Problem Set 3: Networks and Gene Regulation

Ethan Sherbondy

October 29, 2012

1   Bayesian Decision Theory

2   Gibbs Sampling For Motif Discovery

My Gibbs sampling algorithm is modeled after the algorithm described in Chapter 12.2 of *An Introduction to Bioinformatcs Algorithms*. It also makes use of relative entropy to weigh the random choice of a starting position (step 5 as described on page 413). I make use of pseudocounts for my profile matrix. I do **not** use the log-of-odds when calculating my l-mer probabilities, as Clojure handles extremely small decimal numbers fairly well.

My metric for convergence is a value referred to in the code as the *cutoff*. For simplicity, I say the algorithm has converged when cutoff rounds go by without an increase in the maximum-probability l-mer. I've arbitrarily set the cutoff to be 20 rounds in the current implementation, but it can easily be adjusted. If I had more time, it would be fun to explore using simulated annealing instead of my naive approach.

Despite using relative entropy, my solution seems to perform poorly when the AT/GC ratio is non-uniform. This suggests to me that I am not making proper use of the relative-entropy equation.

Regardless, here are my results: