

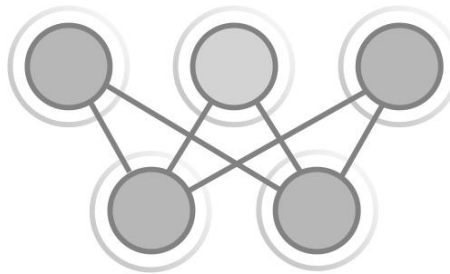
Prof.dr.sc. Bojana Dalbello Bašić

Fakultet elektrotehnike i računarstva
Zavod za elektroniku, mikroelektroniku, računalne i inteligentne sustave

www.zemris.fer.hr/~bojana
bojana.dalbello@fer.hr

Strojno učenje

Uvod





"Nothing is as practical as a good theory"

- **Predavanja**
 - Prof. dr. sc. Bojana Dalbelo Bašić
- **Projekti SU / Laboratorij računarske znanosti**
 - Frane Šarić, dipl. ing.
 - Artur Šilić dipl. ing.

<http://www.fer.hr/predmet/su>

- **Predavanja**

- četvrtkom 8 – 11 sati u B1

- **Konzultacije**

- Četvrtkom u 11-12 sati u D339C

- **Bodovanje aktivnosti:**

- **MI1: 20 bodova** (1. međuispit)
- **MI2: 25 bodova** (2. međuispit)
- **ZI: 35 bodova** (završni ispit)
- **LV-Projekti : 20 bodova**

- **+ do 5 bodova za aktivno sudjelovanje u nastavi**

- **LV/Projekti** (grupe po 5 studenata – voditelj grupe)
-program, podaci, dokumentacija, teorijska osnova, analiza rada grupe (detaljan opis doprinosa svakog člana)

Međuispiti:

- svaki ispit (međuispit, završni ispit) obuhvaća svo do tada obrađeno gradivo
- **Prag na završnom ispitu:**
 - $ZI \geq 12$
- **Prag prolaznosti:**
 - ukupno **50 bodova**

- **Nadoknade:**

- samo u **opravdanim slučajevima**
- tajnici ZEMRIS-a potrebno je predati **molbu i valjanu ispričnicu** najkasnije tjedan dana nakon propuštene aktivnosti

- **Međuispiti:**

- održat će se u roku od **14 dana** nakon međuispita
- način izvođenja: **50% pismeno i 50% usmeno**

- Što je strojno učenje?
- Zašto strojno učenje?
- Interdisciplinarnost strojnog učenja
- Vrste učenja
- Primjeri primjene
- Sadržaj predmeta
- Literatura

- SU - razvija se zadnjih 50-tak godina
- Odgovara na pitanje:

Kako napraviti računalne sustave koji automatski
poboljšavaju svoje performanse kroz iskustvo
i
koji su osnovni zakoni postupaka učenja?

- **Strojno učenje** bavi se izgradnjom sustava koji poboljšavaju performanse kroz iskustvo koristeći kriterijsku funkciju
- **Model** (hipoteza) je definiran i izgrađen do razine nepoznatih parametara, *strojno učenje* je programiranje računala da **podešava parametre modela** na temelju danih primjera
- Model može biti:
 - **prediktivni** (predviđa buduće vrijednosti) i
 - **deskriptivni** (sadrži znanje o podacima)

- **Složeni problemi** - ne postoji ljudsko znanje o procesu ili ljudi ne mogu dati objašnjenje o procesu (raspoznavanje govora) – programske implementacije koje nije moguće riješiti na klasičan način!
- Rastuće količine podataka – ima li znanja u njima?
Otkrivanje znanja u skupovima podataka (data & text mining)
- **Sustavi koji se dinamički mijenjaju** – potrebna prilagodba (prilagodba korisničkih sučelja)
- **ALI** ne treba “učiti” da bi se obračunale plaće

- Učenje općenitih modela iz danih podataka
- **Podataka** ima u izobilju – (skladišta podataka), **znanje** je skupo i potrebno.
- 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)

- Izgradnja modela koji je dobra i korisna aproksimacija podataka.

- umjetna inteligencija, računarstvo
- statistika i vjerojatnost (Bayesove metode)
- računalska teorija kompleksnosti (engl. computational complexity theory) - teoretska ograničenja zbog kompleksnosti zadatka učenja, mjerena u terminima računalskih resursa, broja primjera za učenje, broja pogrešaka, itd.
- teorija informacija (mjere entropije, optimalno kodiranje...)
- filozofija (Occam-ova britva – najjednostavnija hipoteza je najbolja)
- psihologija i neurobiologija

Strojno učenje i dubinska analiza podataka (eng. data mining)

Dubinska analiza podataka je područje primjene SU

- **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
- **Web mining:** Tražilice
- .

(Data mining – knowledge discovery in data sets)

- 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
- Raspoznavanje uzoraka: robotika, robotski vid
raspoznavanje govora, raspoznavanje lica,
- ...

- Temeljni pojam u strojnom učenju:

indukcija, generalizacija

- Cilj:

**Naći opće pravilo koje objašnjava podatke
ako je dan uzorak ograničene veličine**

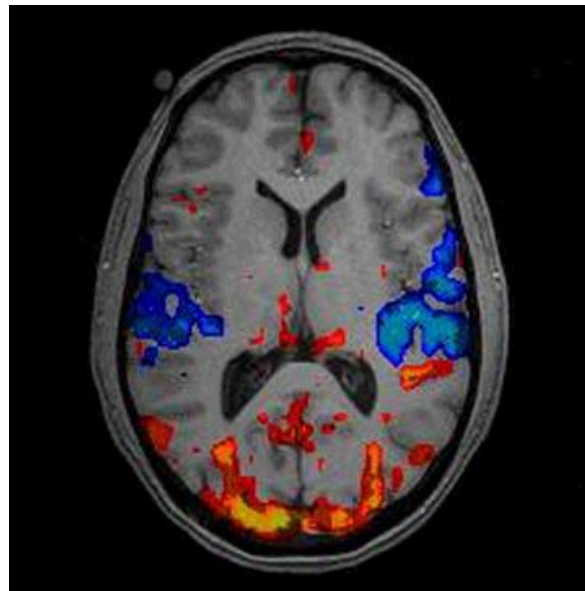
Statistika : Zaključivanje na temelju uzorka

(generalisation->inference ; laerning ->estimation)

- **Računarska znanost:** Učinkoviti algoritmi koji rješavaju
 - Rješavaju optimizacijske probleme
 - Omogućava predstavljanje modela i njegovu evaluaciju u računalu
 - Problemi prostorne i vremenske složenosti

Strojno učenje i kognitivna znanost - fMRI

- Bilježi protok krvi kroz mozak (hemodinamika). Aktivna područja mozga koriste više energije - kisika. Više krvi u određenim područjima zadovoljavaju potrebe aktivnih neurona.
- Oslanjanje na činjenicu da molekule u krvnim stanicama reagiraju u magnetskom polju u ovisnosti o količini kisika.

