EGT2
ENGINEERING TRIPOS PART IIA

Monday 5 May 2025   14.00 to 15.40

Module 3M1

MATHEMATICAL METHODS

Answer all the questions.

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

*Write your candidate number **not** your name on the cover sheet.*

STATIONERY REQUIREMENTS
Write on single-sided paper.

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed.
Engineering Data Books.

10 minutes reading time is allowed for this paper at the start of the exam.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

You may not remove any stationery from the Examination Room.

1 For $x_d \in [0, 1]$, one of D independent measurements taking values in the unit interval, the D-dimensional column vector $\mathbf{x} = [x_1, x_2, \cdots, x_d, \cdots, x_D]^\mathsf{T}$ can be formed. The D-dimensional column vector $\mathbf{w} = [w_1, \cdots, w_D]^\mathsf{T} \in \mathbb{R}^D$ represents the D parameters of a single layer neural network that models data $y \in \mathbb{R}$ such that the approximation of the data is denoted as $\hat{y}(\mathbf{x}, \mathbf{w})$ where the measurements are nonlinearly transformed by the function $\phi(\cdot)$,

$$\hat{y}(\mathbf{x}, \mathbf{w}) = \sum_{d=1}^{D} w_d \phi(x_d)$$

The Integrated Squared Error (ISE) is defined as

$$\int_0^1 \int_0^1 \cdots \int_0^1 |\hat{y}(\mathbf{x}, \mathbf{w}) - y|^2 dx_1 dx_2 \cdots dx_D$$

(a) By defining the elements of the $D \times D$ matrix $\mathbf{C}$, and $D \times 1$ column vector $\mathbf{b}$, prove that the weights $\mathbf{w}$ that minimise the ISE can be identified by minimising the function of the weights

$$f(\mathbf{w}) = \frac{1}{2}\mathbf{w}^\mathsf{T}\mathbf{C}\mathbf{w} - \mathbf{w}^\mathsf{T}\mathbf{b}$$

[20%]

(b) For the case where $D = 2$ and $\phi(\cdot)$ is the identity function such that the neural network approximation is $\hat{y}(\mathbf{x}, \mathbf{w}) = \sum_{d=1}^{2} w_d x_d$, verify that the ISE is globally minimised when the weights take the unique values $w_1 = w_2 = \frac{6y}{7}$. [20%]

(c) Consider the addition of a nonlinear term to the function $f(\mathbf{w})$ such that

$$\varphi(\mathbf{w}) = f(\mathbf{w}) + \mathbf{1}^\mathsf{T}\mathbf{g}(\mathbf{w})$$

where the D-dimensional vector of ones is denoted as $\mathbf{1} = [1, 1, \cdots, 1]^\mathsf{T}$ and $\mathbf{g}(\mathbf{w})$ is an element-wise application of the function $g(\cdot)$ which acts on each component of the vector $\mathbf{w}$ such that $\mathbf{g}(\mathbf{w}) = [g(w_1), g(w_2), \cdots, g(w_D)]^\mathsf{T}$. Derive and justify the necessary and sufficient conditions for a point $\mathbf{w}^* \in \mathbb{R}^D$ to be a strong local minimum of the function $\varphi(\mathbf{w})$. [30%]

(d) For the specific case where $D = 2$, $\phi(\cdot)$ is the identity function, and the nonlinear term is the quartic polynomial, i.e. $g(w_d) = w_d^4$, assess whether $\varphi(\mathbf{w})$ is convex in $\mathbb{R}^2$ and state the implication on the nature of the point $\mathbf{w}^*$. [30%]

2 (a) Consider the family of matrix operator norms

$$\|\boldsymbol{A}\|_{\alpha,\beta} = \max_{\boldsymbol{x} \neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_{\beta}}{\|\boldsymbol{x}\|_{\alpha}}.$$

(i) For a matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$, derive expressions for the $\|\boldsymbol{A}\|_{1,1}$ and $\|\boldsymbol{A}\|_{\infty,\infty}$ norms. [20%]

(ii) Show that $\|\boldsymbol{A}\boldsymbol{B}\|_{\alpha,\beta} \leq \|\boldsymbol{A}\|_{\gamma,\beta}\|\boldsymbol{B}\|_{\alpha,\gamma}$ for any γ. [20%]

(b) For a matrix $\boldsymbol{B}$, if $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$ and $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$, what can you infer about $\boldsymbol{B}\boldsymbol{u}$? [10%]

(c) A matrix $\boldsymbol{P} \in \mathbb{C}^{n \times n}$ is an *orthogonal projection matrix* if $\boldsymbol{P}^2 = \boldsymbol{P}$ (i.e., $\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}$) and $\boldsymbol{P} = \boldsymbol{P}^H$.

(i) Comment on the eigenvalues of $\boldsymbol{P}$ and give possible values for the condition number $\kappa_2(\boldsymbol{P})$. [10%]

(ii) Show that $(\boldsymbol{P}\boldsymbol{x})^H(\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y}) = 0$ and comment on the significance of this result. [10%]

(iii) Using the Cauchy–Schwarz inequality, $|\boldsymbol{u}^H\boldsymbol{v}| \leq \|\boldsymbol{u}\|_2\|\boldsymbol{v}\|_2$, show that $\|\boldsymbol{P}\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_2$. [10%]

(iv) The 'centering' matrix $\boldsymbol{C}$ is a matrix that when applied to a square matrix $\boldsymbol{A}$ subtracts from each column of $\boldsymbol{A}$ the mean of the column. Give an expression for $\boldsymbol{C}$, show that it is an orthogonal projection matrix, and give its rank. [20%]

3 A simple five-stage manufacturing process can be described by a Markov process. At each time, an item is placed in State 1, the start of the process, just before each transition occurs. Completed items are then kept in State 5, the end of the process. Fig. 1 shows the state-diagram for the process.

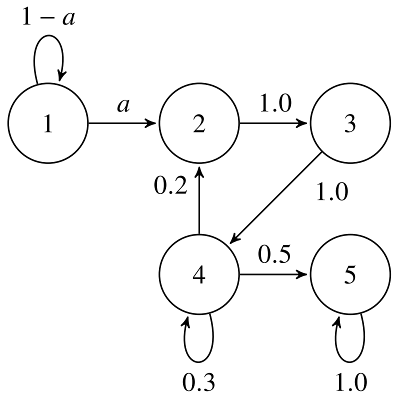

Fig. 1

(a) Write down the transition matrix, **P**, for this process. What is the stationary distribution for this process? [15%]

(b) What is the expected time for an item to first enter State 5? [30%]

(c) Write an expression in terms of **P** for the probability distribution over the states for a single item n steps after it entered the process. [10%]

(d) The process is run for N steps:

(i) Show that the expected number of items in each state can be expressed as

$$\pi(\mathbf{P} - \mathbf{A})\mathbf{B}^{-1}$$

What are π, **A**, and **B**? [30%]

(ii) How does the value of a in Fig. 1 influence the expected number of items in each of the states? [15%]

END OF PAPER