Answer any TWO of the following questions

Q4 (a) Solve the following constraint optimization problem:

$$\begin{array}{rl} \max & f(x,y) = x^2 y \\ \text{subject to} & x^2 + 2y^2 = 1. \end{array}$$

State clearly the steps for finding the optimal solution (x^*, y^*).

(10 marks)

(b) Denote $f_{\max}$ as the maximum value of f that meets the constraint in Q4(a). Draw the contour $x^2 y = f_{\max}$ and the contour $x^2 + 2y^2 = 1$ on a 2-D space with x and y as the horizontal and vertical axes, respectively. Draw a marker '$\times$' at the location corresponding to (x^*, y^*).

(8 marks)

(c) The RBF kernel of a support vector machine (SVM) has the form

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left\{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right\},$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are points on the input space and σ is the width of the RBF kernel. Explain how the value of σ affects the curvature of the decision boundaries of the SVM.

(7 marks)

Q5 (a) Denote $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ and $\mathcal{Y} = \{y_1, \ldots, y_N\}$ as zero-mean training data of a linear regression model:

$$y_i = \boldsymbol{\beta}^\mathsf{T}\mathbf{x}_i + \epsilon_i,$$

where $\boldsymbol{\beta}$ is the parameter vector and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and $\mathbf{x}_i \in \mathbb{R}^D$. Denote $\mathbf{X}$ as an $N \times D$ data matrix comprising N training vectors $\mathbf{x}_i^\mathsf{T}$, where $i = 1, \ldots, N$. The least squared estimate of $\boldsymbol{\beta}$ is given by

$$\beta_{\text{LS}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2,$$

where $\mathbf{y} = [y_1\ y_2\ \cdots\ y_N]^\mathsf{T}$ and $\|\cdot\|_2$ is the L_2-norm. Show that the least squared estimate of $\boldsymbol{\beta}$ is the same as the maximum-likelihood estimate of $\boldsymbol{\beta}$.

(13 marks)

(b) Assume that you need to measure the weight of a truck using a weighbridge. To increase the accuracy, you take a reading from the weighbridge every second and repeat the process for a period of time. However, the weighbridge does not have memory and does not have the capability of computing the mean of these readings. As a smart engineer, you would like to estimate the weight of the truck without the need to write down the readings on a paper. Also, you want to limit the number of readings to be taken without scarifying the precision of the estimated weight. To achieve this goal, denote $\hat{x}_{t|t}$ as the estimated weight of the truck at time step t based on the weighbridge readings $\{z_1, z_2, \ldots, z_t\}$. Also denote $\hat{x}_{t|t-1}$ as the previous estimate (prediction) of the true weight based on readings up to z_{t-1}.

(i) Assume that readings have been taken up to time step t. Given $\hat{x}_{t|t} = \frac{1}{t}\sum_{i=1}^{t} z_i$ and $\hat{x}_{t|t-1} = \frac{1}{t-1}\sum_{i=1}^{t-1} z_i$, show that

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \frac{1}{t}\left(z_t - \hat{x}_{t|t-1}\right). \qquad \text{Eq. Q5}$$

(7 marks)

(ii) Show that Eq. Q5 is a special case of the update formula of a Kalman filter. Show also that the variance of the estimate $\hat{x}_{t|t}$ decreases progressively when the number of measurements t increases.

(5 marks)

Q6 (a) Discuss how deep neural networks (DNNs) and convolutional neural networks (CNNs) can be applied in your research discipline or in a discipline that you are aware of. Explain under what situations DNNs and CNNs perform much better than classical methods such as Gaussian mixture models.

(15 marks)

(b) Discuss and explain the situation(s) in which Gaussian mixture models should **not** be used to model the distribution of data.

(5 marks)

(c) Assume that you have 1,000 samples from ten different species of flowers. Assume also that each species has 100 samples and that each sample is represented by a 1,000-dimensional feature vector. Explain why it is a **bad** idea to use linear discriminant analysis (LDA) to reduce the dimension of the feature vectors. What is the maximum dimension of the LDA-project vectors? Briefly explain your answer.

(5 marks)

— END OF PAPER —