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 23, 2018

Question 1

(a) For a set of 3 variables we must look at 3 kinds of conditional independencies, w.l.o.g (all other permutations are symmetrical):

1. $X_1 \perp X_2$
2. $X_1 \perp X_2 | X_3$
3. $X_1 \perp X_2, X_3$

Type 1 CI - hold

Starting with the first kind of CI, for any values for x_1, x_2 it holds that $q(x_1, x_2) = \frac{1}{6} + \frac{1}{12} = \frac{1}{4}$ and for any $k \in [3]$, and any value for x_k $q(x_k) = \frac{1}{6} + \frac{1}{12} + \frac{1}{6} + \frac{1}{12} = \frac{1}{2}$. Therefore $q(x_1, x_2) = q(x_1)q(x_2)$ which implies that for all i, j: $X_i \perp X_j \in I(q)$.

Type 2 CI - don't hold

We shall show an example for which this CI doesn't hold:
We will show $p(x_1|x_3)p(x_2|x_3) \neq p(x_1, x_2|x_3)$ for the case: $x_1, x_2, x_3 = 0$

$$p(x_1, x_2|x_3) = \frac{p(x_1 = 0, x_2 = 0, x_3 = 0)}{p(x_3 = 0)} = \frac{\frac{1}{12}}{\frac{1}{2}} = \frac{1}{6}$$

$$p(x_1|x_3) = p(x_2|x_3) = \frac{\frac{1}{12} + \frac{1}{6}}{\frac{1}{2}} = \frac{1}{2}$$

$$\frac{1}{4} = p(x_1|x_3)p(x_2|x_3) \neq p(x_1, x_2|x_3) = \frac{1}{6}$$

Type 3 CI - don't hold

We shall show an example for which this CI doesn't hold:
We will show $p(x_1, x_2, x_3) \neq p(x_1)p(x_2, x_3)$ for the case: $x_1, x_2, x_3 = 0$

$$p(x_1 = 0, x_2 = 0, x_3 = 0) = \frac{1}{12}$$

$$p(x_1 = 0) = \frac{1}{2}$$

$$p(x_2 = 0, x_3 = 0) = \frac{1}{12} + \frac{1}{6} = \frac{1}{4}$$

$$\frac{1}{12} = p(x_1 = 0, x_2 = 0, x_3 = 0) \neq p(x_1 = 0)p(x_2 = 0, x_3 = 0) = \frac{1}{8}$$

Overall: $I(q) = \{(X_1 \perp X_2), (X_2 \perp X_3), (X_3 \perp X_1)\}$.

(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 because for all i, j: $X_i \perp X_j \in I_{sep}(G)$ there are no edges in G. because there are no routes from X_1 to X_2, X_3 at all it holds that $X_1 \perp X_2, X_3 \in I_{sep}(G)$. But we know that $X_1 \perp X_2, X_3 \notin I(q)$ a contradiction to $I_{sep}(G) = I(q)$. $\square$

Old Question 1

For the second type of CI: it is 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 $X_i \perp X_j|X_k \notin I(q)$. For the Third kind of CI:

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

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:

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

Equivalently:

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

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 $X_i \perp X_{\bar{S}\setminus i} | X_S \in I_{d-sep}(G)$.

More specifically we will show that: X_i is d-separated from $X_{\bar{S}\setminus i}$ given X_S by showing that there is no active trail from X_i to some node $j \in X_{\bar{S}\setminus i}$ given X_S (definition 6).

There are 3 possible cases for an undirected path between i and j:

- An undirected path between i and j via $k \in Ch(i)$ and some $l \in Pa(k)$. Then because $l \in S$ and l is not the descendent of the child in a v-structure the path is not active.
- An undirected path between i and j via $k \in Ch(i)$ but not via some $l \in Pa(k)$, then $k \in S$ and k is not the child in a v-structure in the path or a descendent of one so the path is not active.
- An undirected path between i and j via $k \in Pa(i)$. In this case k which is in S cannot be the child in a v-structure or a descendent of one so the path is not active.

