COMP 652: Assignment 3

Due on Tuesday, March 31 2015

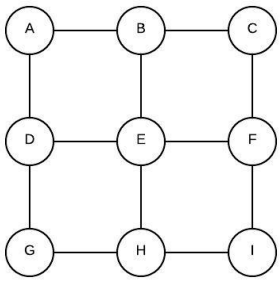
Presented to Dr. Doina Precup

Geoffrey Stanley
Student ID: 260645907

Question 1

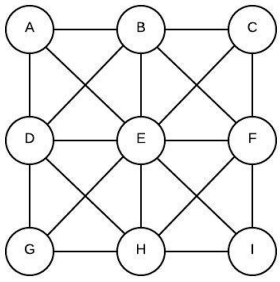
A)

In the 4-neighbor spin glass model the maximal cliques were the edges between each pixel. In an 8-neighbor spin glass model the maximal cliques become clusters of 4 pixels. As such, parametization for a 4 neighbor spin glass model will be as follows:



$$P(E) = \psi(B, E)\psi(D, E)\psi(E, F)\psi(E, H) \tag{1}$$

While the parameters for an 8 neighbor spin glass model will be as such:



$$P(E) = \psi(A, B, D, E)\psi(B, C, E, F)\psi(D, E, G, H)\psi(E, F, H, I) \tag{2}$$

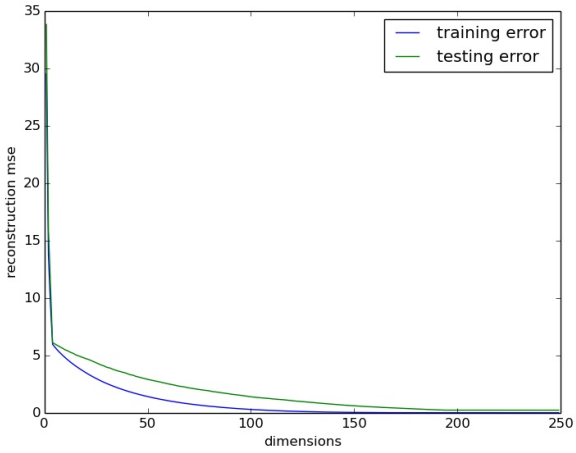
B)

The advantages and disadvantages would be related to a trade off between model precision and computation time.

With the 8 neighbor model more data will be used to infer the value a pixel, this could improve the models accuracy.This is at the expense of having to perform more calculations as well as increasing the potential of overfitting the model.

C)

Question 2



As dimensions are reduced from 250 the reconstruction error is initially quite small but becomes more important as dimensions approach 0. The shoulder of the reconstruction error line is at a dimension of approximately 25.

From this, we can conclude that given this input space we can use PCA to reduce dimensionality of the data to 25 in order to make computation more manageable without losing very much granularity in the feature data.

Question 3

A)

As with standard Hidden Markov Models, Coupled Hidden Markov models will have three categories of parameters. These are the initial probabilities, the transition probabilites and the emission probabilities. Given the system depicted in Figure 1 of assignment 3:

Initial Probabilities:

$$P(s0) \tag{3}$$

$$P(u0) \tag{4}$$

Transition Probabilities:

$$P(s_i|s_{i-1}, u_{i-1}) \tag{5}$$

$$P(u_i|u_{i-1}, s_{i-1}) \tag{6}$$

Emission Probabilities:

$$P(y_i|s_i) \tag{7}$$

$$P(z_i|u_i) \tag{8}$$

B)

In order to compute the joint probability of a sequence of observations a forward algorithm will need to be derived.

$$\begin{aligned}\alpha_t(s_t, u_t) &= P(s_t, u_t, y_{0:T}, x_{0:T}) \\ &= \sum_{t=1}^{T} p(s_t, s_{t-1}, u_t, u_{t-1}, y_{1:T}, x_{1:T}) \\ &= \sum_{t=1}^{T} p(y_t|s_t)p(x_t|u_t)p(s_t|s_{t-1}, u_{t-1})p(u_t|s_{t-1}, u_{t-1})p(s_{t-1}, u_{t-1}, y_{1:T-1}, x_{1:T-1}) \\ &= \sum_{t=1}^{T} p(y_t|s_t)p(x_t|u_t)p(s_t|s_{t-1}, u_{t-1})p(u_t|s_{t-1}, u_{t-1})\alpha_{t-1}(y_{t-1}, x_{t-1})\end{aligned} \tag{9}$$

