The Hong Kong Polytechnic University
Department of Electronic and Information Engineering

Programmes	:	BEng(Hons) in Electronic and Information Engineering (42070/42470) BSc(Hons) in Internet and Multimedia Technologies (42077/42477) BEng(Hons) in Electronic Engineering (42079/42479)
Subject	:	Multimodal Human Computer Interaction Technology Subject Code: EIE4105
Session	:	Semester 1, 2015/16
Date	:	4 Dec 2015
Time Allowed	:	2 hours Time : 12:30 p.m. – 14:30 p.m.
Subject Lecturer	:	Dr M. W. Mak

This question paper has 5 pages including this cover page.

Instructions to Candidates : This question paper has FOUR questions. Answer ALL questions. Each question carries 25 marks.

DO NOT TURN OVER THIS PAGE UNTIL YOU ARE TOLD TO DO SO

Q1 Fig. Q1 shows fifteen 2-dimensional vectors, which are to be clustered by a K-means algorithm with $K = 2$.

a) Draw a flow chart to illustrate the procedure of the K-means algorithm.

(7 marks)

b) Assume that vector $\mathbf{x}_1$ shown in the figure is selected as the initial centroids, i.e., $\boldsymbol{\mu}_1^{\text{init}} = \boldsymbol{\mu}_2^{\text{init}} = \mathbf{x}_1$. Assume also that when a training vector $\mathbf{x}_i$ has identical Euclidean distance to both $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, the K-means program will always assign $\mathbf{x}_i$ to Cluster 1 whose centroid is $\boldsymbol{\mu}_1$.

i) Determine the locations of the centroids after the first iteration of the K-means algorithm.

(5 marks)

ii) Determine the locations of the centroids after the second iteration of K-means.

(5 marks)

You may redraw Fig. Q1 on your answer book, indicating the locations of the centroids $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ after the first and the second iterations.

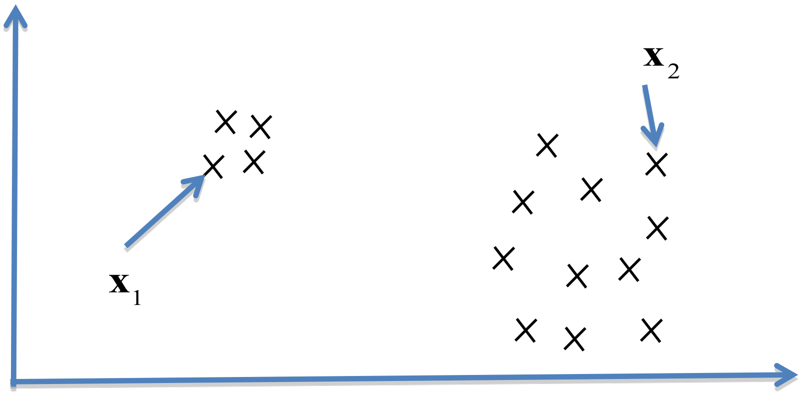


Fig. Q1

c) Assume that a two-mixture Gaussian mixture model (GMM) with parameters $\Lambda = \{\pi_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}_{i=1}^{2}$ is used to approximate the distribution of the samples in Fig. Q1, where $\pi_i, \boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i$ are the mixture coefficient, mean vector, and diagonal covariance matrix of the i-th mixture, respectively.

i) If the expectation-maximization (EM) algorithm is used to estimate the parameters of the GMM, determine the locations of the Gaussian means found by the EM algorithm. You may assume that samples $\mathbf{x}_1$ and $\mathbf{x}_2$ in Fig. Q1 are selected as the initial positions of the Gaussian means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively.

(2 marks)

ii) Determine the value of the mixture coefficient corresponding to the cluster whose mean is $\boldsymbol{\mu}_1$.

(3 marks)

iii) Identify the covariance matrix ($\boldsymbol{\Sigma}_1$ or $\boldsymbol{\Sigma}_2$) whose determinant is larger.

