

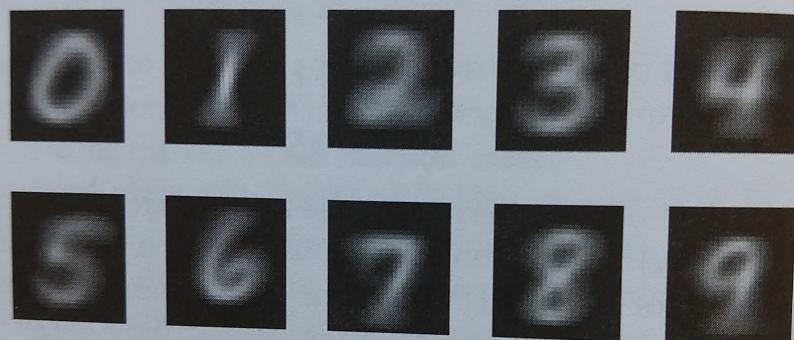
for ranking college football teams and is one factor in determining which teams play in bowl games.)

17. This exercise uses the MNIST database of handwritten digits, which contains a training set of 60,000 numbers and a test set of 10,000 numbers. Each digit in the database was placed in a 28 by 28 grayscale image such that the center of mass of its pixels is at the center of the picture. To load the database, download it from the book's web page and type `load mnist_all.mat` in MATLAB. Type `who` to see the variables containing training digits (`train0,...,train9`) and test digits (`test0,...,test9`). You will find digits intended to train an algorithm to recognize a handwritten 0 in the matrix `train0`, which has 5923 rows and 784 columns. Each row corresponds to one handwritten zero. To visualize the first image in this matrix, type

```
digit = train0(1,:);
digitImage = reshape(digit,28,28);
image(rot90(flipud(digitImage),-1)),
colormap(gray(256)), axis square tight off;
```

Note, the `rot90` and `flipud` commands are used so the digits appear as we write them, which is more noticeable with digits like 2 or 3.

- (a) Create a 10 by 784 matrix `T` whose  $i$ th row contains the *average* pixel values over all the training images of the number  $i - 1$ . For instance, the first row of `T` can be formed by typing `T(1,:) = mean(train0);` Visualize these average digits using the `subplot` command as seen below.



- (b) A simple way to identify a test digit is to compare its pixels to those in each row of `T` and determine which row most closely resembles the test digit. Set `d` to be the first test digit in `test0` by typing `d = double(test0(1,:));`. For each row  $i = 1, \dots, 10$ , compute `norm(T(i,:) - d)`, and determine for which value of  $i$  this is smallest; `d` probably is the digit  $i - 1$ . Try some other digits as well and report on your results.

(A more sophisticated approach called Principal Component Analysis or PCA attempts to identify characteristic properties of each digit, based on the training data, and compares these properties with those of the test digit in order to make an identification.)