CS 194-10, Fall 2011
Assignment 4

1. *Linear neural networks*
The purpose of this exercise is to reinforce your understanding of neural networks as mathematical functions that can be analyzed at a level of abstraction above their implementation as a network of computing elements. It also introduces a somewhat surprising property of multilayer linear networks with *small* hidden layers.

For simplicity, we will assume that the activation function is the same linear function at each node: $g(x) = cx + d$. (The argument is the same (only messier) if we allow different c_i and d_i for each node.)

(a) The outputs of the hidden layer are

$$H_j = g\left(\sum_k W_{k,j} I_k\right) = c \sum_k W_{k,j} I_k + d$$

The final outputs are

$$O_i = g\left(\sum_j W_{j,i} H_j\right) = c\left(\sum_j W_{j,i}\left(c \sum_k W_{k,j} I_k + d\right)\right) + d$$

Now we just have to see that this is linear in the inputs:

$$O_i = c^2 \sum_k I_k \sum_j W_{k,j} W_{j,i} + d\left(1 + c \sum_j W_{j,i}\right)$$

Thus we can compute the same function as the two-layer network using just a one-layer perceptron that has weights $W_{k,i} = \sum_j W_{k,j} W_{j,i}$ and an activation function $g(x) = c^2 x + d\left(1 + c \sum_j W_{j,i}\right)$.

(b) The above reduction can be used straightforwardly to reduce an n-layer network to an $(n-1)$-layer network. By induction, the n-layer network can be reduced to a single-layer network. Thus, linear activation functions restrict neural networks to represent only linear functions.

(c) The original network with n input and output nodes and h hidden nodes has $2hn$ weights, whereas the "reduced" network has n^2 weights. When $h \ll n$, the original network has far fewer weights and thus represents the i/o mapping more concisely. Such networks are known to learn much faster than the reduced network; so the idea of using linear activation functions is not without merit.

2. *ML estimation of exponential model*

(a) $$L(\mathbf{X}; b) = \prod_{i=1}^{N} \frac{1}{b} e^{-\frac{x_i}{b}} = b^{-N} e^{-\frac{1}{b}\sum_{i=1}^{N} x_i}$$

(b) $$l(\mathbf{X}; b) = \log(L(\mathbf{X}; b)) = -N \log(b) - \frac{1}{b} \sum_{i=1}^{N} x_i$$

$$\frac{\partial l}{\partial b}(\mathbf{X}; b) = -\frac{N}{b} + \frac{1}{b^2} \sum_{i=1}^{N} x_i$$

(c) $$\frac{\partial l}{\partial b}(\mathbf{X}; b) = 0 \Rightarrow b = \frac{1}{N} \sum_{i=1}^{N} x_i = \bar{x}$$