

2 Bioinformatics (pl219)

- (a) You are given a table of gene expression data. Each row corresponds to a gene and in the columns there is the gene expression at different time steps (or different experimental conditions). Discuss at least one method, with one example, to identify genes with similar behaviour in time and the method's complexity and limitations. [4 marks]
- (b) Discuss with one example how the stiffness parameter affects soft k -means clustering. [4 marks]
- (c) Describe the advantage of using suffix arrays to find matches in genome sequencing. [4 marks]
- (d) Describe solutions to the problem of 'bubbles' in De Bruijn graphs of genomes. [4 marks]
- (e) Describe opportunities and challenges presented by DNA storage of data including a technique for indexed retrieval. [4 marks]