

CMSC423: Midterm 1 Recap

Spring 2017

The midterm exam will consist of: ~10-15 quick questions (multiple choice, true/false), ~5 short questions, 1 or 2 longer questions.

It will cover the following material (this is not an exhaustive list, there may be material, especially from Chapter 1, that is not listed here, that may be included in the test):

1. Molecular Biology concepts. Most questions will be about term/concept identification, possibly a short question to test your understanding of basic molecular biology processes: e.g., replication, transcription, translation.
2. Programming for Bioinformatics. What are the types of entities commonly encapsulated by data structures in bioinformatics libraries? What is the importance of reproducibility in the analysis of genomics data?
3. Bioinformatics Resources. Short questions to check your ability to identify resources containing specific types of data. E.g., genomic sequences may be found in refseq, sequencing experiments in the Short Read Archive
4. Motif finding. What are transcription factors? What are motifs? What is the motif finding problem (biologically and computationally)? What is a “profile”? What is entropy, and why should we use it to score motifs? Why should we score motifs? How is motif finding an optimization problem? In Gibbs sampling, what is the benefit of randomly selecting a starting point for a given sequence instead of choosing the starting point that maximizes probability?
5. Clustering. What is the general clustering problem (assuming gene expression, or other continuous measurements/features). What is the k-means algorithm? What is the objective function minimized by the k-means algorithm? What is the fuzzy k-means formulation and what is its relationship with the randomized algorithms used in motif finding.
6. Peptide sequencing. What are peptides? What is a peptide’s mass spectrum? What is the cyclospectrum peptide sequencing problem? Given an observed experimental spectrum, and a peptide, what is the difference between scoring as a linear vs. cyclic peptide?
7. General algorithmics. What is a branch and bound algorithm? What is a heuristic algorithm? What is a randomized algorithm? For each of these, what are appropriate ways of analyzing the correctness and performance (running time) of these algorithms.