

Strojno učenje

Uvod u strojno učenje

prof. dr. sc. Bojana Dalbelo Bašić
doc. dr. sc. Jan Šnajder

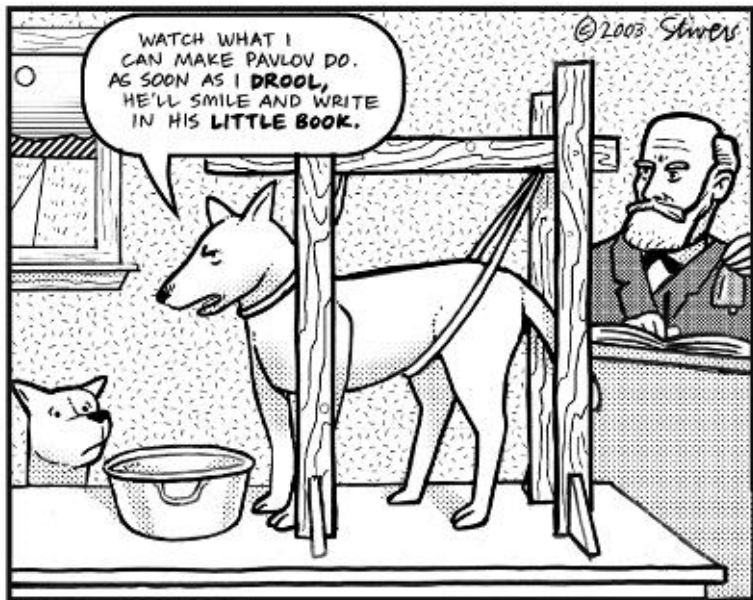
Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

Ak. god. 2011/12.



Nothing is as practical as a good theory.

— Kurt Lewin (1890.–1947.), psiholog



Data Mining Cup 2010

Using the existing characteristics of a customer's initial order, such as order quantity per type of goods, title and delivery weight, a decision must be made on whether to send a voucher worth EUR 5.00. The customers who receive a voucher should be those who would not have decided to re-order by themselves.

IEEE ICDM 2010 DM Competition

Modeling the process of traffic jams formation during morning peak in the presence of roadworks, based on initial information about jams broadcast by radio stations. Input data contain identifiers of road segments closed due to roadworks, accompanied by a sequence of segments where the first jams occurred. The algorithm should predict a sequence of segments where next jams will occur in the nearest future.

ACM KDD Cup 1999

Learn a predictive model (i.e. a classifier) capable of distinguishing between legitimate and illegitimate connections in a computer network.

ACM KDD Cup 2000

Given a set of page views, will the visitor view another page on the site or will the visitor leave? Given a set of page views, characterize killer pages, i.e., pages after which users leave the site. Given a set of page views, characterize which product brand a visitor will view in the remainder of the session.

- 1 Što je strojno učenje?
- 2 Srodna područja
- 3 Pregled postupaka
- 4 Literatura i internetski resursi

Što je strojno učenje?

Strojno učenje (Alpaydin 2009)

Strojno učenje jest programiranje računala na način da **optimiziraju** neki **kriterij uspješnosti** temeljem **podatkovnih primjera** ili prethodnog iskustva.

- Posjedujemo **model** koji je definiran do na neke **parametre**
- Učenje: optimizacija parametara modela temeljem podataka
- Model može biti **predikcijski** ili **deskriptivan**

Zašto strojno učenje?

Barem tri razloga:

- 1 **Složeni problemi** – ne postoji ljudsko znanje o procesu ili ljudi ne mogu dati objašnjenje o procesu (npr. raspoznavanje govora)
 - problemi koje nije moguće riješiti na klasičan algoritamski način (*AI-complete problems*)
- 2 **Ogromne količine podataka** – ima li znanja u njima?
 - otkrivanje znanja u skupovima podataka (engl. *data mining*)
- 3 **Sustavi koji se dinamički mijenjaju** – potrebna prilagodba (npr. prilagodba korisničkih sučelja)

NB: To ne znači da sve probleme treba rješavati strojnim učenjem (npr. program za obračun plaća)

Od podataka do znanja

- Učenje općenitih modela iz podataka: od podataka do znanja
- **Podataka** ima u izobilju (web, tekst, eksperimentalni podatci, skladišta podataka, deep web, dnevnici)

Koliko je ukupno pohranjenih podataka ?

Čovječanstvo je od 1986. godine pohranilo ukupno više od **295 eksabajta** (295×10^{18} bajtova) podataka (*Science Express*, 2011.)



