Strojno učenje

Metode redukcije dimenzionalnosti

Tomislav Šmuc

Redukcija dimenzionalnosti

Problem više-dimenzionalnih prostora ("curse of dimensionality", Bellman,1961)

- Problemi vezani uz multivarijantnu analizu vezani uz povećanje dimenzionalnosti

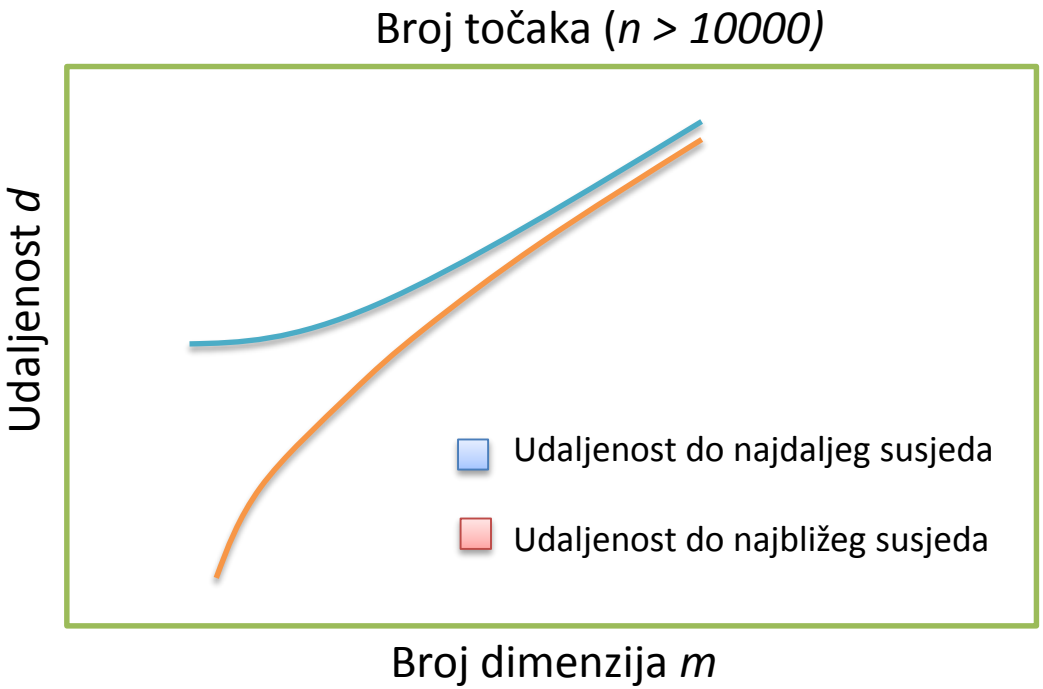
Implikacije više-dimenzionalnosti

- eksponencijalni porast primjera ako želimo sačuvati "gustoću" primjera:
  - uz "sačuvanu" gustoću primjera (N primjera/intervalu) i *m* dimenzija, ukupni broj primjera je N^m
- Eksponencijalno raste i kompleksnost ciljne funkcije:
  - funkcija u višedimenzionalnom vjerojatno će biti puno kompleksnija od one u niže-dimenzionalnom prostoru

Dva pristupa – redukciji dimenzionalnosti

1. Selekcija podskupa varijabli (onih koje su više-informativne)
2. Ekstrakcija manjeg broja novih dimenzija-varijabli - kombiniranjem postojećih

Prokletstvo velikog broja dimenzija



$$\lim_{m\to\infty} E\left(\frac{d_{max}(m) - d_{min}(m)}{d_{min}(m)}\right) \to 0$$

Za neki fiksni broj točaka *n* maksimalna i minimalna udaljenost između tih točaka i neke referentne točke postaje sve manja kako se povećava *m*;

Funkcije udaljenosti gube smisao u visoko-dimenzionalnom prostoru !

CoD - vrijedi uz pretpostavku da su individualne (jedno-dimenzionalne) distribucije nezavisne i uniformno distribuirane…

Realni podaci su u principu drukčijih karakteristika...

Selekcija varijabli

Svrha

- Smanjenje količine podataka (kod problema sa vrlo velikim brojem primjera i brojem varijabli/atributa: 1000-100000+)
- Smanjiti broj varijabli
- Odabrati samo najvažnije varijable/atribute
- Prednosti:
    - Bolji modeli (poboljšanje točnosti)
    - Brže učenje modela; kompatniji modeli (u eksploataciji)
    - Interpretabilniji modeli

	a_1	a_2	...	a_i	a_j	a_{n-1}	a_n
P_1							
P_2							
...							
P_m							

Važne/relevantne varijable

Svrha - realni problemi - primjeri

Klasifikacija teksta

– Varijable ~ 10^5 riječi, parovi riječi (?)
– Tipična praksa: koristi sve riječi - prepusti metodama odabira varijabli da riješe nekorisni višak varijabli
– Treniranje sa svim varijablama (riječima) je preskupo
– Prisutnost irelevantnih varijabli može negativno utjecati na generalizaciju

Klasifikacija tumora na osnovu ekspresije gena [Xing, Jordan, Karp '01]

– 72 pacijenta (primjeri)
– 7130 varijabli (nivoi ekspresije različitih gena)

Dijagnoza bolesti

– Varijable su rezultat (skupih) laboratorijskih testova
– koji bi test trebali napraviti na pacijentu?

Ugrađeni sustavi (Embedded systems) sa limitiranim resursima

– klasifikator mora biti kompaktan:
- npr. prepoznavanje glasa s mobitela
– Predikcija na CPU (4KB ograničenje!)

Neki osnovni pojmovi

Individualna irelevantnost varijable $\boldsymbol{X_i}$ (V^{-i}- podskup bez varijable X_i)

$$P(X_i, Y|\boldsymbol{V}^{-i}) = P(X_i|\boldsymbol{V}^{-i})P(Y|\boldsymbol{V}^{-i}).$$

Dostatni podskup varijabli $\boldsymbol{V} \subset \boldsymbol{X}$

$$P(Y|\boldsymbol{V}) = P(Y|\boldsymbol{X}).$$

Gornji izrazi su u praksi nerealni (znak jednakosti)

- u stvarnosti radi se obično o približno zadovoljenim tvrdnjama (razlika < ε), t.j.:
  - Vjerojatno približno irelevantnim varijablama
  - Minimalnom približno dostatnom podskupu varijabli $\boldsymbol{V}$

Odabir metode

Odabir metode selekcije varijabli ovisi:

- skupu varijabli i ciljnoj varijabli (binarne, kategoričke, kontinuirane)
- samom problemu (kakve su zavisnosti između varijabli/ciljne varijable, linearne/ne-linearne)
- Količini dostupnih podataka (odnosu broja primjera naspram broja varijabli, točnosti podataka - šum)

Osnovna podjela

- **Pojedinačno rangiranje**
  –> nezavisna ocjena svake varijable pojedinačno

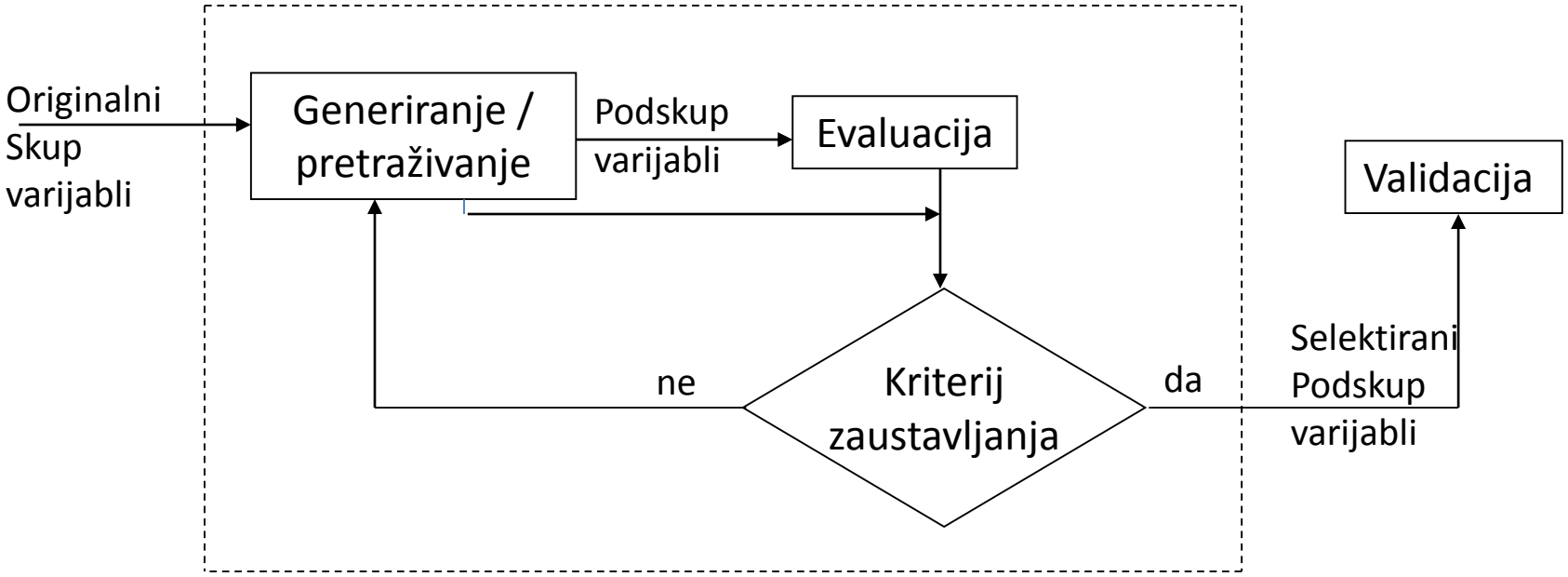
- **Grupno (multivarijantno) rangiranje**
  –> uzima u obzir istovremeno skup varijabli

- **Filter metode**
  –> rangiranje varijabli ili skupa varijabli na bazi indeksa(relevantnosti), nezavisno od algoritma za učenje(klasifikatora)

- **Metode "omotača" (wrapper) metode**
  -> koristi se klasifikator da bi se odredila vrijednost varijabli ili skupa varijabli

- **Ugrađene metode (embedded) ili algoritmi**
  -> istovremeno zajednički se uči i model i selekcija varijabli
  (primjer stabla odlučivanja !)

Generalna shema procesa odabira varijabli



Generiranje/pretraživanje - odabir podskupa (varijable)
Evaluacija - izračunati relevantnost podskupa varijabli.
Kriterij zaustavljanja - odrediti da li je podskup relevantan.
Validacija - nezav. verificiranje odabranog podskupa

Generiranje/pretraživanje:

- odabir podskupa ili varijable za evaluaciju
  - Početak: prazan skup, sve varijable, slučajno generirani podskup.
  - Inkrement : dodavanje, uklanjanje, dod/ukl varijabli
- kategorizacija načina generiranja/pretraživanja

  - Iscrpne (exhaustive, complete)
  - Heurističke
  - Slučajne

Proces: Generiranje

Iscrpno (complete/exhaustive)

- Ispitivanje svih kombinacija podskupova varijabli
  - {a1,a2,a3} => { {a1},{a2},{a3},{a1,a2},{a1,a3},{a2,a3},{a1,a2,a3} }
- Prostor pretraživanja - $O(2^m)$, m - # varijabli
- Optimalni podskup je dohvatljiv
- Nedopustiva složenost za m>>

Heuristički pristup

- Selekcija po određenom principu
  - izbacivanje varijabli
  - kandidati = { {f1,f2,f3}, {f2,f3}, {f3} }
- inkrementalno generiranje podskupova
- Prostor pretraživanja je drastično manji
- Neki od relevantnih podskupova varijabli mogu biti preskočeni !

