HW07: Neural modeling and kernels

Hand in at: `http://www.cs.utah.edu/~hal/handin.pl?course=cs726`. 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`".

1. Explain how to get a hidden unit to compute an AND function and then explain how to combine your AND function with the book's OR function to get XOR. (I.e., define all the weights.)

2. Consider Algorithm 27. What would you have to change in order to do back-propogation of hinge loss and logisitic loss rather than than squared error?

3. Suppose that you have N data points in D dimensions. Suppose you run perceptron for 1 pass over the data set and it makes K updates. How long does this take? (Big-O notation, please: note, it should not actually depend on K.) How long would it take to run if you preprocessed your data with the quadratic feature map? How long for cubic feature map?

   Now, suppose that you run kernelize perceptron over the same data with a linear kernel. How long will this take (it *should* depend on K now, and note that under different feature maps, the numbers K will not be comparable.) What about for quadratic or cubic kernels?