Od podataka do znanja

- **Znanje** je skupo i potrebno
- Cilj: izgradnja modela koji je dobra i korisna **aproksimacija** podataka

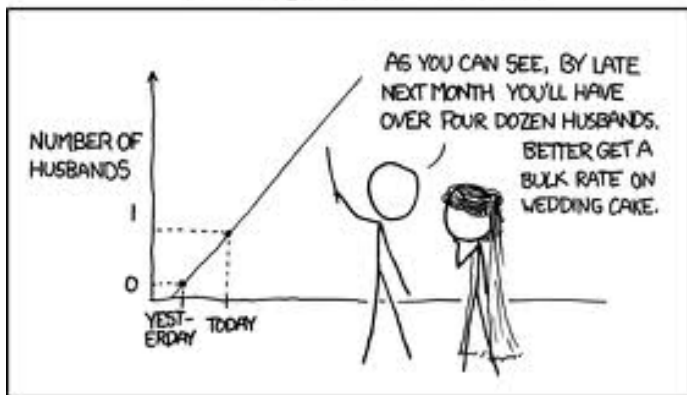
Primjer: Korisničke transakcije mogu objasniti ponašanje korisnika

People who bought "Da Vinci Code" also bought "The Five People You Meet in Heaven" (www.amazon.com)



Model: dobra i korisna aproksimacija!

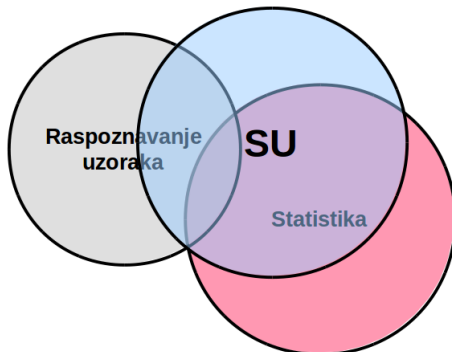
MY HOBBY: EXTRAPOLATING



- 1 Što je strojno učenje?
- 2 Srodna područja
- 3 Pregled postupaka
- 4 Literatura i internetski resursi

- Računarstvo, umjetna inteligencija
- Statistika i vjerojatnost (probabilističke metode)
- Raspoznavanje uzoraka
- Računalna teorija složenosti (teoretska ograničenja zbog složenosti učenja)
- Teorija informacije (mjere entropije, optimalno kodiranje...)
- Filozofija (Occamova britva – najjednostavnija hipoteza je najbolja)
- Psihologija i neurobiologija

Interdisciplinarnost strojnog učenja



- Temeljni pojmovi u strojnom učenju: **indukcija** i **generalizacija**
- Cilj: Za zadani uzorak ograničene veličine, pronaći opće pravilo koje objašnjava podatke
- **Statistika:** zaključivanje na temelju uzorka
 - generalizacija (engl. *generalisation*)
→ **zaključivanje** (engl. *inference*)
 - učenje (engl. *learning*)
→ **procjena** (engl. *estimation*)



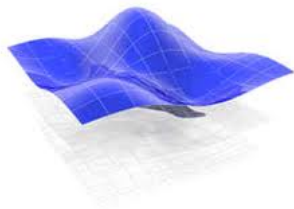
Strojno učenje i umjetna inteligencija

- Intelligentni sustav treba se **prilagođavati okolini**
 - imati sposobnost učenja. Ako može učiti onda može planirati ponašanje u novim situacijama
- Strojno učenje je okosnica Umjetne inteligencije
 - Robotika
 - Robotski vid
 - Raspoznavanje govora
 - Raspoznavanje uzoraka
 - Obrada prirodnog jezika: parsiranje, razrješavanje višeznačnosti, označavanje vrste riječi
 - Pretraživanje informacija: rangiranje, query log mining
 - Umjetne neuronske mreže
 - ...



Računarska znanost:

- Učinkoviti algoritmi koji rješavaju optimizacijske probleme
- Omogućava predstavljanje modela i njegovu evaluaciju u računalu
- Problemi prostorne i vremenske složenosti



Strojno učenje i kognitivna znanost

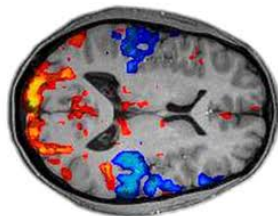
Razumijevanje algoritama strojnog učenja \Leftrightarrow Razumijevanje ljudske sposobnosti (ili ograničenja) učenja

Thought Reading Experiment:

<http://www.cs.cmu.edu/afs/cs/project/theo-73/www/index.html>

- **Funkcijska magnetska rezonancija (fMRI)**

- Bilježi protok krvi kroz mozak: aktivna područja mozga koriste više kisika.
- Oslanjanje na činjenicu da molekule u krvnim stanicama reagiraju u magnetskom polju u ovisnosti o količini kisika



