**Intelligent Systems, 2021/22, written exam, 25 January 2022**
All questions count equally. Literature, electronic and communication devices are not allowed. It is allowed to use one sheets of A4 format paper with notes. You can write your answers in English or Slovene. Duration: 90 minutes.
A possibility to see grading of your exams will take place on Thursday, 27 January 2022, at 11:00, in the office of Prof Robnik Šikonja (2nd floor, room 2.06).

1. What is semi-supervised learning? Describe the difference between supervised and semi-supervised learning. Give an example of when would you use semi-supervised learning.

2. Describe how to evaluate prediction models if the class distribution is imbalanced? Argue when to use classification accuracy, sensitivity and specificity.

3. We are interested in finding the most similar words to a group of words, e.g., we have a list of gene names and want to find names of additional genes that appear in the same contexts. Propose a procedure that will take as an input a set of words and a document and return the set of the most similar words from the given document.

4. In an undirected graph $G = (V, E)$, a clique is a complete (fully connected) subgraph of G. In other words, clique is a subset $V' \subseteq V$ of vertices where each pair of nodes is connected by an edge in E. In the graph G below, there are many cliques, e.g., $\{z, w\}$ or $\{z, u, x\}$ but the maximal clique of size 4 is $\{u, v, y, x\}$. Your task is to find a clique of the maximal size in a given graph using genetic algorithm.
   a) Propose a representation (i.e. a data structure) to be used in solving this problem with genetic algorithms. What will be your fitness function?
   b) Describe and illustrate how your crossover and mutation operators work for the proposed representation.
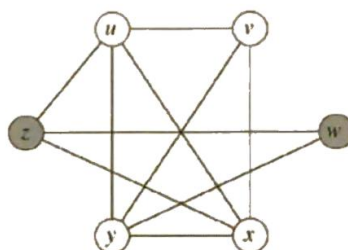
5. Inside your industrial process, some of your control cameras are exposed to high temperature and produce very noisy images of products. Once the products leave the hot area the captured images are of good quality. Your aim is to automatically detect defects on the products as soon as possible from these noisy images before the products leave the hot environment. To remove the noise from images, you decide to apply autoencoders. Construct a sketch of code for two procedures:
   i) a training procedure for the autoencoder, and
   ii) application of the trained model inside the hot industrial process.
   Present the idea, draw the architecture of the two procedures, and comment your code.