10-708:Probabilistic Graphical Models 10-702, Spring 2017

Homework 1

Lecturer: Eric Xing *Name: Yuan Liu(yuanl4)*

1 Bayesian Network Solution

1.1 I-map

Solution: (Koller and Friedman Textbook Page 62)
Let $X_1, ..., X_n$ is a topological ordering of the variables in $\mathbb{X}$ relative to $\mathbb{G}$. We can first use the chain rule to represent $\mathbb{P}$:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i|X_1, ..., X_{i-1})$$

Now, consider one of the factors $P(X_i|X_1, ..., X_{i-1})$. Because $\mathbb{G}$ is a I-map for P, we have

$$(X_i \perp NonDescendants_{x_i}|Pa^{\mathbb{G}}_{x_i}) \in I(\mathbb{P})$$

By the topological ordering assumption, all of X_i's parents are in the set $X_1, ..., X_{i-1}$. Furthermore, none of X_i's descendants can possibly be in the set. Hence,

$$\{X_1, ..., X_{i-1}\} = Pa_{X_i} \cup Z$$

where $Z \subset NonDescendants_{X_i}$. In addition to it, we already know $(X_i \perp NonDescendants_{x_i}|Pa^{\mathbb{G}}_{x_i}) \in I(\mathbb{P})$, so it follows that $(X_i \perp Z|Pa_{x_i})$. Hence we have that

$$P(X_i|X_1, ..., X_{i-1}) = P(X_i|Pa_{X_i})$$

Applying this transformation to all of the factors in the chain rule decomposition, the result follows.

1.2 D-separation

- $B \perp G|A$: True. Since H is unobserved, and $G \rightarrow H \leftarrow F$ is a v-structure, influence cannot flow through H.
- $C \perp D|F$: False. Influence can flow along the path $C \leftarrow B \rightarrow D$.
- $C \perp D|A$: False. Influence can flow along the path $C \leftarrow B \rightarrow D$.
- $H \perp B|C, F$: True. Since F is observed and $H \leftarrow F \leftarrow C$ or $H \leftarrow F \leftarrow E$ are neither v-structure, then H can not flow through F.

1.3 I-Equivalence

- G_1 and G_2 has the same skeleton is a necessary, but not sufficient condition for G_1 and G_2 to be I-equivalent.

Not sufficient part proof: The v-structure and common parent structure have the same skeleton, but they are not I-equivalent.

Necessary part proof: Proof by contradiction. Assume G_1 and G_2 are I-equivalent, but they have different skeleton. In this situation, we can find a trail in one network that does not exist in the other. Assume this trail is $X_1 \rightleftharpoons X_2 \rightleftharpoons ... \rightleftharpoons X_i$ in G_1. Given all the v-structures in this trail, like $A_k \rightarrow B_k \leftarrow C_k$, let all the $\{B_k\}_k$ observed, and left others unobserved. Then $X_1 \perp X_2|\{B_k\}_k$ in G_2, but $X_1 \not\perp X_2|\{B_k\}_k$ in G_1

- G_1 and G_2 has the same skeleton and same v-structures, is a sufficient, but not necessary condition for G_1 and G_2 to be I-equivalent.

Sufficient part proof: (http://www.wisdom.weizmann.ac.il/pgm/Ex/ps3_sol.pdf) First we assume that $(\mathbb{X} \perp \mathbb{Y}|\mathbb{Z}) \in I(G_1)$ and we show that $(\mathbb{X} \perp \mathbb{Y}|\mathbb{Z}) \in I(G_2)$. If two graphs have the same skeleton, then they have the same trails. Lets look on the trails that between $X \in \mathbb{X}$ and $Y \in \mathbb{Y}$ that given $\mathbb{Z}$ is inactive in G_1, and then show this trail in G_2 is in active too. Consider two cases:

i The trail in G_1 is inactive because some of the nodes on the trail that are not in v-structure are observed in $\mathbb{Z}$. Then clearly these nodes also blocks the trail in G_2