- Nema univerzalnog algoritma za učenje! (*no free lunch theorem*)
 - izumljeni su učinkoviti algoritmi koji rješavaju određen tip problema
 - omogućili su bolje teoretsko razumijevanje učenja

Dubinska analiza podataka (engl. *data mining*) ili otkrivanje znanja u skupovima podataka (engl. *knowledge discovery in datasets*) – primjena strojnog učenja na velike baze podataka

- Trgovina: analiza potrošačke košarica, CRM
- Financije: Određivanje kreditne sposobnosti, detekcija zlouporaba kartica
- Proizvodnja: optimizacija, *troubleshooting*
- Medicina: postavljanje dijagnoza
- Telekomunikacije: optimizacija usluga
- Bioinformatika: analiza izražajnosti gena, poravnavanje
- Text mining: klasifikacija teksta, ekstrakcija informacija
- Računalni vid: prepoznavanje lica, praćenje vozila
- ...

- 1 Što je strojno učenje?
- 2 Srodna područja
- 3 Pregled postupaka**
- 4 Literatura i internetski resursi

Vrste strojnog učenja

Nadzirano učenje (engl. *supervised learning*)

Podatci su u obliku (ulaz,izlaz). Cilj učenja jest pronaći preslikavanje $\hat{y} = f(x)$ s ulaza na izlaz

- Ako je y je diskretna vrijednost: **klasifikacija**
- Ako je y kontinuirana vrijednost: **regresija**

Nenadzirano učenje (engl. *unsupervised learning*)

Dani su podaci bez ciljne vrijednosti. Cilj nenadziranog učenja jest pronaći pravilnosti u podacima

- **grupiranje** (engl. *clustering*)

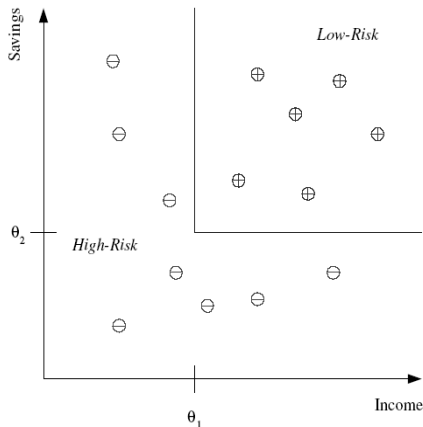
Podržano/ojačano učenje (engl. *reinforcement learning*)

Učenje optimalne strategije na temelju pokušaja s odgođenom nagradom

- **Predviđanje:** na temelju ulaznih vrijednosti predvidjeti buduće
- **Ekstrakcija znanja:** učenje lako tumačivih moela
- **Sažimanje:** model koje objašnjava podatke umjesto podataka
- **Otkrivanje ekstremnih vrijednosti:** iznimke koje nisu pokrivene modelom (npr. zlouporaba)
- **Upravljanje:** upravljački ulazi dobiveni regresijom

Primjer klasifikacije: Analiza kreditne sposobnosti

- Klasifikacija u svrhu **predviđanja**
- Cilj: Razlikovanje između grupa klijenata niskog rizika i visokog rizika na temelju podataka o njihovom prihodu i uštedjevini
- Diskriminacijska funkcija:
IF prihod $> \theta_1$ AND
uštedjevina $> \theta_2$ THEN
nizak rizik ELSE visok rizik

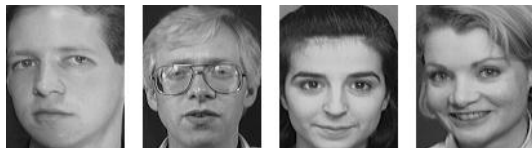


Primjer klasifikacije: raspoznavanje lica

- Cilj: prepoznati lice osobe unatoč promjenama u pozi, osvjetljenju, frizuri, šminki te okluzijama (naočale, brada)
- Skup podataka za učenje:



- Budući podatci:



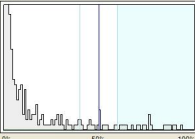
- Baze lica: <http://www.face-rec.org/databases/>

Primjer klasifikacije: kategorizacija novinskih članaka

Filter: One category view

Showing category 'POLITIKA|IZBORI|Parlamentarni izbori' from collection 'Collection0002'

Label distribution



0% 50% 100%

Show documents

Labeled: 0 0 5

Unlabeled: 4954 18 15

Manual label	Auto output	Auto rule	Id
Yes	0.531	0	97273
Yes	0.508	0	96049
Yes	0.899	0	98242
Yes	0.874	0	94597
No	0.855	0	95792
	0.844	0	94920
	0.778	0	97720
	0.769	0	97458
	0.767	0	94764
	0.766	0	96628
	0.750	0	97631
	0.733	0	96532
	0.715	0	96476
	0.704	0	96606
	0.676	0	97643
	0.646	0	93932
	0.636	0	94729
	0.627	0	97640

Article

Publisher: Glas Slavonije Date: 20.06.2005.

