1. Derive the modified delta learning rule for the case where there is an output non-linearity of the form $\sigma(u) = 1/(1 + e^{-u})$, with $u = \sum_i w_i x_i$.

2. Download the 2D data apples.mat and oranges.mat from the class website, and train a single neuron to discriminate between the apples and oranges. Try this for both a linear neuron and one with a sigmoidal output nonlinearity. (Use +1/-1 as the category assignments in the linear case, and 1/0 in the non-linear case.) You can use the script lab2s.m on the website to help visualize the convergence to the solution during learning, but you will have to write the code for simulating the network and learning yourself.

3. Now augment the data above with the additional datasets apples2.mat and oranges2.mat. As you can see from plotting out the combined data, the problem of discriminating the apples from the oranges is no longer linearly separable, so we must use a multilayer network for this problem. Train a two-layer network (using backprop) to learn the discrimination. You can use the script lab2m.m on the website to help visualize the solution, as we did in class. Try also adding a momentum term to see if it helps with convergence.

4. Consider the following pattern discrimination task:

```
  1 0 1
  0 1 0
```

First verify that this problem is not linearly separable—i.e., try training a single layer network to see if it can solve the problem. Then try a two-layer network. How many hidden units are needed? What representation is learned by the hidden units in order to solve this problem?