# ADVANCED MACHINE LEARNING
# EXERCISE #1 SOLUTION

Uri Avron [uriavron@gmail.com] [308046994]
Jonathan Somer [jonathan.somer@gmail.com] [307923383]
Matan Harel [matan.harel.mh@gmail.com] [302695721]

March 22, 2018

# Question 1

**(a)** Due to simmetry, for any pair $i, j$ it holds that $q(x_i, x_j) = \frac{1}{6} + \frac{1}{12} = \frac{1}{4}$ and for any $k \in [3]$ $q(x_k) = \frac{1}{2}$. Therefore $q(x_i, x_j) = q(x_i)q(x_j)$ which implies $X_i \perp X_j \in I(q)$. Moreover, it is now clear that for any permutation $(i, j, k)$ of $[3]$ it holds that

$$\begin{aligned} q(x_i|x_k)q(x_j|x_k) &= \frac{1}{4} \\ &\neq \begin{cases} \frac{1}{6} & x_1 \oplus x_2 \oplus x_3 = 0 \\ \frac{1}{3} & x_1 \oplus x_2 \oplus x_3 = 1 \end{cases} = 2q(x_1, x_2, x_3) = q(x_i, x_j|x_k) \end{aligned}$$

which implies $X_i \perp X_j | X_k \notin I(q)$, and

$$\begin{aligned} q(x_i) &= \frac{1}{2} \\ &\neq \begin{cases} \frac{1}{3} & x_1 \oplus x_2 \oplus x_3 = 0 \\ \frac{2}{3} & x_1 \oplus x_2 \oplus x_3 = 1 \end{cases} = 4q(x_1, x_2, x_3) = q(x_i|x_j, x_k) \end{aligned}$$

which implies $X_i \perp X_j, X_k \notin I(q)$. so overall we get $I(q) = \{X_1 \perp X_2, X_2 \perp X_3, X_3 \perp X_1\}$.

**(b)** No, proof by contradiction. Let us assume there exists a DAG $G$ where $I_{LM}(G) = I(q)$, so $X_i \perp X_j \in I_{LM}(G)$ which means $Pa(i) = \emptyset$ and $ND(i) = \{X_j\}$ but this is true for any $i, j \in \{(1, 2), (2, 3), (3, 1)\}$ and hence the contradiction is obvious (there are no parents but every node has one decendant for example). $\square$

**(c)** No, proof by contradiction. Let us assume there exists an undirected graph $G$ where $I_{sep}(G) = I(q)$, so $X_i \perp X_j \in I_{sep}(G)$ which means there are no edges in $G$ but in this case for example $X_1 \perp X_2, X_3 \in I_{sep}(G)$ where $X_1 \perp X_2, X_3 \notin I_{sep}(G)$ and hence $I_{sep}(G) \neq I(q)$, a contradiction. $\square$

## Question 2

We are given a positive distribution $p(w, x, y, z)$. It is known that:

1. We are given - $(X \perp Y|Z, W)$, equivalently: $p(x|y, z, w) = p(x|z, w)$
2. We are given - $(X \perp W|Z, Y)$, equivalently: $p(x|w, z, y) = p(x|z, y)$

From 1,2 we have: $p(x|z, w) = p(x|z, y)$ (*)

**We will prove that** $(X \perp Y, W|Z)$**, that is:** $p(x|y, w, z) = p(x|z)$**:**

From 1, we have $p(x|y, z, w) = p(x|z, w)$. Multiplying both sides by $p(y|z, w)$ gives:

$$p(y|z, w)p(x|y, z, w) = p(y|z, w)p(x|z, w)$$

Equivalently:

$$p(x, y|w, z) = p(y|w, z)p(x|w, z)$$

Multiplying both sides by $p(w|z)$ gives:

$$p(w|z)p(x, y|w, z) = p(w|z)p(y|w, z)p(x|w, z)$$

Summing over $y$ gives:

$$p(x, w|z) = p(w|z)p(x|w, z)$$

From (*):

$$p(x, w|z) = p(w|z)p(x|y, z)$$

Summing over $w$ gives:

$$p(x|z) = p(x|y, z)$$

From 2:

$$p(x|z) = p(x|w, z, y) \Rightarrow X \perp Y, W|Z$$

□

## Question 3

We define $S = Pa(i) \cup Ch(i) \cup \bigcup_{j \in Ch(i)} Pa(j)$ and argue that $S$ is the Markov Blanket for some Bayesian network $p$ on $G$. We first show that $X_i \perp X_{\bar{S}\setminus i} | X_S \in I_{d-sep}(G)$. Let $k \in X_{\bar{S}\setminus i}$. We have 3 cases:

- If there is an undirected path between $i$ and $j$ via $k \in Ch(i)$ and $l \in Pa(j)$, then $l \in S$ is not the child in this v-stracture.
- If there is an undirected path between $i$ and $j$ via $k \in Ch(i)$ but not via $Pa(j)$, then $k \in S$ and $k$ is not in a v-stracture in the path.
- If there is an undirected path between $i$ and $j$ via $k \in Pa(i)$, then$k \in S$ is not in a v-stracture.

Overall, there is no active trail between $X_i$ and $X_{\bar{S}\setminus i}$ given $X_S$ and we get $X_i \perp X_{\bar{S}\setminus i} | X_S \in I_{d-sep}(G) \subseteq I(p)$ by definition and Proposition 2.3.1. We now have left to show the minimalism of $S$. If we remove $k \in Pa(i) \cup Ch(i)$, it is clear that $X_i$and $X_k$are dependant given $X_S$ and hence $X_i \perp X_{\bar{S}\setminus i} | X_S \notin I(p)$. If we remove $k \in \bigcup_{j \in Ch(i)} Pa(j)$, we have an active trail $i$ and $k$ given $X_S$ (via the mutual child) and hence $X_i \perp X_{\bar{S}\setminus i} | X_S \notin I_{d-sep}(G)$. We know that except for measure zero set, all Baseyian networks on $G$ will hold $I(p) = I_{d-sep}(G)$, and therefore in any case of removal we get$X_i \perp X_{\bar{S}\setminus i} | X_S \notin I(p)$ so $S$ is indeed minimal. $\square$

## Question 4

**(a)** By Theorem 4, it's suffices to show that $I_{pair}(G) \subseteq I(p)$. Indeed, if $X_i \perp X_j | X_{V\setminus\{i,j\}} \in I_{pair}(G)$ then by definition of $G$ we get $ij \notin E$ and hence $X_i \perp X_j | X_{V\setminus\{i,j\}} \in I(p)$ since otherwise edge $ij$ was supposed to be added to $G$. Therefore $I_{sep}(G) \subseteq I(p)$. $\square$

**(b)** Towards contradiction, let us assume that $G$ is not a minimal I-map for $p$. Therefore, there exists edge $ij$ such that it can be removed and still $I_{sep}(G) \subseteq I(p)$. On one hand, be the definition of $G$, we have $X_i \perp X_j | X_{V\setminus\{i,j\}} \notin I(p)$. On the other hand, since now $ij \notin E$, we get $X_i \perp X_j | X_{V\setminus\{i,j\}} \in I_{pair}(G)$, and hence $X_i \perp X_j | X_{V\setminus\{i,j\}} \in I_{sep}(G)$ so therefore $I_{sep}(G) \nsubseteq I(p)$, a contradiction. We thus conclude $G$ is indeed minimal I-map for $p$. $\square$

# Question 5

Obviously, $I_{sep}(G) = \{X_1 \perp X_3 | X_{2,4},\, X_2 \perp X_4 | X_{1,3}\}$. First, we show that $X_1 \perp X_3 | X_{2,4} \in I(p)$. Becuase $p(x_3 = 0) = 1/2$, the following tables show $p(x_1 | x_{2,3,4}) = p(x_1 | x_{2,4})$ and hence $X_1 \perp X_3 | X_{2,4} \in I(p)$:

Each cell is the probability of $p(x_1|x_{2,3,4})$:

| $x_1 \| x_{2,3,4}$ | 0,0,0 | 1,0,0 | 0,1,0 | 1,1,0 | 0,0,1 | 1,0,1 | 0,1,1 | 1,1,1 |
|---|---|---|---|---|---|---|---|---|
| 0 | 1/2 | 0 | undefined | 0 | 1 | unedfined | 1 | 1/2 |
| 1 | 1/2 | 1 | undefined | 1 | 0 | undefined | 0 | 1/2 |