Vlada namjerava zabraniti koalicije prije izbora? ČELNICI VLADE RAZMATRAJU PROMJENE IZBORNIH PRAVILA

Vlada namjerava zabraniti koalicije prije izbora? Neslužbeno, HDZ za sljedeće izbore namjerava predložiti ili zabranu predizborno koaliciranja, ili povišenje praga na osam posto za koaliciju dvije, odnosno 11 posto za koaliciju tri ili više stranaka. ZAGREB - Vlada posljednjih dana intenzivno razmatra mogućnost izmjena izbornog zakona kojima bi se onemogućile manipulacije kakve su se pojavile nakon nedavnih lokalnih izbora, potvrdio nam je visoki Vladin dužnosnik. Kako se neslužbeno doznaje, HDZ za sljedeće parlamentarne i lokalne izbore namjerava predložiti ili zabranu predizborno koaliciranja, ili povišenje praga na osam posto za koaliciju dvije, odnosno 11 posto za koaliciju tri ili više stranaka. Zabranu koaliciranja prije izbora značila bi prihvatanje njemačkog modela, a povišenje izbornog praga povratka na izborna pravila po kojima su se u Hrvatskoj izbori održavali u devedesetima. Neslužbeno se može čuti da u Vladu stižu signali iz Europske komisije da konačno zakonski onemogući cirkuse s kupovanjima vijećnika i zastupnika, kojih smo bili svjedoci posljednjih tjedana. Vlada vjerojatno ipak neće brzati s izmjenama izbornih zakona zbog toga jer do redovitih parlamentarnih izbora ima dvije i pol, a do lokalnih čak četiri godine. No, HDZ-u je vjerojatno u interesu da se sljedeći parlamentarni izbori, ako oni budu i prijevremeni, održe po novim pravilima. Visoki Vladin dužnosnik otkriva nam da Banski dvori posljednjih dana dobivaju mnogo zahtjeva s terena, i to ne samo od HDZ-a nego i ostalih stranaka, da se utvrde nova pravila i onemogući lakdijpa na budućim izborima.

TEŠKO IZVEDIVA ZABRANA

Promjena izbornih pravila ponajprije bi pogodovala takozvanim velikim strankama, primjerice HDZ-u i SDP-u, jer bi pokupili bardo glasova koji se rasprše na male stranke, umjesto da nakon izbora pokušavaju privoljeti na suradnju njihove, često nepouzdate, zastupnike i vijećnike kako bi osigurali većinu.

Zastupnik SDP-a Mato Arlović kaže kako bi zabrana predizborno koaliciranja bila teško izvediva, ali i da je u svojoj biti suprotna ustavnim odredbama koje jamče višestranački sustav. Stoga Arlović smatra da ne treba zabranjivati predizborne koalicije, ali da svakako treba otežati njihovo formiranje. - Trebalo bi ići na proporcionalni sustav, ali onaj koji bi omogućio građanima da mogu glasovati za ljude i za njihov redoslijed na listi, a ne samo za stranačke liste. Da o tome iko će na kraju biti u Saboru odlučuju birači, a ne stranačka vodstva, kaže Arlović. Što se izbornih zakona tiče, izjavio je nakon što je 13. prosinca Arlović rekao da će poslije izbora biti teško izvesti zabranu koaliciranja.

URL:
 Etem id: -

Relative impacts

Feature	Impact
izbor	2.091
stranka	1.251
predizboran	1.033
izboran	0.904
parlamentarna	0.579
koalicija	0.515
izboriti	0.432
zastupnik	0.275
vlada	0.243
hdz	0.209
birač	0.155
jednica	0.120
glasovati	0.108
glas	0.105
namjeravati	0.097
izici	0.097
pravilo	0.092
onemogućiti	0.072
pravi	0.068
politički	0.058
manipulacija	0.054
sdp	0.050
taština	0.048
jamčiti	0.040
promjena	0.035
sabor	0.034
osigurati	0.033
vodstvo	0.029
čelnik	0.028
zabrana	0.028
signal	0.025
ustavan	0.022
kazati	0.019
odlučivati	0.019
kupovanje	0.016
kupovan	0.016
čuti	0.015
sljedeći	0.014
stizati	0.012
matr	0.011

Save Cancel

KTLab KTN indexer (<http://ktlab.fer.hr/en/products/63-ktn-indexing-system>)

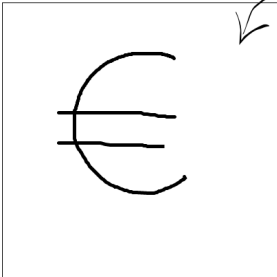
Primjer klasifikacije: Raspoznavanje simbola L^AT_EX-a

Detexify² - LaTeX symbol classifier

classify

symbols

blog



clear

What is this?