Slučajno generiranje

- Slučajni odabir varijable (Probabilistički pristup)
- Optimalni podskup zavisi od broja pokušaja (~ ovisi o resursima)

Proces: Evaluacija

Evaluacija

- Odredi važnost generiranog podskupa varijabli za klasifikacijski problem

  R_v = J(podskupa varijabli)

  if (R_v > best_R_v) => best_R_v = R_v

- Osnovni tipovi evaluacijskih funkcija.
  - udaljenost

    (euklidska udaljenost, Manh. Udaljenost, sl.)
  - informacijske mjere

    (entropija, Infogain - porast informacije, sl.)
  - zavisnost između varijable i ciljne varijable

    (Pearsonov korelacijski koeficijent)
  - konzistentnost podskupa

    (minimalni konzistentan broj varijabli)

  } Filter metode

  - pogreška klasifikatora

  } "wrapper"

Proces: Evaluacija

Udaljenost

- $z^2 = x^2 + y^2$
- Selektirati one varijable koje podupiru “bliskost” primjera iste klase
- Primjeri iste klase morali bi biti međusobno bliži u smislu udaljenosti nego primjeri različitih klasa

Informacijske mjere

- Entropija – mjera sadržaja informacije
- Info-gain varijable : (kao kod stabla odlučivanja)
  IG(V) = I(p,n) - E(V)
  IG(V) = prije grananja – suma po svim čvorovima poslije grananja
- Odaberi A <- if IG(A) > IG(B).

Mjere zavisnosti

- Korelacija **između varijable i ciljne varijable**

$$C(j) = \frac{|\sum_{i=1}^{m}(x_{i,j} - \bar{x}_j)(y_i - \bar{y})|}{\sqrt{\sum_{i=1}^{m}(x_{i,j} - \bar{x}_j)^2 \sum_{i=1}^{m}(y_i - \bar{y})^2}} ,$$

- Zavisnost **između prediktorskih varijabli = nivo redundancije**
  - ako je neka varijabla zavisna o drugoj, tada je i redundantna

Konzistentnost

- Nekonzistentni primjeri - ako imaju iste vrijednosti varijabli - različite klase

Primjer	f1	f2	klasa
P1	a	b	C1
P2	a	b	C2

- Odaberi {f1,f2}
  => ako u trening setu nema primjera kao u gornjoj tablici
- min-feature = traži se najmanji podskup koji je konzistentan

Proces: Evaluacija

Klasifikacijska pogreška

- Samo u “wrapper” metodi

  evaluacija = pogreška_klasifikatora(podskup varijabli)
  if (error_rate < predefinirani threshold) select the feature subset

- selekcija varijabli – nije generalna (zavisi o konkretnom klasifikatoru), ali poboljšava točnost konkretnog klasifikatora na danom problemu
- Računalno skup pristup

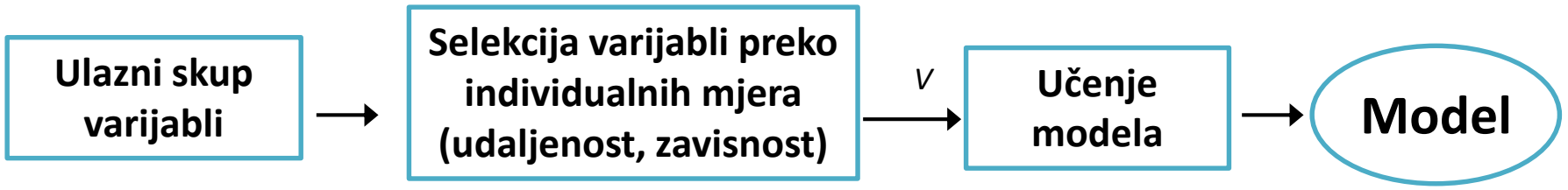
Proces: Evaluacija

metoda	Primjenjivost/ općenitost	Vremenska složenost	Točnost
Udaljenost	Da	Mala	*
Informacijske mjere	Da	Mala	*
Zavisnost	Da	Mala	*
Konzistentnost	Da	Srednja	*
Klasifikacijska točnost	ne	Velika	Visoka

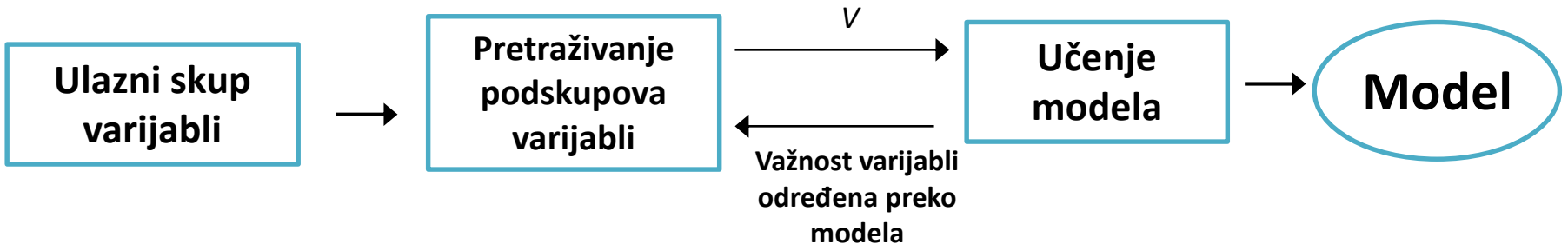
Primjenjivost – koliko su rezultati generalni (primjenjivi) kod različitih klasifikatora
Točnost – koliko je točan konačni (klasifikacijski) model
(*) Točnost ovisi od konkretne kombinacije metode selekcije i algoritma za klasifikaciju

Filter, “wrapper” i ugrađene metode

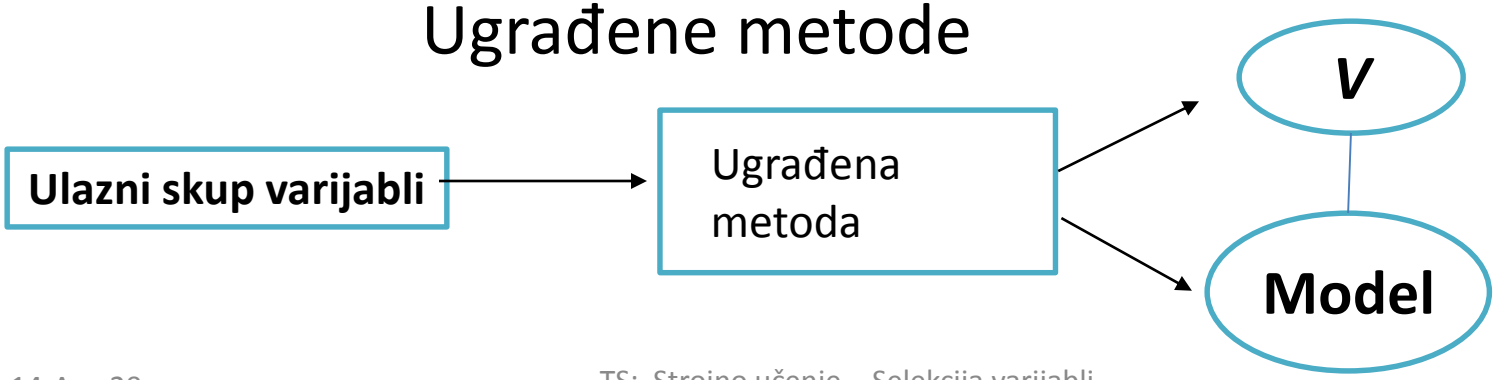
Filter



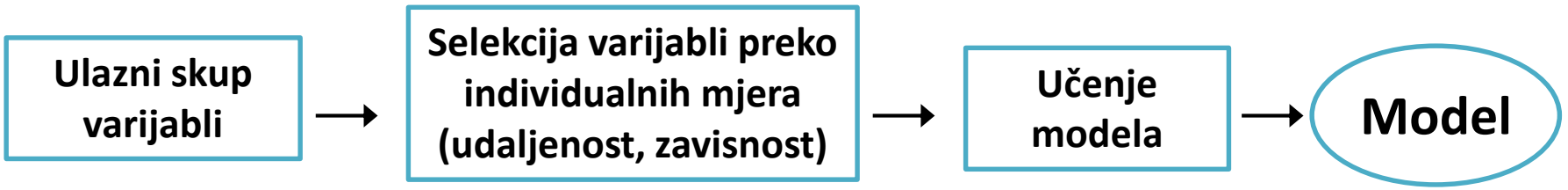
“Wrapper”



Ugrađene metode



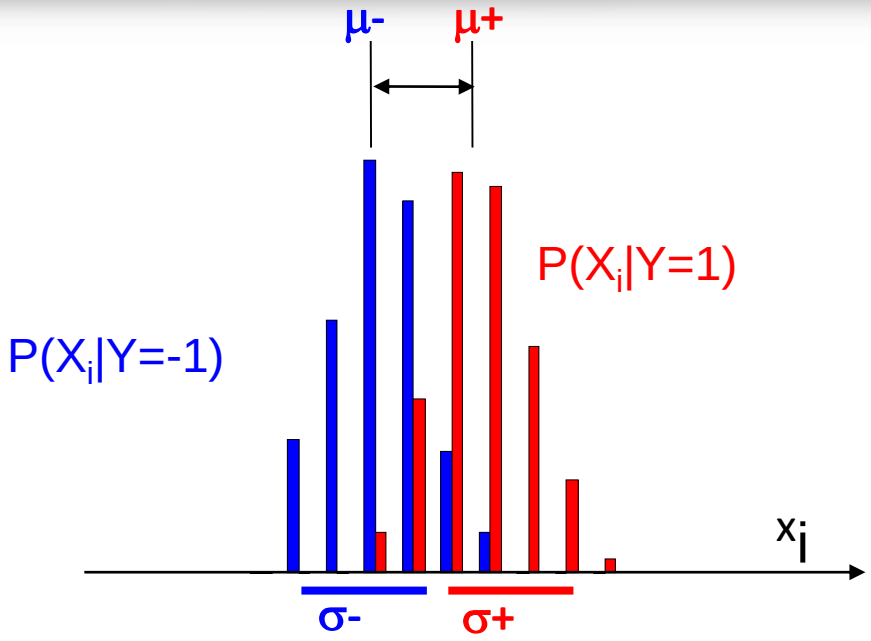
Filter



- Varijable se evaluiraju individualno a najboljih *v* se selektira i kasnije koristi u učenju modela

- Evaluacijske mjere**: korelacija, uzajamna informacija, t-test,** itd.