ii Otherwise, all nodes on the trail that are not in a v-structure are not observed (not in $\mathbb{Z}$), but then for some v-structure V_{i-1}, V_i, V_{i+1} on the trail, none of the of V_i are observed. That is for every node V such that there is a directed path in G_1 from V_i to V, all the nodes on the path are not observed. Consider such a directed path from such a V_i to some V in G_1, then in G_2 this trail must also be directed the same, from V_i to V, because other wise it introduces a v-structure that is not in G_1(either on this path itself, or with respect to $V_{i-1} \rightarrow V_i$), and clearly all it's nodes are not observed too. There fore this path also inactivates the trail between $\mathbb{X}$ and $\mathbb{Y}$(given $\mathbb{Z}$) in G_2

Thus, $(\mathbb{X} \perp \mathbb{Y}|\mathbb{Z}) \in I(G_1)$ implies that every trail between any pair $X \in \mathbb{X}$ and $Y \in \mathbb{Y}$ that given $\mathbb{Z}$ is inactive in G_1, and showed for each such a trail that it is inactive in G_2. Since the two graphs have the same trails, then $(\mathbb{X} \perp \mathbb{Y}|\mathbb{Z}) \in I(G_2)$. Hence $I(G_1) \subset I(G_2)$, and by symmetry we also get that $I(G_2) \subset I(G_1)$, therefore $I(G_1) = I(G_2)$

Not necessary part proof: Consider complete graphs over a set of variables. Recall that a complete graph is one to which we cannot add additional arcs without causing cycles. Such graphs encode the empty set of conditional independence assertions. Thus, any two complete graphs are I-equivalent. Although they have the same skeleton, they invariably have different v-structures.

2 Independence & Equivalence Testing Solution

2.1 D-separation and Independence

(1) First, prove that D-separation implies conditional independence, or mathematically:

$$desp(X, Y|Z) \Rightarrow X \perp Y|Z$$

Solution (Koller and Friedman Textbook Page 137) Let $U = X \cup Y \cup Z$. Let $U^* = U \cup Ancestors_U$. Let $G_{U^*} = G^+[U]$ be the induced graph over U^*, and let H be the moralized graph $M[G_{U^*}]$. Let P_{U^*} be the Bayesian network distribution defined over G_{U^*} in the obvious way: the CPD for any variable in U^* is the same as in original distribution P_B. Because U^* is upwardly closed, all variables used in these CPDs are in U^*.
Now, consider an independence assertion $(X \perp Y|Z) \in I(G)$; we want to prove that $P \models (X \perp Y|Z)$. By definition, if $(X \perp Y|Z) \in I(G)$, we have that $desp(X, Y|Z)$. It follows that $sep_H(X, Y|Z)$, and hence that $(X \perp Y|Z) \in I(H)$. P_{U^*} is a Gibbs distribution over H, and hence, $P_{U^*} \models (X \perp Y|Z)$. We know $P_{U^*}(U^*)$ is the same as $P_B(U^*)$. Hence, it follows that $P_B \models (X \perp Y|Z)$

(2) Prove the completeness of the D-separation property.

(3) Provide a pseudo-code and describe in detail an algorithm for testing D-separation.

Solution (Koller and Friedman Textbook Page 75) We use the following definition of active trail to find all the nodes that can be reached by X.

- Whenever we have a v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, then X_i or one of its descendants are in Z.
- no other node along the trail is in Z.

So we first need to find all the ancestors of Z, and put them into A.(Using broad first search)
Then we use broad-first search to find all the linked node.

2.2 I equivalence

(1) Proof: The two BNs, G and G', are I-equivalent if both graphs have the same set of trails and a trail is active in G if and only if it is active in G'.

Solution: First we assume that $(\mathbb{X} \perp \mathbb{Y}|\mathbb{Z}) \in I(G_1)$ and we show that $(\mathbb{X} \perp \mathbb{Y}|\mathbb{Z}) \in I(G_2)$.

Lets look on the trail that between $X \in \mathbb{X}$ and $Y \in \mathbb{Y}$ that given $\mathbb{Z}$ is inactive in G_1, and because of the assumption this trail in G_2 is inactive too.

