configuration j et N_{ijk} le nombre de cas où les parents de i ont la configuration j et i a la valeur k.
Donner la formule générale du score BIC en fonction des paramètres N, N_{ij}, N_{ijk}, q_i et r_i.

$$\mathrm{Score}_{\mathrm{BIC}}(G, D) = \log_2 L(\Theta^{MV}, G : D) - \frac{1}{2} \cdot \mathrm{Dim}(G) \cdot \log_2 N$$

On sait également que

$$\mathrm{Dim}(G) = \sum_{i=1}^{n} ((r_i - 1) \cdot q_i)$$

et $\log_2 L(\Theta^{MV}, G : D) = -N \cdot H(G, D)$ avec :

$$H(G, D) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} -\frac{N_{i,j,k}}{N_{i,j}} \cdot \log_2 \left(\frac{N_{i,j,k}}{N_{i,j}}\right)$$

Ce qui donne au final :

$$\mathrm{Score}_{\mathrm{BIC}}(G, D) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} \frac{N \cdot N_{i,j,k}}{N_{i,j}} \cdot \log_2 \left(\frac{N_{i,j,k}}{N_{i,j}}\right) - \frac{1}{2} \cdot \sum_{i=1}^{n} ((r_i - 1) \cdot q_i) \cdot \log_2 N$$

Dans un second temps, simplifier au maximum cette formule pour le cas de ce réseau bayésien et de ces n variables **binaires**.

Ici, $n = 3$, $r_i = 2$, $q_i = 2^{p_i}$, $N = 4$, enfin les variables valent 0 ou 1, on change donc l'indice k. Ce qui donne :

$$\mathrm{Score}_{\mathrm{BIC}}(G, D) = 4 \cdot \sum_{i=1}^{3} \sum_{j=1}^{2^{p_i}} \sum_{k=0}^{1} \frac{N_{i,j,k}}{N_{i,j}} \cdot \log_2 \left(\frac{N_{i,j,k}}{N_{i,j}}\right) - \sum_{i=1}^{3} 2^{p_i}$$

Q 1.3 (6 points) Soit la base de donnée suivante :

A	B	C
0	0	0
0	1	1
1	0	1
1	1	0

Q 1.3.1 (2 points) Calculer le score BIC du graphe G^0 sans aucun arc.

Dans G^0, puisqu'aucune variable n'a de parent, $\forall i, p_i = 0$ et j varie de 1 à 1. $\forall i, j, k$, $N_{i,j,k}$ correspond au nombre de fois où la variable i prend la valeur k et $N_{i,j} = N$.
Dans la base, on trouve alors $\forall i$, $N_{i,j,0} = N_{i,j,0} = 2$ et $N_{i,j} = 4$.
D'où :

$$\begin{aligned}\mathrm{Score}_{\mathrm{BIC}}(G^0, D) &= 4 \cdot \sum_{i=1}^{3} \sum_{j=1}^{1} \sum_{k=0}^{1} \frac{2}{4} \cdot \log_2 \left(\frac{2}{4}\right) - \sum_{i=1}^{3} 1 \\ &= 4 \cdot 3 \cdot 2 \cdot \frac{1}{2} \cdot (-\log_2 2) - 3 \\ &= -12 - 3 = -15\end{aligned}$$

Q 1.3.2 (2 points) Calculer le score BIC du graphe G' où il n'existe qu'un arc. Le choix de cet arc est-il important ?

Dans G', il existe un arc. Supposons que cet arc soit $A \to B$. Alors :

- $\forall i, r_i = 2$
- $p_A = p_C = 0$ et $p_B = 1$
- pour A : $\left.\begin{array}{l} N_{A,1,0} = 2 \\ N_{A,1,1} = 2 \end{array}\right\} N_{A,1} = 4$
- pour C : $\left.\begin{array}{l} N_{C,1,0} = 2 \\ N_{C,1,1} = 2 \end{array}\right\} N_{C,1} = 4$
- pour B : $\left.\begin{array}{l} N_{B,0,0} = 1 \\ N_{B,0,1} = 1 \end{array}\right\} N_{B,0} = 2$ et $\left.\begin{array}{l} N_{B,1,0} = 1 \\ N_{B,1,1} = 1 \end{array}\right\} N_{B,1} = 2$