Univarijantne metode selekcije: primjer



- Normalna distr. klasa, σ^2 nepoznato - procjena iz podataka kao σ^2_p.
- Nulta hipoteza H_0: $\mu+ = \mu-$
- **T - test**:

$$t = (\mu^+ - \mu^-)/(\sigma_p \sqrt{1/m^+ + 1/m^-}) \propto Student(m^+ + m^- - 2d.f.)$$

Problemi Univarijantnog filter pristupa

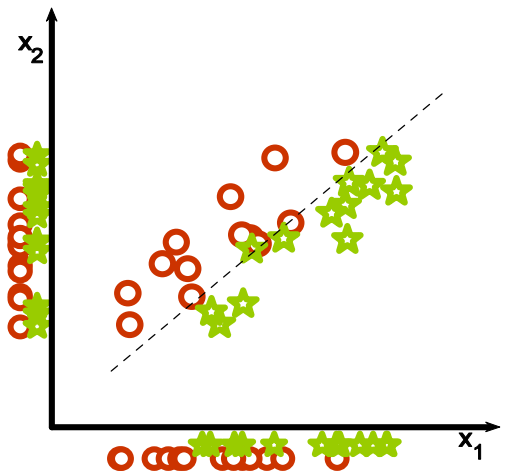
- **Redundancija odabranih varijabli** => varijable su odabrane nezavisno, ne kontrolira se donose li dodatnu informaciju u skup

- **Interakcije između varijabli** ne mogu se eksplicitno uključiti u određivanje podskupa varijabli. Individualno nevažne varijable, mogu biti važne u interakciji!

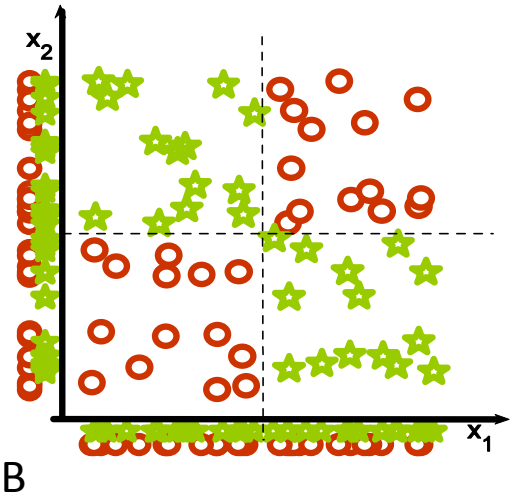
- **Zanemarena je važnost (specifičnost) klasifikacijskog algoritma**: neke filter metode su prikladne za određene klasifikatore, a za neke nisu.

Problemi univarijantnog filter pristupa

Guyon-Elisseeff, JMLR 2004



A



B

Prediktivna snaga varijabli - kada se promatraju zajedno

A) x_2 – irelevantna varijabla sama za sebe; relevantna u kombinaciji s x_1

B) Dvije varijable koje su individualno irelevantne - postaju relevantne u kombinaciji

Multivarijantni filter pristup

Relief algoritam [generiranje=heurističko, evaluacija=udaljenost].

- Osnove
  - svaka varijabla dobiva kumulativno težinu koja se određuje preko primjera iz trening skupa
  - varijable sa težinom iznad zadane vrijednosti *T* se selektiraju u odabrani skup varijabli
- Određivanje težine varijabli
  - princip => primjeri koji pripadaju istoj klasi trebali bi biti bliže negoli primjeri različite klase
  - **bliski-pogodak** (near-hit) primjer = najbliži primjer iste klase
  - **bliski-promašaj** (near-miss) primjer = najbliži primjer suprotne klase
  - **update mehanizam za težine** => $W = W - d(X,nearhit)^2 + d(X,nearmiss)^2$

Relief algoritam

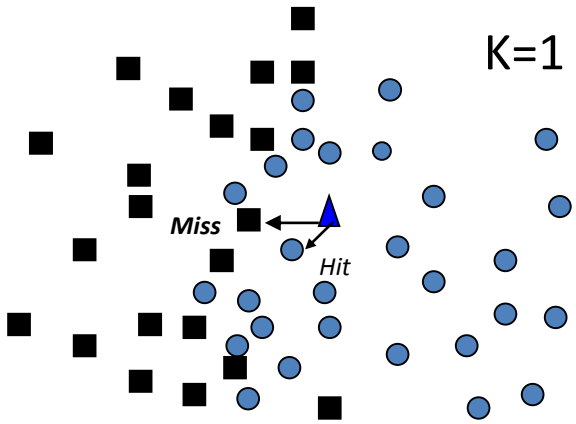
1. odabrani_podskup = {}

2. inicijaliziraj težine varijabli $w_i = 0$ (i=1,M)

3. za i = 1 to N % N - # primjera
   uzmi jedan primjer X iz trening skupa D.
   pronađi ***bliski-pogodak*** H = primjer iz D za kojeg je d(X,H) udaljenost najmanja & X.class=H.class
   pronađi ***bliski promašaj*** M = primjer iz D za kojeg je d(X,M) udaljenost najmanja & X.class<>M.class
   osvježi težine svih varijabli:
   $w_i = w_i - d(X,H)^2 /N + d(X,M)^2 /N$

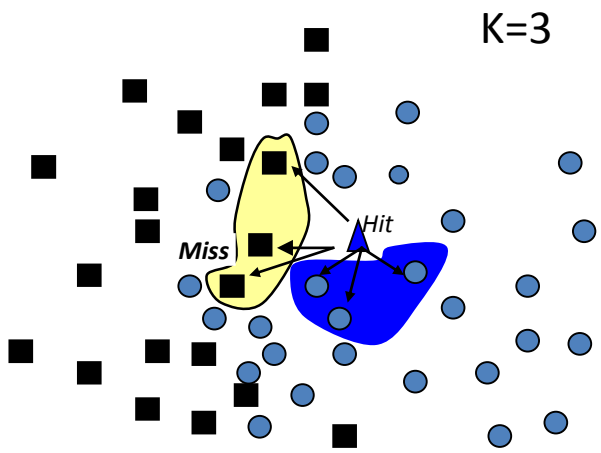
4. za j = 1 to M (npr. 2)
   if w_j >= T, dodaj v_j u odabrani skup varijabli

Multivarijantni filter pristup



Relief algoritam Kira & Rendell, 1992

Relief algoritam – slučajno uzorkovanje primjera i lociranje najbližeg primjera iste i suprotne klase. Faktori vezani uz svaku dimenziju se mijenjaju u zavisnosti o udaljenosti primjera iste i suprotne klase.



ReliefF algoritam
(više-klasni problemi, otporan na šum)
Robnik-Sikonja and I. Kononenko, 2003

Multivarijantni „wraper” pristup

- Korištenje ansambl algoritama za određivanje **važnosti varijabli** (i selekciju)

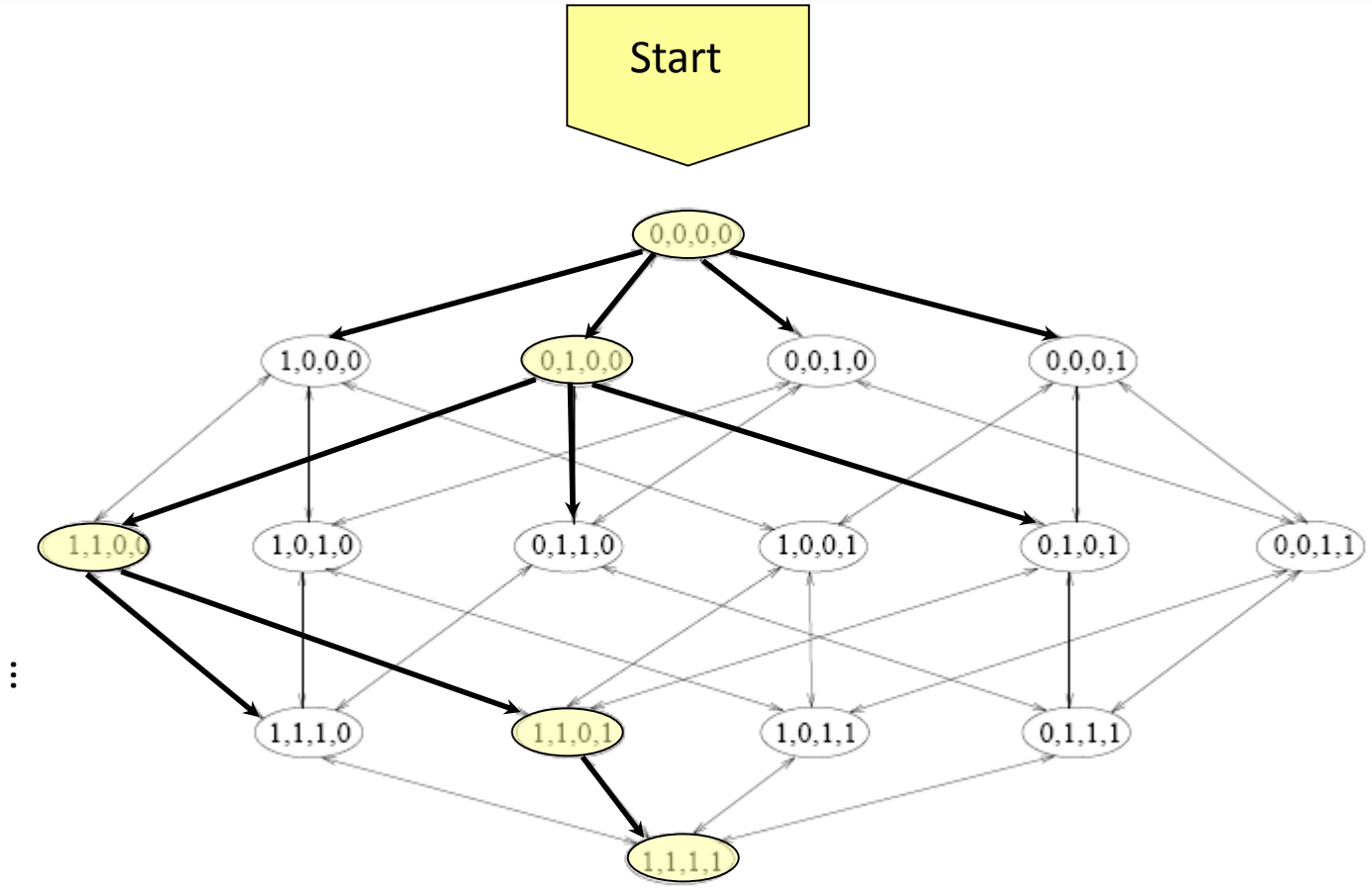
  Primjer: Random Forest

  – Važnost varijabli ~ relativni doprinos odlučivanju algoritma

  (~ točnosti algoritma)

