Assignment 3: Comparing Models

Unsupervised Learning

Zoubin Ghahramani

Due: Mon Nov 15, 2004

1. Download the data file called `geyser.txt` from the course web site. This is a sequence of 295 consecutive measurements of two variables from Old Faithful geyser in Yellowstone National Park: the duration of the current eruption in minutes (to nearest 0.1 minute), and the waiting time until the next eruption in minutes (to nearest minute).

   25 % Examine the data by plotting the variables within and between consecutive time steps. E.g. `plot(geyser(1:end-1,1),geyser(2:end,1),'o');`. Discuss and justify based on your observations what kind of model might be most appropriate for this data set: e.g. a mixture of Gaussians, a hidden Markov model, a linear dynamical system, etc.

2. Download the data file called `data1.txt` from the course web site. This is a data set of 100 observations of 3-dimensional vectors. Examine this data set and answer the following questions. Clearly justify your answers with arguments and plots. Unjustified answers will not get credit.

   10 % Is this a time series? Justify your answer.

   15 % What kind of probabilitic model would you use to model this data? Why? Some choices are probabilistic PCA, factor analysis, mixtures of Gaussians, hidden Markov models, linear dynamical systems, nonlinear dynamical systems. Support your answer with plots and arguments.

3. Consider a data set consisting of the following string of 160 symbols from the alphabet {`A`, `B`, `C`}:

   ```
   AABBBACABBBACAAAAAAAAABBBACAAAABACAAAAAAABBBBACAAAAAAAA
   AAAABACABACAABBACAAABBBBACAAABACAAAABACAABACAAABBACAAAA
   BBBBACABBACAAAAAABACABACAAABACAABBBACAAAABACABBACA
   ```

   15 % Look *carefully* at the above string. Having analysed the string, describe an HMM model for it. Your description should include the number of states in the HMM, the transition matrix including the values of the elements of the matrix, the emission matrix including the values of its elements, and the intial state probabilities. You need to provide some description/justification for how you arrived at these numbers. I am **not** expecting you to code the HMM algorithm—you should be able to answer this question just by examining the sequence carefully.

   10 % Do the same thing (build and justify an HMM model) for the following string of 60 symbols from the alphabet {`A`, `B`, `C`, `D`, `E`, `F`}:

   ```
   CEFBAFCEFBAFBAFCEFCADCAFBEFCEFBEFBEFCAFCEDCAFBEDCADBAFCAFCEF
   ```

4. In the automatic speech recognition community, HMMs are sometimes trained by using the Viterbi algorithm instead of the forward–backward algorithm. In other words, in the E step of EM (Baum–Welch), instead of computing the expected sufficient statistics from the posterior distribution over hidden states: $p(\mathbf{s}_{1:T}|\mathbf{y}_{1:T}, \theta)$, the sufficient statistics are computed using the single *most probable* hidden state sequence: $\mathbf{s}^*_{1:T} = \arg\max_{\mathbf{s}_{1:T}} p(\mathbf{s}_{1:T}|\mathbf{y}_{1:T}, \theta)$.

   15 % Is this algorithm guaranteed to converge? To answer this you might want to consider the proof for the EM algorithm and what happens if we constrain $q(s)$ to put all its mass on one setting of the hidden variables. Support your arguments.

   10 % If it converges, will it converge to a maximum of the likelihood? If not, will it oscillate? Support your arguments.

5. (**bonus** +10 %) Consider the following nonlinear system:

$$\begin{aligned} z_{t+1} &= 0.9\, z_t - \frac{z_{t-1}^2}{2(z_{t-1}^2 + 1)} + e \\ y_t &= z_t + v \end{aligned}$$

   where $e \sim N(0, 1)$ and $v \sim N(0, 0.1)$. Using Matlab, generate 300 data points from this system starting with $z_1 = z_2 = 0$.

   Write it as a nonlinear state-space model: i.e. using a state vector $\mathbf{x}_t$ which depends only on the previous state $\mathbf{x}_{t-1}$ and noise; and with y_t depending on $\mathbf{x}_t$ and noise.

6. (**bonus** +5 %) With enough state variables, can you model the dynamics in the above system perfectly with a linear dynamical system? Argue why or why not.

7. (**bonus** +5%) Fred says that if you give him a linear-Gaussian state-space model he can convert it into an equivalent one with the same number of states but with the covariance of the state-noise $Q = I$, the identity matrix. Is he right? Justify your answer.