Assignment 1

Methods of PCA, MDS and Isomap

DD2434 Advanced Machine Learning
Filip Bergentoft, bergento@kth.se

qqq: Add figures to problem 1

2.1 Dependencies in a Directed Graphical Model

Question 2.1.1:

In the graphical model of Figure **??**, *is* $\mu_{r,c} \perp \mu_{r,c+1}$*?*

Answer: No, $\mu_{r,c}$ and $\mu_{r,c+1}$ are dependent.

Question 2.1.2:

In the graphical model of Figure **??**, *is* $X_{r,c} \perp X_{r,c+1} | \{mu_{r,c}, \mu_{r,c+1}\}$*?*

Answer: Yes, $X_{r,c}$ and $X_{r,c+1}$ are d-separated and thus independent.

Question 2.1.3:

Give a minimal set of variables A *such that* $X_{r,c} \perp \mu_0 | A$

Answer: The minimal set of variables are given by $A = \{\mu_{r,c}, \mu_{r-1,c}, \mu_{r+1,c}, \mu_{r,c-1}, \mu_{r,c+1}\}$

Question 2.1.4:

In figure **??**, *is* $Z \perp X | C$ *where* $Z = \{Z_m^n : n \in [N], m \in [M]\}$, $X = \{X_m^n : n \in [N], m \in [M]\}$ *and* $C^n = \{C^n : n \in [N]\}$*?*

Answer: No, Z and X are not d-separated given C and are thus dependent.

Question 2.1.5:

In figure **??**, *is* $A \perp e | B$ *where* $A = \{A_{i,j}^k : k \in [K], i, j \in [I]\}$, $e = \{e_{i,r}^k : k \in [K], i \in [I], r \in [R]\ \}$ *and* $B = \{Z_m^n : m \in [M], m \text{ odd}\} \cup \{X_m^n : m \in [M], m \text{ even}\}$*?*

Answer: No, A and e are not d-separated given B and are thus dependent.

Question 2.1.6:

In figure ??, give a minimal set of variables B such that $A \perp X|B$ where $A = \{A^k_{i,j} : k \in [K], i, j \in [I]\}$ and $X = \{X^n_m : n \in [N], m \in [M]\}$

Answer: The minimal set of variables are given by

$$B = \{Z^n_m : n \in [N], m \in [M]\} \cup \{C^n : n \in [N]\}$$

2.2 Likelihood of a Tree Graphical Model

Question 2.2.1:

Implement a dynamic programming algorithm that, for a given T, Θ and β computes $p(\beta|T, \Theta)$.

T is a binary tree with a vertex set $V(T)$ and a leaf set $L(T)$. For each vertex $v \in V(T)$ there is an associated random variable $X_v \in [K]$ with a corresponding CPD $\theta_v = p(X_v|x_{pa(v)})$ which is a categorical distribution. β is defined as the set of values of all leafs in T such that $\beta = \{x_l : l \in L(T)$.

In order to compute $p(\beta|T, \Theta)$ we need to find an expression that can be used for dynamic programming, I.E. splitting up the full problem into smaller subproblems. By looking at the definition of s in equation (1)

$$s(u, i) = p(X_{Observed \cap \downarrow u}|X_u = i) \tag{1}$$

and letting the root node of the tree being denoted by r, one can use that if u is chosen as the root r we get the following expression

$$s(r, i) = p(X_{Observed \cap \downarrow r}|X_r = i) = \left\{ X_{Observed \cap \downarrow r} = \beta \right\} = p(\beta|X_r = i, T, \Theta)$$

We can then marginalise this using Bayes' theorem in the following manner

$$p(\beta|T, \Theta) = \sum_i p(\beta, X_r = i|T, \Theta) = \sum_i p(\beta|X_r = i, T, \Theta)p(X_r = i) \tag{2}$$

$$= \sum_i s(r, i)p(X_r = i) \tag{3}$$

Using that T is a binary tree and thus if v, w are children to a node u then

$$
\begin{aligned}
s(u,i) &= p(X_{Observed \cap \downarrow u} | X_u = i) \\
&= p(X_{Observed \cap \downarrow v} | X_v = i) p(X_{Observed \cap \downarrow w} | X_w = i) \\
&= \Big(\sum_j s(v,j) p(X_v = j | x_u = i) \Big) \Big(\sum_j s(w,j) p(X_w = j | x_w = i) \Big)
\end{aligned}
\tag{4}
$$

A special case is when the node u is a leaf node, then the following holds

$$
s(u,i) = \begin{cases} 1, & X_u = i \\ 0, & otherwise \end{cases} \tag{5}
$$

Equation (3) can then be computed using dynamic programming by starting at the leaf nodes using equation (5) and then traversing up the nodes in the tree to the root using equation (4) one level at a time and storing the achieved probabilities s along the way.

Question 2.2.2:

Apply your algorithm to the graphical model and data provided separately.

The following likelihoods were achieved when applying my implementation of the dynamic programming algorithm on the given trees.

Tree sample:	0	1	2	3	4
Small tree	0.016	0.015	0.011	0.007	0.041
Medium tree	$4.336 \cdot 10^{-18}$	$3.094 \cdot 10^{-20}$	$1.050 \cdot 10^{-16}$	$6.585 \cdot 10^{-16}$	$1.488 \cdot 10^{-18}$
Large tree	$3.288 \cdot 10^{-69}$	$1.109 \cdot 10^{-66}$	$2.522 \cdot 10^{-68}$	$1.242 \cdot 10^{-66}$	$3.535 \cdot 10^{-69}$