Thus, $(\mathbb{X} \perp \mathbb{Y}|\mathbb{Z}) \in I(G_1)$ implies that every trail between any pair $X \in \mathbb{X}$ and $Y \in \mathbb{Y}$ that given $\mathbb{Z}$ is inactive in G_1, and showed for each such a trail that it is inactive in G_2. Since the two graphs have the same trails, then $(\mathbb{X} \perp \mathbb{Y}|\mathbb{Z}) \in I(G_2)$. Hence $I(G_1) \subset I(G_2)$, and by symmetry we also get that $I(G_2) \subset I(G_1)$, therefore $I(G_1) = I(G_2)$

(2) Prove that a minimal active trail may only contain a triangle of the following form:

Algorithm 1 D-separation

Input: Bayesian network graph: G; Source variables: X; Evidence variables: Z;
Output: Reachable Set R

//Step one: find all the ancestors of Z
$L \leftarrow Z$ //Node to be visited
$A \leftarrow \emptyset$ //Ancestors of Z
//Broad first search
while $L \neq \emptyset$ **do**
  Select some Y from L
  $L \leftarrow L - \{Y\}$
  if $Y \notin A$ **then**
    $L \leftarrow L \cup Pa_Y$
  end if
  $A \leftarrow A \cup \{Y\}$
end while
  //Step two:traverse active trails starting from X
$L \leftarrow \{(X, \uparrow)\}$ //(Node,direction) to be visited
$V \leftarrow \emptyset$ //(Node, direction) marked as visited
$R \leftarrow \emptyset$ //Nodes reachable via active trail
while / **do**/Select some (Y,d) from L
  $L \leftarrow L - (Y, d)$
  if $(Y, d) \notin V$ **then**
    if $Y \notin Z$ **then**
      $R \leftarrow R \cup \{Y\}$ // Y is reachable
    end if
    $V \leftarrow V \cup (Y, d)$ // Mark(Y,d) as visited
    if $d = \uparrow$ and $Y \notin Z$ **then** //Trail up through Y active if Y not in Z
      for each $Z \in Pa_Y$ **do** //Y's parents to be visited from bottom
        $L \leftarrow L \cup \{(Z, \uparrow)\}$
      end for
      for each $Z \in Ch_Y$ **do** //Y's children to be visited from top
        $L \leftarrow L \cup \{(Z, \downarrow)\}$
      end for
    else if $d = \downarrow$ **then**
      if $Y \notin Z$ **then** //Downward trails to Y's children are active
        for each $Z \in Ch_Y$ **do**
          $L \leftarrow L \cup \{(Z, \downarrow)\}$ //Y's children to be visited from top
        end for
      end if
      if $Y \in A$ **then** //v-structure trails are active
        for each $Z \in Pa_Y$ **do**
          $L \leftarrow L \cup \{(Z, \uparrow)\}$ // Y's parents to be visited from bottom
        end for
      end if
    end if
  end if
end while
Return R

Solution Let first consider the trail: $X_{i-1} \rightleftharpoons X_i \rightleftharpoons ... \rightleftharpoons X_{i+1}$. There are four following situations:

- $X_{i-1} \rightarrow X_i \rightarrow X_{i+1}$.Because these three variables form a triangle. There are two sub-cases in this situation:
  * $X_{i-1} \rightarrow X_{i+1}$ In this situation, $X_{i-1} \rightarrow X_{i+1}$ will be an active trail, which breaks the minimal assumption.
  * $X_{i-1} \leftarrow X_{i+1}$ In this situation, there will be a cycle among these three variables.
- $X_{i-1} \leftarrow X_i \leftarrow X_{i+1}$.Because these three variables form a triangle. There are two sub-cases in this situation:
  * $X_{i-1} \rightarrow X_{i+1}$ In this situation, there will be a cycle among these three variables.
  * $X_{i-1} \leftarrow X_{i+1}$ In this situation, $X_i \rightarrow X_{i-1} \leftarrow X_{i+1}$ is a v-structure. So there must be $X_{i-2} \leftarrow X_{i-1}$ to make the whole trail active. It follows $X_{i-2} \leftarrow X_{i-1} \leftarrow X_{i+1}$ is active, which breaks the minimal assumption.
- $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$.Because these three variables form a triangle. There are two sub-cases in this situation:
  * $X_{i-1} \rightarrow X_{i+1}$ In this situation, $X_{i-2} \rightleftharpoons X_{i-1} \rightarrow X_{i+1}$ will be an active trail, which breaks the minimal assumption.
  * $X_{i-1} \leftarrow X_{i+1}$ In this situation, $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ is a v-structure, so there must be $X_{i-2} \leftarrow X_{i-1}$. In this situation $X_{i-2} \leftarrow X_{i-1} \leftarrow X_{i+1}$ breaks the minimal assumption.
- $X_{i-1} \leftarrow X_i \rightarrow X_{i+1}$ This is the situation we want to proof.

(3) Two Bayes networks, G and G', are I-equivalent if and only if G and G' have the same skeletons and the same set of immoralities.

Solution:

Part one: Prove that if G and G' have the same skeletons and the same set of immoralities then they are I-equivalent. So we need to prove if $(X \perp Y|Z) \in I(G)$, then $(X \perp Y|Z) \in I(G')$.
Proof by contradiction.
Assume $(X \perp Y|Z) \in I(G)$ and $(X \perp Y|Z) \notin I(G')$. So there is a minimal active trail between X and Y given Z in G', but blocked in G.For example $X_{i-1} \rightleftharpoons X_i \rightleftharpoons X_{i+1}$ is blocked in G. There are two situations it is blocks:

Situation one: In X_{i-1}, X_i, X_{i+1}, X_i is observed and they form cascade or common parent structure in G, while they form v-structure in G'.Because they forms v-structure in G' and the trail is minimal, there is no link between X_{i-1} and X_{i+1}(Problem 2's result). Thus this v-structure is a immorality in G'. So it should also be a immorality in G. This is a contradiction.

Situation two: X_{i-1}, X_i, X_{i+1} are all unobserved and they form a v-structure in G, while in G' they don't form a v-structure. Thus X_{i-1}, X_i, X_{i+1} can not form a immorality, and there will be a edge between X_{i-1} and X_{i+1}. It violates the assumption that X_{i-1} and X_{i+1} is blocked in G.

Part two: Prove that two networks G and G' that induce the same conditional independence assumptions must have the same skeleton and the same immoralities. By Problem 1.3 we already proved that G and G' have the same skeleton, thus we only need to prove they have the same immoralities.
Proof by contradiction:
Two networks G and G' are I-equivalent, have the same skeleton but different immoralities. For example,X_{i-1}, X_i, X_{i+1} forms a immorality in G, but they form a cascade or common parent structure in G'. Given X_i observed so the trail $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ is active in G, where in G' this trail is inactive. This situation violates the Proposition 1, we proved.

Algorithm 2 I-equivalent

Input: Bayesian network graph's direct edge set (X,Y): G1, G2;
Output: Whether they are I-equivalent: Flag

```
function isIEquivalent(G1,G2)
//Whether two graphs have the same skeleton
    if isSameSkeleton(G1,G2)=False then
        return False
    end if
//Get immoralities of each graph
    M1 ← getImmoralities(G1)
    M2 ← getImmoralities(G2)
    if M1 != M2 then
        return False
    end if
    return True
end function

// Whether two graphs have the same skeleton
function isSameSkeleton(G1,G2)
    if G1.size()!=G2.size() then
        return False
    end if
    for (X, Y) ∈ G1 do
        if (X, Y) ∉ G2 and (Y, X) ∉ G2  then
            return False
        end if
    end for
    return True
end function

//Output the immoralities in the graph
function getImmoralities(G)
//Get the node Set V
    V ← ∅
    for (X, Y) ∈ G do
        V ← V ∪ {X, Y}
    end for//Iteration all there node triple to find the immoralities
    M ← ∅ //The set of all the immoralities
    for X ∈ V do
        for Y ∈ V do
            for Z ∈ V do
                if X,Y,Z are distinct and (X, Y) ∈ G and (Z, Y) ∈ G then
                    if (X, Z) ∉ G and (Z, X) ∉ G  then
                        M = M ∪ (X, Y, Z)
                    end if
                end if
            end for
        end for
    end for
    return M
end function
```

