Machine Learning Module

Week 8

Laboratory Exercise, Week 8

Cluster Analysis

Mark Girolami
girolami@dcs.gla.ac.uk
Department of Computing Science
University of Glasgow

February 27, 2006

1 Cluster Analysis

1. Download all the files associated with this weeks laboratory session and run the two demos `wk8_demo_1.m` & `wk8_demo_2.m`.

2. In demo one the task of segmenting an image is attempted using K-means clustering. For the image of the `wee_dog.jpg` and the image of `water_lillies.jpg` it is clear what value K should be set to. Examine the impact that the value of K has on the segmentation of the images.

3. How variable are the eventual solutions from K-means for each of these two images? Can you give an explanation for what you observe?

4. In demo two the inability of standard K-means clustering to cluster data with nonlinear feature dependencies is highlighted. This problem is resolved using the kernel-based K-means algorithm. However, the value of the kernel function parameter is going to be important. Vary the kernel parameter values and observe the impact that this has on the visual quality of the clusterings.

5. Over multiple runs of K-means do you see a relationship between the figure of merit $\mathcal{E}_K$ and the 'subjective' visual quality of the clustering? The image segmentation problem using the 'wee dog' may be interesting here.