Department of Computer Science and Engineering
Indian Institute of Technology Madras
CS5691: Pattern Recognition and Machine Learning
Problem Sheet

Date: 14th November, 2023

1. Consider a multilayer feedforward network with two hidden layers. The number of nodes and the activation functions for the nodes in each of the layers are given below. Every node in a layer is connected to all the nodes in the next layer.

Layer	Number of nodes	Node index	Activation function
Input layer	d	i	$\phi_i(a_i) = a_i = x_i$
Hidden layer 1	J	j	$\phi_j^{h_1}(a_j^{h_1}) = tanh(\beta_1 a_j^{h_1})$
Hidden layer 2	K	k	$\phi_k^{h_2}(a_k^{h_2}) = \frac{1}{1+e^{-\beta_2 a_k^{h_2}}}$
Output layer	L	l	$\phi_l^o(a_l^o) = \frac{e^{(a_l^o)}}{\sum_{m=1}^{L} e^{(a_m^o)}}$

Pattern mode of the error back propagation method with a learning rate parameter η is used for training the network. The instantaneous cross-entropy error function is used in learning. For an example belonging to class l, **derive the expressions** for the change of weights Δw_{kl}^o, Δw_{km}^o with $m \neq l$, and $\Delta w_{jk}^{h_2}$, and **give the expression** for the change of weight $\Delta w_{ij}^{h_1}$. Clearly define the terms used in these expressions.

2. Consider the following two separating hyperplanes H_1 and H_2 in a 2-dimensional space:

$$H_1 : 3x_2 - 4x_1 - 6 = 0$$
$$H_2 : 4x_2 - 3x_1 - 8 = 0$$

Let $\mathbf{x}_m = [x_{m1} \quad x_{m2}]^t = [0 \quad 1]^t$ be the nearest training example. Determine the margin for H_1 and H_2. Let H_1^c and H_2^c be the canonical separating hyperplanes of H_1 and H_2 respectively. Give the equations of H_1^c and H_2^c. Consider another training example $\mathbf{x}_n = [x_{n1} \quad x_{n2}]^t = [\quad -3 \quad -1]^t$. Assume that $\mathbf{x}_m$ and $\mathbf{x}_n$ belong to the same class. Determine the value of the slack variable for $\mathbf{x}_n$ for each of the following cases: (i) H_1^c is the optimal soft-margin hyperplane, and (ii) H_2^c is the optimal soft-margin hyperplane.

3. The equation of the optimal canonical separating hyperplane for two linearly non-separable classes in a 3-dimensional space is given by $x_3 + x_2 + 2x_1 - 3 = 0$. For each of the following training examples of the negative class, determine whether it is a support vector or not. If it is a support vector, then determine whether it is a bounded support vector or an unbounded support vector. Justify your answer.

(a) $x_1 = 0, x_2 = 0, x_3 = 0$

(b) $x_1 = 1, x_2 = 1, x_3 = 1$

(c) $x_1 = 2, x_2 = 0, x_3 = 0$

(d) $x_1 = 0.5, x_2 = 0.5, x_3 = 0.5$

(e) $x_1 = 1, x_2 = 1, x_3 = 2$

(f) $x_1 = 0, x_2 = 2, x_3 = 1$

4. The cost function to be minimized in order to find a soft-margin hyperplane for a support vector machine for two linearly non-separable classes is as follows:

$$\frac{1}{2}\mathbf{w}^t\mathbf{w} + \frac{N_1 C}{N}\sum_{n=1}^{N_1}\xi_n + \frac{N_2 C}{N}\sum_{n=N_1+1}^{N}\xi_n$$

The constraints are as follows:

$$\begin{aligned} y_n(\mathbf{w}^t\mathbf{x}_n + w_0) \geq 1 - \xi_n, &\quad n = 1, 2, ..., N \\ \xi_n \geq 0, &\quad n = 1, 2, ..., N \end{aligned}$$

Here C is the trade-off parameter. The terms $\xi_n, n = 1, 2, ..., N$ are the slack variables. The number of training examples for Class 1 and Class 2 is N_1 and N_2, respectively, with $N = N_1 + N_2$.

Give the primal form of the Lagrangian objective function. Obtain the optimality conditions using the derivatives of the primal form of the Lagrangian objective function with respect to the primal variables $\mathbf{w}, w_0$ and ξ_n. Derive the dual form of the Lagrangian objective function and give the associated constraints to be satisfied. Give the KKT conditions. State how the optimal value of w_0 can be determined.

5. Consider a parent node N in a decision tree. Out of 100 training examples that reach the node N, 50 examples belong to class C_1, 30 to class C_2 and 20 to class C_3. The number of examples of each class that reach the left child node N_L and the right child node N_R for each of two queries Q_1 and Q_2 is given in the table below

Query	N_L			N_R		
	C_1	C_2	C_3	C_1	C_2	C_3
Q_1	30	20	10	20	10	10
Q_2	20	10	5	30	20	15

Determine the query that will used to split the node N.