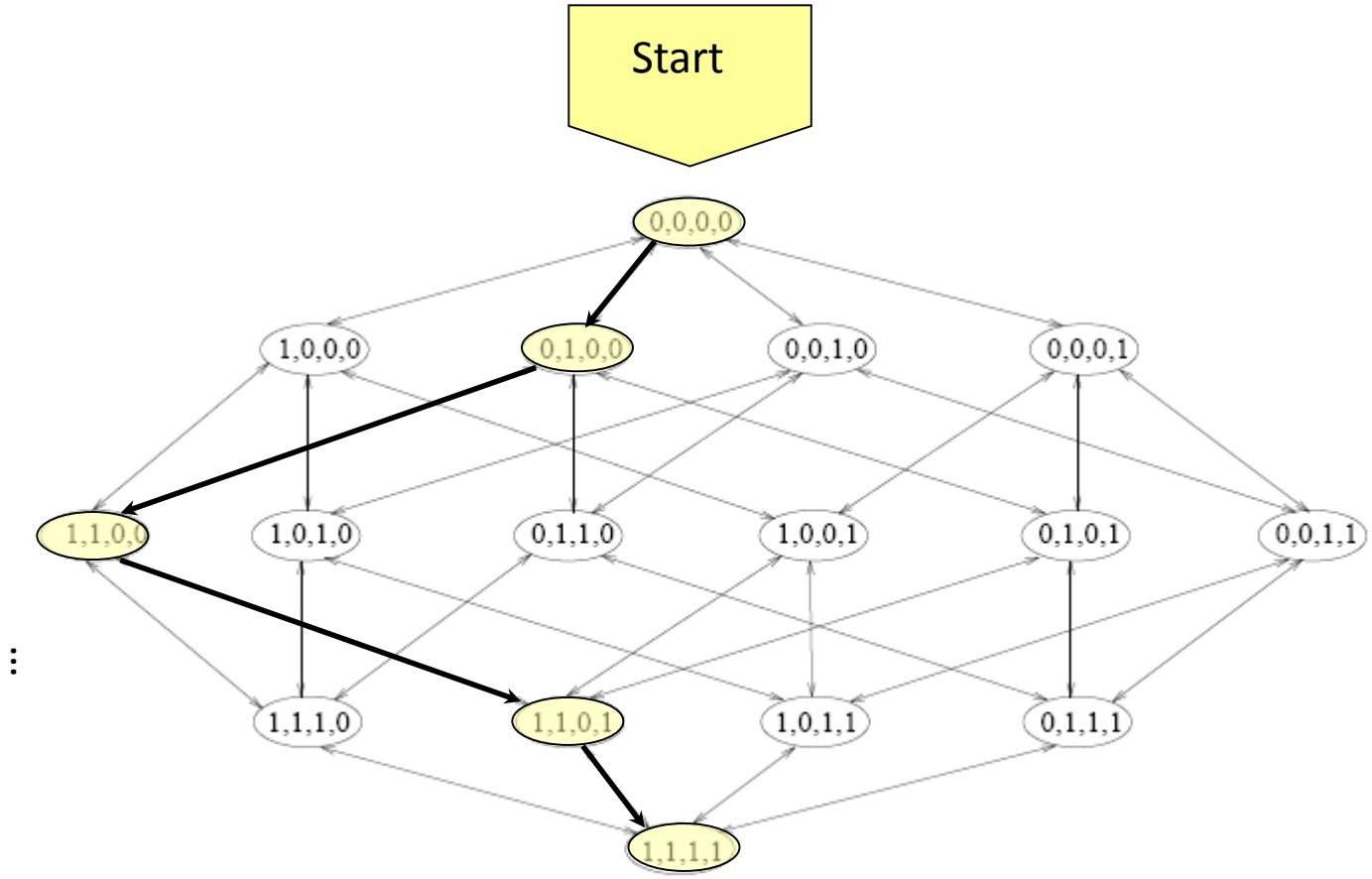
  – RF (MDA - Mean Decrease Accuracy) za varijablu v_i *:*
    - Mjeri se na „OOB => out-of-bag” primjerima (posebno za svako stablo u „šumi”)
    - točnost klasifikatora se prvo mjeri na stablima koja uključuju varijablu v_i u svom modelu, i to **sa originalnim vrijednostima v_i za OOB primjere => acc_{real}**
    - Napravi se permutacija vrijednosti varijable v_i po OOB primjerima
    - točnost klasifikatora se zatim mjeri na stablima koja uključuju varijablu v_i u svom modelu, i to **sa permutiranim vrijednostima v_i za OOB primjere => acc_{perm}**
    - **MDA** = **acc_{real} - acc_{perm}**
    - Varijable se potom rangiraju prema **MDA**
    - Važnije varijable – veći MDA

Pretraživanje podskupova– wrapper pristup



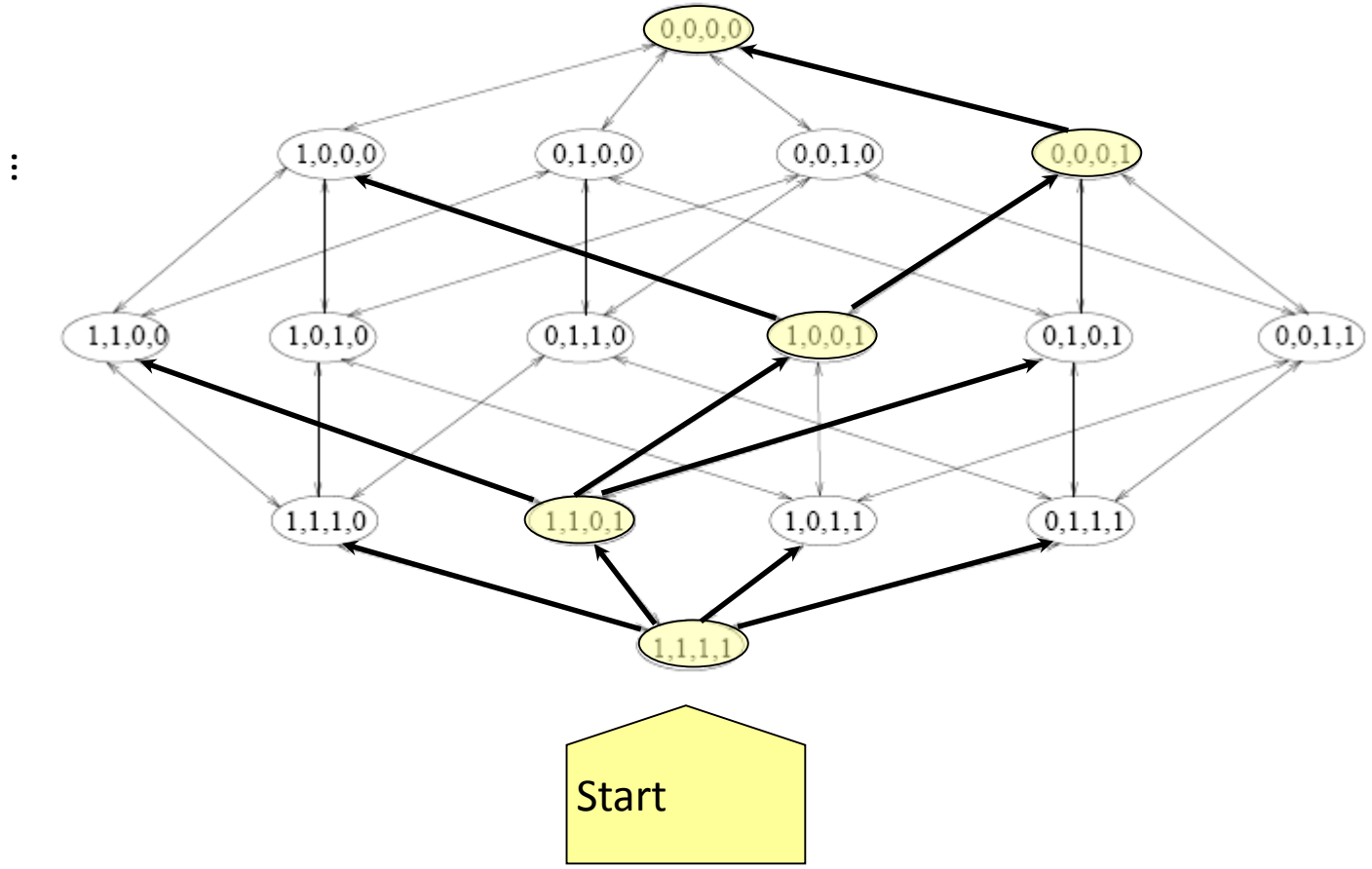
Sekvencijalno **pretraživanje unaprijed** (**SFS: Sequential Forward Selection**)

Pretraživanje podskupova - ugrađeni (embedded) pristup



Vođeno pretraživanje: ne razmatraju se alternativni putevi
Primjer.: stabla odlučivanja !

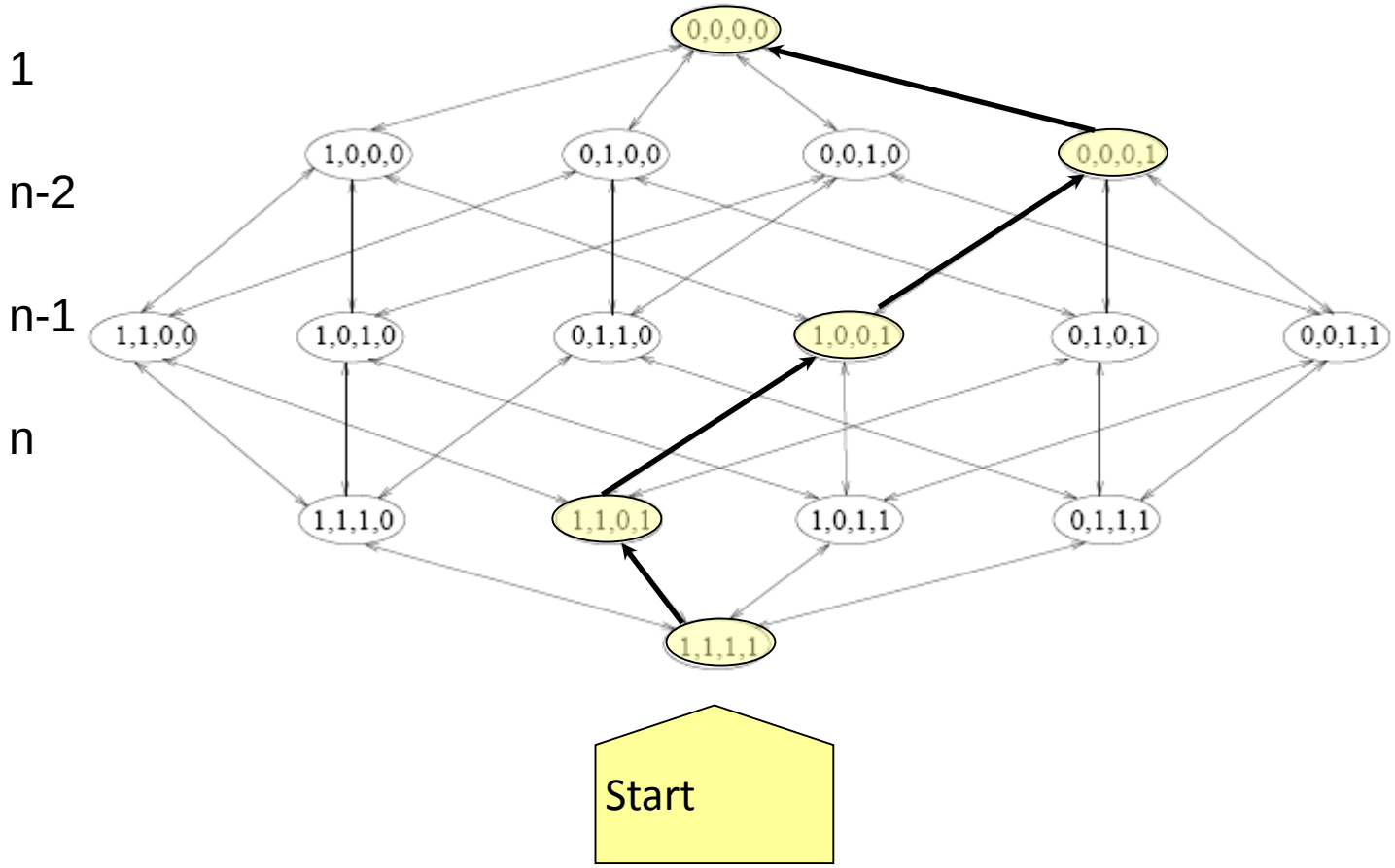
Pretraživanje podskupova – wrapper pristup



