University of TEHRAN

Probabilistic Graphical Models

Data Analysis Project # 3

Hesam Montazeri
Fahimeh Palizban
Naser Elmi

Ordibehesht 18, 1398
(May 8, 2019)

Regulatory motif finding

Identifying regulatory motifs is a fundamental problem in the field of bioinformatics. Expectation maximization and Gibbs sampling are popular motif discovery algorithms. The goal of this project is that you gain deeper familiarity with these algorithms and their applications in motif discovery. In this project, you should perform the following tasks:

1. Write the mathematical derivations for soft-assignment and hard-assignment EMs for motif discovery.
2. Implement soft-assignment and hard-assignment EM algorithms as well as a Gibbs sampler in R.
3. Investigate how each algorithm performs on four datasets uploaded to the course website (as a single .RData file). The length of motif is 4 for the first dataset and 9 for other datasets. For each dataset, report position probability matrix, learned motifs and their locations on input sequences.

Submit your report as a single PDF file in addition to your code to naser.elmi@ut.ac.ir and fahimeh-palizban@ut.ac.ir by Khordad 4, 1398.