Name, SURNAME and ID ⇒

Middle East Technical University
Department of Computer Engineering

CENG 734

Advanced Topics in Bioinformatics

Fall '2008-2009

Final Exam

- **Duration: 60** minutes.
- **Exam:**
  - This is a **closed book**, **closed notes** exam. The use of any reference material is strictly forbidden.
- **About the exam questions:**
  - The points assigned for each question are shown in parenthesis next to the question.
  - For *True-False* type questions, put your results in the boxes provided.
- **This exam consists of *5* pages including this page. Check that you have them all!**
- **GOOD LUCK !**

Question 1	
Question 2	
Question 3	
Question 4	
Total ⇒	

1 **(20 pts)** ☐

For the following 10 statements, indicate whether the statement is *true* or *false* by marking the corresponding box with **T** or **F**, respectively (2 points each).

i. Obtaining a human individual's genome sequence in one shot is not possible with today's technology, therefore the genome sequence is obtained as a set of overlapping fragments. ☐

ii. A *gene duplication* event in evolution produces (at least) two copies of a gene in the same organism. ☐

iii. MAMMOTH is a structural alignment tool and its scoring function for a pair of proteins is positive semi-definite. ☐

iv. The number of secondary structure elements in a protein chain is always less than its sequence length in terms of number of amino acids. ☐

v. The number of genes is always larger than the number of columns in a constant bi-cluster. ☐

vi. Two genes, which show similar behavior in a small set of experimental conditions, show similar behavior in all possible conditions, because they are functionally related. ☐

vii. The ordering of experimental conditions is important in a time-series microarray experiment. ☐

viii. Interacting proteins should have equal number of amino acids. ☐

ix. In order to use support vector machines for protein-protein interaction prediction, you have to represent each protein as a fixed-length numerical vector. ☐

x. Bayesian Learning is an **un-supervised** learning method. ☐

2 **(20 pts)**

What is bioinformatics? Describe in your own words.

3 **(20 pts)**

What is the biggest promise of bioinformatics in the upcoming years? Which problem or area do you think bioinformatics will help most? Describe briefly.

4 **(40 pts)**

A biologist friend of yours is interested in a gene that she thinks is related to lung cancer. As her computer scientist friend who took CEng 734 she needs your help. What advice would you give her? Which data sources would you recommend? Would you recommend her any computational techniques to help increase her knowledge related to this gene? Feel free to comment on anything that may be related to this question.