Tutorial 7 - Markov Networks

COMP9418 – Advanced Topics in Statistical Machine Learning

Lecturer: Gustavo Batista

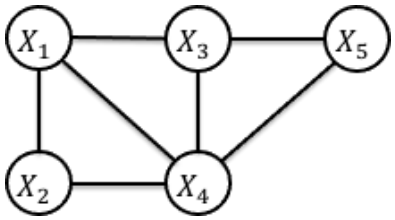
Lecture: Markov Networks

Topic: Questions from lecture topics

Last revision: Thursday 29th October, 2020 at 10:38

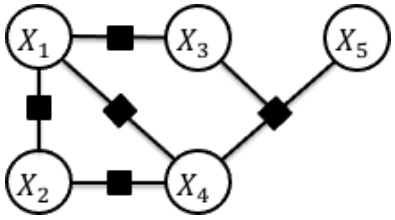
Question 1

Consider the following Markov network in the form of an undirected graph.



a. Determine the cliques in the graph.
b. Express the joint probability as a product of clique potentials.
c. Assuming all the variables are binary, how many parameters are there in this model?

Now, consider the same Markov network expressed as a factor graph.



Answer the same questions (a,b and c) for this model.

Answer

a. The Markov network expressed as an undirected graph can lead to more than one factorisation. For instance, we can interpret the graph as a pairwise Markov network. In this case, the cliques are $X_1 - X_2$, $X_1 - X_3$, $X_1 - X_4$, $X_2 - X_4$, $X_3 - X_4$, $X_3 - X_5$, and $X_4 - X_5$. Another factorisation considers the network as composed of maximal cliques. In this case, we have a smaller number cliques: $X_1 - X_2 - X_4$, $X_1 - X_3 - X_4$, and $X_3 - X_4 - X_5$. In the remaining of this exercise, we will adopt this last interpretation.

b. For maximal cliques, we have the following factorisation: $\phi_1(X_1, X_2, X_4)\phi_2(X_1, X_3, X_4)\phi 3(X_3, X_4, X_5)$.

c. For maximal cliques, each clique requires 2^3 parameters, totaling 24 parameters. Note that we cannot reduce the number of parameters required to represent each factor, because they are unnormalised.

Now, the factor graph has a single interpretation.

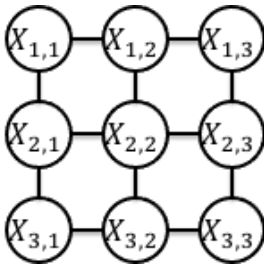
a. The cliques are $X_1 - X_2$, $X_1 - X_3$, $X_1 - X_4$, $X_2 - X_4$, and $X_3 - X_4 - X_5$.

b. The factorisation according to the factor graph is $\phi_1(X_1, X_2)$, $\phi_2(X_1, X_3)$, $\phi_3(X_1, X_4)$, $\phi_4(X_2, X_4)$, and $\phi_5(X_3, X_4, X_5)$.

c. The number of parameters for the factorisation in the previous answer is $4 \times 2^2 + 2^3 = 24$ parameters.

Question 2

Consider the following pairwise Markov network in the form of a grid.



a. Express the pairwise Markov independencies of this graph for missing edges between the nodes $X_{1,1}$, $X_{1,2}$ and $X_{2,2}$.

b. List the nodes that compose the Markov blanket for each of nodes $X_{1,1}$, $X_{1,2}$ and $X_{2,2}$

c. Is $X_{1,1}$ separated from $X_{2,2}$ given $X_{1,2}$? If not, which minimal set of nodes is necessary to observe to guarantee the separation between these nodes?

