COMP 652: Machine Learning - Assignment 3

Posted Wednesday March 18, 2015
Due Wednesday Tuesday March 31, 2015

1. [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.

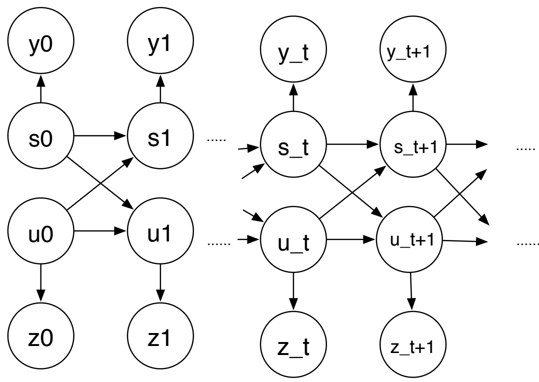
2. [20 points] **PCA**

   Consider the data set available in file hw4pca.dat; 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. Explain what you see and what are the implications for choosing dimensionality of the data.

3. [45 points] **Coupled Hidden Markov Models**

   We discussed in class the Hidden Markov Model (HMM), which has hidden states and visible observations. The Coupled Hidden Markov model (CHMM) is a similar kind of graphical model: we have several hidden Markov models running in parallel, and their states interact. This model is quite useful, for example, when you try to parse video, and you consider the observations as being sound and visual data, respectively.

   Consider a system with two HMMs, depicted in Figure 1:



Here, s_i and u_i are the states of the two coupled HMMs, y_i and z_i are the observations coming from the two chains, and the two chains interact in the way depicted in the picture.

(a) [5 points] Specify what are the parameters of this model.

(b) [10 points] Derive an algorithm for computing the joint probability of a sequence of observations $(y_0, z_0), (y_1, z_1) \ldots (y_T, z_T)$.

(c) [10 points] Derive a forward-backward algorithm that computes the "belief state" at time step t given a sequence of observations up to time step T.

(d) [10 points] Suppose that instead of the chains being coupled at every time step, the coupling only happens every k time steps (on time step 0, k $2k$ etc). For $k = 1$, you get the same model as above. If k is fairly large compared to the length of sequences, the chains are called **loosely coupled**. Describe how your model and the inference algorithms change in this case.

(e) [10 points] Suppose that you observe several sequences of two time series and you know that they come from a loosely coupled HMM; you know the number of possible states for each individual chain, but you do not know k. Describe a learning algorithm for this problem.