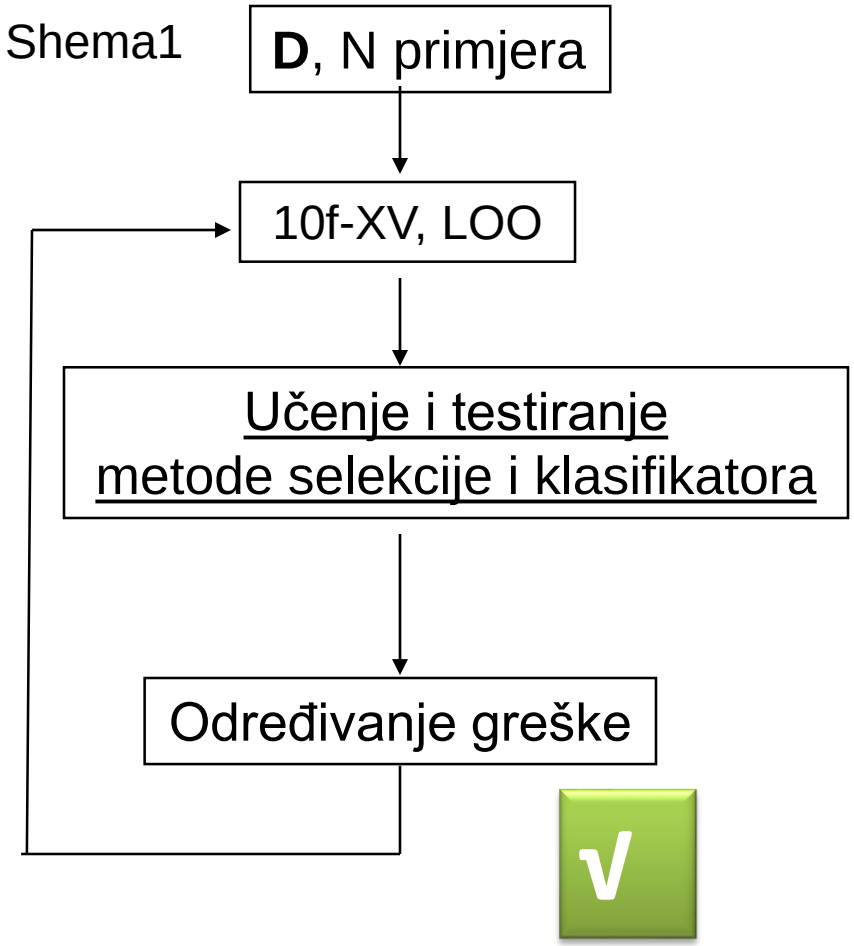
Sekvencijalno **pretraživanje unatrag** (SBS - **Sequential Backward Selection**)

Pretraživanje podskupova – embedded(ugrađeni) pristup

Vođeno pretraživanje: ne uzimaju se u obzir alternativni putevi
Primjer: “rekurzivna eliminacija varijabli bazirana na težinama (SVM)”
RFE-SVM.

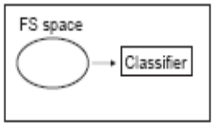
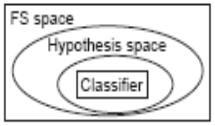
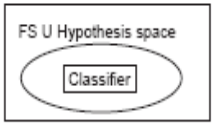


Procjena greške: XV - dvije sheme



Taksonimija(e) metoda selekcije varijabli

Table 1. A taxonomy of feature selection techniques. For each feature selection type, we highlight a set of characteristics which can guide the choice for a technique suited to the goals and resources of practitioners in the field.

	Model search	Advantages		Disadvantages	Examples
Filter		Univariate	Fast Scalable Independent of the classifier	Ignores feature dependencies Ignores interaction with the classifier	Chi-square Euclidean distance t-test Information gain, Gain ratio [6]
		Multivariate	Models feature dependencies Independent of the classifier Better computational complexity than wrapper methods	Slower than univariate techniques Less scalable than univariate techniques Ignores interaction with the classifier	Correlation based feature selection (CFS) [45] Markov blanket filter (MBF) [62] Fast correlation based feature selection (FCBF) [136]
Wrapper		Deterministic	Simple Interacts with the classifier Models feature dependencies Less computationally intensive than randomized methods	Risk of over fitting More prone than randomized algorithms to getting stuck in a local optimum (greedy search) Classifier dependent selection	Sequential forward selection (SFS) [60] Sequential backward elimination (SBE) [60] Plus q take-away r [33] Beam search [106]
		Randomized	Less prone to local optima Interacts with the classifier Models feature dependencies	Computationally intensive Classifier dependent selection Higher risk of overfitting than deterministic algorithms	Simulated annealing Randomized hill climbing [110] Genetic algorithms [50] Estimation of distribution algorithms [52]
Embedded		Interacts with the classifier Better computational complexity than wrapper methods Models feature dependencies		Classifier dependent selection	Decision trees Weighted naive Bayes [28] Feature selection using the weight vector of SVM [44, 125]

Literatura

• Guyon & Elisseeff “An Introduction to Variable and Feature Selection”,
J Mach Learn Res, 3 (2003), 1157-1182

• Feature Extraction, Foundations and Applications, I. Guyon et al, Eds. Springer,
2006.

- WEKA:
  - Preprocessing algorithms/Filters/Attribute/Supervised (...Infogain, ReliefF)
- R:
  - *Fselector (reliefF, rf), caret(RFE), randomForest...*

Redukcija dimenzionalnosti – ekstrakcija manjeg broja varijabli

....kombiniranjem/maprianjem originalnih varijabli

Formulacija:

- Treba pronaći za prostor $x_i \in R^N$ mapiranje $y=f(x):R^N \rightarrow R^M$ *uz* $M<(<)N$, tako da za neki transformirani primjer (vektor) većina informacije, ili strukture ostane sačuvana

- U principu bi optimalna mapiranja trebala biti nelinearnog karaktera

- Tradicionalno: najčešće su korištene linearne transformacije

Kad su podaci => matrica uzoraka

- Vrlo tipično za strojno učenje – podaci u formi mjerenja, **vektori svojstava** ili uzorci
- Vektori svojstava (m-dimenzionalni Euklidski prostor)

$$\mathbf{x_i} \in \mathcal{X} \subseteq \mathbb{R}^m, \quad i = 1, \ldots, n$$

- **Matrica uzoraka**

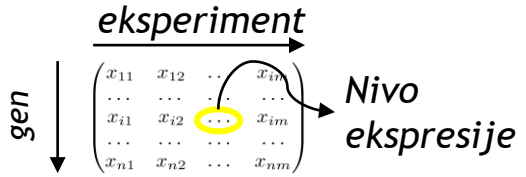
$$\mathbf{X} \in \mathbb{R}^{n \times m}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \cdots \\ \mathbf{x}_i' \\ \cdots \\ \mathbf{x}_n' \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{im} \\ \cdots & \cdots & \cdots & \cdots \\ x_{i1} & x_{i2} & \cdots & x_{im} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

i-ti uzorak

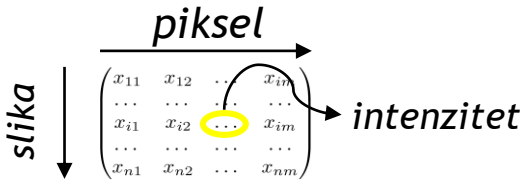
Primjeri matrice uzoraka

- **Mjerenja**
  nivo ekspresije
  - i: primjer, gen
  - j: mjerenje – eksperiment (bolesti)

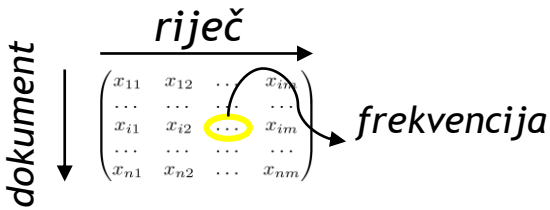
$$\mathbf{X} \in \mathbb{R}^{n \times m}$$



- **Digitalne slike**
  vektori (nijanse sivog)
  - i: slika
  - j: piksel intenzitet na lokaciji $j=(k,l)$

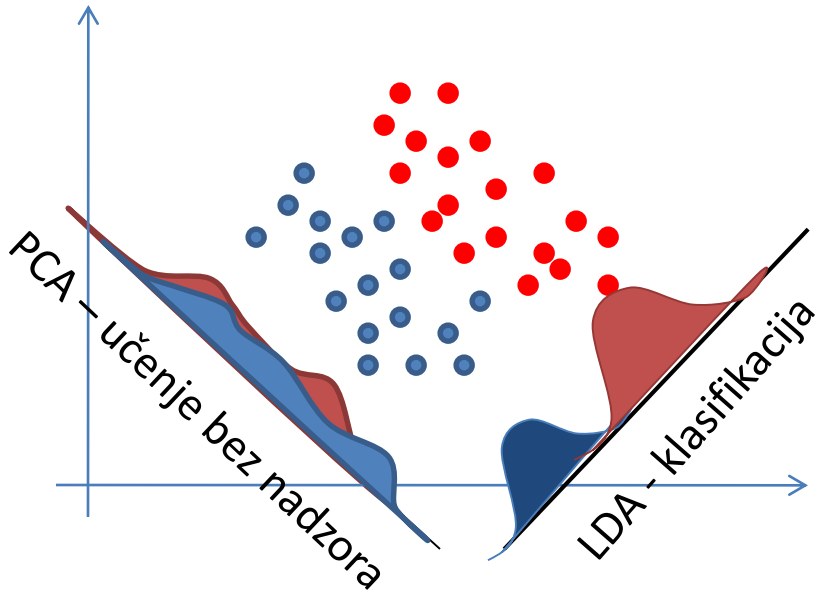


- **Korpus dokumenata**
  tzv. bag-of-words reprezentacija
  - i: dokument
  - j: riječ iz riječnika



Metoda osnovnih komponenata
(PCA - Principal Component Analysis)

PCA - Principal Component Analysis (metoda osnovnih komponenti)



Projekcija zavisi od funkcije cilja koja se optimira

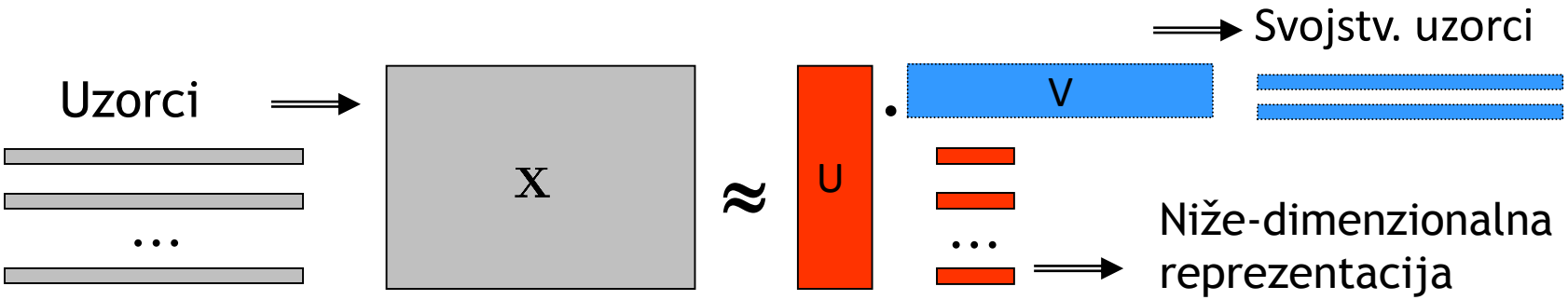
PCA – funkcija cilja je da reprezentacija primjera u niže-dimenzionalnom prostoru mora biti što točnija (sačuvanje udaljenosti izm. Primjera)

LDA – funkcija cilja je da reprezentacija primjera u niže-dimenzionalnom prostoru ima što bolja klasifikacijska svojstva (razdvajanje klasa)

Redukcija dimenzija - PCA

Pearson, 1901: **PCA = Ortogonalna linearna projekcija s minimalnom greškom – u smislu najmanjih kvadrata**

- **Osnovna ideja PCA:** redukcija dimenzionalnosti (*m* ➔ p: p<<m)
- **Pretpostavka (uobičajena):** *m* je velik broj zavisnih varijabli (koreliranih)
  - Intuitivno – želimo zadržati što je više moguće originalnih odnosa/varijacije između podataka/uzoraka
- **Postupak**: transformacija u novi skup (međusobno nekoreliranih) varijabli koje su i rangirane tako da **one prve u rangu zadržavaju najveći dio varijacije koja je prisutna u svim originalnim varijablama**.



