

## Intelligent Systems, 2018/19, written exam, preparatory sample, DD Month 2019

All questions count equally. Literature, electronic and communication devices are not allowed. It is allowed to use up to 10 sheets of A4 format paper with notes. You can write your answers in English or Slovene. Duration: 90 minutes.

Oral exam for students who wish to improve their grade and have achieved at least 50% of points in written exam, will take place on DD Month 2019 at HH<sup>mm</sup>, in the office of Prof Robnik Šikonja (2<sup>nd</sup> floor, room 2.06).

- 1) A logical formula in the 3CNF form is represented as a conjunction of disjunctive terms where each disjunctive term is composed of three disjunctively connected literals. A literal is a logical variable in nonnegated or negated form, i.e.  $X_i$  or  $\bar{X}_i$ . An example of 3CNF formula is  $f = (\bar{X}_3 \vee X_2 \vee X_7) \wedge (X_1 \vee \bar{X}_2 \vee X_3) \wedge (X_2 \vee X_5 \vee X_3) \wedge (\bar{X}_1 \vee \bar{X}_4 \vee \bar{X}_6)$ . A disjunctive term is satisfied if at least one of its three literals is *true*. The problem of maximal logical formula satisfiability tries to find an assignment of logical variables  $X_1$  to  $X_n$  (with values *true* or *false*) so that a maximal number of disjunctive terms is satisfied. For example, an assignment  $X_1 = \text{true}, X_2 = \text{false}, X_3 = \text{true}, X_4 = \text{true}, X_5 = \text{false}, X_6 = \text{false}, X_7 = \text{true}$  satisfies all four disjuncts in the above formula.
  - a. Propose a representation (i.e. a data structure) for solving maximal logical formula satisfiability problem with genetic algorithm.
  - b. Describe how crossover and mutation operators work for the proposed representation and simulate one step of crossover and mutation for a problem where there are  $n = 10$  variables and the formula contains 7 disjunctive terms.
- 2) Describe the difference between
  - a. supervised and unsupervised learning,
  - b. classification and regression.
- 3) Explain why ensemble methods work and explain possible advantages of a random forest compared to a single classification tree.
- 4) Present a sketch of code for selection of the best set of features for kNN model. Assume that initially you start with an empty set of features and that you evaluate your models using a random 70:30 split to training and evaluation data set. Present the main idea of your algorithm in a few sentences. Comment your code.

- 5) In the tic-tac-toe game, two opponents interchangeably mark the board spaces with circles and crosses. The opponent who first marks three spaces in one row, column, or diagonal is the winner. A part of the problem space is illustrated on the figure.

Suppose, you want to design a system that will play the game using the reinforcement learning.

- a. Describe what representation you would use to represent the value function.
- b. Describe how to set the rewards and explain why your approach would be successful.
- c. Chose a suitable reinforcement learning algorithm for this problem and illustrate how the state values are propagated in the problem space.

