

# I-SUNS: Zadanie č.2

## VIACVRSTVOVÝ PERCEPTRÓN A REGRESORY

Vo vybranom programovacom jazyku implementujte program, bude vedieť odporúčať cenu prenájmu bytu na jednu noc. K dispozícii mať rôzne ukazovatele o ponukách na stránkach AirBnB a vašou úlohou je na základe nich natrénovať modely, ktorý budú vedieť odhadnúť vhodnú cenu nových ubytovaní. **Prosím začnite prečítaním celého zadania vrátane spresnení, ktoré sú uvedené na konci.**

Čas odovzdania je určený časom vloženia do AIS. Deadline pre získanie 8 bodov je **08.11.2019 o 07:59** pre cvičenie o 08:00 a **08.11.2019 o 09:59** pre cvičenie o 10:00. Každý týždeň omeškania je penalizovaný stratou dvoch bodov.

## Dáta

Dáta sú ukladané v *.csv* súbore, kde prvý riadok je popis stĺpcov a nasledujú číselné hodnoty pre jednotlivé ubytovania (označené identifikačným číslom). Pár poznámok:

- Dáta nájdete na adrese [InsideAirBnb](#). Môžete si sami zvoliť, na ktoré mesto budete vaše modely trénovať.
- Dáta nie sú zozbierané bezchybne - v textových popisoch sa vyskytujú čiarky, nové riadky (nie je to teda validné csv) a údaje nemusia byť kompletne. Venujte preto zvýšenú pozornosť príprave dát.
- Pri vašich modeloch berte do úvahy:
  - polohu ubytovania - pomocou geografických súradníc alebo mestskej časti (alebo oboch).
  - údaje o prenajímateľovi - odpovede, počet iných ponúk, hodnotenia a pod.
  - veľkosť bytu - počet izieb/kúpeľní/postelí, slovný popis, počet hostí a pod.
  - vybavenosť bytu - typ postelí, iné položky vybavenia
  - iné údaje - napr. obsadenosť.
- Stretnete sa aj so dátami s kategrickou hodnotou. Vtedy je potrebné buď ich prekódovať na jeden číselný vstup alebo na viacero číselných vstupov (pomocou one-hot encoding).
- Samotnú cenu ubytovaniu popisuje (zrejme) kombinácia stĺpcov *price*, *guest\_included*, *extra\_people* a *accomodates*.

# Úlohy

1. **Načítajte dáta a pripravte ich na spracovanie.** Načítajte dáta z .csv do vášho programu, zakódujte kategorické (slovné) hodnoty, rozdeľte ich na vstupné a výstupné hodnoty, pripravte si trénovaciu/testovaciu/(validačnú) množinu, normalizujte dáta. Vyberajte množiny tak, aby boli výpovedné (nemôže sa vám napríklad stať, že v testovacej množine nie je kategória reprezentovaná) **1b**
2. **Natrénуйте viacvrstový perceptrón aby predpovedal cenovú kategóriu, do ktorej ubytovanie patrí.** Svoje modely si overte na testovacej množine (POZOR - trénovanie dáta z testovacej množiny v žiadnom prípade neovplyvňujú). Vhodne vyberte parametre akými sú parameter rýchlosti učenia, dávkovanie a zastavovacia podmienka. **1b** Vyskúšajte aspoň dve rôzne architektúry (počet neurónov/vrstiev/aktivačné funkcie a pod.). Výsledky porovnajte. **1b** V dokumentácii zobrazte priebeh trénovania na grafoch. **1b**
3. **Natrénуйте regresný klasifikátor, aby predpovedal cenu ubytovania.** Vyberte si regresný klasifikátor (RBF sieť, SVR, KRR a pod.) a natrénуйте ho na vašej trénovacej množine (rovnakej ako v úlohe 2). Overte úspešnosť modelu (treba si vybrať spôsob evaluácie - MSE,  $R^2$  a pod) . **1b** Skúste zmeniť parametre pri vašom učení (počet funkcií, kernel a pod.) za účelom zlepšenia úspešnosti. Výsledky porovnajte. **2b**

## Nepovinné úlohy

- Použite aj ďalšie súbory zo stránky [InsideAirBnb](#) a:
  - predpovedajte cenu ubytovania pre konkrétny časový interval v roku (deň/týždeň/mesiac). **1b**
  - rozumne zohľadnite v trénovaní aj popisné hodnoty (čiže textové polia). **2b**

## Poznámky, spresnenia, odkazy

- Zadanie má tri časti:
  1. vytvorenie kódu a spracovanie vzoriek
  2. napísanie dokumentácie
  3. osobné odovzdanie na cvičení

Aby bolo zadanie považované za odovzdané je potreba spraviť každú časť. Ne-  
podceňte 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).

- Je potrebné, aby bolo zadanie pred cvičením, na ktorom sa chystáte odovzdávať,  
nahraté v AIS v prislúchajúcom mieste odovzdania.
- Zadanie bude obodované na cvičení po prezentácii pred cvičiacim. **Pochopenie  
použitých metód a funkcií sa chápe ako prirodzená súčasť zadania a  
neschopnosť zodpovedať na otázky o týchto metódach je penalizované  
stratou bodov z danej časti**, a to aj v prípade, že kód je funkčný a správny.
- Dobre čítajte dokumentáciu metód, ktoré používate - napr. kolikorozmerný vstup  
očakávajú - upravte podľa toho vstupy.
- Nie ste hodnotení na základe úspešnosti vašich modelov, ale pri zlých výsledkoch je  
očakávaná aspoň snaha ich zlepšiť a pochopenie, prečo tomu tak bolo.
- Môže vám pomôcť:
  - [Machine learning guide](#).
  - [Stanford CS231n - Module 1: Neural Networks](#)
  - [Scikit - neural networks for classification](#).
  - [Cheatsheet ML tips and tricks](#)
  - [Scikit- svr](#)
  - [medium.com - how to evaluate regression models](#).