Anyone who works with LaTeX knows how time-consuming it can be to find a symbol in [symbols-a4.pdf](#) that you just can't memorize. Detexify is an attempt to simplify this search.

How do I use it?

Just draw the symbol you are looking for into the square area above and look what happens!

Did this help?

Hosting Detexify costs money and if it helps you may consider helping to pay the hosting bill.



\$727.41 Raised!

€

Score: 0.0521765839556664
`\usepackage{ marvosym }`
`\EURtm`
textmode

€

Score: 0.0521765839556664
`\usepackage{ marvosym }`
`\EURcr`
textmode

€

Score: 0.0597098127426264
`\usepackage{ marvosym }`
`\EUR`
textmode

€

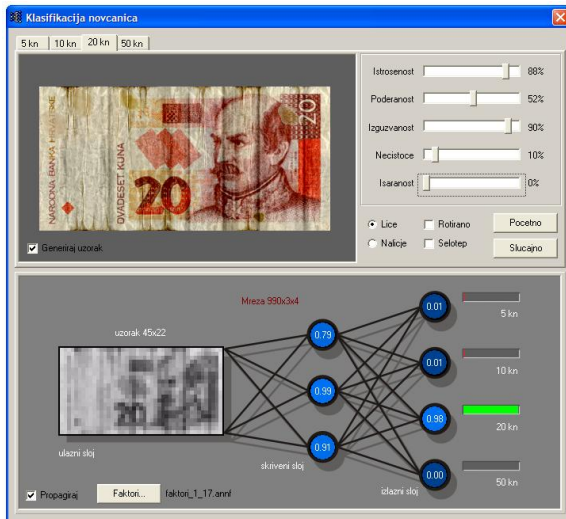
Score: 0.0600260193436406
`\usepackage{ marvosym }`
`\EURhv`
textmode

€

Score: 0.0705110677528292
`\usepackage{ marvosym }`
`\EURdig`
textmode

Detexify LaTeX symbol classifier (<http://detexify.kirelabs.org/>)

Primjer klasifikacije: Raspoznavanje novčanica



Neuronska mreža i generator uzoraka (<http://www.zemris.fer.hr/predmeti/su/seminari/>)

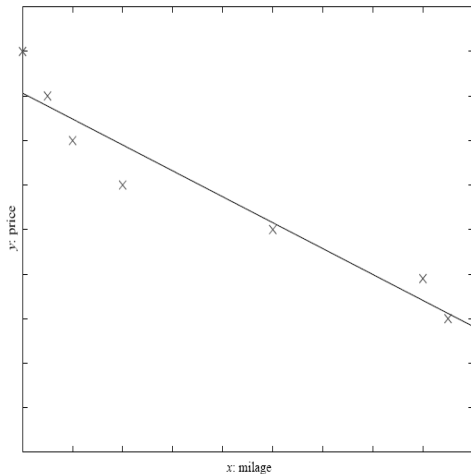
- Klasifikacija novinskih dokumenata u rubrike
- Detekcija neželjenih poruka e-pošte (engl. *spam detection*)
- Predviđanje kretanja dionica
- Određivanje smisla višeznačne riječi
- Raspoznavanje dlanova u svrhu autentikacije
- Automatsko dodjeljivanje ključnih riječi nekom dokumentu
- Medicinska dijagnostika (od simptoma do dijagnoze ili obrnuto)
- Prepoznavanje vrste plesa na temelju ritma
- Predviđanje ishoda nogometnih utakmica
- Klasificiranje mentalnog zdravlja autora teksta
- Predviđanje ocjene filma na temelju ocjena gledatelja

- Bayesov klasifikator
- Stroj s potpornim vektorima (engl. *support vector machine*, SVM)
- Stabla odluke (engl. *decision trees*)
- Algoritam k-najbližih susjeda
- Perceptron
- Neuronske mreže (višeslojni perceptron)
- Skriveni Markovljev model
- Logistička regresija
- ...

