

Indian Statistical Institute
Semester Examination: 2025-2026

M. Tech. (CrS) 2nd year

Subject: Machine Learning for Security

Date: 17/11/2025

Maximum Marks: 100

Time: 3 Hours

Note: Notations used are as explained in the class.

1. Let the domain of the inputs for a learning problem be $\mathcal{X} = \mathbb{R}$. Consider using hypotheses of the following form:

$$h_{\theta}(x) = 1\{\theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_d x^d \geq 0\},$$

and let $\mathcal{H} = \{h_{\theta} : \theta \in \mathbb{R}^{d+1}\}$ be the corresponding hypothesis class. What is the VC dimension of $\mathcal{H}$? Justify your answer. [15]

2. Let a sequence of training examples $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})$ be given and they are linearly separable. Formulate a dual optimization problem for the maximum margin classifier. [15]
3. Let $K : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ be given. Then prove that if K is a valid (Mercer) kernel, then for any $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$, the corresponding kernel matrix is symmetric positive semi-definite. [10]
4. Summarize Backpropagation algorithm in a fully connected feedforward neural network. Derive $\frac{\partial J(\Theta)}{\partial \theta_{ij}^{(l)}}$ and $\delta_i^{(l)}$ to be used in the backpropagation algorithm. [8 + 12 = 20]
5. Illustrate with an example that the cost function $J(\Theta)$ for a multi layer perceptron neural network is not a convex function. Why do we not use the value zero as the initial value of $\theta_{ij}^{(l)}$. [5 + 5 = 10]
6. What is principal components analysis (PCA) and why is it used? Prove that, if we wish to find a 2-dimensional subspace with which to approximate the data (which maximizes the variance of the data), we should choose the first 2 principal eigenvectors of Σ (covariance matrix of the data). [5 + 10 = 15]
7. Summarize k -means clustering algorithm. Is the algorithm guaranteed to converge? [8 + 7 = 15]