Because the equation listed above does not contain the initial probability we need to compute it seperately:

$$\alpha_0(s_0, u_0) = p(y_0, z_0, s_0, u_0) = p(s_0)p(y_0|s_0)p(u_0)p(z_0|u_0) \tag{10}$$

Summing the equations above we obtain the joint probability:

$$p(y_0, z_0, y_1, z_1, ..., y_T, z_T) = \sum_{i=0}^{T} \alpha_i(s_i, u_i) \tag{11}$$

C)

The forward-backward algorithm is equal to the product between the forward and the backward algorithm. In the previous question I derived the forward algorithm. So, what is remaining is to derive the backward algorithm:

$$\begin{aligned}\beta_t(s_t, u_t) &= p(y_{t+1:n}|s_t, u_t)p(z_{t+1:n}|s_t, u_t) \\ &= \sum_{k=0}^{T-1} p(y_{t+1:n}, z_{t+1:n}, s_{t+1}, u_{t+1}|s_t, u_t) \\ &= \sum_{k=0}^{T-1} p(y_{t+2:n}|s_{t+1}, u_{t+1})p(z_{t+2:n}|s_{t+1}, u_{t+1})p(y_{t+1}|s_{t+1})p(z_{t+1}|u_{t+1})p(s_{t+1}|s_t, u_t)p(u_{t+1}|s_t, u_t) \\ &= \sum_{k=0}^{T-1} \beta_{t+1}(s_{t+1}, u_{t+1})p(y_{t+1}|s_{t+1})p(z_{t+1}|u_{t+1})p(s_{t+1}|s_t, u_t)p(u_{t+1}|s_t, u_t)\end{aligned} \tag{12}$$

The probability at time T will be equal to 1 :

$$\beta_T(s_T, u_T) = 1 \tag{13}$$

Summing the equations above we obtain te backward algorithm:

$$p(y_{t+1:n}|s_t, u_t)p(z_{t+1:n}|s_t, u_t) = \sum_{i=0}^{T} \beta_i(s_i, u_i) \tag{14}$$

We can now derive the forward-backward algorithm as the product between the forward and the backward algorithm:

$$p(s_t, u_t|y_{0:T}, x_{0:T}) = \left(\sum_{i=0}^{T} \alpha_i(s_i, u_i)\right)\left(\sum_{i=0}^{T} \beta_i(s_i, u_i)\right) \tag{15}$$

D)

The transition probabilities as well as the inference algorithms used will be conditional on whether $t \mod k = 0$. More precisely, for transition probabilities:

$$\begin{cases} P(s_i|s_{i-1}, u_{i-1}) & \text{if } i \mod k = 0 \\ P(s_i|s_{i-1}) & \text{otherwise} \end{cases} \tag{16}$$

$$\begin{cases} P(u_i|u_{i-1}, s_{i-1}) & \text{if } i \mod k = 0 \\ P(u_i|u_{i-1}) & \text{otherwise} \end{cases} \tag{17}$$

And the computation for the inference algorithms will be similar. For the forward algorithm:

$$\begin{cases} \sum_{t=1}^{T} p(y_t|s_t)p(x_t|u_t)p(s_t|s_{t-1}, u_{t-1})p(u_t|s_{t-1}, u_{t-1})\alpha_{t-1}(y_{t-1}, x_{t-1}) & \text{if } t \mod k = 0 \\ \sum_{t=1}^{T} p(y_t|s_t)p(x_t|u_t)p(s_t|s_{t-1})p(u_t|u_{t-1})\alpha_{t-1}(y_{t-1}, x_{t-1}) & \text{otherwise} \end{cases} \tag{18}$$

And for the backward algorithm:

$$\begin{cases} \sum_{t=0}^{T-1} \beta_{t+1}(s_{t+1}, u_{t+1})p(y_{t+1}|s_{t+1})p(z_{t+1}|u_{t+1})p(s_{t+1}|s_t, u_t)p(u_{t+1}|s_t, u_t) & \text{if } i \mod k = 0 \\ \sum_{t=0}^{T-1} \beta_{t+1}(s_{t+1}, u_{t+1})p(y_{t+1}|s_{t+1})p(z_{t+1}|u_{t+1})p(s_{t+1}|s_t)p(u_{t+1}|u_t) & \text{otherwise} \end{cases} \tag{19}$$

E)

The object of the learning algorithm should be to find the K whose joint probability is the greatest. As such, for each K in the the range 0 to n The joint probability will have to be calculated. Once this is done, the K whose joint probability is the higest should be used as the most accurate model.