Regresija – primjer

Cilj: Predviđanje cijene rabljenih automobila

- x – atributi automobila (prijeđeni km)
- y – cijena
- $\hat{y} = f(x|w_0, w_1)$
- model:
 $f(x) = w_1x + w_0$
- w_0, w_1 – parametri modela



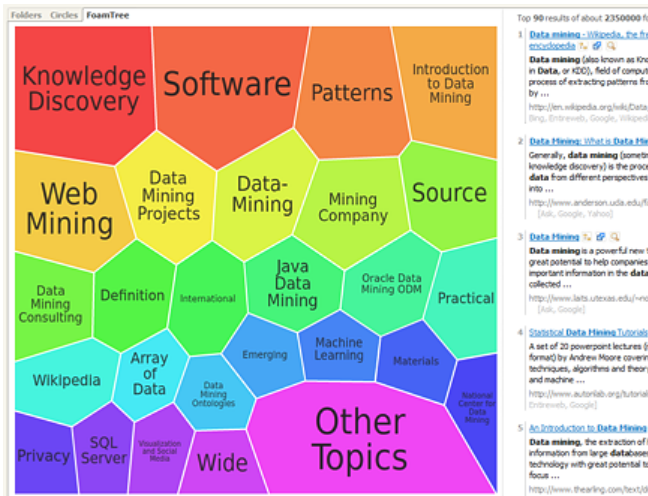
Nenadzirano učenje – primjene

- Dani su podaci, bez ciljne vrijednosti – **neoznačeni podaci** (engl. *unlabeled data*)
- Cilj nenadziranog učenja jest naći pravilnosti u podacima
- Tipične primjene:
 - Eksplorativna dubinska analiza podataka
 - Marketing: segmentacija korisnika
 - Biologija: grupiranje biljaka ili životinja prema njihovim značajkama
 - Text mining: grupiranje sličnih dokumenata
 - Pretraživanje informacija: grupiranje sličnih rezultata
 - Bioinformatika: grupiranje DNA-mikropolja
 - Obrada slike: sažimanje slike grupiranjem sl. elemenata sličnih boja

Grupiranje (engl. *clustering*)

Razvrstavanje podataka u grupe tako da slični podatci budu u istoj grupi, a različiti podatci u različitim grupama.

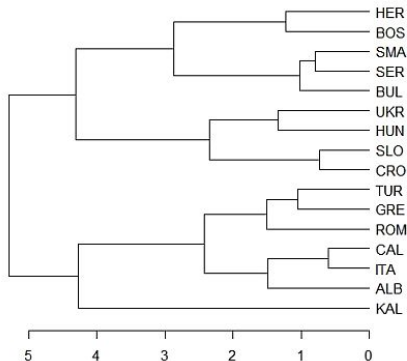
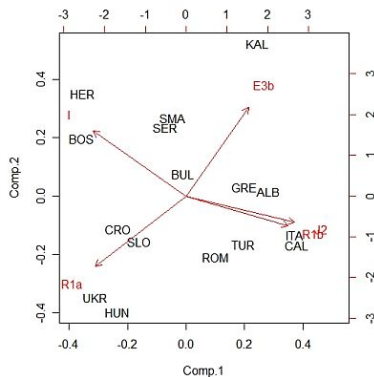
Primjer: grupiranje rezultata pretraživanja



Carrot² – Open Source Search Results Clustering Engine (<http://project.carrot2.org/>)

Primjer: grupiranje haploskupina (evolucijska biologija)

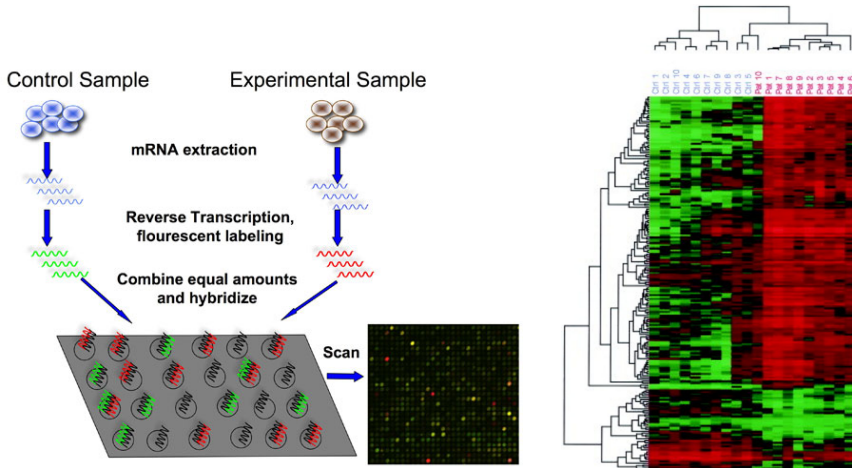
Haploskupine – nasljedno, polovično genetičko obilježje, korisno za analizu genetičkog podrijetla populacija