(3 marks)

Q2 (a) Assume that you are given a training dataset comprising the images of *noisy* handwritten digits ('0' – '9') with size 28 x 28 pixels. Assume also that each digit in the dataset has 5,000 samples. Fig. Q2 shows 5 samples of the digit '0'.

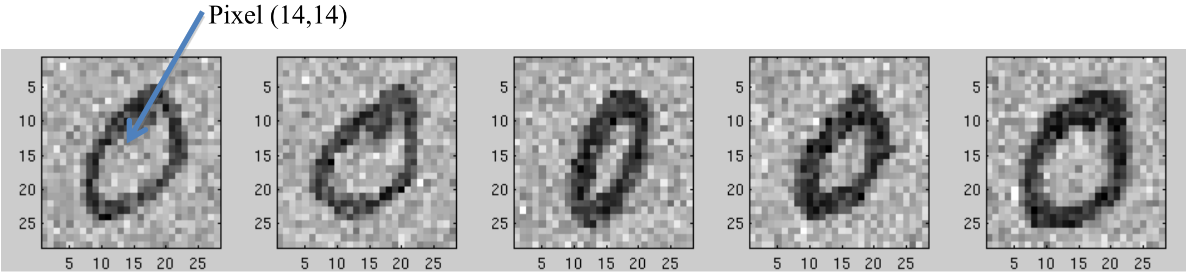


Fig. Q2

i) Assuming that none of the central pixels (14, 14) in the 5,000 training samples contains the strokes of Digit '0', sketch the histogram of the central pixel as indicated by the arrow shown in Fig. Q2. You may ignore the actual values of the x- and y-axes of the histogram.

(4 marks)

ii) Explain how you would train a Gaussian classifier for classifying the ten digits.

(4 marks)

iii) Given the prior information that even digits (including '0') occur twice as often as odd digits, how would you use the Gaussian classifier in Q2(a)(ii) to classify a test image $\mathbf{x}$ of size 28 x 28 pixels? Your answer may include the prior probabilities of even and odd digits and the likelihoods of $\mathbf{x}$ given the Gaussian model of each digit, i.e., $\{p(\mathbf{x} \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\}_{i=0}^{9}$.

(8 marks)

iv) If the number of training samples per digit reduces to 100, can the Gaussian classifier still be able to recognize the ten digits? Briefly explain your answer.

(4 marks)

(b) A Gaussian classifier has been developed to classify $\mathbf{x} = [x_1 \;\; x_2]^T$ into two classes: Class 1 and Class 2. The means ($\boldsymbol{\mu}_k$) and covariance matrices ($\boldsymbol{\Sigma}_k$) in the Gaussian classifier are as follows:

$$\text{Class 1: } \boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } \boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\text{Class 2: } \boldsymbol{\mu}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Assume that both classes have equal prior probabilities, i.e., $P(C_1) = P(C_2)$. Mark the locations of the class means and sketch the decision boundary of the two classes. You may assume that the horizontal and vertical axes in your diagram correspond to x_1 and x_2, respectively.

(5 marks)

Q3(a) Assume that you are given a training dataset comprising the facial images of ten persons, where each person has 20 facial images with different poses. Assume also that each image has size 360 x 260 pixels.

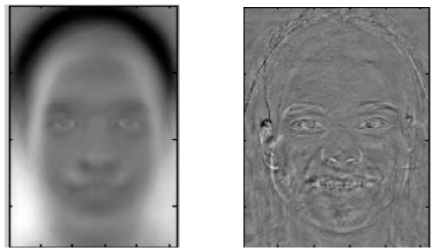
i) Describe how you would build a support vector machine (SVM) classifier to recognize these ten persons based on the raw pixels of their facial images. You may draw a diagram to illustrate the architecture of the SVM classifier.

(10 marks)

ii) Select the appropriate kernel (linear, polynomial or RBF) for your classifier. Explain your selection.

(5 marks)

(b) Fig. Q3 shows two of the eigenfaces obtained from the training set in Q3(a). These eigenfaces have the largest and the smallest eigenvalues among all of the eigenfaces that can possibly be obtained from this dataset.