3 Exact Inference

3.1 Variable Elimination

$$\begin{aligned}\alpha_{t+1}(j) = p(z_{t+1} = j|x_{1,..,t+1}) &\propto \sum_i p(z_{t+1} = j|z_t = i)p(x_{t+1}|z_{t+1} = j)p(z_t = i|x_{1,..,t}) \\ &\propto p(x_{t+1}|z_{t+1} = j)\sum_i p(z_{t+1} = j|z_t = i)\alpha_t(i)\end{aligned}$$

$$\begin{aligned}\beta_{t-1}(j) = P(x_{t,...,T}|z_{t-1} = j) &\propto \sum_i P(x_t|z_t = i)P(z_t = i|z_{t-1} = j)P(x_{t+1,...,T}|z_t = i) \\ &\propto \sum_i p(x_t|z_t = i)p(z_t = i|z_{t-1} = j)\beta_t(i)\end{aligned}$$

Then because

$$p(z_t = j|x_1, ..., T) \propto \alpha_t(j)\beta_t(j)$$

we can get

$$p(z_t = j|x_1, ..., T) = \frac{\alpha_t(j)\beta_t(j)}{\sum_i \alpha_t(i)\beta_t(i)}$$

3.2 Gaussian Belief Propagation

(1) Proof: $N(x|\mu_1, \lambda_1^{-1}) \times N(x|\mu_2, \lambda_2^{-1}) = CN(x|\mu, \lambda^{-1})$

Solution

$$\begin{aligned}N(x|\mu_1, \lambda_1^{-1}) \times N(x|\mu_2, \lambda_2^{-1}) &= exp(-\frac{\lambda_1(x-\mu_1)^2}{2}) \times exp(-\frac{\lambda_2(x-\mu_2)^2}{2}) \\ &= exp(-\frac{1}{2}(\lambda_1(x-\mu_1)^2 + \lambda_2(x-\mu_2)^2)\end{aligned}$$

$$\begin{aligned}\lambda_1(x-\mu_1)^2 + \lambda_2(x-\mu_2)^2 =&(\lambda_1+\lambda_2)x^2 - 2(\lambda_1\mu_1+\lambda_2\mu_2)x + \lambda_1\mu_1^2 + \lambda_2\mu_2^2 \\ =&(\lambda_1+\lambda_2)[x^2 - \frac{2(\lambda_1\mu_1+\lambda_2\mu_2)}{\lambda_1+\lambda_2}x + (\frac{\lambda_1\mu_1+\lambda_2\mu_2}{\lambda_1+\lambda_2})^2] - \frac{(\lambda_1\mu_1+\lambda_2\mu_2)^2}{\lambda_1+\lambda_2} \\ &+\lambda_1\mu_1^2 + \lambda_2\mu_2^2 \\ =&(\lambda_1+\lambda_2)[x - \frac{\lambda_1\mu_1+\lambda_2\mu_2}{\lambda_1+\lambda_2}]^2 + \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}(\mu_1-\mu_2)^2\end{aligned}$$

So we can get

$$\lambda = \lambda_1 + \lambda_2$$
$$\mu = \frac{\lambda_1\mu_1+\lambda_2\mu_2}{\lambda_1+\lambda_2}$$
$$C = \sqrt{\frac{\lambda_1\lambda_2}{2\pi(\lambda_1+\lambda_2)}}exp(-\frac{1}{2}\frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}(\mu_1-\mu_2)^2)$$

(2) Solve for $m(x_t)$. Hint First solve for the messages passed into Node t's neighbor nodes from their neighbors excluding t, then use these messaged to represent $m(x_t)$

