

Domáca úloha č. 2

2-INF-150: Strojové učenie, Zima 2009

Termín: 9.11.2009, M163 (pod dvere)

1. Lagrangeova veta.

- Nájdite minimum funkcie $f(x, y, z) = x^2 + y^2 + z^2$ za podmienky $x - 2y + 3z = 4$.
- Nájdite minimum funkcie $f(x, y, z) = 2x^2 + 4y^2 + z^2$ za podmienky $4x - 8y + 2z = 10$. Viete povedať niečo aj o maxime $f(x)$ funkcie za tej istej podmienky?

2. Excelentné červené vína. Dostali ste za úlohu vybrať k slávnostnej večeri dobré červené víno. Keďže sa vo vínach vôbec nevyznáte, rozhodli ste sa ísť na to vedecky a aplikovať na tento problém vaše najnovšie poznatky zo strojového učenia.

V UCI repository na adrese <http://archive.ics.uci.edu/ml/datasets/Wine+Quality> nájdete data set "Wine Quality" (konkrétne nás zaujíma data set červených vín). Tento data set obsahuje pre každý druh červeného vína 11 kvantitatívnych atribútov a hodnotenie kvality každého vína. *Excelentnými* sú pre nás vína, ktorých kvalita je hodnotená aspoň stupňom 7. Ostatné vína budeme nazývať *bežnými*.

V každej časti zdokumentujte svoj postup a ukážte kusy kódu, ktoré ste použili na riešenie jednotlivých podúloh.

- Dáta náhodným spôsobom rozdeľte na 5 približne rovnako veľkých skupín. Prvé tri skupiny použijete v častiach b-d ako tréningovú množinu, posledné dve skupiny použijete na testovanie. Koľko excelentných vín sa nachádza v každej skupine?
- Natrénujte a otestujte SVM pre rozlišovanie excelentných vín od bežných a vhodným spôsobom vyhodnoťte jeho úspešnosť. Použite lineárny kernel (obyčajný skalárny súčin). Použite softvér SVM-light.
- Do SVM-light doprogramujte Gaussovský kernel so štandardnou odchýlkou 1. Opäť vyhodnoťte jeho úspešnosť na tých istých dátach.
- Naprogramujte taktiež klasifikáciu pomocou rozhodovacích stromov a ID3 algoritmu. Navrhnite vhodný spôsob diskretizácie atribútov a doplňte ďalšie potrebné detaily v algoritme podľa vlastného uváženia. Ukážte, ako vyzerá vami natrénovaný rozhodovací strom a vyhodnoťte úspešnosť vašej implementácie.
- Porovnajte úspešnosť častí b,c a d pomocou metódy holdout testing (konkrétne 5-fold testovanie pomocou skupín dát z časti a).