GMM and speaker recognition

(1) What is a multivariate Gaussian probability density function (PDF)? Your definition should be formal and include the relevant mathematical expressions.

(2) State Bayes' theorem in terms of observation vector y and speaker model S and explain its role in speaker recognition.

(3) What is a Gaussian Mixture Model (GMM)? As in part (1), your definition should be formal and include the relevant mathematical expressions.

(7) Consider a sequence of T one-dimentional data values $y=y_1,\ldots,y_T$ to be modeled by Gaussian distribution. Describe the principle of estimation of the parameters of a Gaussian distribution by using the maximum-likelihood technique. Derive the mathematical formulas for estimation of the parameters of a Gaussian distribution.

(8) Consider a sequence of one-dimensional data Y = y_1, y_2, y_3, y_4 be modeled by a two-component GMM. Starting from the following initial GMM parameters and training data Y = 2,1,-1,-2 .

P(Y|GMM component 1)= N(1,1) , P(Y|GMM component 2)= N(-1,1), P(C1)=P(C2)=0.5.

(i) Fill in the following table by calculating the posterior probabilities of each Gaussian component given training data points Y.

	y_1	y_2	y_3	y_4
$P(C_1\|Y)$				
$P(C_2\|Y)$				

(ii) Use EM algorithm to re-estimate the GMM parameters (one iteration). Hint: use posterior probability matrix in part (i) above.

(10) In a simple speaker recognition system, a speaker S is represented by a 4-component Gaussian mixture PDF b_{SD} given by:

$$b_{SD}(y_t) = 0.2b_1(y_t) + 0.3b_2(y_t) + 0.1b_3(y_t) + 0.4b_4(y_t)$$

Where b_1, b_2, b_3, b_4, are all multivariate Gaussian PDFs and y_t is an acoustic vector. Let $Y = \{y_1, y_2, y_3\}$ be a sequence of acoustic feature vectors which represents a sample of speech which is claimed to have been spoken by speaker S. The probability for each vector and each PDF is given in the following table:

	b_1	b_2	b_3	b_4
y_1	0.03	0.07	0.04	0.12
y_2	0.02	0.06	0.05	0.09
y_3	0.03	0.06	0.03	0.11

Table 1

(i) Write down the class conditional probability of the data Y given the speaker S.

(ii) Assuming the prior probability of speaker S is 0.1 and assuming that the system will accept a speaker if the posterior probability of speaker S given the data Y is greater than 0.5, will this speaker be accepted or rejected? Justify your answer.

HMM

A Hidden Markov Model (HMM) M has four emitting states: s_1, s_2, s_3, and s_4 each of which can emit the symbol a, b or c. The state output probabilities for these states are given by:

$$p(a \mid s_1) = 0.6 \quad p(b \mid s_1) = 0.2 \quad p(c \mid s_1) = 0.2$$
$$p(a \mid s_2) = 0.2 \quad p(b \mid s_2) = 0.7 \quad p(c \mid s_2) = 0.1$$
$$p(a \mid s_3) = 0.2 \quad p(b \mid s_3) = 0.4 \quad p(c \mid s_3) = 0.4$$
$$p(a \mid s_4) = 0.1 \quad p(b \mid s_4) = 0.1 \quad p(c \mid s_4) = 0.8$$

The initial state probability vector is given by $\pi = (1,0,0,0)$ and the state transition probability matrix A is given by:

$$A = \begin{bmatrix} 0.5 & 0.2 & 0.3 & 0 \\ 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Let y be the observation sequence $ababbcc$

(1) Draw a state-time trellis corresponding to the model M and the observation sequence y.

(2) What quantity is computed by using the Viterbi algorithm?

(3) Use the Viterbi algorithm to calculate the probability P(y,S'|M) and the optimal state sequence S'.

(4) What is meant be 'connected speech recognition'? Explain how to construct a connected speech recognition system from an isolated word speech recognition system. (Also draw an example of the state-time trellis diagram).

(5) Consider the sequence of observations Y = $y_1, \ldots, y_T$ to be modeled by an HMM whose states are associated with a single Gaussian distribution with a diagonal covariance matrix. Describe the principle of estimation of the parameters of the Gaussian distributions associated with HMM states. Also give formulas for estimating the parameters of the Gaussian distribution associated with each state, if each observation is associated uniquely with a state.