



FIȘA DISCIPLINEI

1. Date despre program

1.1 Instituția de învățământ superior	Universitatea “Alexandru Ioan Cuza” din Iași
1.2 Facultatea	Facultatea de Informatică
1.3 Departamentul	Departamentul de Informatică
1.4 Domeniul de studii	Informatică
1.5 Ciclul de studii	Licență
1.6 Programul de studii / Calificarea	Informatică/Licențiat în Informatică

2. Date despre disciplină

2.1 Denumirea disciplinei							
2.2 Titularul activităților de curs		CONF. DR. LIVIU CIORTUZ					
2.3 Titularul activităților de seminar		Conf. dr. L. Ciortuz, Conf. dr. M. Breăban, Lect. dr. A. Ignat					
2.4 An de studiu	II I	2.5 Semestru	1	2.6 Tip de evaluare	E	2.7 Regimul disciplinei*	OB

* OB - Obligatoriu / OP - Opțional

3. Timpul total estimat (ore pe semestru și activități didactice)

3.1 Număr de ore pe săptămână	4	din care: 3.2 curs	2	3.3 seminar/laborator	2
3.4 Total ore din planul de învățământ	56	din care: 3.5 curs	28	3.6 seminar/laborator	28
Distribuția fondului de timp					ore
Studiu după manual, suport de curs, bibliografie și altele					28
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					28
Pregătire seminarii/laboratoare, teme, referate, portofolii și eseuri					28
Tutoriat					-
Examinări					4
Alte activități					-
3.7 Total ore studiu individual					56
3.8 Total ore pe semestru					130
3.9 Număr de credite					5

4. Precondiții (dacă este cazul)

4.1 De curriculum	
4.2 De competențe	Cunoștințe de bază din matematica de liceu (studiul funcțiilor, geometrie elementară), precum și analiză matematică, probabilități și algoritmică din facultate

5. Condiții (dacă este cazul)



5.1 De desfășurare a cursului	
5.2 De desfășurare a seminarului/ laboratorului	Prezența la seminar este obligatorie. Prezența la curs este recomandată călduros.

6. Competențe specifice acumulate

transversale	C1. Capacitatea de a înțelege, a implementa și utiliza algoritmi de bază din domeniul învățării automate.
Competențe profesionale	CT1. Abilitatea de utilizare a instrumentelor matematice (analiză matematică, probabilități și statistică) pentru proiectarea și analiza algoritmilor de bază din învățarea automată. CT2. Capacitatea de a lucra în paradigma „data-driven programming”: explorarea datelor, crearea automată de modele, evaluarea modelelor.

7. Obiectivele disciplinei (din grila competențelor specifice acumulate)

Obiectivul general	Înșușirea conceptelor de bază din domeniul învățării automate; înțelegerea și folosirea unor algoritmi de bază din acest domeniu.
Obiectivele specifice	O1. Înțelegerea algoritmilor de bază pentru crearea arborilor de decizie (ID3), clasificare bayesiană (Bayes Naiv și Bayes Corelat), învățarea bazată pe memorare (k-NN) și clusterizare (clust. ierarhică, K-means și EM/GMM). Aplicarea acestor algoritmi pe seturi relativ mici de date.



8. Conținut

8.1		Metode de predare	Observații (ore și referințe bibliografice)
1.	Noțiuni introductive de Învățare Automată. Tipuri de algoritmi de învățare automată: clasificare, clusterizare, regresie, ranking, selecția trăsăturilor, sisteme de recomandare.	Expunere	2
2.	Recapitularea noțiunilor fundamentale de probabilități și statistică: evenimente aleatoare, funcție de probabilitate, independența evenimentelor aleatoare; variabile aleatoare (discrete și continue), independența variabilelor aleatoare. Medii și varianțe. Distribuții probabiliste uzuale. Proprietăți de bază.	Expunere	2
3.	Elemente de bază din teoria informației: entropie, entropie condițională, câștig de informație, entropie relativă, cross-entropie. Proprietăți de bază. Arbori de decizie. Algoritmul ID3.	Expunere	2
4.	Variante / extensii ale algoritmului ID3. Evaluarea algoritmilor de clasificare: acuratețe, eroarea la antrenare, eroarea la [cross-]validare, precizie, recall etc. Overfitting. Strategii de pruning pentru arbori de decizie.	Expunere	2
5.	Metode ansambliste bazate pe arbori de decizie: Algoritmul AdaBoost. [Bagging, Random Forests.]	Expunere	2
6.	Tipuri de ipoteze în învățarea automată: ipoteze ML și ipoteze MAP. Clasificare bayesiană. Algoritmii Bayes Naiv și Bayes Corelat/Optimal.	Expunere	2
7.	Calculul ratei medii a erorii pentru clasificatori de tip bayesian.	Expunere	2
8.	Recapitulare.	Examen parțial	2
9.	Metode de estimare a parametrilor unor distribuții probabiliste: MLE și MAP. Aplicare la distribuția gaussiană: cazul uni-variabil vs. cazul multi-variabil. Clasificare bayesiană de tip gaussian. Legătura dintre clasificarea bayesiană și regresia logistică.	Expunere	2
10.	Învățare bazată pe memorare. Algoritmul k-NN. Clasificare automată de tip "eager" vs. clasificare automată de tip "lazy".	Expunere	2
11.	„Blestemul marilor dimensiuni” în învățarea bazată pe memorare. Relația dintre rata erorii asimptotice a algoritmului 1-NN și rata erorii algoritmului Bayes Corelat.	Expunere	2
12.	Învățare supervizată (clasificare) vs. învățare nesupervizată (clusterizare). Tipuri de algoritmi de clusterizare. Clusterizare ierarhică aglomerativă și Clusterizare ierarhică divizivă. Tipuri de funcții de similaritate (single-link, complete-link, average-linkage).	Expunere	2



13.	Clusterizare neierarhică cu asignare „hard” a instanțelor la clustere: algoritmul K-means. Monotonia criteriului de „distorsiune” (J) la aplicarea algoritmului K-means.	Expunere	2
14.	Mixturi de distribuții gaussiene (GMM). Algoritmul EM pentru rezolvarea de GMM (clusterizare neierarhică cu asignare „soft” a instanțelor la clustere).	Expunere	2

Bibliografie**Referințe principale:**

Tom Mitchell. Machine Learning. McGraw Hill. 1997.

