E1 213 Pattern Recognition and Neural networks
Problem Sheet 7

1. Consider a joint density of five random variables given by (under the usual abuse of notation)

$$f(x1, x2, x3, x4, x5) = f(x1)f(x2|x1, x3)f(x3)f(x4|x3)f(x5|x2, x3)$$

Represent this as a directed graphical model. Use D-separation theorem to answer the following. Are there any random variables here which are unconditionally independent? Are $x2$ and $x4$ independent conditioned on $x3$? Are $x1$ and $x3$ independent conditioned on $x2$.

2. Consider a directed graphical model over four nodes. The nodes x_1 and x_2 have no parents; x_3 has both x_1 and x_2 as parents; and x_4 has only x_3 as parent. Draw the directed graph. Write down the joint distribution given by this graph. From this joint distribution, show that x_1 and x_2 are independent but they are not, in general, conditionally independent given x_4. Verify that this is what is given by D-separation theorem.

3. In a graphical model, given a node x_i, we define its markov blanket, MB_i as the smallest set of nodes that satisfies

$$f(x_i|x_j, j \neq i) = f(x_i|\text{MB}_i).$$

We have seen what the markov blanket is in a directed graphical model. What would it be in an undirected graphical model?

4. Consider a convolutional neural network. Suppose the input is a 31×31 image with three color channels. (Thus the input layer is of size/volume $31 \times 31 \times 3$). The network has two convolutional layers. The first one uses fifty filters. The filter size is 5×5 and we use a stride of 2. We use no zero padding. What would be the size of first convolutional layer? After the first convolutional layer we use a pooling layer with filter size 2×2 and stride 2 so that image size gets halved in both directions. The pooling layer is followed by the second convolutional layer. For the second convolutional layer we use zero padding so that the size does not change because of the convolutional layer. In this convolutional layer we use hundred filters each of size 3×3 with stride of one. What is the total number of weights that we need to learn for the two convolutional

layers? After the second convolutional layer, we have a pooling layer to halve the size of the image and then we have two fully connected layers. The first one has 20 nodes and the next one has 5. What are the total number of weights that we need to learn in this network.

5. Consider a Hopfield network with 4 nodes and the outputs taking values in $\{0, 1\}$. Suppose the weight matrix is given by

$$W = \begin{bmatrix} 0 & 1 & -1 & -1 \\ 1 & 0 & -1 & -1 \\ -1 & -1 & 0 & 1 \\ -1 & -1 & 1 & 0 \end{bmatrix}$$

Suppose the biases for all nodes are equal to 1. Let the initial values of the nodes be [1 1 1 0]. What will be the final values of nodes at convergence? Suppose you started the network with [0 0 1 1]. What would the network converge to?

6. Consider a Boltzmann machine with two visible and two hidden nodes. Draw the undirected graphical model for this. Write a general expression for the joint distribution represented by this graph. Now suppose this is an RBM. Then what would be the graph and the joint distribution?

7. Suppose $S = \{0,\ 1\}^n$ and let π be a a mass function over it given by $\pi(x) = b_x/Z,\ x \in S$ where Z is the normalizing constant. Suppose we want to generate samples from this distribution. We can use MCMC technique for this. Does the Boltzmann machine give us some idea of what proposal distribution to use if we want to employ Hastings-Metropolis algorithm? Now let $E(x) = \ln(b_x),\ x \in S$. Suppose $E(x)$ is quadratic in the components of x. (Note that x is a n-dimensional binary vector). Is there any we can make use of this in our sampling algorithm?