

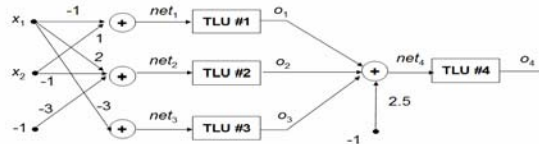
# Neural Networks

## Final Exam

January 16, 2003, 9:10 a.m. - 11:10 a.m.

**Note:** You have to answer the questions with supporting explanations if needed.

- Explain and compare the following three learning rules: (10%)  
 (i) Perceptron Learning Rule, (ii) Delta Learning Rule, (iii) Widrow-Hoff Learning Rule.
- A feedforward network using bipolar discrete neurons is shown in the figure. Please find the segment of the  $x_1, x_2$  plane for which  $o_4=1$ . (10%)



- A minimum-distance linear classifier has the weight and augmented patterns as:  

$$w = [2 \quad -1 \quad 2]^T, \quad y = [x_1 \quad x_2 \quad 1]^T$$
 (i) Find the equation of the decision surface in the pattern space. (5%)  
 (ii) Compute the new weight vector if the two class prototype points are:  

$$x_1 = [2 \quad 5]^T, \quad x_2 = [-1 \quad -3]^T$$
 (iii) Sketch the decision surface for (i) and (ii) (5%)
- A functional link network based on the tensor model has been trained with the extended input vectors:

$$\text{Class 1: } x_1 = [6 \quad 12 \quad 72]^T, \quad x_2 = [10 \quad 8 \quad 80]^T, \quad x_3 = [1 \quad -2 \quad -2]^T, \quad x_4 = [-3 \quad -5 \quad 15]^T,$$

$$\text{Class 2: } x_5 = [8 \quad -8 \quad -64]^T, \quad x_6 = [-7 \quad 5 \quad -35]^T.$$

- Are the training patterns in the original space linearly separable? (5%)
- Show the functional network with weights that satisfy the required classification. (5%)
- The error back-propagation training algorithm has been extensively used in multilayer feedforward networks. What is the deriving criterion for this algorithm? (4%)  
 Write down (and explain) possible methods that can help to speedup or correct the training procedure. (6%)
- Perform three steps of the discrete-time learning algorithm of the ART1 network for  $n=9$  and  $M=3$ , assuming that the training vectors are:

$$x_1 = [1 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 1]^T,$$

$$x_2 = [1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1]^T,$$

$$x_3 = [1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1]^T.$$

Select an appropriate vigilance threshold  $\rho$  such that all patterns are clustered separately, when the sequence  $x_1, x_2$ , and  $x_3$  is presented. (5%) Also, please compute the final weights  $W$  and  $V$ . (10%)

7. Compare the linear associative memory with Hopfield's Autoassociative memory in terms of the inputs and outputs, recording and storage constraints, and scalability. (15%)
8. The principal component analysis (PCA), linear discriminant analysis (LDA), and heteroscedastic discriminant analysis (HDA) are three approaches for discriminative feature extraction or dimension reduction. Try to explain the spirits, merits and drawbacks (or constraints) of these three approaches in your words. (15%) (Heteroscedastic: a set of statistical distributions having different variances)