Franklin Hu, Sunil Pedapudi
CS 194-10 Machine Learning
Assignment 4

1. Linear neural networks

   (a) Suppose we have a three layer linear neural network with one input layer $\mathbf{x}$, one hidden layer $\mathbf{h}$, and one output layer $\mathbf{y}$. Each layer can be expressed as a vector of the values of the nodes in that layer. For example, $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$. Assume that each neural node has its own set of weights w_i where i is the node index. We can express the value of the a particular output in terms of the hidden layer:

   $$y_k = c_k \cdot \mathbf{w_k} \cdot \mathbf{h}$$

   Similarly, we can express the value of each node in the hidden layer in terms of the inputs.

   $$h_j = c_j \cdot \mathbf{w_j} \cdot \mathbf{x}$$

   Now, we can see that the output layer can simply be written in terms of the inputs without the hidden layer.

   $$\begin{aligned}
   \mathbf{y} &= \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} \\
   &= \begin{pmatrix} c_{k_1} \cdot \mathbf{w_{k_1}} \cdot \mathbf{h} \\ c_{k_2} \cdot \mathbf{w_{k_2}} \cdot \mathbf{h} \\ \dots \\ c_{k_n} \cdot \mathbf{w_{k_n}} \cdot \mathbf{h} \end{pmatrix} \\
   &= \begin{pmatrix} c_{k_1} \cdot \mathbf{w_{k_1}} \cdot \begin{pmatrix} c_{j_1} \cdot \mathbf{w_{j_1}} \cdot \mathbf{x} \\ c_{j_2} \cdot \mathbf{w_{j_2}} \cdot \mathbf{x} \\ \dots \\ c_{j_n} \cdot \mathbf{w_{j_n}} \cdot \mathbf{x} \end{pmatrix} \\ c_{k_2} \cdot \mathbf{w_{k_2}} \cdot \begin{pmatrix} c_{j_1} \cdot \mathbf{w_{j_1}} \cdot \mathbf{x} \\ c_{j_2} \cdot \mathbf{w_{j_2}} \cdot \mathbf{x} \\ \dots \\ c_{j_n} \cdot \mathbf{w_{j_n}} \cdot \mathbf{x} \end{pmatrix} \\ \dots \\ c_{k_n} \cdot \mathbf{w_{k_n}} \cdot \begin{pmatrix} c_{j_1} \cdot \mathbf{w_{j_1}} \cdot \mathbf{x} \\ c_{j_2} \cdot \mathbf{w_{j_2}} \cdot \mathbf{x} \\ \dots \\ c_{j_n} \cdot \mathbf{w_{j_n}} \cdot \mathbf{x} \end{pmatrix} \end{pmatrix}
   \end{aligned}$$

2. ML estimation of exponential model

3. ML estimation of noisy-OR model