

A félévben vett anyag (a vizsga anyagának) vázlatos ismertetése

Gépi tanulás (VIMIM136)

2013/2014 2. (tavaszi) félév

Ez csupán segítség kíván lenni, nem a tananyag tételes, kimerítő felsorolása! Senki se hivatkozzon erre, ha a vizsgán olyasmit kérdezek, ami előadáson szerepelt, de itt nem szerepel!

A vizsgán mind elméleti kérdések, mind egyszerű feladatok lesznek.

A vizsga - egy rövid beugró után, amely a jegyet nem befolyásolja (csak GO/NOGO) - szóbeli, mindenki kap egy elméleti kérdést és egy feladatot, egyiket az anyag első feléből, másikat a második feléből. Nem feltétlenül az előadások kronológiai sorrendje mentén osztottam két részre az anyagot, kisebb-nagyobb sorrendcserék lehetnek. Létszámkorlát nincs. A vizsgára bocsáthatóság feltétele a zárthelyi (vagy pótzh. vagy pót-pótzh.) elégséges szintű megírása.

Az anyag *legnagyobb része* megtalálható a Russel-Norvig könyv 17. és 18. és 21. fejezetében (2005-ös kiadás), továbbá a Neurális hálók könyv (2006-os kiadás) 2. és 9. fejezetében. (Pl. a hibaarány-komplexitás alapon történő metszés tudtommal egyikben sincs benne úgy, ahogy az előadáson vettük.) Igyekeztem az egyes kérdéskörökhöz odairni, hogy hol található meg róluk a legtöbb infó, remélem nem rontottam el. (Az irodalmak felsorolását lásd a Gépi tanulás/Jegyzetek alatt.)

Ahol nem hangsúlyozom külön, ott az előadáson elhangzott anyag elegendő.

Vizsgatételek

ELSŐ RÉSZ

1. A tanulás három alapvető fajtája (ellenőrzött, nemellenőrzött, megerősítéses). Az ellenőrzött tanulás alapvető blokkvázlata. Mit nevezünk induktív tanulásnak? Mikor realizálható egy tanulási probléma?
2. Döntési fa tanulása példák alapján. Osztályozásra, illetve regresszióra használt döntési fák felépítése. A döntési fa tanulás pszeudókódja. Az attribútum-teszt választás az információnyereség kritérium alapján.
3. A tanuló algoritmus teljesítményének becslése: a tanulási görbe kétféle értelmezése. A tanítás során elkövetett tipikus hiba: a kukucskálás. A zaj és túlilleszkedés. Keresztvalidáció.
4. A döntési fák metszése szignifikancia teszt, illetve hibaarány-komplexitás kompromisszum alapján.
5. Hiányzó adatok problémájának kezelési lehetőségei. Nagyon sok értékkel rendelkező attribútumok okozta nehézség – az információ nyereség arány. Folytonos bemeneti attribútumok felhasználása.
6. A tanulás számítási elmélete. A szükséges példamennyiségre adható alsó korlát (levezetés). (Russel-Norvig 18. fejezet) Az ERM konzisztenciája, a szükséges mintaszámra adható másik korlát (Neurális hálózatok könyv 2.3.1 fejezetében).
7. Moduláris eszközök kialakítása a feladat dekompozíciója alapján. (Neurális hálózatok könyv 9.1).
8. Hipotézishalmaz együttes tanulása, a pontos, de különböző megoldások integrálásának követelménye. (ld. P. Sollich – A. Krogh: Learning with Ensembles. How Overfitting Can Be Useful, vagy a Neurális hálózatok könyv 9.2 fejezetében)
9. MOE struktúra. Az egyes szakértők paramétervektorainak tanítása. Az asszociált zaj szerepe a tanulásban. (Van róla anyaga a Neurális hálózatok könyv 9.2.3 fejezetében, de némiképp más hangsúlyokkal vettük.)
10. MOE struktúra: a kapuzó hálózat paramétereinek tanítása. Nemlineáris kapuzó hálózatok.
11. Nemellenőrzött tanulás. Az EM algoritmus alapgondolata. Az EM algoritmus bemutatása kevert Gauss eloszlások esetén.
12. A mintahalmaz komplexitásának jelentősége, a jellemzés – előadáson tanult – módszerei.

MÁSODIK RÉSZ

- 13.** Nemellenőrzött tanulás. A k-átlagképző (k-means) algoritmus bemutatása.
- 14.** Szekvenciális döntések (Russel-Norvig könyv 17. fejezet: 17.1-17.2-17.3)
- 15.** Az értékiteráció és az eljárás mód iterációs algoritmus (Russel-Norvig könyv 17. fejezet: 17.1-17.2-17.3)
- 16.** Passzív megerősítéses tanulás. (Russel-Norvig könyv 21.2 fejezet)
- 17.** Aktív megerősítéses tanulás. (Russel-Norvig könyv 21.3 fejezet)
- 18.** Cselekvésérték függvény tanulása, általánosítás a megerősítéses tanulásban. (Russel-Norvig könyv 21.3 vége, 21.4)
- 19.** A turbózás (boosting) eljárás, konkrétan az Adaboost eljárást tanultuk. (A neurális könyv – 9.3 fejezet – mellett ajánlom a Gépi Tanulás tárgy honlapja /Jegyzetek alá feltöltött Freund-Shapire: A Short Introduction to Boosting cikket is).
- 20.** Generatív és diszkriminatív klaszterezés. Féligellenőrzött tanulás, alapötlete, mindig alkalmazható-e. Identifikálhatóság problémája. Öntanulás, együttes tanulás.
- 21.** Féligellenőrzött klaszterezési eljárások – a k-átlagképző eljárás bemutatva. (Megtalálható pl.: Sugato Basu: Semi-supervised Clustering: Probabilistic Models, Algorithms and Experiments, PhD Thesis, Univ. of Texas, 2005. A tárgy honlapjáról (/jegyzet) elérhető)
- 22.** Féligellenőrzött tanulás. Gráf-alapú módszerek, a mincut eljárás és módosítása.
- 23.** Féligellenőrzött regresszió, a tanult megoldás (COREG) pszeudókódja. (lásd pl. a tantárgy tanszéki jegyzetek oldalán:
SemiSupervisedRegressionWithCoTraining_Zhou.pdf)
- 24.** Féligellenőrzött tanulás. Mit jelent, hogy egy eljárás transzduktív? Transzduktív svm. Indukció gráf-alapú módszereknél.