Dieneks' Anthropology Blog, <http://dienekes.blogspot.com/2005/08/haplogroup-frequency-correlations-in.html>

Primjer: grupiranje DNA-mikropolja (bioinformatika)

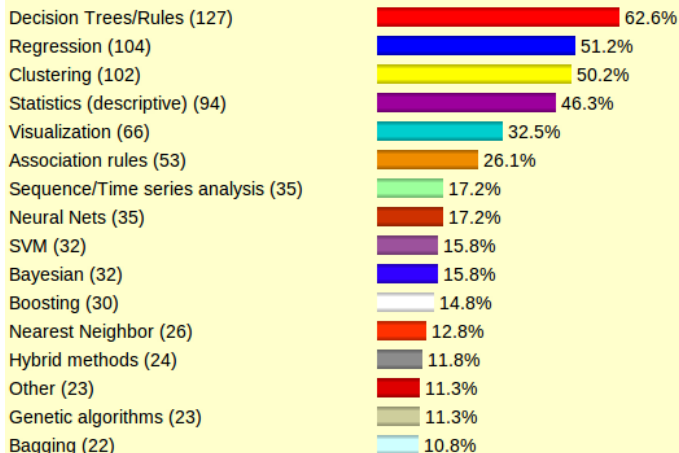
Cilj: grupiranje gena sa sličnom izražajnošću
(slična izražajnost – slična funkcionalnost)



- Algoritam k-srednjih vrijednosti
- Algoritam maksimizacije očekivanja
- Hijerarhijsko aglomerativno grupiranje
- DBSCAN

Algoritmi – popularnost (dubinska analiza podataka)

Data mining/analytic methods you used frequently in the past 12 months: [203 voters]



Izvor: KDnuggets polls 2007 (http://www.kdnuggets.com/polls/2007/data_mining_methods.htma)

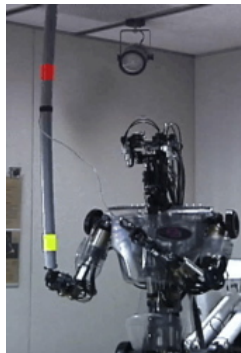
- Učenje strategije na temelju serije izlaza
- Nema nadziranog učenja – samo odgođena nagrada
- Problem dodjeljivanja nagrade (engl. *credit assignment problem*)
- Tipične primjene:
 - Igranje igara
 - Robotika i upravljanje
 - Višeagentski sustavi



Primjer: upravljanje humanoidnim robotom

Cilj: generiranje upravljačkih akcija za humanoidnog robota

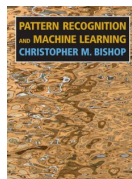
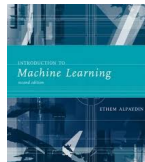
- problem: 7 ili više stupnjeva slobode (npr. ruka)
- prostor stanja ima 21 ili više dimenzija



TU Darmstadt: Intelligent Autonomous Systems (<http://www.robot-learning.de/Research/ReinforcementLearning>)

- 1 Što je strojno učenje?
- 2 Srodna područja
- 3 Pregled postupaka
- 4 Literatura i internetski resursi

- Ethem Alpaydin: *Introduction to Machine Learning*, MIT Press, 2009.
- Christopher Bishop: *Pattern Recognition and Machine Learning*, Springer, 2007.
- Tom Mitchell: *Machine Learning*, McGraw-Hill, 1997.



- Stephen Marsland *Machine Learning: An Algorithmic Perspective*, Chapman and Hall/CRC, 2009.
- Duda, Hart, Stork: *Pattern Classification*, Wiley-Interscience, 2000.
- Hastie, Tibshirani, Friedman: *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2003.
- Witten, Frank, Hall: *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2011.
- Daphne Koller: *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009.
- Japkowicz & Shah: *Evaluating Learning Algorithms: A Classification Perspective*, CUP, 2011.

