COMP 652: Machine Learning - Assignment 3

Posted Tuesday, March 15, 2016
Due Tuesday March 29, 2016

1. [15 points] **Properties of conditional independence**

   Recall that the conditional independence property that $A \perp\!\!\!\perp B | C$ is a statement that the conditional probability distribution satisfies
   $$p(A|B,C) = p(A|C)$$

   (a) [5 points] Show that, if $A \perp\!\!\!\perp B | C$, then $p(A,B|C) = p(A|C)p(B|C)$

   (b) [5 points] Show that conditional independence is symmetric:
   $$A \perp\!\!\!\perp B | C \Leftrightarrow B \perp\!\!\!\perp A | C$$

   (c) [5 points] Show the decomposition property:
   $$A \perp\!\!\!\perp (B \cup D) | C \Rightarrow A \perp\!\!\!\perp B | C \text{ and } A \perp\!\!\!\perp D | C$$

2. [10 points] **Arc reversal in Bayes nets**

   Give an example of a Bayes net in which reversing an arc does not modify the conditional independence relations modelled by the graph structure. Given an example of a Bayes net in which reversing the arc modifies the conditional independence relations.

3. [15 points] **Sigmoid Bayes nets**

   Suppose we have a Bayes net over a set of binary random variables. Each node is parameterized as:
   $$P(X_i|X_{\pi_i}) = \sigma\left(\sum_{j \in \pi_i} w_{ij} X_j\right)$$
   where W_{ij} are real-valued weights and σ is the sigmoid function. Suppose we have such a network with two layers: a layer H of hidden nodes, and a layer V of visible nodes. Give a gradient-based learning rule for W_{ij}, in such a way as to maximize the likelihood of a given set of data, in which only the visible variables V are recorded.

4. [35 points] **Markov Random Fields**

   Consider the 2D spin glass model we discussed in the lecture.

   (a) [10 points] Suppose that instead of connecting pixels in a 4-neighborhood, we want to connect them in an 8-neighborhood. Describe what the parameters of the undirected graphical model will be.

(b) [10 points] Suppose that we want to use such a model to capture natural scenes in images. Describe the advantages and disadvantages of this model compared to connecting a pixel only to 4 neighbours.

(c) [15 points] For the 2D Ising model connected as in class, write a Gibbs sampling algorithm, assuming that potentials are represented using linear energy functions and that evidence can be injected along the leftmost edge of the model. Assume the model is an $n \times n$ lattice.

5. [25 points] **PCA**

   Consider the data set available in file hw3pca.txt; each row represents an instance and the columns represent features. You should split the data into 80% representing the training set and 20% to test the representation. Perform PCA on the data and plot the reconstruction error as a function of the number of dimensions, both on the training set and on the test set, as well as the fraction of the variance accounted for obtained by looking at the top eigenvalues. Explain what you see and what are the implications for choosing dimensionality of the data.