

Domáca úloha č. 7

2-AIN-150, Zima 2018

Termín: 8.1.2019, 23:59, moodle.uniba.sk/fmfi

Skôr ako sa pustíte do riešenia domácej úlohy, oboznámte sa so všeobecnými pokynmi, ktoré sú priložené na konci tohto dokumentu. Riešenia, ktoré odovzdáte, musia byť vaše vlastné. Neopisujte a nesnažte sa nájsť riešenia v literatúre alebo na internete!

Učenie s pomocou neoznačených dát

Uvažujme binárnu klasifikáciu, kde naše tréningové dáta obsahujú 200 príkladov s 2400 atribútmi. Ako bonus ale máme navyše ďalších 10000 príkladov, u ktorých nepoznáme výstup, ale vieme, že sú z rovnakej distribúcie ako tréningové dáta. Vašou úlohou je:

- Pomocou neoznačených dát natréňovať vhodnú transformáciu (napr. takú, ktorá zníži počet atribútov).
- Na transformované dáta použiť vhodný klasifikátor.
- Cross-validáciou overiť presnosť (čokoľvek nad 75% je dobré).
- A keď už máte vybraný postup, tak oklasifikovať nové testovacie dáta.

Za veľmi vysokú presnosť na testovacích dátach (nad 90%), resp. použitie zaujímavých metód, môžete získať až 5 bonusových bodov.

Vo svojich programoch môžete používať knižnice, ktoré robia základnú matiku, maticové operácie, rátajú inverzné matice, systavy lineárnych rovníc, numericky/symbolicky derivujú. Explicitne máte **povolené** použiť knižnicu scikit-learn a jej obdoby v iných jazykoch.

Pokyny pre Python V balíku je súbor `template.py`, v ktorom doprogramujte funkciu `magic(X, y, U, T)`, kde (X, y) sú označené dáta, U sú neoznačené dáta a T sú testovacie dáta. Vaša funkcia by mala vrátiť klasifikáciu pre testovacie dáta. Program sa spúšťa príkazom `python template.py <vstupný súbor>`.

Pokyny pre iné jazyky Napíšte podobnú funkciu ako v Pythone a vhodne ju okomentujte a otestujte. Vstupný súbor obsahuje najprv 200 riadkov, z ktorých každý obsahuje 2401 čísel (prvých 2400 je riadok z matice X , posledné je očakávaný výstup). Ďalších 10000 riadkov obsahuje neoznačené dáta, každý riadok má 2400 čísel. Posledných 500 riadkov obsahuje testovacie dáta (každý riadok 2400 čísel).

Všeobecné pokyny

Úlohy odovzdávajte do moodlu. **Do kódu vložte stručný komentár o vami naprogramovanej metóde.**

Ideálne odovzdávajte domáce úlohy v Pythone (doprogramujte požadované funkcionality zo zadania). Pokiaľ chcete použiť iný jazyk, môžete, ale musíte zároveň naprogramovať aj réžiu okolo (načítanie, výpis, ...). Bolo by ale vhodné, aby som váš program vedel rozbehať pod linuxom bez nutnosti inštalácie komerčných programov (t.j. overte si, či váš Matlabový kód ide spustiť v Octave, C# či funguje pod Monom, ...).