Solution: (Gaussian Belief Propagation:Theory and Application https://arxiv.org/pdf/0811.2518.pdf)

$$N(\mu_{s/t}, \lambda_{s/t}) \propto \phi_s(x_s) \prod_{k \in N(s)/t} m_{ks}(x_s)$$

$$\lambda_{s/t} = \lambda_{ss} + \sum_{k \in N(s)/t} \lambda_{ks}$$

$$\mu_{s/t} = \lambda_{s/t}^{-1}(\lambda_{ss}\mu_{ss} + \sum_{k \in N(s)/t} \lambda_{ks}\mu_{ks})$$

Where λ_{ks} and μ_{ks} are the precision and mean of the message $m_{ks}(x_s)$, and $\lambda_{ss} \triangleq A_{ss}$, $\mu_{ss} \triangleq \frac{b_s}{A_{ss}}$.
Next we calculate $m_{st}(x_t)$

$$\begin{aligned}
m_{st}(x_t) &= \int_{x_s} \phi_{s,t}(x_s, x_t)\phi_s(x_s) \prod_{k \in N(s)/t} m_{ks}(x_s) \\
&\propto \int_{x_s} exp(-x_s A_{st} x_t) exp(-\lambda_{s/t}(x_s^2/2 - \mu_{s/t} x_s)) \\
&= \int_{x_s} exp(-\lambda_{s/t} x_s^2/2 + (\lambda_{s/t}\mu_{s/t} - A_{st}x_t)x_i) \\
&\propto exp((\lambda_{s/t}\mu_{s/t} - A_{st}x_t)^2/(2\lambda_{s/t})) \\
&\propto N(\mu_{st}, \lambda_{st})
\end{aligned}$$

Where μ_{st} and μ_{st} are the precision and mean of the message $m_{st}(x_t)$.

$$\mu_{st} = \frac{\lambda_{s/t}\mu_{s/t}}{A_{st}}$$

$$\lambda_{st} = -\frac{A_{st}^2}{\lambda_{s/t}}$$

Then we can calculate $m(x_t)$

$$m(x_t) = \phi_t(x_t) \prod_{s \in N(t)} m_{st}(x_t) \propto N(\mu_t, \lambda_t^{-1})$$

$$\lambda_t = \lambda_{tt} + \sum_{s \in N(t)} \lambda_{st}$$

$$\mu_t = \lambda_t^{-1}(\lambda_{tt}\mu_{tt} + \sum_{s \in N(t)} \lambda_{st}\mu_{st})$$

Where $\lambda_{tt} = A_{tt}$, $\mu_{tt} = \frac{b_t}{A_{tt}}$

4 Undirected Graphical Models

4.1 Ising Model & Boltzmann Machine

Solution:

$$
\begin{aligned}
exp(\sum_{(i,j)\in E} W_{ij}x_ix_j - \sum_{i\in V} u_ix_i) &\propto exp(\sum_{(i,j)\in E} 4W_{ij}(\frac{x_i}{2}+\frac{1}{2})(\frac{x_j}{2}+\frac{1}{2}) - \sum_{(i,j)\in E} W_{ij}x_i - \sum_{(i,j)\in E} W_{ij}x_j - \sum_{i\in V} u_ix_i) \\
&\propto exp[\sum_{(i,j)\in E} 4W_{ij}(\frac{x_i}{2}+\frac{1}{2})(\frac{x_j}{2}+\frac{1}{2}) - \sum_{i\in V}(u_i + 2\sum_{j\in N(i)} W_{ij})x_i] \\
&\propto exp[\sum_{(i,j)\in E} 4W_{ij}(\frac{x_i}{2}+\frac{1}{2})(\frac{x_j}{2}+\frac{1}{2}) - \sum_{i\in V} 2(u_i + 2\sum_{j\in N(i)} W_{ij})(\frac{x_i}{2}+\frac{1}{2})]
\end{aligned}
$$