- Hastie, Tibshirani, Friedman: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*
<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>
- MacKay: *Information Theory, Inference, and Learning Algorithms*
<http://www.inference.phy.cam.ac.uk/mackay/itila/book.html>
- Sutton & Barto. *Reinforcement Learning: An Introduction*
<http://webdocs.cs.ualberta.ca/~sutton/book/ebook/>
- Rasmussen & Williams: *Gaussian Processes for Machine Learning*
<http://www.gaussianprocess.org/gpml/chapters/>
- Barber: *Bayesian Reasoning and Machine Learning*
<http://www.cs.ucl.ac.uk/staff/d.barber/brml>
- Nilsson: *Introduction to Machine Learning*
<http://ai.stanford.edu/~nilsson/mlbook.html>

- International Conference on Machine Learning (ICML)
<http://www.machinelearning.org/icml.html>
- European Conference on Machine Learning (ECML)
ECML11: <http://www.ecmlpkdd2011.org/>
- Neural Information Processing Systems (NIPS)
<http://nips.cc/Conferences/>
- Uncertainty in Artificial Intelligence (UAI)
<http://www.auai.org/>
- Computational Learning Theory (COLT)
<http://www.learningtheory.org/>
- International Joint Conference on Artificial Intelligence (IJCAI)
<http://ijcai.org/>
- International Conference on Neural Networks (Europe)
ICANN11: <http://www.cis.hut.fi/icann2011/>

- Journal of Machine Learning Research (www.jmlr.org)
- Machine Learning (www.springer.com/computer/ai/journal/10994)
- Neural Computation
- Neural Networks
- IEEE Transactions on Neural Networks
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Annals of Statistics
- Journal of the American Statistical Association
- ...

- MetaOptimize QA
Strojno učenje, NLP, AI, IR, vizualizacija i analiza podataka
<http://metaoptimize.com/>
- CrossValidated QA
Statistika, dubinska analiza i vizualizacija podataka
<http://stats.stackexchange.com/>
- KD nuggets
<http://www.kdnuggets.com/>
- Data Mining Cup
<http://www.data-mining-cup.de/en/>
- Machine Learning Summer School (MLSS)
<http://www.mlss.cc/>

- Videlectures.net
videlectures.net/Top/Computer_Science/Machine_Learning/
- Andrew Ng (Stanford): Machine Learning lectures
academicearth.org/courses/machine-learning
- **Stanford's Machine Learning online course**
www.ml-class.org/

Programski alati

- Weka (GPL)

www.cs.waikato.ac.nz/ml/weka



- Rapid Minner (AGPL)

rapid-i.com



- Orange (GPL)

www.ailab.si/orange



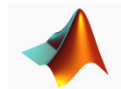
- R (GPL)

www.r-project.org

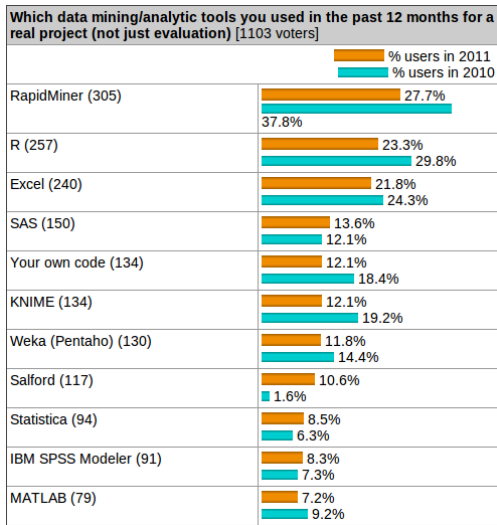


- Matlab

www.mathworks.com/products/matlab



Programski alati – popularnost






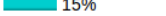
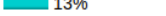
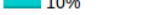



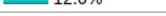



Izvor: KDnuggets polls (<http://www.kdnuggets.com/2011/05/tools-used-analytics-data-mining.html>)

Programski jezici – popularnost

What languages you used for data mining / data analysis?

What programming languages you used for data mining / data analysis in the past 12 months? [570 voters]

R (257)	 45%
SQL (184)	 32%
Python (140)	 25%
Java (139)	 24%
SAS (121)	 21%
MATLAB (83)	 15%
C/C++ (73)	 13%
Unix shell/awk/gawk/sed (59)	 10%
Perl (45)	 7.9%
Hadoop/Pig/Hive (35)	 6.1%
Lisp (4)	 0.7%
Other (70)	 12.0%
None (7)	 1.2%

Izvor: KDnuggets polls (<http://www.kdnuggets.com/2011/05/tools-used-analytics-data-mining.html>)

- UCI Repository
<http://www.ics.uci.edu/~mlearn/MLRepository.html>
- UCI KDD Archive
<http://kdd.ics.uci.edu/summary.data.application.html>
- Statlib
<http://lib.stat.cmu.edu/>
- Weka datasets
http://www.cs.waikato.ac.nz/ml/weka/index_datasets.html
- Delve
<http://www.cs.utoronto.ca/~delve/>