

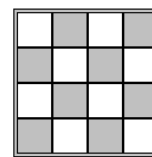
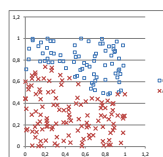
Предмет: **Вовед во препознавање на облици**

Домашна работа 2 – **DT, k-NN**

Задача:

Да се спроведат експерименти со обука и проценка а точноста на следниве класификатори:

1. Дрво за одлучување
2. k Nearest neighbor класификатор



на следниве податочни множества:

- А. синтетичко (генерирано) за двокласен проблем опишан со функцијата $y = \frac{\sin(6x)}{6} + 0.6$ во опсегот $0 \leq x \leq 1$; $0 \leq y \leq 1$ (сè што е под кривата е една класа, сè што е над – друга класа).
- Б. синтетичко (генерирано) за двокласен проблем опишан како 4×4 шаховска табла во опсегот $[0,1] \times [0,1]$ (белите полиња се едната класа, црните другата)
- В. За множеството iris (<https://archive.ics.uci.edu/ml/datasets/Iris>) (да се користи 10-fold cross-validation)
- Г. За множеството seismic-bumps (<https://archive.ics.uci.edu/ml/datasets/seismic-bumps>) (да се користи 10-fold cross-validation)
- Д. За множеството Letter Recognition (<https://archive.ics.uci.edu/ml/datasets/Letter+Recognition>) (да се подели 2/3 за обука, 1/3 за тест)

За синтетичките множества (под А. и Б.) да се спроведат најмалку 3 експерименти со различен број на примероци за обука и тестирање (на пр. 100/50, 1000/300 и 10000/3000). Примероците да се генерираат случајно (со скрипта, програма, Excel табела ... – да се приложат).

За сите податочни множества да се изведат експерименти (1) кај kNN за 1, 3 и 7 најблиски соседи, а (2) кај Decision Tree со различни параметри (confidence factor, pruning, ... зависно од пакетот/библиотеката која ја користите) - најмалку 3 експерименти.

За сите спроведени експерименти да се наведе: платформата (Weka, Matlab, друг пакет, оригинален изворен код, ... кој сте го користеле), поставките (параметрите со кои е обучуван класификаторот), и резултатите од тестирањата. Приложете документ (текст 2-3 страни) во кој во неколку реченици се објаснети подготовките и експериментите кои сте ги спровеле, се наведени резултатите (табела / графикон) и согледувањата од спроведените експерименти при менувањето на поставките за секое податочно множество и класификатор поодделно.

Документот со резултатите и согледувањата, како и евентуалните скрипти за генерирање на синтетичките примероци, како и самите генерирани податоци за А. и Б. да бидат ставени во директориум со име Homework2_<indeks>_<Prezime>_<Ime> и спакуван (zip или rar) се поставуваат на [Moodle сајмот на курсот](#) не подоцна од назначеното време.

Рок за испраќање: 02.12.2018 23:54