

Lab assignment: Neural Network

A. Basic assignments:

1. Implement perceptron training algorithm (PTA). Run it on *NAND*, *XOR*, *5-input palindrome*, *5-input majority*, *5-input parity* problems. Observe convergence when it happens. Look for cycle when it does not. Record the convergence time (# iterations) as function of the initialization point.
2. Give perceptrons for recognizing digits 0-9. Assume 7-segment display. Each perceptron K ($K=0\dots9$) outputs 1 when K is input, else outputs 0.
3. Implement backpropagation (BP) on feedforward neural n/w (FFNN). Give FFNNs for all the above problems including digit recognizer. Choose the learning rate judiciously. Study convergence time, local minima, saturation, effect of initialization, effect of learning rate and momentum factor. 1 and 0 decisions are based on the output being above the high water mark or being below the low water mark.

B. Applications:

- a. Give a neural network for recognizing the sentiments of tweets. Download tweets, do feature engineering on them. A naïve feature vector is the set of words in the tweets. Collect all the words in the tweets, sort them, remove duplicates. Each tweet will be represented by a 1/0 vector depending on the presence/absence of the word in the tweet. We will supply you with some sentiment marked tweets. You will have to annotate some. All the annotated tweets will be used by all the groups.
- b. Download any classification benchmark data from ML repositories (Look up, e.g., University of California at Irvine). Train and test FFNN on such data. Of particular note is a classic problem called IRIS data. Read up on the internet. For any classifier IRIS and MONK serve as benchmark data. You should surely show results on these two data sets.
- c. Apply FFNN to IR. Download TREC data sets. Apply the procedure of a. to classify documents into relevant and irrelevant sets.

For part B, it is essential that you perform N-fold cross validation. Typically N is 5. That means you divide your classification data into 5 partitions. Use 4 partitions for training the neural n/w and the remaining for measuring accuracy.