

I-SUNS: Zadanie č.3

SÚBOROVÉ UČENIE

Vo vybranom programovacom jazyku implementujte program, ktorý bude klasifikovať vesmírne objekty (hviezdy, galaxie, kvazary). K dispozícii sú súbory s tréningovými a testovacími dátami, ktoré dohromady obsahujú údaje o takmer 15.000 vesmírnych objektoch, ktoré boli zozbierané mapovaním oblohy.

Čas odovzdania je určený časom vloženia do AIS. Deadline pre získanie 8 bodov je pred vašim cvičením v 8. týždni semestra (**12/13.11.2020**).

Dáta

Dáta pochádzajú z verejnej databázy [Sloan Digital Sky Survey](#). Spolu s popismi stĺpcov sú dostupné v AIS.

Úlohy

1. Načítajte dáta a pripravte ich na tréning. Pri niektorých príznakoch máte napísané, že nie sú určite vhodné ako vstupy, ostatné môžete zvážiť. Ak si to vaše modely vyžadujú - normalizujte.
2. **Klasifikujte vesmírne objekty** - Natrénujte súborový klasifikátor, ktorý správne identifikuje typ vesmírneho objektu. Snažte sa nájsť dobré parametre pre natréningovanie modelu. Analyzujte aspoň jeden slabý klasifikátor. **2b** Natrénujte ešte jeden klasifikátor podľa vlastného výberu. **1b** Vyhodnoťte a analyzujte dosiahnuté výsledky. **1b**
3. **Odhadnite polohu vesmírnych objektov** - Natrénujte súborový regresor, ktorý určí Karteziánske súradnice *x_coord*, *y_coord*, *z_coord* jednotlivých vesmírnych objektov (bod [0, 0, 0] predstavuje našu galaxiu, presnejšie naše Slnko). Snažte sa nájsť dobré parametre pre natréningovanie modelu. **2b** Natrénujte ešte jeden regresor podľa vlastného výberu. **1b** Vyhodnoťte a analyzujte dosiahnuté výsledky. **1b**

Nepovinné úlohy

- Natrénujte súborové učenie založené na metóde *stacking*. **3b**

- Analyzujte vstupné stĺpce - ako korelujú s výstupmi (medzi sebou), ako sa menia predpovede pri vynechaní stĺpca a pod. **1b**

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.