- **PCA = Latentna struktura**: niže-dimenzionalni podprostor
- **Dekompozicija prostora** ➔ svojstvene vrijednosti, svojstveni vektori

PCA - Principal Component Analysis (metoda osnovnih komponenti)

Uz dani skup točaka $\mathbf{x} \in \mathbf{R}^m$ želimo projicirati svaki **x** u *d<m* (niže-dimenzionalni) podprostor sa $\mathbf{z}=[z_1,z_2,\ldots,z_d] \in \mathbf{R}^d$ tako da je:

$$\mathbf{x} = \sum_{i=1}^{m} z_i u_i$$

- vektori $\mathbf{u}_i$ zadovoljavaju kriterij ortonormalnosti:

$$\mathbf{u}_i^{\mathrm{T}} \mathbf{u}_j = \delta_{ij} \qquad \text{(Kronecker delta)}$$

Mi želimo koristiti d<m. Ostali koeficijenti b_i i $\mathbf{u}_i$ omogućavaju aproksimaciju $\tilde{\mathbf{x}}$

$$\tilde{\mathbf{x}} = \sum_{i=1}^{d} z_i \mathbf{u}_i + \sum_{i=d+1}^{m} b_i \mathbf{u}_i$$

PCA - Principal Component Analysis (metoda osnovnih komponenti)

Za svaki **x** – greška koju činimo zbog redukcije dimenzionalnosti je:

$$\mathbf{x} - \tilde{\mathbf{x}} = \sum_{i=d+1}^{m} (z_i - b_i)\mathbf{u}_i$$

Želimo naći $\mathbf{u}_i$ koeficijente b_i, i vrijednosti z_i **s** najmanjom greškom

Za čitav skup podataka - uz relaciju ortonormalnosti vrijedi:

$$E_d = \frac{1}{2}\sum_{k=1}^{n} \left\| \mathbf{x}^k - \tilde{\mathbf{x}}^k \right\|^2 = \frac{1}{2}\sum_{k=1}^{n}\sum_{i=d+1}^{m} \left\| z_i^k - b_i^k \right\|^2$$

PCA - Principal Component Analysis (metoda osnovnih komponenti)

Izvod - minimizacija E_d po **u,** vodi na koncu do:

$$\mathbf{C}\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

Gdje je **C** matrica kovarijance

$$\mathbf{C} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}^k - \bar{\mathbf{x}})(\mathbf{x}^k - \bar{\mathbf{x}})^{\mathrm{T}}$$

A vektori $\mathbf{u}_i$ su svojstveni vektori matrice **C.** Konačno, E_d je minimalno za slučaj kad se iz rekonstrukcije **odbace svojstveni vektori čije su svojstvene vrijednosti najmanje**:

$$E_d = \frac{1}{2}\sum_{i=d+1}^{m}\lambda_i$$

Redukcija dimenzija – PCA preko SVD

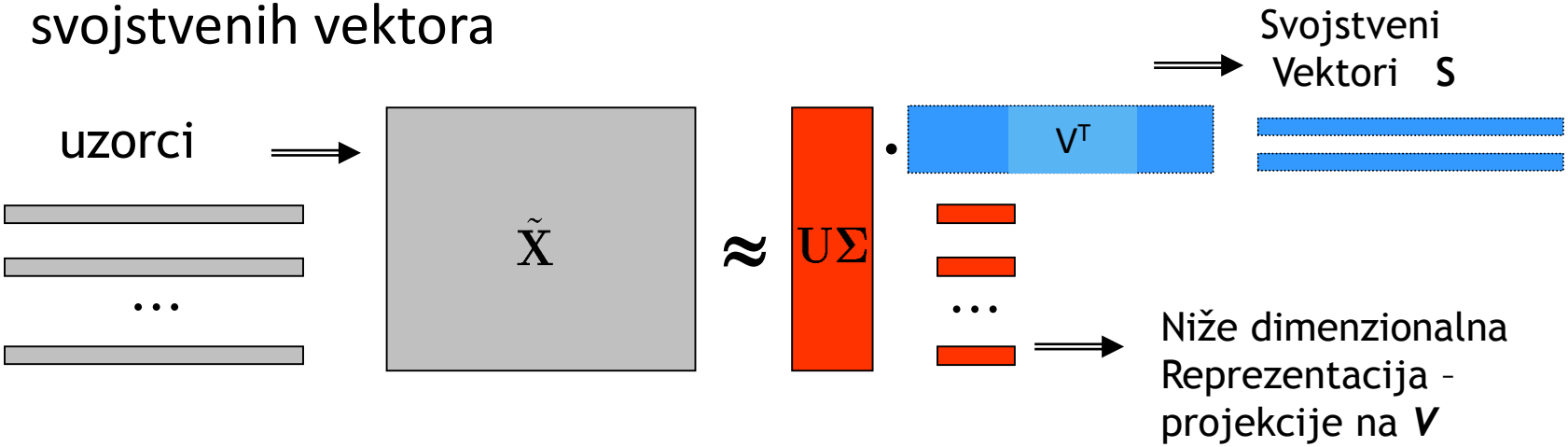
- SVD matrice uzoraka se može koristiti za izračunavanje PCA

$$\tilde{\mathbf{X}} = \mathbf{U}\sum\mathbf{V}^T \Rightarrow$$

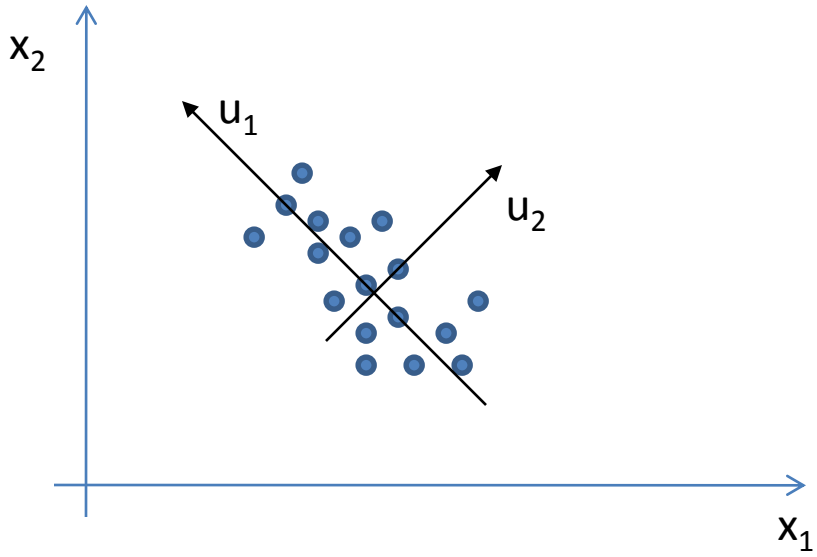
$$S = \frac{1}{n}\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} = \frac{1}{n}(\mathbf{V}\sum\mathbf{U}^T)(\mathbf{U}\sum\mathbf{V}^T) = \frac{1}{n}\mathbf{V}\sum^2\mathbf{V}^T$$

Redovi $\mathbf{V}$ su svojstveni vektori $\mathbf{S}$

- $\tilde{\mathbf{X}}V = \mathbf{U}\sum$ ➔ PC težine ➔ unutarnji produkt uzoraka i svojstvenih vektora



PCA - Principal Component Analysis (metoda osnovnih komponenti)



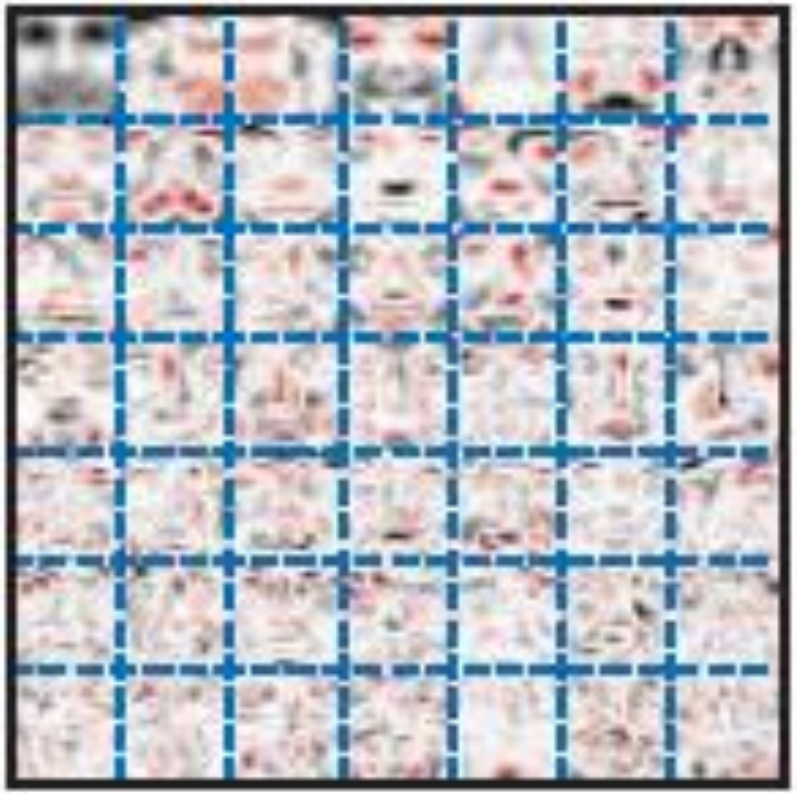
Projekcija $\mathbf{x}^k$ na $\mathbf{u}_1$ i $\mathbf{u}_2$ daje komponente transformiranog vektora $\mathbf{z}^k$

Što dobijamo ovakvom projekcijom ?
Koje su primjene PCA ?
Koji su problemi PCA ?

Redukcija dimenzija – kako izgledaju PC?