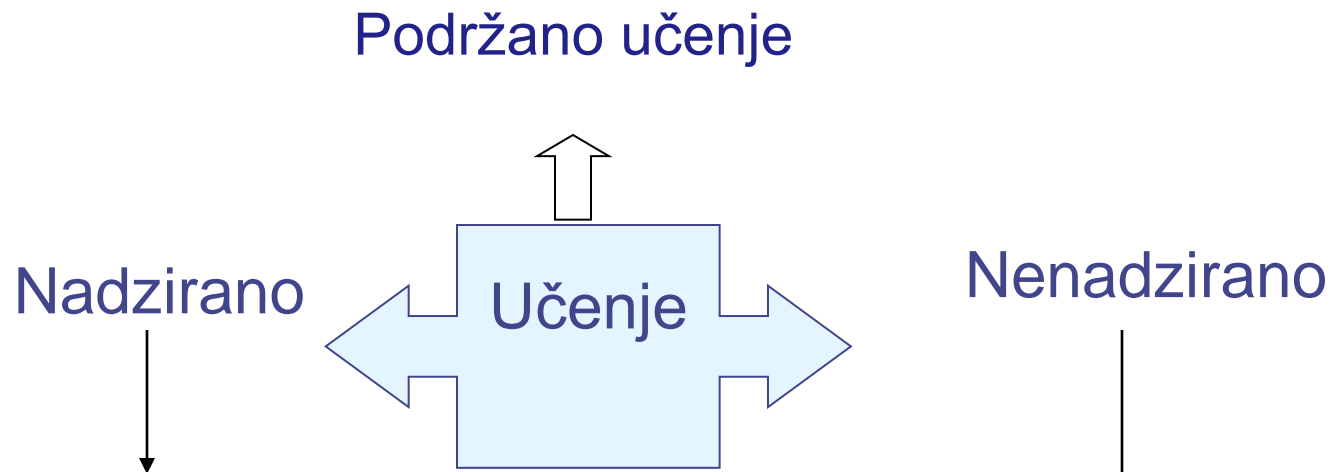


- Razumijevanje algoritama za strojno učenje
<->

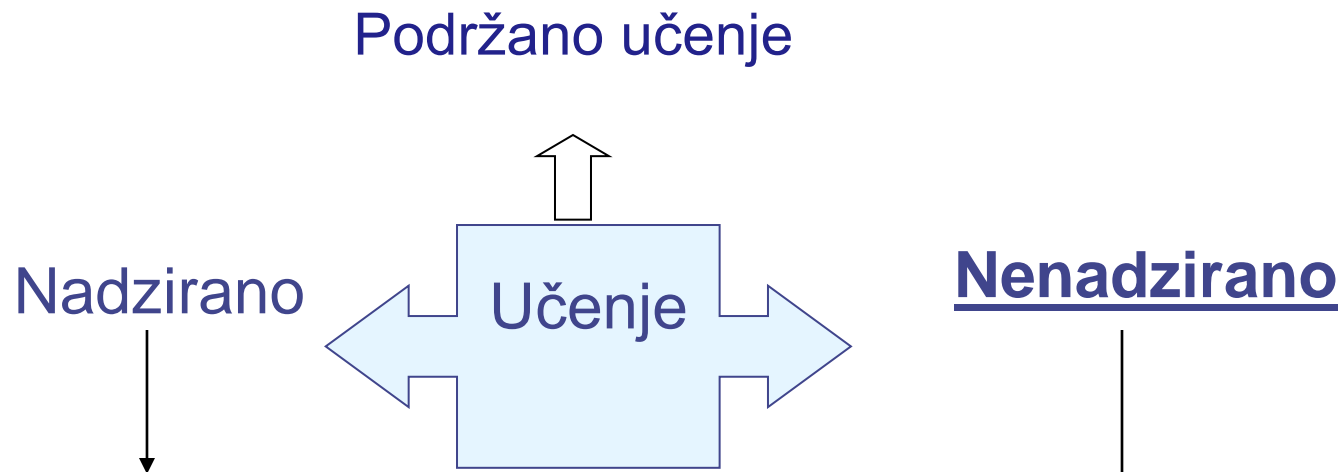
Razumijevanje ljudske sposobnosti (ili ograničenja) za učenje.

- <http://www.cs.cmu.edu/afs/cs/project/theo-73/www/index.html>
- ***Nema univerzalnog algoritma za učenje!*** – ipak izumljeni su efikasni algoritmi koji rješavaju određen tip problema (+ bolje teoretsko razumijevanje učenja).

