What observations can you make about the machine's hidden value encodings of the input values?

After 5000 epochs of learning were done, the outputs of the hidden units converged. The neural network was supposed to learn the identity function. Therefore, we expected the hidden units to generate a unique encoding for a unique input. For example, the hidden unit encoding for input 00000001 might be 001. However, we discovered that the hidden unit encodings for the inputs 00001000 and 10000000 were the same. This suggested that our neural network had not been able to learn the identity function. We verified this fact by checking the predictions of the learner after it was fit with the training data.

One time, we tried to use 4 hidden units and the neural network learned the identity function successfully. After going through this procedure, we realized that our neural network did not incorporate bias terms in the edge weights between neurons. We conducted independent research and found that neural networks typically include "hidden" bias nodes (see figure 1). We realized that since our learner functioned well with 4 hidden units, this would probably be one way to fix our learner. Although this may not be the exact procedure discussed in class and specified within the instructions, we would rather implement a learner that correctly learns the target function than one that fails to do so.

After implementing the bias nodes, the learner was actually able to correctly learn the identity function. After 5000 epochs, each input value was matched with a unique encoding of hidden input units. It makes sense that the three hidden units are adequate to represent the identity function in $\mathbb{R}^8$ because three binary bits are sufficient to represent eight different values.

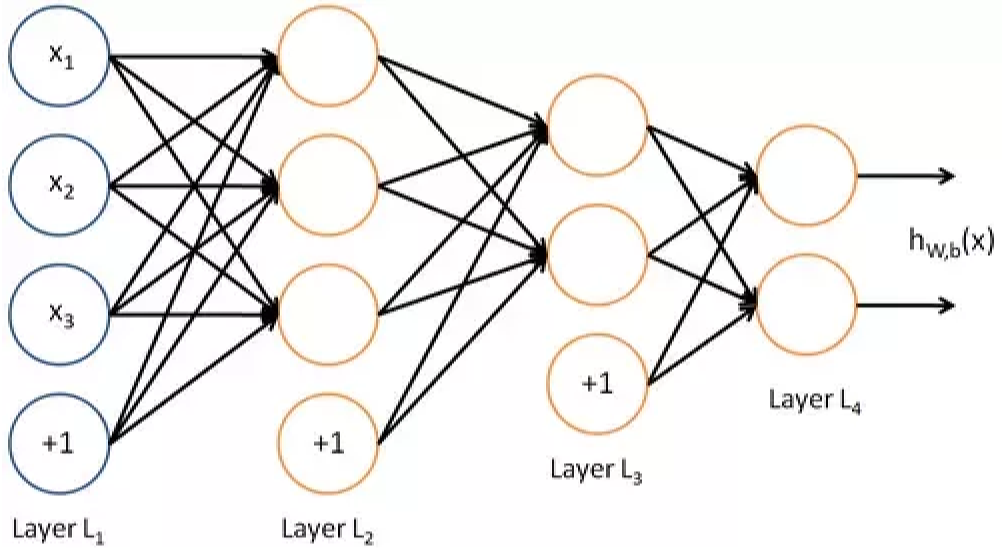

Figure 1: a graphic depiction of an ANN with bias units