On remarque que, par symétrie de la base D, on aurait globalement les même valeurs quelque soit l'arc de G'.
D'où enfin :

$$\begin{aligned}\mathrm{Score}_{\mathrm{BIC}}(G', D) = 4 \cdot & \Bigg(\overbrace{\frac{2}{4} \cdot \log_2 \left(\frac{2}{4}\right) + \frac{2}{4} \cdot \log_2 \left(\frac{2}{4}\right)}^{A,1} + \overbrace{\frac{2}{4} \cdot \log_2 \left(\frac{2}{4}\right) + \frac{2}{4} \cdot \log_2 \left(\frac{2}{4}\right)}^{C,1} \\ & + \overbrace{\frac{1}{2} \cdot \log_2 \left(\frac{1}{2}\right) + \frac{1}{2} \cdot \log_2 \left(\frac{1}{2}\right)}^{B,0} + \overbrace{\frac{1}{2} \cdot \log_2 \left(\frac{1}{2}\right) + \frac{1}{2} \cdot \log_2 \left(\frac{1}{2}\right)}^{B,1} \Bigg) \\ & - (1 + 1 + 2) \\ = & -16 - 4 = -20\end{aligned}$$

Q 1.3.3 (2 points) Calculer le score BIC du graphe G^* : $A \to B \leftarrow C$.

Dans G^*, on a :

- $\forall i, r_i = 2$
- $p_A = p_C = 0$ et $p_B = 2$
- pour A : $\left.\begin{array}{l} N_{A,1,0} = 2 \\ N_{A,1,1} = 2 \end{array}\right\} N_{A,1} = 4$
- pour C : $\left.\begin{array}{l} N_{C,1,0} = 2 \\ N_{C,1,1} = 2 \end{array}\right\} N_{C,1} = 4$
- pour B :

$\left.\begin{array}{l} N_{B,00,0} = 1 \\ N_{B,00,1} = 0 \end{array}\right\} N_{B,00} = 1$ et $\left.\begin{array}{l} N_{B,01,0} = 0 \\ N_{B,01,1} = 1 \end{array}\right\} N_{B,01} = 1$

$\left.\begin{array}{l} N_{B,11,0} = 1 \\ N_{B,11,1} = 0 \end{array}\right\} N_{B,11} = 1$ et $\left.\begin{array}{l} N_{B,10,0} = 0 \\ N_{B,10,1} = 1 \end{array}\right\} N_{B,10} = 1$

D'où enfin :

$$\begin{aligned} \mathrm{Score}_{BIC}(G^*, D) &= 4 \cdot \Bigg(\overbrace{\frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right) + \frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right)}^{A,1} + \overbrace{\frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right) + \frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right)}^{C,1} \\ &\quad + \overbrace{\frac{0}{1} \cdot \log_2\left(\frac{0}{1}\right) + \frac{1}{1} \cdot \log_2\left(\frac{1}{1}\right)}^{B,10} + \overbrace{\frac{0}{1} \cdot \log_2\left(\frac{0}{1}\right) + \frac{1}{1} \cdot \log_2\left(\frac{1}{1}\right)}^{B,01} \\ &\quad + \overbrace{\frac{1}{1} \cdot \log_2\left(\frac{1}{1}\right) + \frac{0}{1} \cdot \log_2\left(\frac{0}{1}\right)}^{B,00} + \overbrace{\frac{1}{1} \cdot \log_2\left(\frac{1}{1}\right) + \frac{0}{1} \cdot \log_2\left(\frac{1}{1}\right)}^{B,11} \Bigg) \\ &\quad - (1 + 1 + 4) \\ &= -8 - 6 = -14 \end{aligned}$$

Q 1.4 (1 point) On propose, comme algorithme d'apprentissage de la strucutre, un algorithme gourmand qui part de G^0 et qui, utilisant le score BIC, améliore sa structure incrémentalement par ajout d'arcs successifs. Que pouvez-vous dire de cet algorithme dans le cas de notre réseau bayésien et de notre base de données ?
Qu'en concluez-vous ?

Tout séquence de graphes trouvés par cet algorithme, issue de G^0 doit passer par un graph G' pour atteindre le graphe G^*. Or le score BIC passerait de -15 à -20 avant de remonter à -14. C'est impossible dans le cadre d'un algorithme gourmand dont la seule opération est l'ajout d'arc.
Un tel algorithme gourmand pour un tel graphe et une telle base de données n'est donc pas optimal : il est piégé dans un maximum local.
Issu de [1]

Références

[1] Milan Studenỳ, *Probabilistic Conditional Independence Structures*, 2005, Springer, Information Science & Statistics,