Thus we can get:

$$
\begin{aligned}
x_i' &= \frac{x_i}{2}+\frac{1}{2} \in \{0,1\} \\
W_{ij}' &= 4W_{ij} \\
u_i' &= 2(u_i + 2\sum_{j\in N(i)} W_{ij})
\end{aligned}
$$

4.2 Determinantal Point Process

1 Give a representation of the DPP distribution in the canonical exponential form as a binary MRF.

$$
\begin{aligned}
P(\mathcal{Y}\subset\mathcal{X}) = \frac{det(S_{\mathcal{Y}})}{det(S+I)} &= \frac{e^{log(det(S_{\dagger}))}}{det(S+I)} = \frac{e^{log(det(S_{\mathcal{Y}}))\delta(\mathcal{Y},\mathcal{Y})}}{det(S+I)} \\
&= \frac{e^{\sum_{\tilde{\mathcal{Y}}\in\mathcal{X}} log(det(S_{\tilde{\mathcal{Y}}}))\delta(\tilde{\mathcal{Y}},\mathcal{Y})}}{det(S+I)} \\
&= \frac{1}{Z}e^{\sum_{\tilde{\mathcal{Y}}\in\mathcal{X}} \theta_{\tilde{\mathcal{Y}}} f_{\tilde{\mathcal{Y}}}(\mathcal{Y})}
\end{aligned}
$$

Where

$$
\begin{aligned}
\theta_{\tilde{\mathcal{Y}}} &= log(det(S_{\tilde{\mathcal{Y}}})) \\
f_{\tilde{\mathcal{Y}}}(\mathcal{Y}) &= \delta(\tilde{\mathcal{Y}},\mathcal{Y})
\end{aligned}
$$

2 Is Boltzmann machine on n variable equivalent to an DPP on a set of n elements? In other words, for a DPP with some kernel matrix S, does there always exit W and u such that the corresponding Boltzmann distribution is equivalent to the one given by DPP? How about the opposite?

No, the Boltzmann machine on n variables is not equivalent to an DPP on a set of n elements. For example, the Boltzmann machine can be written like

$$
P(Y=y) = \frac{1}{Z}exp\{\sum_i w_iy_i + \sum_{i<j} w_{ij}y_iy_j\}
$$

Let consider the situation that $n=3$. We can check a simple case, let $det(S+I)=Z$,$w_1=w_2=w_3=0$. In this situation, we can calculate $exp\{\sum_i w_iy_i + \sum_{i<j} w_{ij}y_iy_j\}$ and $det(S_{\mathcal{Y}})$ individually.

By solving this, we can get

$$2S_{12}S_{23}S_{13} + S_{12}^2 S_{23}^2 S_{13}^2 = S_{12}^2 S_{13}^2 + S_{12}^2 S_{23}^2 + S_{13}^2 S_{23}^2 \tag{1}$$

Now we can let $w_{12} = w_{13} = log\frac{4}{3}, w_{23} = 0$. Then by the $5_{th} - 7_{th}$ rows of the table, we can know $S_{12} = S_{13} = 0.5, S_{23} = 0$. In this situation, it contradicts with (1).

In the other direction, obviously, we can find a positive semi-definite kernel matrix S, which doesn't satisfy (1). Then this S can not be transfered to Boltzmann machine.

$y_1y_2y_3$	$exp\{\sum_i w_i y_i + \sum_{i<j} w_{ij} y_i y_j\}$	$det(S_{\mathcal{Y}})$
000	1	1
100	1	S_{11}
010	1	S_{22}
001	1	S_{33}
110	$e^{-w_{12}}$	$S_{11}S_{22} - S_{12}^2$
101	$e^{-w_{13}}$	$S_{11}S_{33} - S_{13}^2$
011	$e^{-w_{23}}$	$S_{22}S_{33} - S_{23}^2$
111	$e^{-w_{12}-w_{13}-w_{23}}$	$S_{11}S_{22}S_{33} + 2S_{12}S_{23}S_{13} - S_{11}S_{23}^2 - S_{22}S_{13}^2 - S_{33}S_{12}^2$