- Učenje pravila (*eng. learning associations*)
- Nadzirano učenje (*eng. supervised learning*)
 - Klasifikacija
 - Regresija
- Nenadzirano učenje (*eng. unsupervised learning*)
- Podržano učenje (*eng. reinforcement learning*)



- Podaci su u dani u obliku (x, y) tj. ulazna vrijednost x , ciljna vrijednost y . Cilj u nadziranom učenju jest učenje ulazno izlaznog preslikavanja $y = f(x)$
- Dani su podaci, bez ciljne vrijednosti. Cilj nenadziranog učenja je naći pravilnosti u podacima

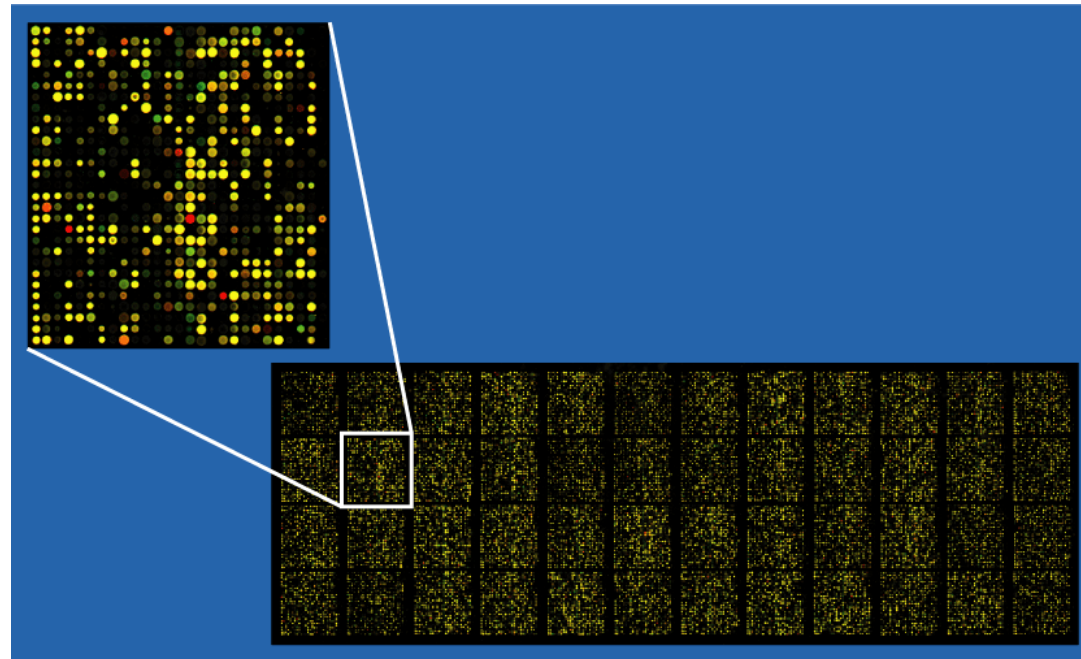


- Podaci su u dani u obliku (x, y) tj. ulazna vrijednost x , ciljna vrijednost y . Cilj u nadziranom učenju jest učenje ulazno izlaznog preslikavanja $y = f(x)$
- Dani su podaci, bez ciljne vrijednosti. Cilj nenadziranog učenja je naći pravilnosti u podacima

Nenadzirano učenje

- Dani su podaci, bez ciljne vrijednosti. **Cilj nenadziranog učenja je naći pravilnosti u podacima**
- Postoje pravilnosti u strukturi ulaznih podataka tako da se neki podaci pojavljuju češće od drugih
- Izradnja općenitih modela koji procjenjuju “što se obično dešava” naziva se procjena gustoće (*eng. density estimation*)
- **Primjer: grupiranje podataka** (*eng. clustering*) – grupiranje sličnih objekata

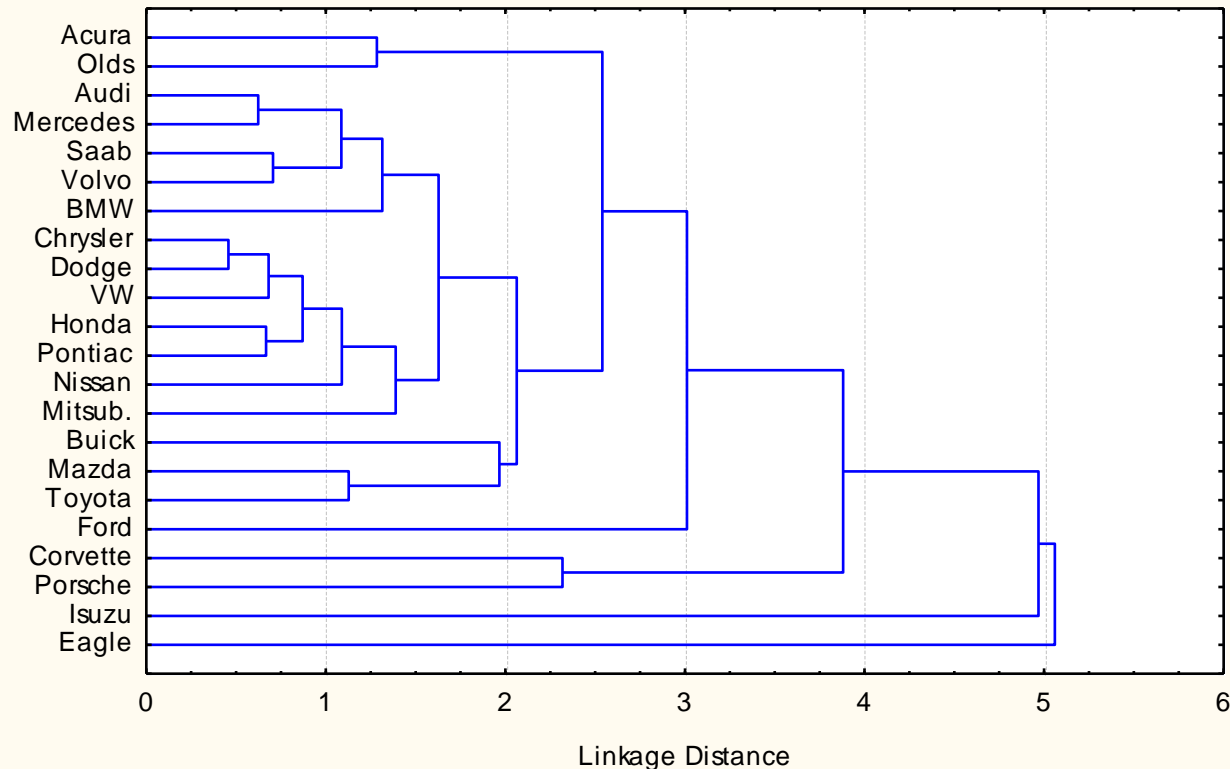
- Nema outputa
- Primjeri
 - Segmentacija korisnika/klijenata (CRM)
 - Kompresija slike
 - Bioinformatika