Each cell is the probability of $p(x_1|x_{2,4})$:

| $x_1 \| x_{2,4}$ | 0,0 | 1,0 | 0,1 | 1,1 |
|---|---|---|---|---|
| 0 | 1/2 | 0 | 1 | 1/2 |
| 1 | 1/2 | 1 | 0 | 1/2 |

$X_2 \perp X_4 | X_{1,3} \in I(p)$ is showed in the same way. Therefore we have $I_{sep}(G) \subseteq I(p)$. To end our proof, we now show that $p$ is not a Markov network with respect to $G$. Towards a contradiction, let us assume that $p$ is indeed a Markov network with respect to $G$. We know that $\mathcal{C}(G) = \{\{X_1, X_4\}, \{X_1, X_2\}, \{X_3, X_4\}, \{X_2, X_3\}\}$. Hence, by the definition of Markov network, $p(\mathbf{x})$ can be written as:

$$p(\mathbf{x}) = Z^{-1} \prod_{c \in \mathcal{C}(G)} \phi_c(x_c) = Z^{-1} \phi_{12}(x_{1,2}) \phi_{23}(x_{2,3}) \phi_{34}(x_{3,4}) \phi_{41}(x_{4,1})$$

Let us inspect the following equations:

$$p(0,0,0,0) = \phi_{12}(0,0)\phi_{23}(0,0)\phi_{34}(0,0)\phi_{41}(0,0) = Z/8 \tag{1}$$
$$p(0,0,1,1) = \phi_{12}(0,0)\phi_{23}(0,1)\phi_{34}(1,1)\phi_{41}(1,0) = Z/8 \tag{2}$$
$$p(1,1,1,0) = \phi_{12}(1,1)\phi_{23}(1,1)\phi_{34}(1,0)\phi_{41}(0,1) = Z/8 \tag{3}$$
$$p(0,0,1,0) = \phi_{12}(0,0)\phi_{23}(0,1)\phi_{34}(1,0)\phi_{41}(0,0) = 0 \tag{4}$$

From (4), we know that $\phi_{12}(0,0) = 0 \vee \phi_{23}(0,1) = 0 \vee \phi_{34}(1,0) = 0 \vee \phi_{41}(0,0) = 0$, but if any one of those is true, at least one out of (1), (2), (3) will be equal to $0 \neq Z/8$, a contradiction. We therefore conclude that $p$ is not a Markov network with respect to $G$. $\square$

## Question 6

First, w.l.o.g we assume that $\forall ij \in E$, $i < j$. Proof by induction on the size of $V$. For $|V| = 1$ the claim holds trivially. Let us assume the claim holds $\forall k \in [n-1]$ and we now prove for $n$. We know that:

$$p(x_{[n]}) = p(x_n | x_{[n-1]}) p(x_{[n-1]}) \tag{5}$$

From the inductive hypothesis:

$$p(x_{[n-1]}) = \prod_{i=1}^{n-1} p(x_i) \prod_{\substack{ij \in E \\ i,j \in [n-1]}} \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \tag{6}$$

By Theorem 4.1 (if $p$ factorizes according to G the $I_{sep}(G) \subseteq I(p)$) and the fact that $I_{LM}(G) \subseteq I_{sep}(G)$:

$$p(x_n | x_{[n-1]}) = p(x_n | x_{Nbr(n)}) = \frac{p(x_n, x_{Nbr(n)})}{p(x_{Nbr(n)})}$$

$G$ is a tree a graph, hence there are no circles and therefore, w.l.o.g $Nbr(n) < n - 1$. So, again by the inductive hypothesis:

$$p(x_n, x_{Nbr(n)}) = \prod_{i \in Nbr(n) \cup \{n\}} p(x_i) \prod_{\substack{ij \in E \\ i,j \in Nbr(n) \cup \{n\}}} \frac{p(x_i, x_j)}{p(x_i)p(x_j)}$$

$$p(x_{Nbr(n)}) = \prod_{i \in Nbr(n)} p(x_i) \prod_{\substack{ij \in E \\ i,j \in Nbr(n)}} \frac{p(x_i, x_j)}{p(x_i)p(x_j)}$$

$$\Rightarrow p(x_n | x_{[n-1]}) = p(x_n) \prod_{in \in E} \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \tag{7}$$

By (5), (6), (7) we get:

$$p(x_n) = \prod_{i=1}^{n} p(x_i) \prod_{ij \in E} \frac{p(x_i, x_j)}{p(x_i)p(x_j)}$$

$\square$