



Студијски програм: Мастер академске студије информатике			
Назив предмета: МАШИНСКО УЧЕЊЕ 2			
Статус предмета: Изборан на Наука о подацима			
Број ЕСПБ: 6			
Услов: Уписан одговарајући семестар			
Циљ предмета Оспособљавање студената за разумевање и практичну примену концепата ненадгледаног машинског учења.			
Исход предмета Савладано градиво омогућава студенту да: <ul style="list-style-type: none"> • Разуме кључне појмове ненадгледаног машинског учења. • Примени поступак избора и евалуације оптималних модела хијерархијске и нехијерархијске кластеризације. • Ефикасно примени фундаменталне алгоритме кластеризације у решавању проблема средње сложености. • Да познаје, оптимално бира и примењује посебене алгоритме кластеризације као што су Self-Organizing Map, Robust PCA, DBSCAN... 			
Садржај предмета <i>Теоријска настава</i> Области примене. Концепти машинског учења. Ненадгледано учење. Мере сличности и дистанце (Similarity and Distance). Методе најближих суседа. k-NN (k-Nearest Neighbors). Кластеризација. Припремање података за кластеризацију. Скалирање. Нормализација. Пондерисање. Кодирање категоријских података. Нетипичне тачке. Групе - кластери. Центар групе. Центроид групе. EM - Expectation-Maximization Algorithm. Јенсенова неједнакост. Модели Гаусових смеша (Gaussian Mixture Models). Одстојање између центара групе. Матрица сличности (similarity coefficients matrix). Шема агломерације (agglomeration schedule). Агломеративно кластерованье (bottom-up). Груписање раздвајањем (divisive clustering / top-down). Дендограми. Rescaled distance. Дијаграми леденица (icicle diagrams). Метод повезивања (linkage method). Минимално (nearest-neighbor), максимално (farthest-neighbor) и просечно одстојање (average linking). Метод варијансе (variance method). Ward процедура. Метода центроида (centroid method). Нехијерархијска кластеризација. k-Means. Методе секвенцијалног прага (sequential threshold). Непараметарски естиматори: Histogram, Kernel, k-Nearest Neighbor. Паралелни праг. Методе оптимизације. k-Medoids. Просечно унутар-кластерско растојање. Методе за одређивање оптималног броја кластера. Интерпретација кластера. Сингуларна декомпозиција матрица. Остале методе кластеризације: Self-Organizing Map, Robust PCA, isomap, Spectral Clustering. Хеуристике кластеризације. Методе засноване на густини (DBSCAN). Методе засноване на дубини. Правила придруживања. Recommender systems. <i>Практична настава</i> Имплементација решења у R окружењу (пакети: fpc, dbscan, RAD, vegan, speccalt, MASS, mclust, reccomenderlab, arules...). Рад на вежбама подразумева примену знања у решавању задатака у домену ненадгледаног машинског учења.			
Литература <ol style="list-style-type: none"> 1. Aggarwal, Charu C., and Chandan K. Reddy, eds. Data clustering: algorithms and applications. CRC press, 2013. 2. Ethem Alpaydın, Introduction to Machine Learning, Third Edition, The MIT Press Cambridge, Massachusetts, London, England, 2014. 3. Berry M. A., Linoff, S. G., Data Mining Techniques For Marketing, Sales, and Customer Relationship, Management; Second Edition, Wiley Publishing, Inc., Indianapolis, Indiana, 2004. 4. Jasna Soldic-Aleksić, Biljana Chroneos Krasavac, Kvantitativne tehnike u istraživanju tržišta, Primena SPSS računarkog paketa, Centar za izdavačku delatnost Ekonomskog fakulteta u Beogradu, Beograd, 2009. 			
Број часова	активне наставе	Теоријска настава:	Практична настава:
		2	1 + 1
Методе извођења наставе Проблемски-оријентисана настава, практична настава и вежбе уз софтверску подршку, самостални рад студената и консултације.			
Оцена знања (максимални број поена 100)			
Предиспитне обавезе	70 поена	Завршни испит	30 поена
активност у току предавања		писмени испит	20
практична настава	20	усмени испит	10
колоквијум-и	20		
семинар-и	30		