

CMSC423: Midterm 2 Recap

Spring 2017

The midterm exam will consist of: ~6-10 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 the motif finding chapter, that is not listed here, that may be included in the test):

1. Biological sequence comparison: Why do we need algorithms that find inexact alignments between DNA or aminoacid sequences? Why are exact matching algorithms not sufficient for biological questions?
2. Sequence Assembly. The Hamiltonian and Eulerian approaches to sequence assembly. The Eulerian cycle finding algorithm.
3. Inexact Alignment. Dynamic programming algorithms: Global alignment, Local alignment. Linear gap penalties, affine gap penalties. The probabilistic interpretation of scoring matrices (as log odds of two probabilistic models). How to do global alignment with linear space complexity?
4. How would you design an EM algorithm to solve the motif finding problem?