CFCS1: Assignment 2

February 16, 2010

Handed out:	10 February 2010
Due date:	22 February 2010

Introduction

This assignment relates to the task introduced in Assignment 1: that of classifying vowel sounds from human speech. This time, we consider just two, similar vowels:

Number	Vowel name	Examples
1	ey	bait, hate
2	ay	bite, height

We have 300 recordings of each vowel. From each recording, we have again extracted the two most relevant features for distinguishing between vowel sounds. To load the samples into Matlab, save the file `a2_data.mat` to your working directory, and use the the command `load a2_data.mat`. This will create two matrix variables, each with 2 rows and 300 columns:

- `samples1`, corresponding to samples of 'ey'
- `samples2`, corresponding to samples of 'ay'

Each **column** corresponds to a single vowel sample.

Cognitive scientists have realised that is possible to use a *matrix transformation* on the samples so that it is only necessary to store a **single feature** for each sample (reducing each sample to one dimension) . This reduces storage requirements, and makes automatic vowel classification simpler to implement. For this assignment, you will use Matlab to find, and apply such a transformation. The process of reducing the number of features for each sample is known as *dimensionality reduction.*

All the transformations you will use in this assignment can be expressed as 2x2 matrices. To apply a transform **M** to all the samples of a vowel 1, for example, use:

```
TransformedSamples1 = M * samples1;
```

1 Simple dimensionality reduction

1. In order to visualise the data you have been given, use Matlab to display the data in a scatter plot, as in Assignment 1. Use the colour blue for group 1 and black for group 2.

[2]

Initially it is suggested to carry out dimensionality reduction on the data without applying any transformation first.

2. Write down a simple matrix that can be applied to the samples to set the **second** feature to zero for every sample.

[2]

3. Apply this matrix to all the samples for each vowel. Plot the new, dimensionality-reduced samples on a new scatter plot. How easy is it to distinguish between the two vowels after the reduction has been applied?

[4]

2 Transforming the samples

2.1 Analysing the samples

Your first scatter plot shows how the samples for each vowel are clustered together. You can see how the shapes of the two clusters are approximately the same. You will use information about the shape of the clusters to devise a transform that will allow the samples to be easily separated in just one dimension.

The centres of the two clusters have been computed from the samples. They are:

```
m1 = [ 0.502          m2 = [  0.401
       0.145 ]                -0.155 ]
```

The shape of the clusters can be described by a *covariance matrix* (this will be introduced later in lectures – you do not need to know what it is for this assignment). This has been computed from the samples and is given by

```
C = [  0.20  -0.55
      -0.55   2.00 ]
```

4. Show that the vectors

   ```
   u = [ 0.4813        v = [ -0.1354
         0.1354 ]              0.4813 ]
   ```

   are both eigenvectors of the matrix **C**. What are the two corresponding eigenvalues? (To avoid rounding errors, use the exact values of **u** and **v** which have been loaded into Matlab from `a2_data.mat`).

   [5]

5. Copy the function `plotv` from Lab Practical 1 into your current directory , and use it to plot the two eigenvectors on your scatter plot, starting at `m1`, and also starting at `m2`. Explain the significance of the eigenvectors.

   [3]

2.2 Constructing the transform

Consider a matrix transforming the vector **u** onto the x-axis (the vector `[ 1; 0]`) and the vector **v** onto the y-axis (the vector `[ 0; 1]`). This transform will change the shape of the two clusters. You will find the matrix in two steps:

6. Write down a matrix that transforms `[ 1; 0]` onto **u** and `[ 0; 1]` onto **v**. (Easy!)

   [2]

7. The matrix you need is the inverse of the above matrix. On paper, showing all your calculations, find the inverse of your matrix from Question 6. (You can check your answer in Matlab.)

   [5]

8. Apply the final matrix to both sets of samples. Plot the transformed samples on a new scatter plot.

   [2]

You should be able to see from your plot that the two vowels can now be separated using just the new feature 1 (the x-dimension).

To submit

For the assignment, you should submit:

- All the scatter plots you produced.
- Written answers to the required questions.
- Printouts of any Matlab code you wrote.

Please hand in a hard copies of the required material by 4:00pm on the due date to the Informatics Teaching Organisation, Level 4, Appleton Tower. If you have questions regarding the assignment, please contact Yansong Feng at `yansong.feng@ed.ac.uk`.