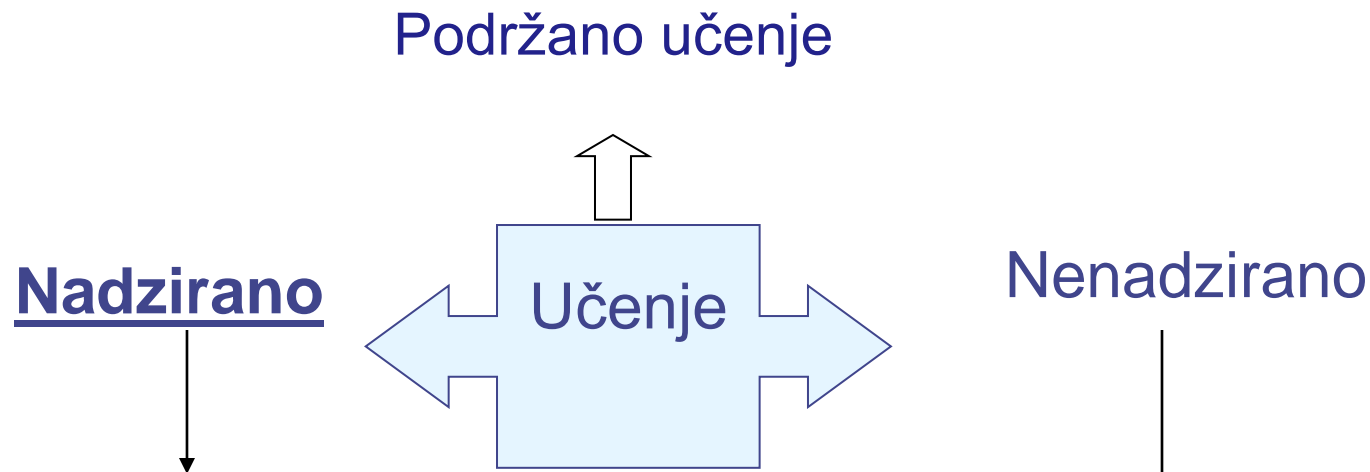


Performance, fuel economy, and approximate price for various automobiles

	PRICE - Approximate Price	ACCEL- ERATIO N - Acceler- ation	BRAKIN G - Breakin g from 80 mph	HANDLI NG - Road holding index	MILEAG E - Miles per gallon
Acura	-0,521	0,477	-0,007	0,382	2,079
Audi	0,866	0,208	0,319	-0,091	-0,677
BMW	0,496	-0,802	0,192	-0,091	-0,154
Buick	-0,614	1,689	0,933	-0,210	-0,154
Corvette	1,235	-1,811	-0,494	0,973	-0,677
Chrysler	-0,614	0,073	0,427	-0,210	-0,154
Dodge	-0,706	-0,196	0,481	0,145	-0,154
Eagle	-0,614	1,218	-4,199	-0,210	-0,677
Ford	-0,706	-1,542	0,987	0,145	-1,724
Honda	-0,429	0,410	-0,007	0,027	0,369
Isuzu	-0,798	0,410	-0,061	-4,230	1,067
Mazda	0,126	0,679	-0,133	0,500	-1,724
Mercedes	1,051	0,006	0,120	-0,091	-0,154
Mitsub.	-0,614	-1,003	0,084	0,382	0,718
Nissan	-0,429	0,073	-0,007	0,263	0,997
Olds	-0,614	-0,734	0,409	0,382	2,114
Pontiac	-0,614	0,679	0,536	0,145	0,195
Porsche	3,454	-2,215	-0,296	0,618	-1,026
Saab	0,588	0,679	0,246	0,263	0,021
Toyota	-0,059	1,218	0,228	0,736	-0,851
VW	-0,706	-0,128	0,102	0,382	0,195
Volvo	0,219	0,612	0,138	-0,210	0,369

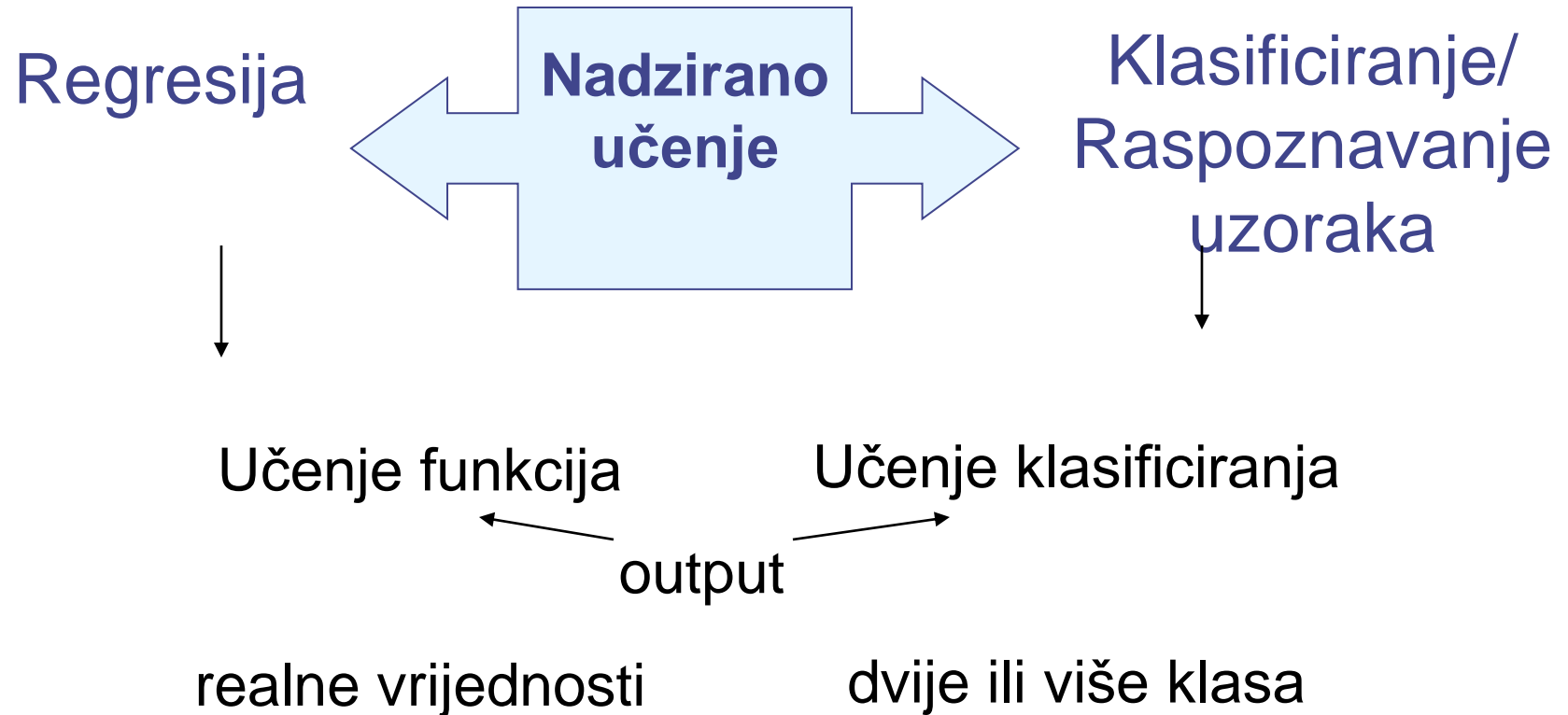
Tree Diagram for 22 Cases
Unweighted pair-group average
Euclidean distances





