

## HW06: Probabilistic and neural modeling

Hand in at: <http://www.cs.utah.edu/~hal/handin.pl?course=cs726>. Remember that only PDF submissions are accepted. We encourage using L<sup>A</sup>T<sub>E</sub>X 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. Optimize the following constrained problem using a log transform and the method of Lagrange multipliers (you may assume all  $x$ s are non-negative):

$$\min_{\theta} \prod_d \theta_d^{x_d} \quad \text{subj. to } \|\theta\|_2 = 1 \quad (1)$$

2. What is the difference between assuming that features are independent and assuming that features are independent given the label?
3. Suppose you wanted to derive a probabilistic model for regression that corresponded to *absolute* penalties on predictions (rather than the quadratic penalties in the standard model). What noise distribution would you have to assume (as opposed to the Gaussian assumption)? (Hint: the answer can be found elsewhere in the chapter in a different context.)