Problem Set 5

MAS 622J/1.126J: Pattern Recognition and Analysis

Due Monday, 6 November 2006

[Note: All instructions to plot data or write a program should be carried out using either Python accompanied by the `matplotlib` package or Matlab. Feel free to use either or both, but in order to maintain a reasonable level of consistency and simplicity we ask that you do not use other software tools.]

Problem 1: Parameter Learning by Estimation and Maximization

Consider data, $D = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ * \end{pmatrix} \right\}$, sampled from a two-dimensionaal (separable) distribution, $p(x_1, x_2) = p(x_1)p(x_2)$, with

$$p(x_1) = \begin{cases} \frac{1}{\theta_1} e^{-x_1/\theta_1} & \text{if } x_1 \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad p(x_2) = \begin{cases} \frac{1}{\theta_2} & \text{if } 0 \leq x_2 \leq \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

and a missing feature value, $*$.

a Start with an initial estimate, $\underline{\theta}^0 = \begin{pmatrix} 5 \\ 7 \end{pmatrix}$, and analytically calculate the estimate, $Q(\underline{\theta}, \underline{\theta}^0)$—the *estimate* step of the EM algorithm.

b Find the $\underline{\theta}$ that maximizes your $Q(\underline{\theta}, \underline{\theta}^0)$—the *maximization* step of the EM algorithm.

Problem 2: Baum-Welch algorithm and discrete HMMs

Download the datasets from the course webpage. The datasets consist of training and testing sequences belonging to two classes. Implement the Baum-Welch algorithm for training a discrete HMM.

a Train two fully connected HMMs each with **one*** hidden node (one HMM for each class of data) and transition probabilities.

i Implement the Viterbi algorithm to decode each test sequence using both HMMs. Show the log probability of each test sequence using each HMM.

ii Compute the recognition accuracy on the entire test set.

iii List the output probabilities and state transition probabilities of each HMM.

iv State the threshold you are using and the maximum number of iterations.

v Include a complete listing of your source code.

b Repeat this problem (a) replacing **one*** with **three**.

c Repeat this problem (a) replacing **one*** with **five**.