EE6310 Image and Video Processing, Fall 2021

Indian Institute of Technology Hyderabad

Homework 4, Assigned 19.10.2021, Due **11:59 pm on 26.10.2021**

You should be writing this code not because a teacher assigned it to you, you should be writing it because you have a desire to learn, and wanting to learn is the biggest advantage you can have!
– inspired by a quote by Cory Althoff.

1. Implement the back-propagation algorithm to learn the weights of a multi layer perceptron with 2 input nodes, 2 hidden nodes and 1 output node. Train your network to learn the following binary operations: (30)

   (a) XOR

   (b) AND

   (c) OR

   Instructions:

   - Note that the input has two elements: $\mathbf{x} = [x_1, x_2]^T$, i.e., $p = 2$. Each element is a binary digit. The label is one element which is also binary.
   - Like in our analysis in class, use squared error to measure performance.
   - Use the sigmoid non-linearity: $\sigma(x) = \frac{1}{1+e^{-x}}$.
   - Generate training samples from the truth table of these logical operations. Specifically, perturb the inputs x (binary digits) with mild Gaussian noise (noise variance $\sigma^2 \ll 1$). For e.g., in the XOR operation $0 \oplus 1 = 1$. So, when you perturb your input with a small amount of noise you will effectively have a sample point for which $0.01 \oplus 1.02 = 1$. *Add noise only to the input and not the labels*.
   - Generate N samples that will be partitioned into N_{Tr} training samples and $N_{Te} = N - N_{Tr}$ validation (or testing) samples. A general rule of thumb is $N_{Tr} = 0.8N$, $N_{Te} = 0.2N$. Experiment with the number of samples N and see how its affects performance. Specifically, choose $N = 1000, 2500, 5000$.
   - Partition the training set into mini batches of size m to expedite training. When all the mini batches (i.e., all the N_{Tr} training samples) are passed through the network once, it is called an *epoch*. Choose 100 epochs for training the model.
   - Choose a learning rate $\gamma = 0.05$. You are encouraged to try out different learning rates and observe its effect on training. For reporting however, choose $\gamma = 0.05$.
   - For each binary operation (XOR, AND, OR), do the following:
     - Save your training loss (MSE) for each epoch.
     - At the end of each epoch, compute the testing loss by passing the M testing samples through the model.
     - At the end of 100 epochs, plot the training loss and testing loss versus epochs (on the same plot). Use different colours for training and testing loss curves.
     - At the end of 100 epochs, print the parameters of your model.
     - Print your observations in terms of the rate of convergence, overfitting, underfitting, and any other findings.