- Primjena na slike lica
  - Uzorak ⟹ intenzitet sivog po pikselima slike
- Eigenfaces

Prosječno lice



[Lee & Seung 1999]

⟸ osnovne komponente

PCA - primjene

Redukcija dimenzija – kao priprema podataka za druge algoritme ili analizu podataka (tipično za probleme vezane uz prepoznavanje slika (lica))

Otkrivanje niže-dimenzionalnih struktura u više-dimenzionalnom prostoru (tzv. data manifolds)

Vizualizacija podataka, interpretacija - 2D-3D grafovi podataka

Eliminacija šuma iz podataka - otkrivanje i eliminacija outlier-a

Ograničenja

Linearnost

-> postoje nelinearne varijante (kernel PCA)...

Dekoreliranost osnovnih komponenti nije i nezavisnost

-> metoda nezavisnih komponenti (ICA)

Aditivni model (ograničenja na predznak koeficienata (tzv. Loadings)

-> NMF metoda (nonnegative matrix factorization)

Ne-Negativna Matrična Dekompozicija

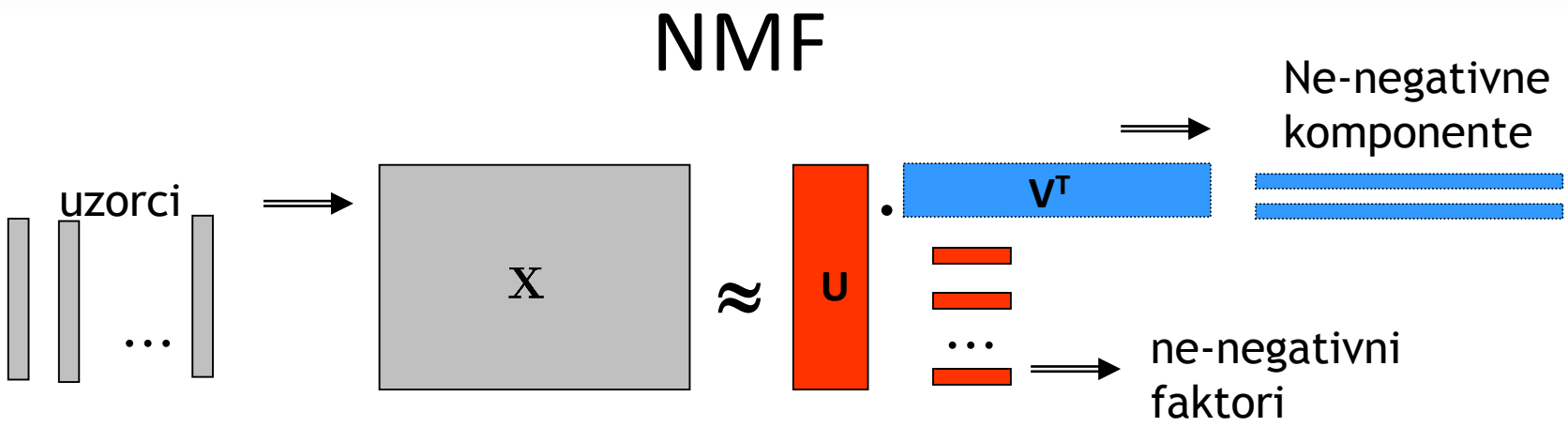
(NMF - Non-Negative Matrix Decomposition)

NMF

- Aproksimativna (niži rang) matrična dekompozicija s ne-negativnim faktorima

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{U}\mathbf{V}^T, \quad \mathrm{U} \in \Re_{\geq o}^{m\,x\,q}, \mathrm{V}^{\mathrm{T}} \in \Re_{\geq o}^{q\,x\,n}$$

- Motivacija
  - Primjena za ne-negativne matrice (matrice uzoraka sa ne-negativnim mjerenjima)
  - Prethodno znanje – latentni faktori su pozitivni !
  - Efekt faktora – kumulativni/aditivni (nema poništavanja efekata zbog negativnih doprinosa!)



- **Podaci**: Matrica uzoraka
- **Latentna struktura**: niže-dimenzionalni podprostor definiran **ne-negativnim baznim vektorima**
- **Dekompozicija**: ne-konveksni opt. problem

Redukcija dimenzionalnosti: NMF

$$\hat{\mathbf{X}} = \mathbf{U}\mathbf{V}^T$$

- **Kvadratna pogreška (Frobeniusova norma)**

$$E(\mathbf{X}, \hat{\mathbf{X}}) = \sum_{i=1}^{n}\sum_{i=1}^{m}(x_{ij} - \hat{x}_{ij})^2 = \left\|\mathbf{X} - \hat{\mathbf{X}}\right\|_F$$

- **Na bazi Kullback-Leibler divergencije**
  (tipično se koristi za udaljenosti dviju distribucija)

$$E_{div}(\mathbf{X}, \hat{\mathbf{X}}) = \sum_{i=1}^{n}\sum_{i=1}^{m}(x_{ij}\log\frac{x_{ij}}{\hat{x}_{ij}} - x_{ij} + \hat{x}_{ij})$$

Uz uvjet $\sum_{i,j}\hat{x}_{ij} = const.$

Redukcija dimenzionalnosti: NMF

Metoda/algoritam za faktorizaciju matrica (pozitivnih) u 2 ne-negativne matrice:

$$\mathbf{X} = \mathbf{U}\mathbf{V}^T \qquad \begin{bmatrix} x_{11} & x_{12} \ldots . x_{1n} \\ x_{21} & x_{22} \ldots . x_{2n} \\ \ldots\ldots \\ \ldots\ldots \\ x_{m1} & x_{m2} \ldots . x_{mn} \end{bmatrix} = \begin{bmatrix} u_{11}\, u_{12} \ldots . u_{1k} \\ u_{21}\, u_{22} \ldots . u_{2k} \\ \ldots \\ u_{m1}\, u_{\mathrm{m}2} \ldots . u_{\mathrm{m}k} \end{bmatrix} . \begin{bmatrix} v_{11}\, v_{12} \ldots . v_{1n} \\ v_{21}\, v_{22} \ldots . v_{2n} \\ \ldots \\ v_{k1}\, v_{k2} \ldots . v_{kn} \end{bmatrix}$$

- Minimiziraj $$J = \frac{1}{2}\left\|\mathbf{X} - \mathbf{U}\mathbf{V}^T\right\|_F$$
- Gdje je $\|\mathbf{X}\|_F \Rightarrow \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij}|^2}$ - Frobeniusova norma
- **+ uvjet da vrijede ograničenja na ne-negativnost: $u_{ij} \geq 0$ i $v_{ij} \geq 0$!!**

➔ vodi na Lagrange-ovu funkciju !!

Lagrangeova funkcija

$$L = J + tr(\alpha \mathbf{U}^T) + tr(\beta \mathbf{V}^T)$$

Uz Derivaciju L =0

$$\frac{\partial L}{\partial \mathbf{U}} = -\mathbf{XV} + \mathbf{UV}^T\mathbf{V} + \boldsymbol{\alpha}$$

$$\frac{\partial L}{\partial \mathbf{V}} = -\mathbf{X}^T\mathbf{U} + \mathbf{VU}^T\mathbf{U} + \boldsymbol{\beta} \quad \Rightarrow \quad \text{i uz Kuhn - Tuckerove uvjete:} \quad \alpha_{ij}u_{ij} = \beta_{ij}v_{ij} = 0 \quad \Rightarrow$$

$$(\mathbf{XV})_{ij}u_{ij} - (\mathbf{UV}^T\mathbf{V})_{ij}u_{ij} = 0$$

$$(\mathbf{X}^T\mathbf{U})_{ij}v_{ij} - (\mathbf{VU}^T\mathbf{U})_{ij}v_{ij} = 0$$

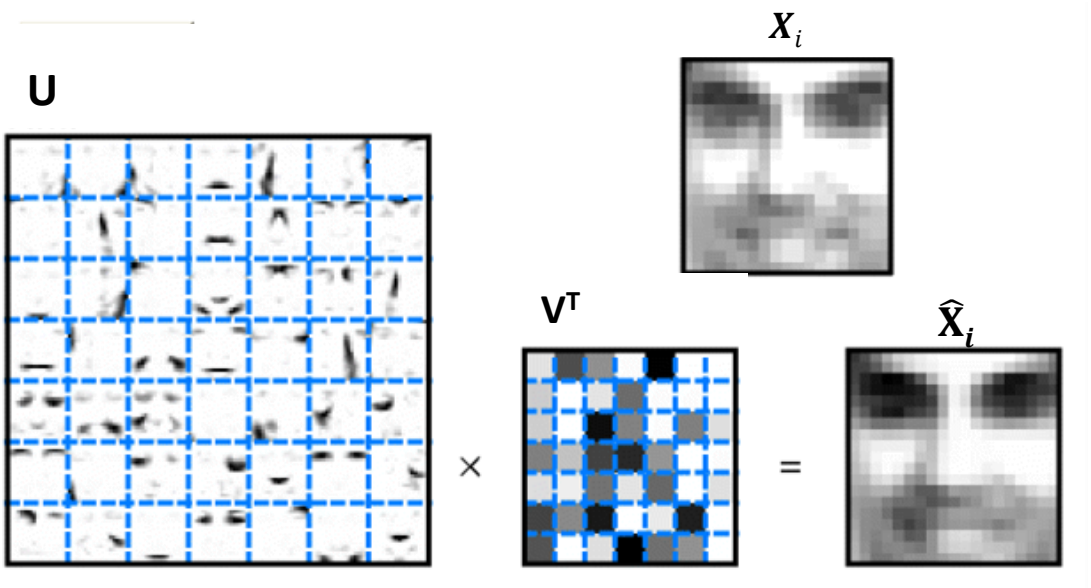
Gornji izrazi vode na iterativni algoritam za „osvježavanje” u_{ij} i v_{ij}

$$u_{ij} \leftarrow u_{ij}\frac{(\mathbf{XV})_{ij}}{(\mathbf{UV}^T\mathbf{V})_{ij}}$$