(a) (b)

Fig. Q3

i) What is the maximum number of eigenfaces (with non-zero eigenvalues) that you can possibly obtain from this dataset?

(2 marks)

ii) Which of the images in Fig. Q3 corresponds to the eigenface with a larger eigenvalue? Briefly explain your answer.

(4 marks)

iii) Describe how the top-five eigenfaces (eigenvectors with the top-5 eigenvalues) can be used for face recognition.

(4 marks)

Q4(a) Restricted Boltzmann machines (RBMs) are important elements of deep neural networks.

i) What is the purpose of RBMs in training deep neural networks (DNNs) and how RBMs achieve this purpose?

(6 marks)

ii) Why RBMs are important for DNNs?

(4 marks)

(b) Fig. Q4 shows the architecture of the DNN-HMM hybrid for continuous speech recognition. Given that the outputs of a DNN can be considered as the posterior probability of phone states $\{s_1, s_2, \ldots, s_K\}$ of the HMM, describe how the outputs of the DNN can be used to compute the likelihood of the acoustic sequence X. ***Hints:*** use the Bayes' rule.

(5 marks)

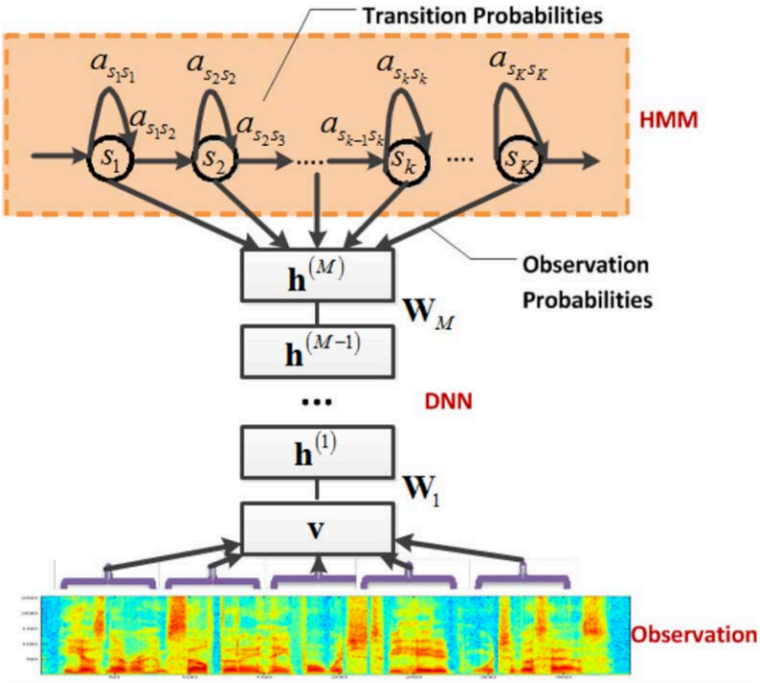


Fig. Q4

(c) The universal background model (UBM) plays an important role in classical GMM-UBM speaker verification and the state-of-the-art iVector/PLDA speaker verification.

i) Describe the data and identify the algorithm for training the UBM.

(4 marks)

ii) Given a sequence of acoustic vectors $X^{(s)}$ of an enrollment utterance from a client speaker s, we may construct his/her Gaussian mixture model (GMM) using the maximum a posteriori (MAP) adaptation as follows:

$$\boldsymbol{\mu}_j^{(s)} = \boldsymbol{\alpha}_j E_j\{X^{(s)}\} + (1 - \boldsymbol{\alpha}_j)\boldsymbol{\mu}_j^{(\text{ubm})}, \qquad j = 1, \ldots, M$$

where M is the number of mixtures, and $E_j\{X^{(s)}\}$ and $\boldsymbol{\mu}_j^{(\text{ubm})}$ are the sufficient statistic and mean vector for the j-th Gaussian of the UBM, respectively. Discuss the value of $\boldsymbol{\alpha}_j$ when the enrollment utterance is very long and when the enrollment utterance is very short.

(6 marks)

-- END OF PAPER --