

- Q1 (a) Assume that you are going to construct a Gaussian classifier for identifying the gender from facial images. Assume also that the covariance matrices in the classifier are diagonal. What will be the minimum number of training samples per class for a valid classifier? State the requirement(s) for these training samples so that the Gaussian classifier can be used for classification. Briefly explain your answer.

(5 marks)

- (b) Explain why a Gaussian mixture model is not appropriate for modeling the outcomes of rolling a dice.

(5 marks)

- (c) A Gaussian mixture model (GMM) based classifier is trained to classify handwritten digits. The decision is based on the Bayes theorem:

$$P(C_k|\mathbf{x}) = \frac{P(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})},$$

where \mathbf{x} is a vector representation of a test image, C_k denotes the k -th class, $P(C_k)$ is the prior probability of the k -th class, $p(\mathbf{x}|C_k)$ is the conditional likelihood, and $p(\mathbf{x})$ is the marginal likelihood.

- (i) How many GMMs are required for the classifier? How would you make use of these GMMs and the Bayes theorem to perform the classification of handwritten digits? Explain your answer.

(4 marks)

- (ii) Explain the role played by $P(C_k)$ in the classifier.

(6 marks)

- (iii) Explain the role played by $p(\mathbf{x})$. Is it necessary to compute $p(\mathbf{x})$ to make a classification decision? Briefly explain your answer.

(5 marks)

- Q2 (a) Discuss the situations in which linear support vector machines (SVMs) are more appropriate than nonlinear SVMs for classification. (5 marks)
- (b) Assume that you are given 100 speech recordings with unknown genders and that you want to identify the genders of these recordings. Assume also that you have a database comprising the speech recordings of 2,500 male and 2,500 female speakers. Each recording is represented by a 500-dimensional i-vector. You use the 5,000 i-vectors extracted from the database to train a linear support vector machine (SVM) to classify their genders. After training, you apply the SVM to identify the gender of the 100 recordings with unknown genders. Discuss the appearance in terms of feminineness and masculinity of the support vectors on
- (i) the correct side of the decision boundary and (4 marks)
 - (ii) the wrong side of the decision boundary. (4 marks)
- (c) A dataset comprises 1,000 facial images of size 100×200 . What is the maximum number of valid eigenfaces that could be obtained from this dataset? Briefly explain your answer and state the assumption(s) that you have made. (5 marks)
- (d) You are given a dataset comprising the images of 10,000 handwritten digits, where each digit has 1,000 samples. Assume that the image size is 28×28 . Discuss how you will use linear discriminant analysis (LDA) and support vector machines (SVMs) to construct a handwritten digit recognizer. What are the pros and cons of using LDA in this application? (7 marks)

- Q3 (a) Discuss the purposes of the pooling operation in convolutional neural networks. What will be the consequence if the pooling operation is omitted? (5 marks)
- (b) Discuss the purposes of the convolutional filters in the first hidden layer of a convolutional neural networks. Why do we need more than one convolutional filter in this layer? (5 marks)
- (c) Explain the purposes of the fully connected layers in a convolutional neural network. Suggest the activation function for the output layer and explain why such function is often used in DNNs or CNNs for classification applications. (5 marks)
- (d) Explain why the cross-entropy loss of a deep neural network has multiple local minima. (5 marks)
- (e) In isolated-word speech recognition, we may use one hidden Markov model (HMM) for each English word. Explain why the number of states in the HMMs depends on the word being modeled. How would you determine the number of states for each HMM? (5 marks)

- Q4 (a) Explain why the dimension of an i-vector is independent of the duration of its corresponding utterance. You may use the factor analysis model of the GMM-supervectors and the formulation of i-vectors to answer this question.
(10 marks)
- (b) In biometric authentication, when the false acceptance rate (FAR) rises, the false rejection rate (FRR) will drop. On the other hand, when the FRR rises, the FAR will drop. Explain why there is always a tradeoff between FAR and FRR. How would you set the decision threshold if security is a major concern?
(5 marks)
- (c) Acoustic modeling plays an important role in large vocabulary continuous speech recognition (LVCSR).
- (i) Explain the purpose of the acoustic models in LVCSR. Describe how the phone-specific acoustic models can be used for acoustic modeling if you are given a short utterance comprising a few words.
(6 marks)
- (ii) GMM-HMM and DNN-HMM are two of the common acoustic models for LVCSR. State the advantages of these machine learning models for acoustic modeling.
(4 marks)