Liviu Ciortuz, Alina Munteanu, Elena Bădărău. Exerciții de învățare automată. Editura Universității “Al.I.Cuza”, Iași, 2017.

Referințe suplimentare:

Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning. Springer, 2009.

Christopher Manning, Heinrich Schutze, Foundations of Statistical Natural Language Processing, MIT Press, 2000.

	8.2 Seminar / Laborator	Metode de predare	Observații (ore și referințe bibliografice)
1.	Sisteme/programe de învățare automată: câteva studii de caz.	Demo-uri	2
2.	Exerciții recapitulative de probabilități și statistică.	Lucru la tabla, lucru individual, metode interactive, programe demonstrative, teste.	2
3.	Exerciții pentru calcularea de entropii, câștiguri de informație, demonstrare unor proprietăți. Exerciții cu arbori de decizii. Aplicarea algoritmului ID3. Calculul erorilor la antrenare și validare încrucișată cu “metoda leave-one-out”.	Idem	2
4.	Aplicarea algoritmului ID3 folosind attribute numerice continue: exerciții. Aplicarea diverselor strategii posibile de pruning pe arbori de decizie: exerciții.	Idem	2
5.	Aplicarea algoritmului AdaBoost: exerciții.	Idem	2
6.	Aplicarea algoritmilor Bayes Naiv și Bayes Corelat: exerciții.	Idem	2
7.	Exerciții de calcul pentru rata erorilor pentru diverși algoritmi de clasificare automată.	Idem	2
8.	Recapitulare.	Idem	2
9.	Estimare MLE și MAP: exerciții. Clasificare bayesiană de tip gaussian. Legătura cu regresia logistică: exerciții.	Idem	2



10.	Aplicarea algoritmului k-NN: exerciții. Arbori KD.	Idem	2
11.	Aplicarea algoritmilor de clusterizare ierarhică aglomerativă și divizivă: exerciții.	Idem	2
12.	Aplicarea algoritmului K-means: exerciții.	Idem	2
13.	Exerciții de estimare a parametrilor pentru diverse tipuri de distribuții probabiliste.	Idem	2
14.	Exerciții de aplicare a algoritmului EM/GMM.	Idem	2

Bibliografie

Liviu Ciortuz, Alina Munteanu, Elena Bădărău. Exerciții de învățare automată. Editura Universității "A.I.Cuza", Iași, 2016, 2018.

9. Coroborarea conținutului disciplinei cu așteptările reprezentanților comunității, asociațiilor profesionale și angajatorilor reprezentativi din domeniul aferent programului

Conținutul disciplinei este coroborat cu tematica concursurilor pentru ocuparea posturilor în companiile IT.

10. Evaluare

Tip activitate	10.1 Criterii de evaluare	10.2 Metode de evaluare	10.3 Pondere în nota finală (%)
10.4 Curs	<ul style="list-style-type: none">- aplicarea corectă a algoritmilor de învățare automată predați, în funcție de diverse cerințe specifice;- calitatea formulării răspunsurilor;- estimarea parametrilor unor distribuții probabiliste;- demonstrarea unor proprietăți teoretice legate de chestiunile de mai sus.	2 examene scrise (sapt. 8, resp. 15-16), eventual teste scrise (inopinate).	60% (12 de puncte la fiecare din cele două examene parțiale).
10.5 Seminar	<ul style="list-style-type: none">- aplicarea concretă a algoritmilor ID3, Bayes Naiv/Corelat, k-NN, clusterizare ierarhică, K-means și EM/GMM pe dataset-uri didactice;	Răspunsuri la tablă, teste scrise.	30% (6 puncte pentru fiecare jumătate de semestru)



	<ul style="list-style-type: none">- estimarea parametrilor unor distribuții probabiliste;- demonstrarea unor proprietăți teoretice legate de chestiunile de mai sus;- elaborarea corectă și eficientă a unor implementări cu caracter demonstrativ.		
--	---	--	--

10.6 Standarde minime de performanță

Pentru promovare trebuie să îndeplinite simultan următoarele criterii:

- Minim 2 (din 6) puncte acumulate la seminar pe fiecare jumătate de semestru; minim 4 (din 12) puncte la fiecare examen parțial.
- Absențele la seminar se penalizează cu câte 0,1 puncte începând de la a doua absență încolo. (Depunctarea se aplică la fiecare jumătate de semestru, înainte de intrarea în examenul parțial.)

Nota: $1/4 \times (4 + \text{sumă punctajelor la examenele parțiale} + \text{punctajele de la seminar})$.
(Nota minimă: 5, după rotunjire)

Observație: Informații mai detaliate despre desfășurarea acestui curs se găsesc în fișierele <http://profs.info.uaic.ro/~ciortuz/SLIDES/ml0.pdf> (vedeți secțiunea Administrativă) și <http://profs.info.uaic.ro/~ciortuz/ML.2018f.what-you-should-know.pdf>.

Data completării
20 septembrie 2018

Titular de curs
Conf.univ.dr. Liviu Ciortuz

Titular de seminar
Conf.univ.dr. Liviu Ciortuz

Data avizării în departament

Director de departament