01.112 Machine Learning
Homework 5

Joel Huang, 1002530

December 13, 2018

1 Bayesian Network Parameters

1.1 Free parameter count

The number of free parameters is given by:

$$\mathcal{P} = \sum_{i=1}^{d} (r_i - 1) \prod_{j \in Pa_i} r_j \tag{1}$$

Assuming variables take values from $\{1, 2, 3\}$:

Parent set	$r_i - 1$	$\prod_j r_j$	params
$Pa_1 = \{\}$	2	3^0	2
$Pa_2 = \{x_1\}$	2	3^1	6
$Pa_3 = \{x_2\}$	2	3^1	6
$Pa_4 = \{x_3\}$	2	3^1	6
$Pa_5 = \{x_4\}$	2	3^1	6
$Pa_6 = \{\}$	2	3^0	2
$Pa_7 = \{x_5\}$	2	3^1	6
$Pa_8 = \{x_5\}$	2	3^1	6
$Pa_9 = \{x_6, x_7, x_8\}$	2	3^3	54
$Pa_10 = \{x_9\}$	2	3^1	6
$Pa_11 = \{x_{10}\}$	2	3^1	6

Total number of free params:

$$\mathcal{P} = 2(2) + 8(6) + 54 = 106 \tag{2}$$

Assuming variables take values from $\{1, 2, 3, 4\}$:

Parent set	$r_i - 1$	$\prod_j r_j$	params
$Pa_1 = \{\}$	3	4^0	3
$Pa_2 = \{x_1\}$	3	4^1	12
$Pa_3 = \{x_2\}$	3	4^1	12
$Pa_4 = \{x_3\}$	3	4^1	12
$Pa_5 = \{x_4\}$	3	4^1	12
$Pa_6 = \{\}$	3	4^0	3
$Pa_7 = \{x_5\}$	3	4^1	12
$Pa_8 = \{x_5\}$	3	4^1	12
$Pa_9 = \{x_6, x_7, x_8\}$	3	4^3	192
$Pa_10 = \{x_9\}$	3	4^1	12
$Pa_11 = \{x_{10}\}$	3	4^1	12

Total number of free params:

$$\mathcal{P} = 2(3) + 8(12) + 192 = 294 \tag{3}$$

1.2 Markov blanket

Markov blanket consists of parents, spouses (parents of children) and children.

$$MB_{x_1} = \{x_2\} \tag{4}$$

$$MB_{x_2} = \{x_5, x_6, x_8, x_9\} \tag{5}$$

1.3 Bayes' ball algorithm

First case: X_1 and X_6 are independent. While X_6 can be reached through X_7 or X_8 through open gates, X_6 cannot be reached without passing through a closed gate.

Second case: X_1 and X_6 are dependent. X_6 can be reached through X_8 through open gates, and X_6 can be reached through X_9 passing through open gates back from X_{10} (Since X_{10} is now given), back to X_9, then to X_6.

1.4 Conditional probability without independence

$$\begin{aligned} P(X_3 = 2 \,|\, X_4 = 1) &= \frac{P(X_3 = 2, X_4 = 1)}{P(X_4 = 1)} \\ &= \frac{0.7 \times 0.5}{0.3 \times 0.1 + 0.7 \times 0.5} \\ &= 0.921 \end{aligned} \tag{6}$$

1.5 Conditional probability with independence

Assuming independence for X_2 and X_11 as from the probability tables, we can simplify the expression:

$$\begin{aligned} &P(X_5 = 2 \,|\, X_2 = 1, X_3 = 1, X_{11} = 2) \\ &= \frac{P(X_3 = 1, X_5 = 2)}{P(X_3 = 1)} \\ &= \frac{0.3 \times (0.1 \times 0.5 + 0.9 \times 0.4)}{0.3} \\ &= 0.41 \end{aligned} \tag{7}$$

2 Comparing graphs using the Bayesian Information Criterion

$$BIC = l(D; \hat{\theta}, G) - \frac{log(m)}{2} dim(G) \tag{8}$$

No such case exists. We need to prove $BIC_{G_1} > BIC_{G_2}$. Since the log likelihood of G_1 can be proven to be equal to G_2, and the number of free parameters must be equal, then there is no sample set that can result in the BIC score of G_1 exceeding G_2 strictly. The likelihood proof is as follows:

$$\begin{aligned} l(D; \hat{\theta}, G_1) = P(X_1 = x_1)\cdot \\ P(X_2 = x_2 \mid X_1 = x_1) \cdots \\ P(X_8 = x_8 \mid X_7 = x_7) \\ = P(X_1 = x_1)\cdot \\ \frac{P(X_2 = x_2, X_1 = x_1)}{P(X_1 = x_1)} \cdots \\ \frac{P(X_8 = x_8, X_7 = x_7)}{P(X_7 = x_7)} \end{aligned} \tag{9}$$

$$\begin{aligned} l(D; \hat{\theta}, G_2) = P(X_8 = x_8)\cdot \\ P(X_7 = x_7 \mid X_8 = x_8) \cdots \\ P(X_1 = x_1 \mid X_2 = x_2) \\ = P(X_8 = x_8)\cdot \\ \frac{P(X_7 = x_7, X_8 = x_8)}{P(X_8 = x_8)} \cdots \\ \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)} \end{aligned} \tag{10}$$

We notice that $P(X_1 = x_1)$ appears both in the numerator and denominator for G_1, and $P(X_8 = x_8)$ for G_2. We cancel both and find that $l(G_1)$ is indeed equal to $l(G_2)$. Therefore, $BIC_{G_1} = BIC_{G_2}$ for all sample sets.