Machine Learning — *CS 689, Spring 2015*

HW08: Neural networks

Hand in via moodle at: `https://moodle.umass.edu/course/view.php?id=20836`. Remember that only PDF submissions are accepted. We encourage using LaTeX to produce your writeups. See `hw00.tex` for an example of how to do so. You can make a `.pdf` out of the `.tex` by running "`pdflatex hw00.tex`". You'll need mydefs.sty and notes.sty which can be downloaded from the course page.

1. Create a neural network with only one hidden layer (of any number of units) that implements $(A \vee \neg B) \oplus (\neg C \wedge \neg D)$. Draw your network, and show all weights of each unit. For simplicity use the step function as the link function, i.e., $f(x) = 1[x \geq 0]$. Assume that you have access to a bias feature at each layer.

2. Suppose you initialized all the weights to zero in a two layer neural network and ran back-propagation. Explain why this will lead to a network where all the hidden nodes are identical, i.e., produce identical outputs. How can you avoid this problem?

3. The theorem due to Håstad showed that computing the parity function using a two layer network requires nodes that is exponential in the number of inputs N.

   (a) Show that with $\log(N)$ hidden layers one can compute the parity function using a binary tree of XOR nodes. How many nodes are there in this tree?

   (b) Decision trees can also compute boolean functions. What is the smallest depth of a decision tree that can compute the parity function? Explain your answer.