

A TANTÁRGY ADATLAPJA

1. A képzési program adatai

1.1 Felsőoktatási intézmény	Babeş–Bolyai Tudományegyetem
1.2 Kar	Matematika és Informatika Kar
1.3 Intézet	Magyar Matematika és Informatika Intézet
1.4 Szakterület	Informatika
1.5 Képzési szint	Mesteri
1.6 Szak / Képesítés	Informatikai modellek optimalizálása

2. A tantárgy adatai

2.1 A tantárgy neve (ang)	Stochastic Data Modelling Sztochasztikus Adatelemzési módszerek (Modelarea Stohastică a datelor)						
2.2 Az előadásért felelős tanár neve	Csató Lehel						
2.3 A szemináriumért felelős tanár neve	Csató Lehel						
2.4 Tanulmányi év	2	2.5 Félév	1	2.6. Értékelés módja	vizsga	2.7 Tantárgy típusa	kötelező – szak

3. Teljes becsült idő (az oktatási tevékenység féléves óraszám)

3.1 Heti óraszám	3	melyből: 3.2 előadás	2	3.3 szeminárium/labor/praktika	1
3.4 Tantervben szereplő össz-óraszám	42	melyből: 3.5 előadás	28	3.6 szeminárium/labor	14
A tanulmányi idő elosztása:					Óra
A tankönyv, a jegyzet, a szakirodalom vagy saját jegyzetek tanulmányozása					54
Könyvtárban, elektronikus adatbázisokban vagy terepen való további tájékozódás					36
Szemináriumok / laborok, házi feladatok, portfóliók, referátumok, esszék kidolgozása					50
Egyéni készségfejlesztés (tutorálás)					14
Vizsgák					4
Más tevékenységek:					
3.7 Egyéni munka össz-óraszám	158				
3.8 A félév össz-óraszám	200				
3.9 Kreditszám	8				

4. Előfeltételek (ha vannak)

4.1 Tantervi	Nincs
4.2 Kompetenciabeli	Matematikai alapismeretek, a valószínűségi számítás ismerete; Az angol nyelv ismerete, mivel a tantárgy és a bemutatók nyelve az angol;

5. Feltételek (ha vannak)

5.1 Az előadás lebonyolításának feltételei	- Az előadásokhoz video-projektor szükséges. - A példák kifejtéséhez és illusztráció számára tábla szükséges.
5.2 A szeminárium / labor lebonyolításának feltételei	A laboratóriumi órák alatt a diákok a számítógépet, az oktató a táblát használja;

6. Elsajátítandó jellemző kompetenciák

Szakmai kompetenciák	- A gépi tanulás alapfogalmai, - Adatok reprezentációja és algoritmusok felírása, - Modern gépi tanulási algoritmusok használata.
Transzverzális kompetenciák	Feladatmegoldó készségek fejlesztése

7. A tantárgy célkitűzései (az elsajátítandó jellemző kompetenciák alapján)

7.1 A tantárgy általános célkitűzése	- Napjaink számos informatikai kihívására egy válasz a gépi tanulás: olyan algoritmusok megalkotása, melyek – futásuk során – saját felismerési, előrejelzési, vagy osztályozási képességeiket tudják javítani. - A tantárgy során a gépi tanulás fogalmainak a megértésére és az algoritmusok operatív ismeretén lesz a hangsúly, a szükséges matematikai háttér vázolásával.
7.2 A tantárgy sajátos célkitűzései	- A következő témaköröket járjuk körül: Komponensek analízise: Statisztikus modellezés alapjai – sajátértékek és sajátvektorok [9], Főkomponensek [2,4]. Független komponensek és a jelek elválasztása [8]. Becslési módszerek és Bayes-féle modellezés: Becslési módszerek [1, 2, 3], Hierarchikus modellek és a modell paramétereinek a becslése [1,2,4], Bayes-féle modellezés: az a-posteriori és a prediktív eloszlások használata [4, 5]. Rejtett Markov modellek (HMMs) és alkalmazásaik: HMM definíciók [7], Rejtett állapotok becslése [7],HMM alkalmazások (1) beszébfelismerés [7,8],HMM alkalmazások (2) DNS szegmentálás[7,8], Gausz-folyamatokkal történő modellezés: A kernel függvény, funkcionális Bayes-modellek[4,5,6], Közelítések és a KL-távolság használata, a Gausz-folyamatokkal történő modellezés alkalmazásai.

8. A tantárgy tartalma

8.1 Előadás	Didaktikai módszerek	Megjegyzések
1. hét Komponensek analízise I: Statisztikus modellezés alapjai – sajátértékek és sajátvektorok, azok használata adatelemzésnél.	tanári magyarázat, rávezetés, példák, interaktív kódolás.	
2. hét Komponensek analízise II: Főkomponensek [2,4], legnagyobb és legkisebb komponensek használata, példák.	tanári magyarázat, rávezetés, munkáltatás, példák, interaktív kódolás.	
3-4. hét Komponensek analízise III:	tanári magyarázat, rávezetés,	

