Probabilistic Graphical Models

Submit a PDF of your answers to Canvas.

1. Consider a joint distribution over a black and white 3×3 image with pixels $x_1, \ldots, x_9$ as shown below:

x_1	x_4	x_7
x_2	x_5	x_8
x_3	x_6	x_9

Note that $\boldsymbol{x}$ can take on one of 2^9 values in $\{0,1\}^9$. Without further assumption, we would need to estimate 2^9 parameters to estimate $p(\boldsymbol{x})$.

a) Recall that by repeated application of the product rule of probability, a joint distribution $p(\boldsymbol{x})$ can be written as:

$$p(\boldsymbol{x}) = p(x_1) \prod_{i=2}^{n} p(x_i | x_1, \ldots x_{i-1}) \tag{1}$$

Write the expression for the joint distribution $p(\boldsymbol{x})$, $\boldsymbol{x} = [X_1, X_2, \ldots, X_9]^T$ using the factorization in expression (1).

b) Is the factorization unique?

c) Assume that each pixel, conditioned on the values of the 2 pixels immediately to the left/above, is independent of any pixel with a lower index. This *conditional independence* assumption implies, for example

$$p(x_5 | x_1, x_2, x_3, x_4) = p(x_5 | x_2, x_4).$$

Assume that for a pixel on the edge of the image, for example, x_3,

$$p(x_3 | x_1, x_2) = p(x_3 | x_2).$$

Simplify your expression from **(a)** by incorporating the conditional independence assumption.

d) Draw a directed graphical model (i.e, a DAG) that matches the decomposition and conditional independence structure found above.

e) How many parameters do you need to estimate to estimate $p(\boldsymbol{x})$ with the conditional independence assumption?

SOLUTION:

a)

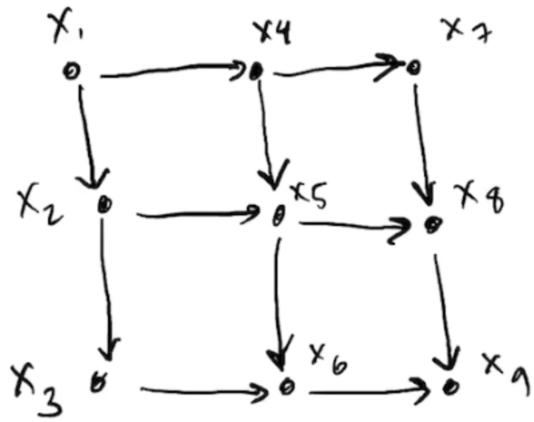
$$\begin{aligned} p(\boldsymbol{x}) &= p(x_1) \\ &\times p(x_2|x_1) \\ &\times p(x_3|x_1, x_2) \\ &\times p(x_4|x_1, x_2, x_3) \\ &\times p(x_5|x_1, x_2, x_3, x_4) \\ &\times p(x_6|x_1, x_2, x_3, x_4, x_5) \\ &\times p(x_7|x_1, x_2, x_3, x_4, x_5, x_6) \\ &\times p(x_8|x_1, x_2, x_3, x_4, x_5, x_6, x_7) \\ &\times p(x_9|x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) \end{aligned}$$

b) No, as we could have just as easily started with any x_j.

c)

$$\begin{aligned} p(\boldsymbol{x}) &= p(x_1) \\ &\times p(x_2|x_1) \\ &\times p(x_3|x_2) \\ &\times p(x_4|x_1) \\ &\times p(x_5|x_2, x_4) \\ &\times p(x_6|x_3, x_5) \\ &\times p(x_7|x_4) \\ &\times p(x_8|x_5, x_7) \\ &\times p(x_9|x_6, x_8) \end{aligned}$$

d) See the figure below.



e) Estimating $p(x_1)$ requires estimating a single parameter, for example. Estimating $p(x_2|x_1)$ requires estimation 2 parameters (one for each state of x_1). p($x_5|x_2, x_4$) requires 4 parameters (one for each value) of x_1, x_2. This gives us $1+2\times4+4\times4 = 25$, which is much less than $2^9 = 512$.