- Podaci su u dani u obliku (x, y) tj. ulazna vrijednost x , ciljna vrijednost y . Cilj u nadziranom učenju jest učenje ulazno izlaznog preslikavanja $y = f(x)$

- Dani su podaci, bez ciljne vrijednosti. Cilj nenadziranog učenja je naći pravilnosti u podacima



- **Predviđanje budućih slučajeva**: Na temelju ulaznih vrijednosti predvidjeti buduće
- **Ekstrakcija znanja** : Pravila su lako razumljiva
- **Kompresija**: Pravilo koje objašnjava podatke umjesto podataka
- **Detekcija ekstremnih vrijednosti**: Iznimke nisu pokrivenne pravilima, e.g., zlouporaba

Regresija

Nadzirano
učenje

Klasificiranje/
Raspoznavanje
uzoraka



Učenje funkcija

Učenje klasificiranja

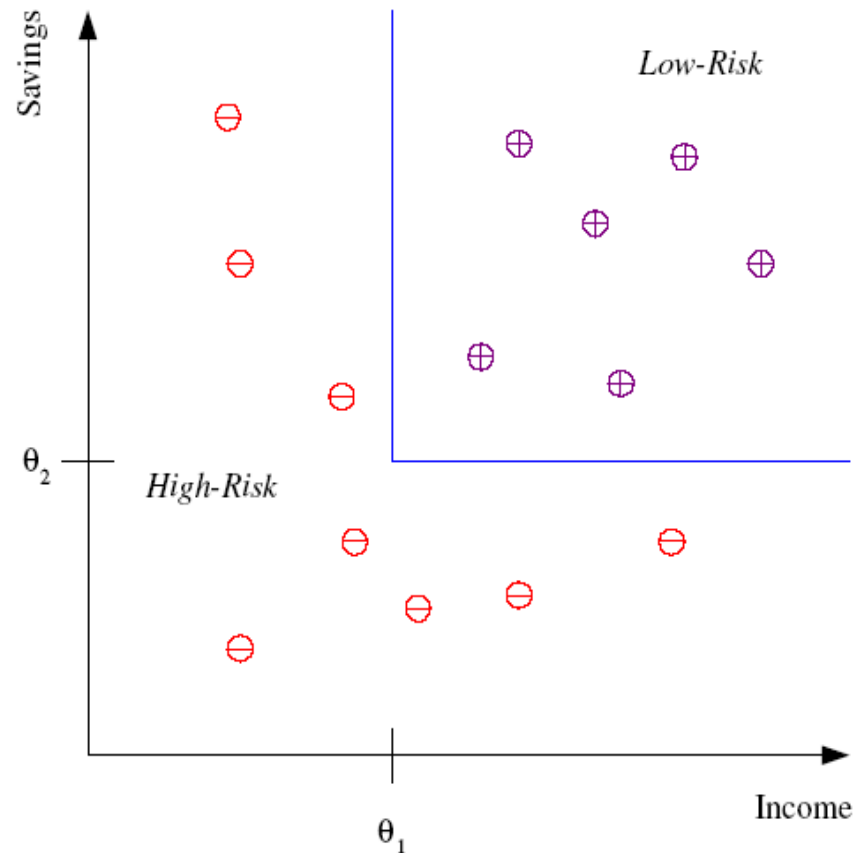
output

realne vrijednosti

dvije ili više klasa

Nadzirano učenje : Klasifikacija

- Primjer: analiza kreditne sposobnosti
- 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

Primjena je PREDVIĐANJE

Nadzirano učenje – klasifikacija -Raspoznavanje lica

Raspoznavanje lica: poza, osvjetljenje, okluzija
(naočale brada), frizure, make-up....

Podaci za učenje



Podaci za testiranje



AT&T Laboratories, Cambridge UK
<http://www.uk.research.att.com/facedatabase.html>

- **Raspoznavanje rukom pisanih znakova:** različiti stilovi pisanja

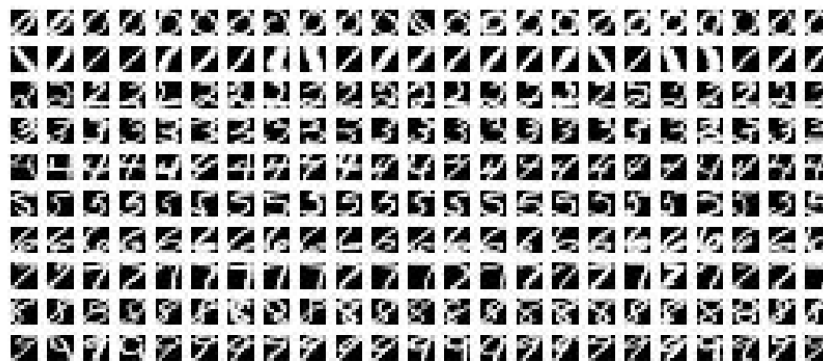


Fig. 3. Images of handwritten digits, normalized for horizontal and vertical scale and translation and sampled on an 8×8 pixel grid. Different writing angles introduce different levels of shearing in each image.

