CALD 10–602
Statistical Approaches to Learning and Discovery

Assignment 3

February 29, 2000

Due 12:00 (noon), Thursday, March 9.

This assignment includes four problems involving the EM algorithm, each selected from Tanner. *You are only required to work three of the four.* Each of the problems has a numerical component, where you are asked to implement EM on a small data set. The data for these problems will be placed into the files /afs/cs/academic/class/10602-s00/-data/ass3.problem*n*.data, so that you don't have to type it in by hand, possibly introducing errors. (This way at least any errors will be common to everyone...).

As for the first assignment, you should write a "stand alone" program in the C or C++ programming language. Pass in your program together with a "Makefile" that will compile it in a standard Unix environment using gnu facilities. Your program and Makefile should be copied to the directory `/afs/andrew/course/10/602/studentir/<id>/ass3` where `<id>` is your andrew or cs userid. Your written solutions can be handwritten and passed in on paper, or they can be formatted and copied to the above directory as the postscript file `ass3.ps`.

1. *Normals with scaled variance.* Problem 4.3.

2. *Lifetimes as incomplete exponential observations.* Problem 4.4.

3. *Mixtures of normals for galaxy clusters.* Problem 4.7.

4. *Capture-recapture counts.* Problem 4.8.