Független komponensek és a jelek elválasztása [8], Példák és programcsomagok használata	munkáltatás, példák, interaktív kódolás.	
5. hét Becslési módszerek és Bayes-féle modellezés I: Becslési módszerek [1, 2, 3], a maximum likelihood módszertől az a-poszteriori becslésen keresztül a teljes eloszlást használó Bayes-becslésig.	tanári magyarázat, rávezetés, munkáltatás, példák, interaktív kódolás.	
6. hét Becslési módszerek és Bayes-féle modellezés II: Bayes-féle modellezés: az a-poszteriori és a prediktív eloszlások használata [4, 5].	tanári magyarázat, rávezetés, munkáltatás, példák, interaktív kódolás.	
7. hét Becslési módszerek és Bayes-féle modellezés III: Hierarchikus modellek és a modell paramétereinek a becslése [1,2,4],	tanári magyarázat, rávezetés, példák, interaktív kódolás.	
8-9. hét Rejtett Markov modellek (HMMs) I: HMM definíciók [7], Rejtett állapotok becslése [7]	tanári magyarázat, rávezetés, példák, interaktív kódolás.	
10. hét Rejtett Markov modellek (HMMs) II: HMM alkalmazások: (1) beszédfelismerés, (2) DNS szegmentálás [7,8]	tanári magyarázat, rávezetés, példák, interaktív kódolás, munkáltatás	
11-12. hét. Gaussz-folyamatokkal történő modellezés I: A kernel függvény, funkcionális Bayes-modellek [4,5,6], Közelítések és a KL-távolság használata	tanári magyarázat, rávezetés, munkáltatás	
13-14. hét Gaussz-folyamatokkal történő modellezés I: A Gaussz-folyamatok alkalmazásai.	tanári magyarázat, rávezetés.	
Könyvészet [1]. Bishop C.M (2006) Pattern Recognition and Machine Learning, Springer Verlag. [2]. Russell S, Norvig P (2003) Artificial Intelligence: A Modern Approach (Second Edition), Prentice Hall. [3]. Mitchell T (1997) Machine Learning, McGraw Hill. [4]. Bernardo J.M, Smith A.F.M (2000) Bayesian Theory, John Wiley & Sons. [5]. MacKay D.J.C (2003) Information Theory, Inference and Learning Algorithms, Cambridge University Press, HTTP: http://wol.ra.phy.cam.ac.uk/mackay/itila/book.html . [6]. Rasmussen C.E, Williams C.K.I (2006) Gaussian Processes for Machine Learning, The MIT Press. [7]. Rabiner L.R, Juang, B.H (1986) An introduction to Hidden Markov models, IEEE ASSP Magazine, pp: 4-15. [8]. Durbin R, Eddy S.R, Krogh A, Mitchison G (1999) Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press. [9]. Hyvärinen A, Karhunen J, Oja E (2001) Independent Component Analysis, Wiley-Interscience. [10]. Barto A. (2002): Statistical Pattern Recognition, John Wiley & Sons.		
8.2 Labor / Gyakorlat	Didaktikai módszerek	Megjegyzések
Labor:		
1. hét A komponens-analízis és annak alkalmazásai Az USPS képek saját-képeinek (eigen-images) a meg-	munkáltatás, demonstráció, példák	

határozása, dekompozíciója, a FASTICA program.		
2. hét A független komponens alkalmazásai A tájképek karakterisztikáinak megtalálása a független-komponens analízissel.	munkáltatás, demonstráció, példák	
3. hét Klaszterező algoritmusok	munkáltatás, demonstráció	
4. hét Prediktív algoritmusok alkalmazása a BOSTON adatokra: mennyibe kerül egy ház?	munkáltatás, demonstráció	Szemináriumi bemutatók, melyeket a diákok tartanak a témáról és a tartalomról történt egyeztetés után.
5. hét Bayes-féle regresszió kódolása	munkáltatás, demonstráció	Szemináriumi bemutatók.
6. hét Gaussz-folyamatok alkalmazása predikcióra	munkáltatás, demonstráció	Szemináriumi bemutatók.
7. hét Kernel-algoritmusok , feladatok ellenőrzése	munkáltatás, demonstráció	Szemináriumi bemutatók.

9. Az episztemikus közösségek képviselői, a szakmai egyesületek és a szakterület reprezentatív munkáltatói elvárásainak összhangba hozása a tantárgy tartalmával.

- A **gépi tanulás**, illetve a sztochaszticitás – az adatok valószínűségi változók megvalósulásaként történő jellemzése nagyon sok előadás alapja, sok mesteri is ilyen címmel működik, például a londoni UCL egyetemen: [MSc in Computational Statistics and Machine Learning](#), oktatják a Stanfordi egyetemen is (Andrew Ng oktatásával: [Machine learning](#)).

10. Értékelés

Tevékenység típusa	10.1 Értékelési kritériumok	10.2 Értékelési módszerek	10.3 Aránya a végső jegyben
10.4 Előadás	Írásbeli vizsga a félév végén	Írásbeli vizsga	60%
10.5 Labor	Programozási feladatok bemutatása és bemutatók	A megoldások pontozása	40%

10.6 A telsscsjesítmény minimumkövetelményei

Az elvárt minimális tudás:

- A gépi tanulás alapfogalmainak az ismerete
- Algoritmusok tervezése és az elemzés lépéseinek a vázolása.

Az átmenő jegy feltételei:

- A vizsgán legkevesebb a pontok felének összegyűjtése (kizáró jellegű).
- A végső pontszám minimálisan 70 (=5; kizáró jellegű).

Kitöltés dátuma

2013.04.21

Előadás felelőse

Csató Lehel

Labor / praktika felelőse

Csató Lehel

Az intézeti jóváhagyás dátuma

.....

Intézetigazgató

Szenkovits Ferenc