CSL 864

	Machine Learning	Major	200 marks	120 minutes

IMPORTANT: Answer Part A (Qs. 1 to 3) on a separate answer BOOK and Part B (Qs. 4 to 7) on a separate one.

Q. 1. A SVM is used to train with M training patterns. The mapping function ϕ transforms an input x to $\phi(x)$. The corresponding Kernel matrix is K, with entries K_{ij}. In the higher dimensional image space, let the centroid of the patterns be denoted by a vector ϕ_M. Determine the magnitude of ϕ_M, i.e. $\|\phi_M\|$

(25 marks)

Q. 2. A self organizing network with N weights uses a modified Oja's rule for updating neuron weights. A set of patterns ξ^k, k = 1, 2, ... is repeatedly presented to the network, where each pattern $\xi^k = \left(\xi_1^k, \xi_2^k, \ldots, \xi_N^k\right)$. The weights are updated using the rule

$$\Delta w_i = \eta\left(V\xi_i - w_i|w|^3\right)$$

Determine the steady state value AND magnitude of the weights.

(25 marks)

Q. 3. Three one dimensional patterns $x^1 = -1$, $x^2 = 0$, and $x^3 = 1$ are drawn from a pattern set. The inputs are mapped to a higher dimensional space, and the corresponding Kernel function is given by $K(x, y) = 1 + 3xy + x^2y^2$.

The kernel matrix corresponding to the training patterns is given by

$$K = \begin{bmatrix} 5 & 1 & -1 \\ 1 & 1 & 1 \\ -1 & 1 & 5 \end{bmatrix}$$

. The eigenvalues of the matrix are given by (0.4384, 4.5616, and 6). The corresponding eigenvectors are given by

$$\begin{bmatrix} 0.2610 \\ 0.9294 \\ 0.2610 \end{bmatrix}, \begin{bmatrix} 0.6572 \\ 0.3690 \\ 0.6572 \end{bmatrix}, \text{ and } \begin{bmatrix} -0.7071 \\ 0 \\ 0.7071 \end{bmatrix}.$$

(a) Determine the projections of the input x = 2 along each of the Kernel Principal Components.

(b) Determine the empirical feature map. Hence determine the image components of x = 2 in the higher dimensional space corresponding to the empirical feature map.

(50 marks)

Q. 4. Consider a concept which is a subset of the d-dimensional space. We have a weak learner which outputs a half-space (i.e., one side of a hyperplane). We use the weak learner to get a strong learner using the boosting algorithm discussed in the class. Give an upper bound on the VC-dimension of the hypothesis produced by the strong learner (the answer may depend on the parameters of the strong learner)? (25 marks)

Q.5. Recall the Johnson-Lindenstraus Lemma. It says that given N points in a d-dimensional space we can project them to a space of dimension $O\left(\frac{\log N}{\varepsilon^2}\right)$ such that distances between every pair of points is preserved within a factor of $(1 \pm \varepsilon)$ with high probability. Now prove the following version of the JL Lemma : we can project the N points to a space of dimension $O\left(\frac{\log N}{\varepsilon^2}\right)$ such that angles between each triplet of points are preserved within a factor of $(1 \pm \varepsilon)$ with high probability. (35 marks)

Q. 6. Consider data D = {(1,1), (3,3), (2, *)} sampled from a two-dimensional distribution $p(x_1, x_2) = p(x_1).p(x_2)$ with $p(x_1) = \begin{cases} \frac{1}{\theta_1}\exp(-x_1/\theta_1) & \text{if } x_1 >= 0 \\ 0 & \text{otherwise} \end{cases}$ and

$p(x_2) = \begin{cases} \frac{1}{\theta_2} & \text{if } 0 <= x_2 <= \theta_2 \\ 0 & \text{otherwise} \end{cases}$. Here * denotes a missing value. We would like to find the parameters of this distribution by the EM method. Start with an initial estimate $(\theta_1, \theta_2) = (2, 4)$ and show calculations for the E-step. Using this, find the next estimate for the parameters by maximizing the expectation of log-likelihood function (M-step). (40 marks)