

CS 322 Homework 5

out: Tuesday 27 March 2007

due: Monday 2 April 2007

SVD detective work

Often one is faced by some multidimensional experimental data in which a lot of measurements are generated from some physical system where the underlying process is simple. For instance, in absorption spectroscopy one might measure absorption spectra (sampled at hundreds of wavelengths) for a number of different samples which are actually just mixtures of a few substances, or one might have camera observations of one moving object from many views. In cases like these the data is redundant, and reducing it to the relevant dimensions is the first task in analyzing it. If we are lucky the redundancy will show up as low-rank structure in the data and we can find the right space for looking at it by using the SVD.

I generated the set of “measurements” in the file `hw5data.txt` by starting with a set of points in \mathbb{R}^k and transforming them into \mathbb{R}^{10} by a linear transformation. Then I contaminated them with a little bit of noise. This closely follows the models often used to analyze experimental data that is expected to be low rank.

Your job is to answer the following questions using the SVD and MATLAB’s plotting tools, and to explain how you arrived at your answer.

1. What, if anything, can you tell about the data by looking at 2D and 3D scatterplots of the measurements against one another?
2. What is the dimension of the space the data started in? What assumptions did you need to decide this?
3. What are the most interesting two axes to project the data on to make a scatterplot? Show the resulting plot.