Probabilistic Graphical Models
Exercise Sheet No. 1

Due Date: November 2nd, 10 am

Hand in: At beginning of tutorial OR at office 629 (Jan-Hendrik) OR by e-mail to (jlange[at]mpi-inf.mpg.de). Begin the subject of your e-mail with [pgm].

Please do not forget to put your name on the submission.

1 MAP-Inference on chain graphs

Points: 4

Let $\mathcal{X}_i = \{0, 1, 2\}$ for $1 \le i \le 5$ and consider the probability distribution $p_\theta(x)$ on $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_5$ that is parameterized on a chain graph as follows

x_i	0	1	2
$\theta_1(x_1)$	-1	-2	1
$\theta_2(x_2)$	-1	1	2
$\theta_3(x_3)$	2	-1	-2
$\theta_4(x_4)$	1	2	1
$\theta_5(x_5)$	-2	1	-1

$$\theta_{i,i+1}(x_i, x_{i+1}) = \begin{cases} -1 & \text{if } x_i = x_{i+1} \\ 1 & \text{else.} \end{cases}$$

Find $\operatorname{argmax}_{x \in \mathcal{X}} p_\theta(x)$ by dynamic programming.

2 Probabilities

Points: 8

Let Ω be a sample space and $p\colon \Omega \to [0, 1]$ a probability distribution. We extend the definition of p to the power set $\mathcal{P}(\Omega)$ via $p(E) = \sum_{x \in E} p(x)$ for any *event* $E \subset \Omega$.

1. For any two events E_1 and E_2, prove that
$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2). \tag{1}$$
2. (Bayes' law) Given the Kolmogorov definition for conditional probabilities
$$p(A \mid B) = \frac{p(A \cap B)}{p(B)}, \tag{2}$$

derive Bayes' law:

$$p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B)}. \tag{3}$$

3. (Law of total probability) Let $E_1, \ldots, E_n$ be mutually disjoint events from the sample space Ω such that $\Omega = \bigcup_{i=1}^n E_i$. Then for any event B from the same space Ω show that

$$p(B) = \sum_{i=1}^n p(B \cap E_i) = \sum_{i=1}^n p(B \mid E_i)p(E_i). \tag{4}$$

4. (Linearity of expectation) For any finite collection of discrete random variables $X_1, \ldots, X_n$ with finite expectations $\mathbb{E}[X_i] = \sum_x x \cdot p[X_i = x]$, show that

$$\mathbb{E}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathbb{E}[X_i]. \tag{5}$$

5. Let X, Y, Z be three disjoint subsets of random variables. We say X and Y are conditionally independent given Z if and only if

$$p_{X,Y|Z}(x, y \mid z) = p_{X|Z}(x \mid z)p_{Y|Z}(y \mid z) \tag{6}$$

Show that X and Y are conditionally independent given Z if and only if the joint distribution for the three subsets of random variables factors in the following form:

$$p_{X,Y,Z}(x, y, z) = h(x, z)g(y, z) \tag{7}$$

(Be careful to prove both directions!)

3 Complexity analysis

Points: 6

Consider the three random variables X, Y, Z all of which are binary.

- How many variable assignments to x, y, z (states) do you need in general to fully specify the joint distribution $p(x, y, z)$?
- How many states are needed if the distribution factorizes in $p(x, y, z) = p(x \mid y)p(y \mid z)p(z)$?
- How many states do you need if the variables are not binary but can take values in $\{1, 2, \ldots, N\}$? Consider both previous cases.
- How many states do you need to specify a distribution over all 8-bit gray-scale images of size $k \times \ell$ pixels? There are random variables $x_1, x_2, \ldots, x_{k\ell}$ with $x_i \in \{0, \ldots, 255\}$ for $i = 1, \ldots k\ell$.
- Suppose the previous distribution is a Gibbs distribution with energy function $E(x) = \sum_i f_i(x_i) + \sum_{i\sim j} f_{i,j}(x_i, x_j)$ where $i \sim j$ means that pixels i and j are adjacent. How many states do you need now?