

# Domáca úloha č. 2

2-INF-150: Strojové učenie, Zima 2017

Termín: 22.11.2017 18:10, M163 (pod dvere) alebo pred prednáškou  
domácu úlohu odovzdávajú písomne (nie elektronicky)

**1. Duálne programy.** V tejto úlohe sa znova pozrieme na problém počítania vzdialenosti bodu od nadroviny. Na prednáške sme túto úlohu riešili pomocou geometrickej úvahy. Keďže už však vieme niečo o duálnych programoch a Lagrangeových multiplikátoroch, ide to aj jednoduchšie.

- a) Ukážte, že Lagrangeovu vetu o silnej dualite z prednášky viete rozšíriť aj na programy s podmienkou typu  $g(x) = 0$  (namiesto  $g(x) \leq 0$ ). Sformulujte tvrdenie a ukážte, že je jednoduchým dôsledkom pôvodnej vety.
- b) Sformulujte problém vzdialenosti bodu  $y$  od nadroviny danej rovnicou  $\langle \theta, x \rangle - b = 0$  ako kvadratický program.  
Poznámka: Vzdialenosť bodu  $y$  od nadroviny je vlastne vzdialenosťou od najbližšieho bodu  $z$ , ktorý na tejto nadrovine leží.
- c) Pomocou metódy Lagrangeových multiplikátorov vytvorte duálny program.
- d) Nájdite riešenie duálneho programu. Viete okrem vzdialenosti určiť aj najbližší bod  $z$  na nadrovine k bodu  $y$ ?

**2. O taxikároch.** Na stránke [http://www.nyc.gov/html/tlc/html/about/trip\\_record\\_data.shtml](http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml) nájdete veľmi podrobné dáta o jazdách taxíkov v New Yorku. Tradičnou (a pomerne nudnou) úlohou by bolo pomocou regresie hľadať závislosť medzi jednotlivými atribútmi a finálnou zaplatenou sumou. Takéto dáta sa však dajú využiť aj všelijakými netypickými kreatívnymi spôsobmi.

- a) Stiahnite si dáta, pozrite si ich štruktúru a na jej základe navrhnite jednu netradičnú úlohu regresie a jednu netradičnú úlohu klasifikácie. (Ideálne vaše úlohy v skutočnosti nebudú nič hovoriť o taxíkoch, ale o iných spoločenských, kultúrnych, či všeobecno-dopravných aspektoch života v New Yorku, či všeobecne.)  
Okrem definície úlohy aj v krátkosti napíšte, prečo si myslíte, že by sa vaša úloha dala pomocou týchto dát úspešne riešiť.
- b) **Bonusová úloha.** Vyberte si jednu z vašich úloh, vyriešte ju a zanalyzujte dáta, ktoré máte k dispozícii. Výsledkom by mal byť popis vášho riešenia, linka na kód (na githube alebo na vašej stránke) a popis záverov, ku ktorým ste na základe vašej analýzy dospeli.