Answer

a. $X_{1,1}$ has an edge to $X_{1,2}$. The same occurs to $X_{1,2}$ and $X_{2,2}$. Therefore, we cannot claim any pairwise Markov independencies for these pairs of nodes. However, $X_{1,1}$ and $X_{2,2}$ are not connected by an edge. Therefore, $X_{1,1} \perp X_{2,2} | X_{1,2}, X_{2,1}, X_{3,1}, X_{1,3}, X_{3,2}, X_{2,3}, X_{3,3}$.

b. The Markov blanket for $X_{1,1}$ is $X_{2,1}$ and $X_{1,2}$. For node $X_{1,2}$, the Markov blanket has three nodes $X_{1,1}$, $X_{2,2}$ and $X_{1,3}$. Finally, for node $X_{2,2}$, the Markov blanket has four nodes $X_{1,2}$, $X_{2,1}$, $X_{2,3}$ and $X_{3,2}$.

c. No. We would need to observe $X_{2,1}$ as well to guarantee a separation between $X_{1,1}$ and $X_{2,2}$.

Question 3

In an image smoothing application, the image is represented by a regular Markov Random Field with a neighbourhood defined as the grid of the previous question.

In this application, we have two kinds of potentials: $\phi_1(X_{i,j}, X_{i',j'})$ represents the correlation between neighboring pixels and $\phi_2(X_{i,j}, Y_{i,j})$ represents the correlation between the filtered image pixel $(X_{i,j})$ and the respective noisy image pixel $(Y_{i,j})$. A simple form of assigning values to these potentials, is using the absolute difference between pixel values, i.e, $\phi(a, b) = |a - b|$.

Also, we define the energy function for an image $\mathbf{X}$ as

$$E(\mathbf{X}) = \sum_{i,j} (\phi_1(X_{i,j}, X_{i+1,j}) + \phi_1(X_{i,j}, X_{i,j+1})) + \lambda \sum_{i,j} \phi_2(X_{i,j}, Y_{i,j})$$

with $\lambda = 4$, giving more weight to the observations. (For simplicity, this equation does not deal with the edge cases on the bottom and right edge of the image. Consider how these should be dealt with).

Suppose we want to smooth a small image represented by a 4×4 grid, where each site can take one of two values, 0 and 1. Given the initial configuration F

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and the observation G,

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

a. Obtain the MAP configuration using the Iterated Conditional Modes (ICM) variant of the stochastic search algorithm. This involves considering each location one by one and checking whether changing that pixel will decrease the total energy.
b. Would you expect the results obtained by the other variants of the stochastic search algorithm to be different? Why?

Answer

a. Scholastic search algorithm uses a series of local changes in the attempt to find a MAP configuration. In this exercise, each change involves a single pixel that has only two possible values 0 and 1. Each change will locally affect the energy function since the impact of the difference in a single pixel can only change the energy function concerning the Markov blanket of that pixel.

   Let's suppose we will scan the image from the left top corner $(1, 1)$ to the right lower corner $(4, 4)$. The first observed pixel with value different of zero is at position $(2, 2)$. The change of $X_{2,2}$ to 1 does not lead to an energy decrease for $\lambda = 4$. The same occurs for pixels in positions $(2, 3)$ and $(3, 2)$. The only pixel that will change value is the one at $(4, 3)$ because it only has three neighbours. Therefore, the MAP configuration is

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

b. Yes. In this case, the stochastic algorithm might accept a change that leads to an increase in energy, or doesn't change the energy. For instance, if we accept a change for pixel $(2, 2)$, such change would result in changes for the neighbouring pixels as well, resulting in MAP configuration that is the same as the observation G.

Question 4

What is the time complexity of the stochastic search algorithm for its different variants?

Answer

The stochastic search algorithm has three main components that influence its efficiency: the number of iterations (m), the number of variables (n) and the cost of computing the energy function (e). Therefore, the complexity is $O(mne)$. However, we can use the following strategies to reduce its complexity:

1. Early abandon: in the case we detect convergence, we can stop the algorithm earlier. This is usually implemented by setting a threshold for changes in the energy function between consecutive iterations.

2. Markov blanket: the energy function usually has a factorisation according to the graph structure. As stochastic search makes local changes, the impact of those changes can be contained to the Markov blanket of the changed variable. Therefore, the computation of the energy function can be reduced to the terms that belong to the blanket.

Finally, the variations do not impact the algorithm complexity significantly.