

HW05: Neural Networks

Hand in at: <http://www.cs.utah.edu/~hal/handin.pl?course=cs5350>. Remember that only PDF submissions are accepted. We encourage using L^AT_EX to produce your writeups. See `hw01.tex` for an example of how to do so. You can make a `.pdf` out of the `.tex` by running “`pdflatex hw05.tex`”.

1 PRML Exercises

1. *5.6
2. 5.7
3. 5.18 (6350 only)

2 Additional Exercises

1. Consider the soft-max function over a vector $\mathbf{a} = \langle a_1, a_2, \dots, a_K \rangle$ defined by:

$$\text{softmax}(\mathbf{a}, k) = \frac{\exp[a_k]}{\sum_j \exp[a_j]} \quad (1)$$

Take, for example, $\mathbf{a} = \langle -1, 2, 1 \rangle$. Compute the soft-max value for each component of \mathbf{a} (please use a calculator and only give a couple of decimal values for each number). Now, scale \mathbf{a} by some “temperature” T ; that is, $(1/T)\mathbf{a} = \langle -1/T, 2/T, 1/T \rangle$. Compute the soft-max value for each component of $T\mathbf{a}$ for $T \in \{10, 2, 1, 0.5, 0.1, 0.01\}$ to verify that as the the temperature decreases, the soft-max function approaches just the max function. Argue (in words, not math) why this happens and why what’s so special about using something like an exponential in the softmax function.