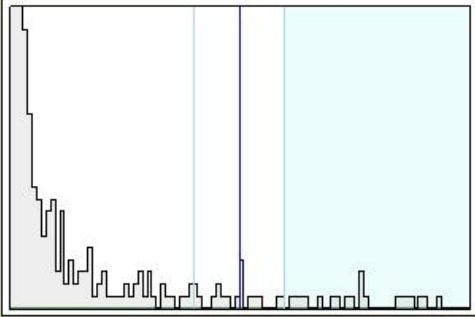
- **Raspoznavanje govora:**
- **Medicinska dijagnostika:** od simptoma do dijagnoze
- **Inteligentna analiza teksta (TM & IR) :** Automatska klasifikacija vijesti, dokumenta, web stranica, detekcija spam-a, automatsko sažimanje dokumenata, automatsko dodjeljivanje ključnih riječi...
- ..

Strojno učenje klasificiranja dokumenata

FilterOne category view

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

Label distribution



Show documents

NegativeUncertainPositive

Labeled: 005

Unlabeled: 49541815

Manual label	Auto output	Auto rule	Id
✓ Yes	0,931	0	97273
✓ Yes	0,908	0	96049
✓ Yes	0,899	0	98242
✓ Yes	0,874	0	94597
	0,869	0	97914
✗ No	0,855	0	95792
	0,844	0	94920
	0,778	0	97730
	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

<<<1/1>>>

SaveCancel

Article

Publisher: Glas SlavonijeDate: 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 predizbornog koaliranja, 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 predizbornog koaliranja, ili povišenje praga na osam posto za koaliciju dvije, odnosno 11 posto za koaliciju tri ili više stranaka. Zabrana koaliranja prije izbora značila bi prihvaćanje 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 onemogućiti 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 danadobivaju mnogo zahtjeva s terena, i to ne samo od HDZ-a nego i ostalih stranaka, da se utvrde nova pravila i onemogućiti lakrdija na budućim izborima.

TEŠKO IZVEDIVA ZABRANA

Promjena izbornih pravila ponajprije bi pogodovala takozvanim velikim strankama, primjerice HDZ-u i SDP-u, jer bi pokupili bar dio 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 predizbornog koaliranja 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 i za ljude i za njihov redoslijed na listi, a ne samo za stranačke liste. Da o tome tko će na kraju biti u Saboru odlučuju birači, a ne stranačka vodstva, kaže Arlović. Što se tiče zabranjivanja koaliranja od osam ili 11 posto, Arlović kaže da će ono

URL: Extent Id: -

Relative impacts

Feature	Impact
izbor	2,091
stranka	1,251
predizboran	1,033
izboran	0,904
parlamenta...	0,579
koalicija	0,515
izboriti	0,432
zastupnik	0,275
vlada	0,243
hdz	0,209
birač	0,155
jedinica	0,120
glasovati	0,108
glas	0,105
namjeravati	0,097
izići	0,097
pravilo	0,092
onemogućiti	0,072
praviti	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
mato	0,011

- Analiza potrošačke košarice:

$P(Y | X)$ vjerojatnost da netko tko kupi X također kupi Y
(X, Y su proizvodi, usluge).

Primjer: $P(\text{čips} | \text{pivo}) = 0.7$

Pravilo: 70% kupaca koji kupe pivo ujedno kupe i čips

Može nas zanimati i $P(X | Y, D)$, gdje su D atributi korisnika

Regresija

**Nadzirano
učenje**

**Klasificiranje/
Raspoznavanje
uzoraka**



Učenje funkcija

Učenje klasificiranja

output

realne vrijednosti

jedna ili više klasa

- Primjer: *Cijena rabljenih automobila*

- x : atributi auta (prijeđeni km)

