Q.P Code: D154249	Total Pages: 3	Name 795932
		Register No.

FIFTH SEMESTER (CUFYUGP) DEGREE EXAMINATION, NOVEMBER 2026

MATHEMATICS

MAT5EJ304 (2)
Machine Learning - I

2024 Admission Onwards

Maximum Time: 2 Hours	**Maximum Marks: 70**

Section A

All Question can be answered. Each Question carries 3 marks (Ceiling: 24 Marks)

1	Let X and Y be two random variables. Define conditional probability and show that $p(X, Y) = p(Y\|X)p(X)$.
2	State the properties of marginal and conditional distributions of a multivariate Gaussian distribution. Are they also Gaussian? Briefly explain.
3	Explain the connection between maximum likelihood and least squares.
4	Explain the difference between the parameter distribution and the predictive distribution in Bayesian linear regression.
5	Define the model evidence or marginal likelihood $p(\mathbf{t}\|M)$. Explain how it is obtained by integrating over the model parameters.
6	Consider the two-class linear discriminant function $y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$. Define the decision boundary and explain the roles of $\mathbf{w}$ and w_0.
7	Define the within-class scatter matrix $\mathbf{S}_W$ and the between-class scatter matrix $\mathbf{S}_B$ for two classes with mean vectors $\mathbf{m}_1$ and $\mathbf{m}_2$.
8	What is the perceptron criterion function? Explain the meaning of the set M of misclassified patterns.
9	Distinguish between batch gradient descent and stochastic gradient descent.
10	If a neural network has W weights and biases, what is the dimension of its Hessian matrix? How many elements does it contain?

Section B

795932

All Question can be answered. Each Question carries 6 marks (Ceiling: 36 Marks)

11	Suppose that the two variables x and z are statistically independent. Show that the mean and variance of their sum satisfies (a) $\mathbb{E}[x+z] = \mathbb{E}[x] + \mathbb{E}[z]$. (b) $\text{var}[x+z] = \text{var}[x] + \text{var}[z]$.
12	Consider a problem in which two boxes contain different proportions of two types of fruit. A box is selected according to a given prior probability and then a fruit is selected from that box. If the selected fruit is known to be an orange, use Bayes' theorem to determine the posterior probability that the fruit came from each box. Clearly identify the prior, likelihood, evidence, and posterior probabilities.
13	Show that, for a linear regression model with Gaussian noise, maximizing the likelihood is equivalent to minimizing the sum-of-squares error function $E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\left\{t_n - \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_n)\right\}^2$. Explain the connection between maximum likelihood and least squares.
14	Consider Bayesian linear regression after observing N data points. Show that the posterior distribution over $\mathbf{w}$ can be used as the prior distribution when a new observation $(\mathbf{x}_{N+1}, t_{N+1})$ is received. Explain why the resulting posterior has the same Gaussian form.
15	Define the within-class scatter matrix $\mathbf{S}_W$ and the between-class scatter matrix $\mathbf{S}_B$ for a two-class problem. Show that Fisher's criterion can be written as $J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{S}_B\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}}$. Explain the significance of the numerator and denominator.
16	Explain the perceptron learning algorithm for a two-class classification problem. State the perceptron criterion $E_P(\mathbf{w}) = -\sum_{n\in M}\mathbf{w}^T\boldsymbol{\phi}_n t_n$ and describe the weight update for a misclassified pattern.
17	Define the Hessian matrix $\mathbf{H}$ of a neural-network error function $E(\mathbf{w})$. If the network has W weights and biases, (a) write the general form of $\mathbf{H}$; (b) state its dimensions; (c) determine the number of elements in $\mathbf{H}$.

18	Consider a neural network with weight vector $\mathbf{w} = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$. The unregularized error is $E(\mathbf{w}) = 10$ and the regularization coefficient is $\lambda = 0.5$. (a) Calculate the weight-decay regularization term $\frac{\lambda}{2}\mathbf{w}^T\mathbf{w}$. (b) Calculate the regularized error $\widetilde{E}(\mathbf{w})$.

795932

Section C

Answer any ONE. Each Question carries 10 marks (1x10=10 Marks)

19	Evaluate the Kullback–Leibler divergence between the two Gaussian distributions $p(x) = \mathcal{N}(x\|\mu, \sigma^2)$ and $q(x) = \mathcal{N}(x\|m, s^2)$. Starting from $D_{\mathrm{KL}}(p\|q) = \int (x) \ln \frac{p(x)}{q(x)}\, dx$, derive the expression for $D_{\mathrm{KL}}(p\|q)$.
20	Consider a feed-forward neural network with a sum-of-squares error function $E_n = \frac{1}{2}\sum_k (y_{nk} - t_{nk})^2$. (a) Define the output-unit error signal $\delta_k = \frac{\partial E_n}{\partial a_k}$. (b) Show that $\delta_k = (y_k - t_k)$ for linear output units. (c) Derive the expression for the derivative of the error with respect to an output-layer weight.

1143