Bi/BE/CS183 2019

Profs. Matt Thomson and Lior Pachter

MIDTERM

Due on Moodle by Thursday February 13 at 12:00pm

Midterm policy: This is a take home midterm which you must complete by yourself. You may not work with others, or use the internet while working on the exam. You may use materials (notes, slides etc.) distributed via Moodle, your homework sets/solutions, code you wrote for homework, and prepared notes of your own. Programming exercises may be completed in any language, and the code (as well as examples demonstrating that it works) must be turned in online with the exam. You have three hours to work on the exam, not counting breaks. The exam must be completed in one sitting.

Problem 1

Give an example of 8 distinct points in R^2 that have the property that they do not have a unique 1-dimensional principal component projection.

Problem 2

Find an optimal alignment (i.e. alignment with maximal score) for the sequences $GATTACA$ and $AATGGACA$ for the scores $match = 1$, $mismatch = -1$, $space = -2$, where the score of an alignment is the sum of the scores for the matches, mismatches and spaces in the alignment.

Problem 3

An open reading frame is a sequence of DNA with length divisible by 3 that has the properties that it starts with ATG and contains only one stop codon in frame, at the very end of the sequence. Write a program that implements an efficient linear time algorithm taking as input a DNA sequence of arbitrary length, and outputs the longest subsequence that is an open reading frame.

Problem 4

Derive the formula for the expected number of cells uniquely barcoded in a droplet single-cell RNA-seq experiment. Formally, consider the problem of assigning N cells each a different barcode from among M barcodes. The probability that a cell is assigned any specific barcode is $\frac{1}{M}$.

Hint: The probability that a barcode ends up in k different cells is given by the binomial distribution, where p is the probability that a cell gets a specific distinguished barcode among the available ones:

$$\mathbb{P}(N = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

$$\binom{N}{k} = \tfrac{N!}{k!(n-k)!}$$