y : cijena

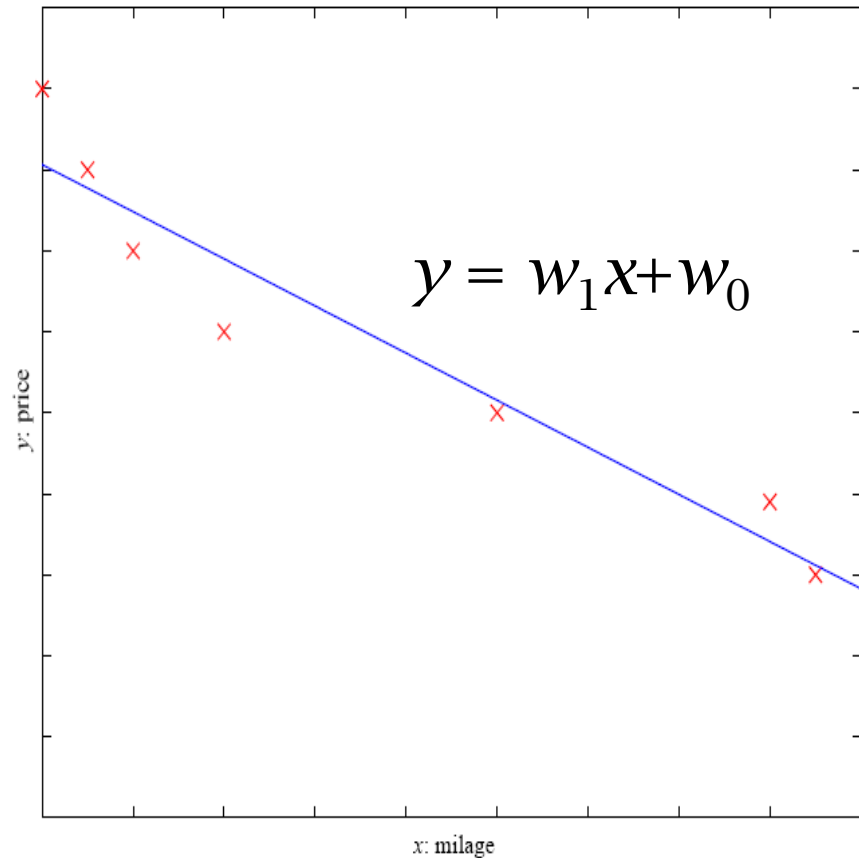
$$y = g(x | \theta)$$

$g()$ model,

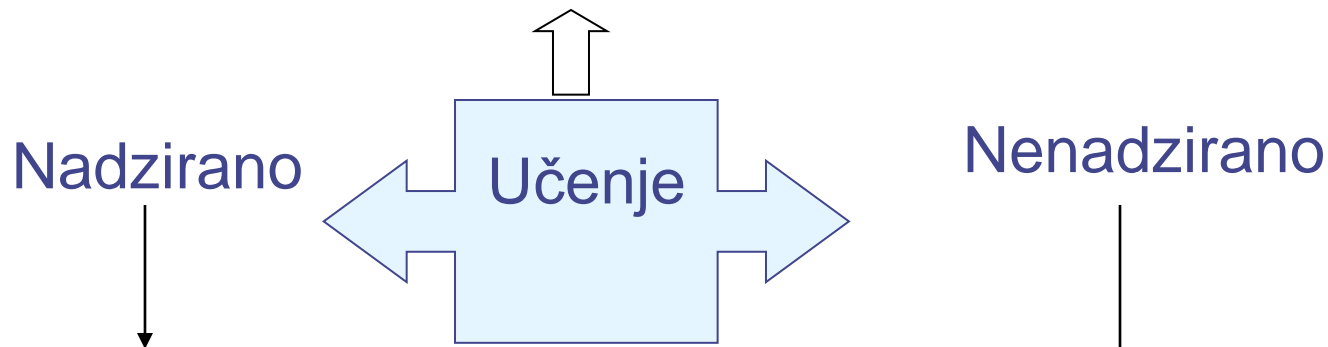
θ parametri – w , w_0
ili

$$y = w_2 x^2 + w_1 x + w_0$$

Problem linearan u
parametrima



Podržano učenje



- Podaci su u dani u obliku (x, y) tj. ulazna vrijednost x , ciljna vrijednost y . Cilj u nadziranom učenju jest učenje ulazno izlaznog preslikavanja $y = f(x)$
- Dani su podaci, bez ciljne vrijednosti. Cilj nenadziranog učenja je naći pravilnosti u podacima

- Učenje strategije na temelju serije izlaza.
- Nema nadziranog učenja samo odgođena nagrada
- Problem dodjeljivanja nagrade (*eng. credit assignment problem*)
- Igranje igara
- Robot u labirintu



Do sada smo govorili o znanju i učenju.

Izazov ...**stroj uči umjetničku interpretaciju**... ?

Primjer: glazbena umjetnost

G. Widmer, Application of Machine Learning to Music Research: Empirical Investigation into Phenomenon of Musical Expression, Machine Learning and Data Mining: Methods and Applications, pp. 269-293.