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 6 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: as p is positive it 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 $p(x_1 | x_{2,3,4})$:

$x_1 \vert x_{2,3,4}$	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 $p(x_1 | x_{2,4})$:

$x_1 \vert x_{2,4}$	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

$G = (E, V)$ is a tree graph. $p(x)$ is a markov network on G. We will show that for any assignment $x_1, .., x_n$ it holds that:

$$p(x_{[n]}) = \prod_{i=1}^{n} p(x_i) \prod_{ij \in E} \frac{p(x_i x_j)}{p(x_i)p(x_j)}$$

Note: we use the notation $x_{[n]}$ to denote: $x_1, .., x_n$

Proof:

We will prove the claim by induction over the size of V.

Base case $|V| = 1$ trivially holds:

$$p(x_1) = \prod_{i=1}^{1} p(x_i) \prod_{ij \in \emptyset} \frac{p(x_i x_j)}{p(x_i)p(x_j)}$$

The case for $|V| = 2$ is similar (and will enable us to assume $n \geq 3$ in all cases henceforth):

$$p(x_1, x_2) = \prod_{i=1}^{2} p(x_i) \prod_{ij \in (1,2)} \frac{p(x_i x_j)}{p(x_i)p(x_j)}$$

Let us assume that the claim holds for any graph where $|V| < n$, and we now prove for $|V| = n$. Let $G = (V, E)$ be some tree graph where $|V| = n$, and $p(x)$ a markov network on this graph. It holds for any distribution that:

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

In any tree graph there exists a leaf. Assume w.l.o.g that x_n is a leaf. That is x_n has only one neighbour. Assume w.l.o.g that this neihbour is x_{n-1}.
By Theorem 4.1 (if p factorizes according to G then $I_{sep}(G) \subseteq I(p)$) and the fact that $I_{LM}(G) \subseteq I_{sep}(G)$:

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

Replacing this in (5) we arrive at:

$$p(x_{[n]}) = p(x_n | x_{n-1}) p(x_{[n-1]}) \tag{6}$$

Note that the sub-graph resulting from removing x_n from G is also a tree as x_n was a leaf so the remaining graph is still connected and has no cycles.

Claim: $p(x_{[n-1]})$ is a markov network on the tree graph G' resulting from G by removing the leaf x_n. Proof:

By definition of markov network and the fact that all cliques in a tree are of size 2 (or else we will have cycles), corresponding exactly to edges:

$$p(x_{[n-1]}) = \sum_{x_n} p(x_{[n-1]}, x_n) = \sum_{x_n} \frac{1}{Z} \prod_{ij \in E} \phi_{ij}(x_i, x_j) = \frac{1}{Z} \prod_{ij \in E \setminus (n,n-1)} \phi_{ij}(x_i x_j) \sum_{x_n} \phi_{n.n-1}(x_n, x_{n-1}) \tag{7}$$

In the last equality we used the fact that all x_i where $i \neq n$ are constants within the sum. We would like to arrange the expression as a product of $\phi_{ij}(x_i, x_j)$ for $ij \in E \setminus (n, n-1)$. To acheive this we will "push" the value of $\sum_{x_n} \phi_{n.n-1}(x_n, x_{n-1})$ into one of the ϕ_{ij}. We need all functions to depend only on their 2 paramaters so we have to choose one where x_{n-1} is one. Recognize that because $n \geq 3$ and x_{n-1} is connected to the rest of the graph, there exists some variable, w.l.o.g x_{n-2}, s.t: $(x_{n-1}, x_{n-2}) \in E$. Now we define:

$$\phi'_{n-1,n-2}(x_{n-1}, x_{n-2}) = \phi_{n-1,n-2}(x_{n-1}, x_{n-2}) \sum_{x_n} \phi_{n.n-1}(x_n, x_{n-1})$$

