



KOOPERÁCIÓ ÉS GÉPI TANULÁS LABORATÓRIUM

Nemellenőrzött és féligellenőrzött tanulás Gyakorlati feladatok

Készítette:

Dr. Pataki Béla
(pataki@mit.bme.hu)

Méréstechnika és Információs Rendszerek Tanszék
Budapesti Műszaki és Gazdaságtudományi Egyetem

2009, november.

KÖVETELMÉNYEK:

- A laborgyakorlat teljesítésére 3 órányi tiszta munkaidő áll rendelkezésre.
- Időre történő, maradéktalan teljesítéshez a segédletek hiánytalan ismeretét feltételezzük.
- A sikeres teljesítéshez a gyakorlat során végzett munka – a futási eredményeket is beleértve – jegyzőkönyv formájában történő dokumentációja és határidőre történő leadása szükséges.

JEGYZŐKÖNYV:

A jegyzőkönyvnek – a megoldók nevéen és neptun-kódján kívül – a következőket kell tartalmaznia minden egyes részfeladatra vonatkozóan:

- KI-MIT oldott meg (feladat felosztása és értelmezése).
- HOGYAN oldották meg (megoldás leírása).
- MIÉRT így oldották meg (megoldás indoklása).
- EREDMÉNYEK összefoglalása (értelmezés és értékelés)

LEADÁS:

A gyakorlat során előállt MATLAB-os anyagokat ZIP-pelve, illetve a jegyzőkönyvet (*PDF* formában) a megadott határidőig (**laborgyakorlati héten péntek éjfélig**) e-mail-hez fájlként csatolva kérjük elküldeni a dkovacs@mit.bme.hu címre.

A levél subject-je „[MScLab1-8]”, a levél szövege pedig rendre a megoldók *neve* és *neptun-kódja* legyen. Több megoldó esetén az email-t a *többieknek is CC-ézni* kell.

1. feladat: Önszerveződő térkép (SOM) vizsgálata

1.a. feladat: Futtassa a SOM_1 programot, és vizsgálja meg az elérhető pontosságot a tanítási lépésszám függvényében 20 és 300 lépés közt (természetesen csak kb. 5 értéknél). A mintavektor-mátrix mérete legyen 4×4 , a szomszédság paraméter legyen 3! Mi az oka, hogy a hiba hirtelen csökken 50, 100 és 150 tanító lépésnél viszont attól kezdve nem csökken ilyen módon?

1.b. feladat: Futtassa a SOM_1 programot, és vizsgálja meg az elérhető pontosságot a mintavektorok számának függvényében 4×4 és 10×10 közt (természetesen csak 5-8 értéknél). A tanulási lépések száma legyen 200!

1.c. feladat: Futtassa a SOM_1 programot, és vizsgálja meg az elérhető pontosságot a módosítandó szomszédság paraméter (Nsz0) függvényében! A mintavektormátrix legyen 6×6 , és a tanulási lépések száma legyen 200!

Az 1.a.-1.c. lépések mindegyikében az eredmény értékelésénél használja fel az otthon megírt homogenitás kijelző eljárását!

2. feladat: K-átlagképző algoritmus féligellenőrzött tanításának vizsgálata

2.a. feladat: Futtassa a KAtlagkepzo programot, és vizsgálja meg az elérhető pontosságot a klaszterszám függvényében 2,3,...,8 klaszterszám esetén. Tisztán nemellenőrzött tanítást végezzen! A tanítási lépések száma legyen 100! Néhány beállításban végezzen több futtatást, és értékelje, hogy mennyiben lehetne egy futtatás eredményére támaszkodni!

2.b. feladat: Futtassa a KAtlagkepzo programot, és vizsgálja meg az elérhető pontosságot a tanítási lépések számának függvényében 10...200 lépésszám közt. Tisztán nemellenőrzött tanítást végezzen! A klaszterek száma legyen 6! Néhány beállításban végezzen több futtatást, és értékelje, hogy mennyiben lehetne egy futtatás eredményére támaszkodni!

2.c. feladat: Az otthon előre megírt program segítségével futtassa le valamelyik beállítást tízszer, és elemezze az elért eredményt!

2.d. feladat: Az otthon előre megírt program segítségével vizsgálja meg, hogy a kimeneti zaj milyen szintjénél érdemes csak az első tanítási ciklusban figyelembe venni az ismert (de zajos) besorolást?

3. feladat: Féligellenőrzött K-átlagképző algoritmus vizsgálata valós adatokon

A vizsgált adatbázis negyedórás felbontásban mért elektromos fogyasztási görbéket tartalmaz. Az áramkereskedelemmel foglalkozó cégek szempontjából fontos adat, hogy mennyire mutatható ki összefüggés egy-egy fogyasztó tevékenysége (pl. iskola, iroda, kereskedelmi egység, üzem stb. és a fogyasztási görbe jellege közt).

Ezért azt vizsgáljuk, hogy a fogyasztási jelleggörbe bizonyos paraméterei alapján klaszterezve a mintákat, mennyire válnak szét az „iroda”, az „üzem” és az „iskola” típusú fogyasztók. A paraméterek az IrodaIskolaUzemParameterek.mat fájlban találhatóak, három a mért adatokból számított paraméter, a negyedik a fogyasztó valós besorolása.

3.a. feladat: Vizsgálja meg külön-külön a 3 paramétert, eltérő színnel kijelezve az egyes osztályokhoz tartozókat!

3.b. feladat: Futtassa a `KAtlagkepzoFogyasztasra` programot a mellékelt `IrodaIskolaUzemParameterek.mat` fájlban található adatvektorokon, vizsgálja meg, hogy milyen klaszterbe sorolási pontosságot tud elérni!. Féligellenőrzött tanítást végezzen 5%, 10%, illetve 20% paraméterekkel!