G. Widmer, Learning Musical Expressions, Machine Learning showcases and success stories, WWW stranice (Studenit, 2001.):

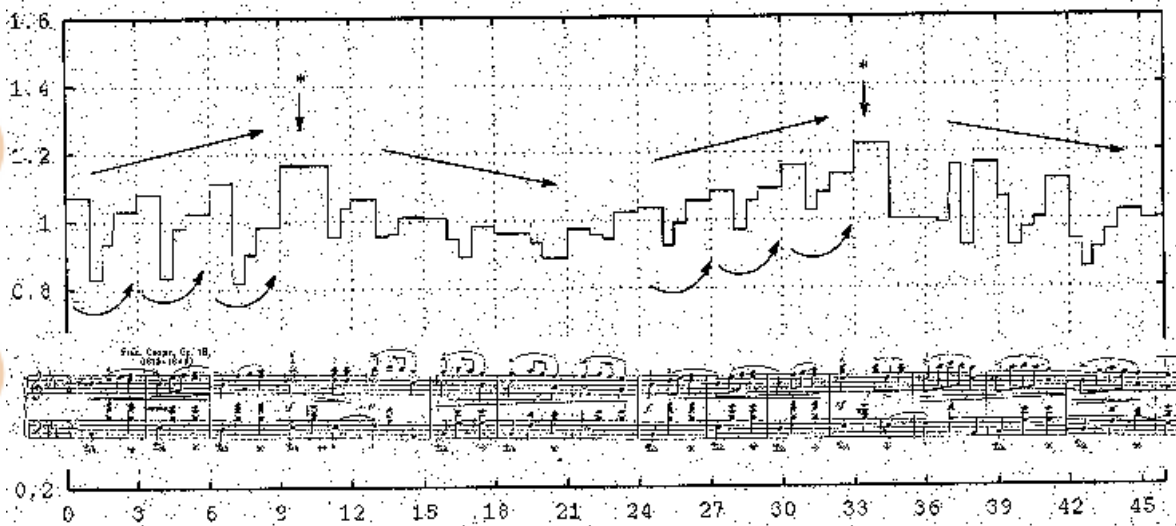
<http://www.cp.jku.at/people/widmer/>

Chopin Waltz op.18, Es dur (početak),

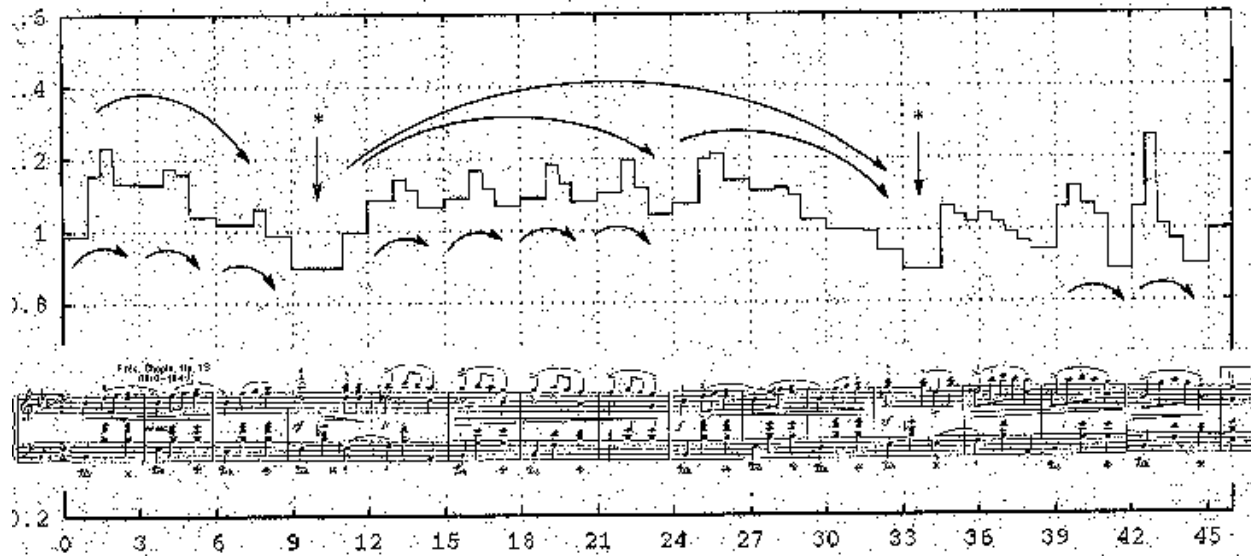
Primjer prije
učenja





Primjer poslije
učenja





Varijacija tempa u izvedbi



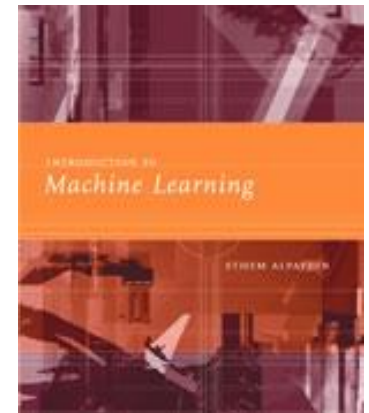
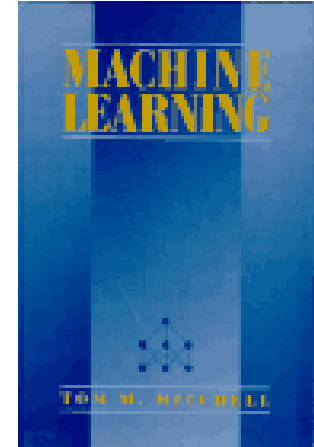
Chopin Waltz
op.18, Es dur
(početak),

Primjer prije
učenja  

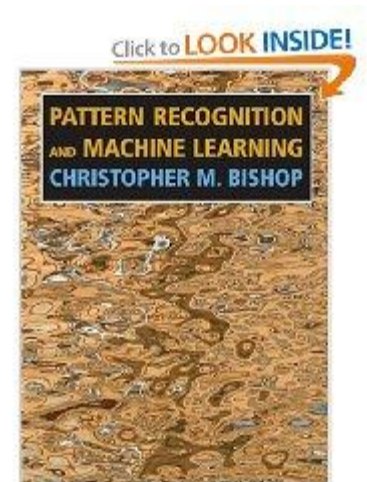
Primjer poslije
učenja  

- Osnovna načela strojnog učenja, Učenje koncepata, Stabla odluke, CART, statističko odlučivanje,
- Bayesove mreže, Parametarske metode, Ocjenjivanje i usporedba klasifikacijskih algoritama, kombiniranje više klasifikatora
- Neparametarske metode, grupiranje, (k-nn), podržano učenje

- Mitchell, Tom: “*Machine Learning*”, McGraw-Hill Comp., 1997.
(Prof. Tom M. Mitchell, Carnegie Mellon University, <http://www.cs.cmu.edu/~tom>)
- Alpaydin, Ethem: “*Introduction to Machine Learning*”, MT Press, 2004.
(<http://www.cmpe.boun.edu.tr/~ethem/i2ml/>)



- Elezović, Neven: “*Statistika i procesi*”, Element, Zagreb 2007.
- Bishop, Christopher: “*Pattern Recognition*” and Machine Learning, Springer, 2007.



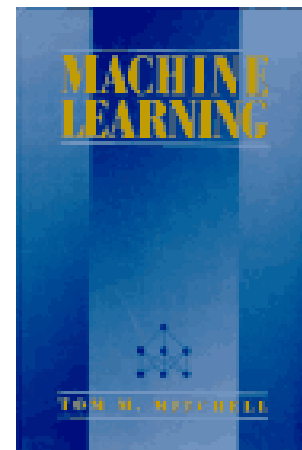
- UCI Repository:
<http://www.ics.uci.edu/~mlearn/MLRepository.html>
- UCI KDD Archive:
<http://kdd.ics.uci.edu/summary.data.application.html>
- Statlib: <http://lib.stat.cmu.edu/>
- Delve: <http://www.cs.utoronto.ca/~delve/>

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- 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
- ...

- International Conference on Machine Learning (ICML)
 - ICML05: <http://icml.ais.fraunhofer.de/>
- European Conference on Machine Learning (ECML)
 - ECML05: <http://ecmlpkdd05.liacc.up.pt/>
- Neural Information Processing Systems (NIPS)
 - NIPS05: <http://nips.cc/>
- Uncertainty in Artificial Intelligence (UAI)
 - UAI05: <http://www.cs.toronto.edu/uai2005/>
- Computational Learning Theory (COLT)
 - COLT05: <http://learningtheory.org/colt2005/>
- International Joint Conference on Artificial Intelligence (IJCAI)
 - IJCAI05: <http://ijcai05.csd.abdn.ac.uk/>
- International Conference on Neural Networks (Europe)
 - ICANN05: <http://www.ibspan.waw.pl/ICANN-2005/>
- ...

Literatura za uvodno predavanje

- Ch1 Introduction



- Ch1 Introduction