For ease of notation, for all other pairs denote $\phi'_{ij} = \phi_{ij}$ We arrive at:

$$\begin{gathered}
\prod_{ij \in E \setminus (n,n-1)} \phi_{ij}(x_i x_j) \sum_{x_n} \phi_{n.n-1}(x_n, x_{n-1}) = \\
\prod_{ij \in E \setminus (n,n-1),(n-1,n-2)} \phi_{ij}(x_i x_j) \cdot (\phi_{n-1.n-2}(x_{n-1}, x_{n-2}) \sum_{x_n} \phi_{n.n-1}(x_n, x_{n-1})) = \\
\prod_{ij \in E \setminus (n.n-1)} \phi'_{ij}(x_i, x_j)
\end{gathered}$$

And for Z

$$\begin{gathered}
Z = \sum_{x_{[n]}} \prod_{ij \in E} \phi_{ij}(x_i, x_j) = \sum_{x_{[n-1]}} \sum_{x_n} \prod_{ij \in E} \phi_{ij}(x_i, x_j) = \\
\sum_{x_{[n-1]}} \prod_{ij \in E \setminus (x_{n-1}, x_n)} \phi_{ij}(x_i, x_j) \sum_{x_n} \phi_{n-1,n}(x_{n-1}, x_n) = \sum_{x_{[n-1]}} \prod_{ij \in E \setminus (x_{n-1}, x_n)} \phi'_{ij}(x_i, x_j)
\end{gathered}$$

Finally arriving at:

$$\begin{gathered}
p(x_{[n-1]}) = \frac{1}{Z} \prod_{ij \in E \setminus (n,n-1)} \phi_{ij}(x_i x_j) \sum_{x_n} \phi_{n.n-1}(x_n, x_{n-1}) = \\
\frac{1}{\sum_{x_{[n-1]}} \prod_{ij \in E \setminus (x_{n-1}, x_n)} \phi'_{ij}(x_i, x_j)} \prod_{ij \in E \setminus (n.n-1)} \phi'_{ij}(x_i, x_j)
\end{gathered}$$

So $p(x_{[n-1]})$ is a markov network on the resulting tree!

We continue with the induction:

By the induction hypothesis, on the Tree resulting from removing x_n from G:

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

$$p(x_n|x_{n-1}) \prod_{i\in[n-1]} p(x_i) \prod_{ij\in E\setminus(n,n-1)} \frac{p(x_i,x_j)}{p(x_i)p(x_j)} =$$

$$\frac{p(x_n,x_{n-1})}{p(x_{n-1})} \prod_{i\in[n-1]} p(x_i) \prod_{ij\in E\setminus(n,n-1)} \frac{p(x_i,x_j)}{p(x_i)p(x_j)} =$$

$$p(x_n) \prod_{i\in[n-1]} p(x_i) \prod_{ij\in E\setminus(n,n-1)} \frac{p(x_i,x_j)}{p(x_i)p(x_j)} \cdot \frac{p(x_n,x_{n-1})}{p(x_n)p(x_{n-1})} =$$

$$\prod_{i\in[n]} p(x_i) \prod_{ij\in E} \frac{p(x_i,x_j)}{p(x_i)p(x_j)}$$

□

Old 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 (as there are no edges in the graph, the markov network $p(x)$ is in the required form). Let us assume the claim holds $\forall k \in [n-1]$ and we now prove for n. Let there be G, some graph with n nodes, and $p(x)$ a markov network on this graph. We know that for any distribution:

$$p(x_{[n]}) = p(x_n|x_{[n-1]})p(x_{[n-1]}) \tag{8}$$

From the inductive hypothesis:

$$p(x_{[n-1]}) = \prod_{i=1}^{n-1} p(x_i) \prod_{\substack{ij\in E\\ i,j\in[n-1]}} \frac{p(x_i,x_j)}{p(x_i)p(x_j)} \tag{9}$$

By Theorem 4.1 (if p factorizes according to G then $I_{sep}(G) \subseteq I(p)$) 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:

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

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

$$\Rightarrow p(x_n | x_{[n-1]}) = p(x_n) \prod_{in \in E} \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \tag{10}$$

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

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

□