NO EXCHANGE OF MATERIAL, CALCULATORS, OR PHONES. THIS EXAM HAS QUESTIONS ON 2 PAGES.

ELL 784 Major : 120 mts Marks: 100

Q. 1. Consider a 3 input XOR gate. If the inputs are denoted by X and the output by Y, find

1) $H[X]$ 3

2) $H[Y]$ 1

3) $H[X|Y]$ 2

4) $H[Y|X]$ 0

5) The mutual information $I[X, Y]$. 1

(20 marks)

Q. 2. Figure 1 shows the decision boundary for a SVM based classifier, in the input space $(x1, x2)$. Determine the

1) the locus of support vectors of class 1

2) the locus of support vectors of class (-1)

3) the Kernel function K(y, z) for two vectors y and z in terms of their components.

4) What is the two-sided margin ? Derive the answer.

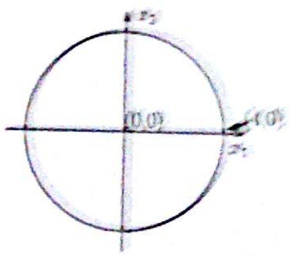

Fig. 1. Locus of support vectors in the input space.

Hint: Think of a simple map ϕ, e.g. a map from 2D to 2D itself. Start by assuming such a map.

(24 marks)

NO EXCHANGE OF MATERIAL, CALCULATORS, OR PHONES. THIS EXAM HAS QUESTIONS ON 2 PAGES.

Q.3. A feedforward network A has W weights. Each weight can assume only m_A values. Network B is identical to network A, except that each weight can assume m_B values. Both networks are trained on a set of N training patterns, so that the training set accuracy is at least 85 %. The probability that ~~loss~~ A (respectively B) makes more than 20% errors on the test set is denoted by p_A (respectively, p_B). The ratio (p_A/pB) is found to be (0.4096). When the number of training patterns for A is doubled to 2N, it is found that p_A is now e^{-10} of its original value. When one of the weights of A is removed, and it is trained with N patterns, p_A falls to 0.25 of its original value (obtained with W weights and N patterns). When one of the weights of B is removed, p_B falls to (1/5)th of its original value.

Find W, mA, mB, and the number of training patterns (N).
Can you estimate the total risk if the confidence level is set at 75% ?

(20 marks)

Q. 4. A single neuron with N weights is presented with patterns from a zero-mean distribution one at a time, and each time, the weights are updated using the rule

$$\Delta w_j = \eta(V\zeta_j - w_j\|w\|^2) \tag{1}$$

$$\text{where } V = \sum_{j=1}^{N} w_j\zeta_j \tag{2}$$

where ζ is the presented pattern, V is the output of the linear neuron, and η is the learning rate.

Derive the expected values of the weights at steady state. Mention any assumptions you make.
What is the norm of the weights at steady state ?
Is there an energy function E so that the steady state weights correspond to a minimum of E ? Derive E if it exists.

(20 marks)

Q. 5. Consider a regression model $y = f_*(x) + \delta$, where δ satisfies $E[\delta] = 0$ and $E[\delta^2] = \sigma^2$. We are given a set of N samples $x^i, i = 1, 2, ..., N$ with corresponding function values y_i at each of them. The K nearest neighbour constructs an estimator

$$\hat{f}_K(x) = \frac{1}{K}\sum_{i=1}^{K} y_{N_i(x)}, \tag{3}$$

where $N_1(x), N_2(x), ..., N_K(x)$ are the K nearest neighbours of x. The performance of the KNN algorithm is given by the expected loss

$$E_{x,y}(y - \hat{f}_K(x))^2 = E_x E_{y|x}(y - \hat{f}_K(x))^2 \tag{4}$$

Let us denote $g(K) = E_{y|x}(y - \hat{f}_K(x))^2$.

Show that $g(K) = \sigma^2 + E_{y|x}(f_*(x) - \hat{f}_K(x))^2$.
Further, obtain a bias-variance decomposition of $E_{y|x}(f_*(x) - \hat{f}_K(x))^2$.
Show that $g(K)$ can be written as an expression involcing only σ, f_*, and K. Derive that expression.

(20 marks)