

I-SUNS: Zadanie č.1

SPRACOVANIE DÁT, NEURÓNOVÉ SIETE I.

Vo vybranom programovacom jazyku implementujte program, ktorý bude odhaľovať ľudí s možnou srdcovou chorobou. K dispozícii je súbor s údajmi o 70.000 pacientoch.

Čas odovzdania je určený časom vloženia do AIS. Deadline pre získanie 8 bodov je pred vašim cvičením v 4. týždni semestra (**15/16.10.2020**).

Úlohy

1. **Načítajte dáta a pripravte ich na spracovanie.** Načítajte dáta z .csv do vášho programu - vyberte vhodnú štruktúru, aby s nimi vedeli vami zvolené algoritmy/modely/knižnice ďalej pracovať.
2. **Dáta predspracujte** - je potrebné zbaviť sa nulových a nezmyselných hodnôt. Takisto je potrebné kódovať textové hodnoty. **1b** Dáta vhodne normalizujte. **1b**
3. **Analyzujte príznaky** - snažte sa zodpovedať otázky - čo je najmocnejší ukazovateľ na prítomnosť srdcovej choroby? Ktoré dáta je vhodné poslať do ďalšej analýzy? **1b**
4. **Natrénуйте binárny klasifikátor.** Zvoľte si neurónovú sieť, ktorú natrénujte tak, aby vám dávala odpoveď áno/nie na prítomnosť srdcovej choroby u pacienta. **1b** Výsledky vyhodnotte a analyzujte. **1b** Nezabudnite vhodne vybrať tréningovú a testovaciu množinu (popr. validačnú).
5. **Natrénуйте regresor.** Vašou výstupnou veličinou tentokrát bude BMI (body mass index). Natrénujte neurónovú sieť. **1b** Porovnajte ju s lineárnym regresorom. **1b** Výsledky vyhodnotte a analyzujte. **1b**

Nepovinné úlohy

- Zopakujte vaše tréningy viackrát pre rozdielne vstupné ukazovatele. Postup odôvodnite a analyzujte výsledky. **1-2b**
- Zopakujte vaše tréningy viackrát pre rozdielne nastavenia tréningovania. Postup odôvodnite a analyzujte výsledky. **1-2b**

Poznámky, spresnenia, odkazy

- Zadanie má dve časti:

1. vytvorenie kódu a spracovanie vzoriek
2. napísanie dokumentácie

Aby bolo zadanie považované za odovzdané je potreba spraviť obe. Nepodceňte dokumentáciu - je potreba sa v nej vyjadriť ku vstupom a výstupom a každej podúlohe v zadaní (aspoň niekoľkými slovami - metódy, výsledky, ak sa to hodí aj obrázky). Takisto (keďže zadanie neodovzdávate osobne) je dobré svoj postup odôvodňovať.

- Je potrebné, aby bolo zadanie pred časom vášho cvičenia nahraté v AIS v prislúchajúcom mieste odovzdania.
- Dobre čítajte dokumentáciu metód, ktoré používate.
- Nie ste hodnotení na základe úspešnosti vašich modelov, ale pri zlých výsledkoch je očakávaná aspoň snaha ich zlepšiť a pochopenie, prečo tomu tak bolo.