Modeling Lab

This week, we're building linear neural networks using either online training (training where the data is not all known at once) using the Widrow-Hoff rule, or batch training, where the data is all known at once. In the latter case, we can solve the linear equation using either Matlab's slash command or via the SVD (form the pseudoinverse).

Classification Problem 1: The Iris Data

In this example, each vector has 4 sample measurements from one of 3 classes of flower. There are 50 flowers in each class, so as a matrix, the inputs are 150×4.

The "targets" tell us which class the flower is, and we have to build this set. As we did in class, each class is a column of the 3×3 identity matrix (we have 3 classes). Thus, the matrix T will be 150×3.

The document online has the instructions for solving the problem using batch training (Matlab's slash command), so the first part of the homework is to implement that training and analyze the results in the confusion matrix that is defined there.

Classification Problem 1(a): Iris Data with Widrow-Hoff

The second part of the homework is to build the linear neural network using the Widrow-Hoff training. You might experiment a bit with the learning rate, but you might start with something small like $\alpha = 0.01$. Finish the problem by plotting your error (`EpochErr` is the output), and then print the confusion matrix.

Classification Problem 2: Breast Cancer Data

The data is given as an $m-$file named `BreastData.m` (so in the command window, to load the data, type `BreastData` (without the $.m$ suffix).

Each data "point" represents nine measurements taken from a breast exam, and there are six target classes:

- Class 1: Carcinoma
- Class 2: Fibro-adenoma
- Class 3: Mastopathy
- Class 4: Glandular
- Class 5: Connective
- Class 6: Adipose

If you would like more information about this data, see "Variability of impedivity in normal and pathological breast tissue", by J. Jossinet. Med. & Biol. Eng. & Comput, 34: 346-350 (1996).

If you download the Matlab $m-$file (this is NOT a binary data file, but a plain text script), you can see a few notes there. To load the data into Matlab, just type `BreastData`

in the command window (or in a script) to load the matrices X and T. Notice that they may not be given to you with the right dimensions (the data may be in columns or in rows), so be sure to check that and use the dimensions appropriate for your algorithm.

Problems 2(a) and 2(b)

Using this data, construct two script files- One using the Widrow-Hoff update (you can use the `WidHoff.m` function on the class website), and one that finds the weights and biases using batch training. Both scripts should "output" the final confusion matrix (just leave the semicolon off of the last computation).

Think about what the numbers mean in terms of the classes we've described, and give a short summary.

Due: Friday (There may be an additional problem on Wednesday).