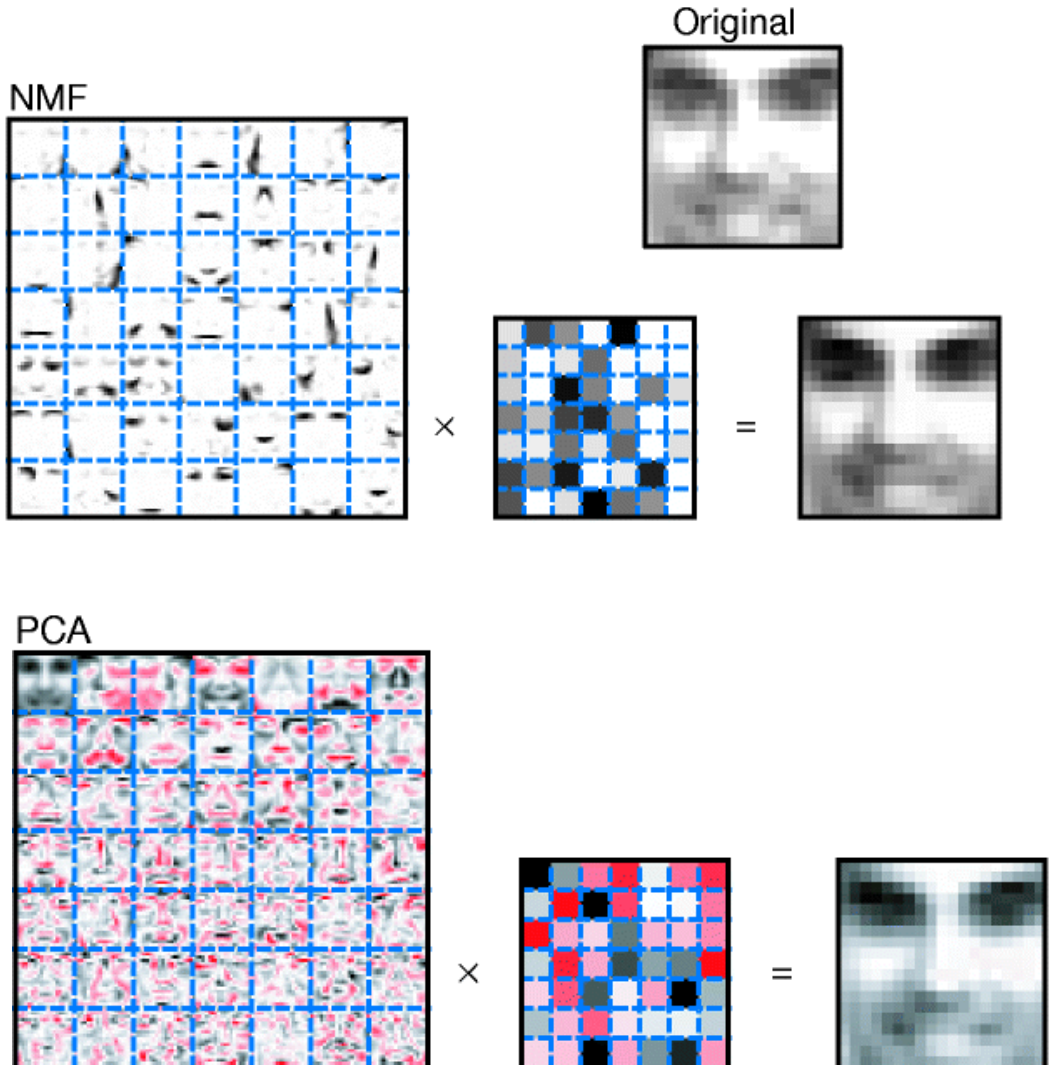
$$v_{ij} \leftarrow v_{ij}\frac{(\mathbf{X}^T\mathbf{U})_{ij}}{(\mathbf{VU}^T\mathbf{U})_{ij}}$$

NMF - Primjena u slikama

- NMF automatska detekcija **dijelova** u slikama ili mjerenjima
- Digitalne slike ➔ matrice uzoraka sa ne-negativnim vrijednostima (intenzitet)
- Faktori korespondiraju sa lokaliziranim uzorcima
- **Eigenfaces** (NMF) ➔ dijelovi lica:



PCA vs NMF



Metoda Nezavisnih Komponenata
(ICA - Independent Component Analysis)

ICA - Independent Component Analysis

Model

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} + \varepsilon \qquad \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Signal / Uzorak

Matrica miješanja

Nezavisne komponente

- Nepoznati su i **A** i **s** (odnosno $\mathbf{W}=\mathbf{A}^{-1}$ => **s=WX)**
- Latentne varijable **s -** ne-Gaussovske (apriorne) distribucije
- **s – međusobno nezavisne**
  (nezavisnost ≠ nekoreliranost – PCA)

ICA –nezavisnost komponenti

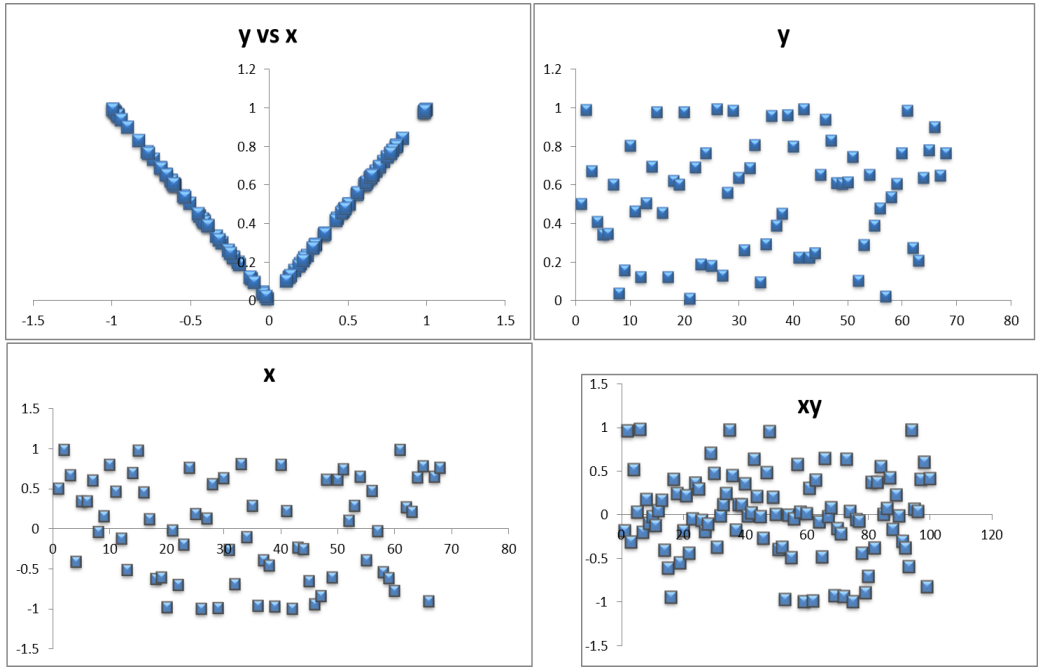
Nezavisnost ≠ nekoreliranost

$x\epsilon[-1,1]$ – uniformna distribucija
$\mathrm{y} = -x\colon\ x\ \leq 0$
$\mathrm{y} = x\colon\ x\ > 0$

$E[\mathrm{x}] = 0$
$E[\mathrm{y}] = 0.5$
$E[\mathrm{xy}] = 0$

$\mathrm{cov}[\mathrm{x},\mathrm{y}] = E[xy] - E[x]E[y]$
$\mathrm{cov}[\mathrm{x},\mathrm{y}] = 0$ => **x,y ne-korelirani**

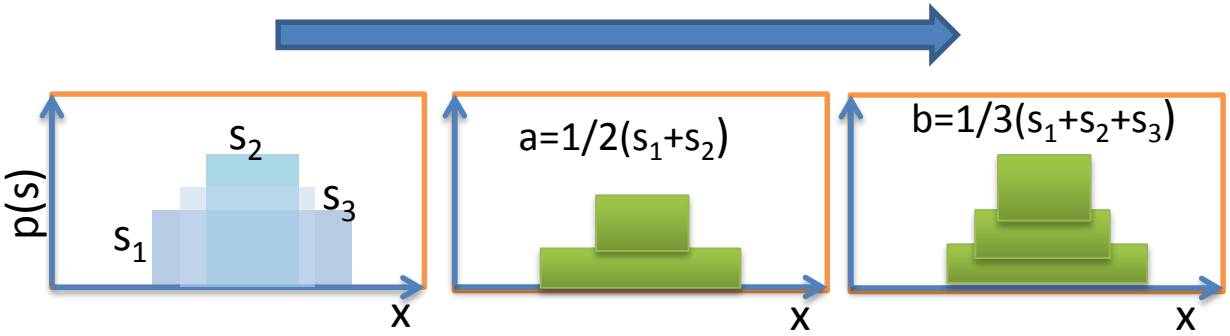
$E[xy|x \leq 0] = -1/3$
$E[xy|x > 0] = +1/3$
$E[y|x] \neq E[y]$ **=> y zavisno o x !**



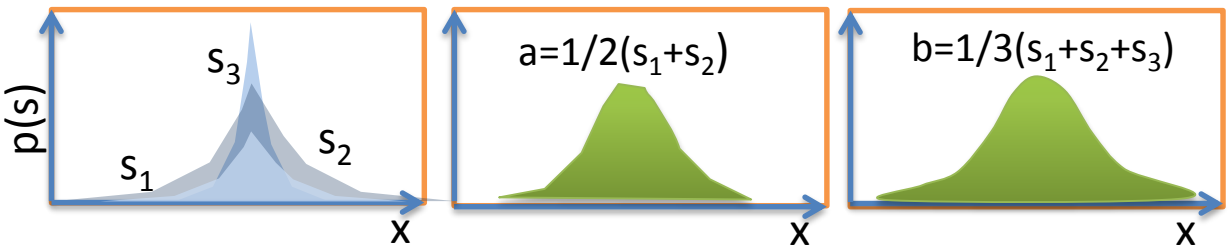
ICA – maksimiziranje ne-gaussovskog ponašanja

- ICA funkcionira za slučajeve kada je **maksimalno jedan od izvora Gaussovskog tipa**, ostali moraju biti **ne-Gaussovske distribucije**
- Linearno miješanje nezavisnih slučajnih varijabli (primjeri A i B) čini ih sve bliže Gaussovskoj raspodjeli (**Centralni granični teorem**)

A)



B)



➔ Za odvajanje komponenti (ne-Gauss.) – treba ići što dalje od Gaussove distribucije

➔ **Mjere razdvajanja**: viši momenti distribucija (3,4 – kurtosis), entropija

ICA – maksimiziranje ne-gaussovskog ponašanja

- Linearno miješanje nezavisnih slučajnih varijabli čini ih "više Gaussijanskim" (Central Limit Theorem)

$$\langle \mathbf{v}, \mathbf{x} \rangle = \langle \mathbf{v}, \mathbf{W}\mathbf{s} \rangle = \langle \mathbf{W}'\mathbf{v}, \mathbf{s} \rangle = \langle \mathbf{z}, \mathbf{s} \rangle, \ \mathbf{z} = \mathbf{W}'\mathbf{v}$$

Vektor težina nad varijablama u originalnom prostoru

Inducirane kombinacije težina Za nezavisne komponente

- Cilj: naći kombinaciju težina koja je maksimalno „ne-Gaussijanska"

Redukcija dimenzionalnosti: ICA – kontrastne funkcije

- Mjere odstupanja od Gaussove distribucije **minimiziramo tzv. kontrastne funkcije**
- **„šiljatost"** (**kurtosis** - kumulant 4. reda):

$$K(z) = E[z^4] - 3(E[z^2])^2$$

> 0: **super Gausijanska d.**
< 0: **sub Gausijanska d.**

- **Negativna entropija (negentropy)**

$$J(z) = H[z_{gauss}] - H[z]$$

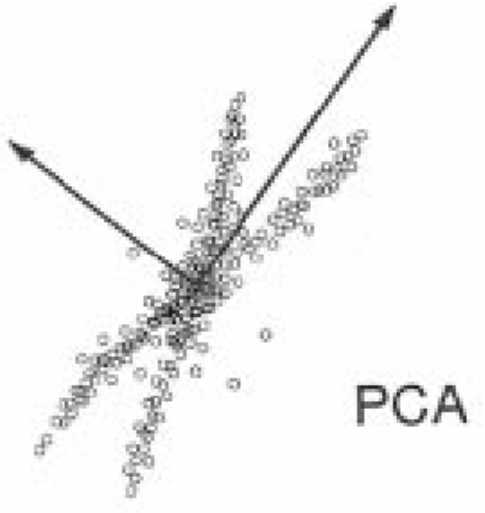
Gaussian iste Varijance kao distr. x

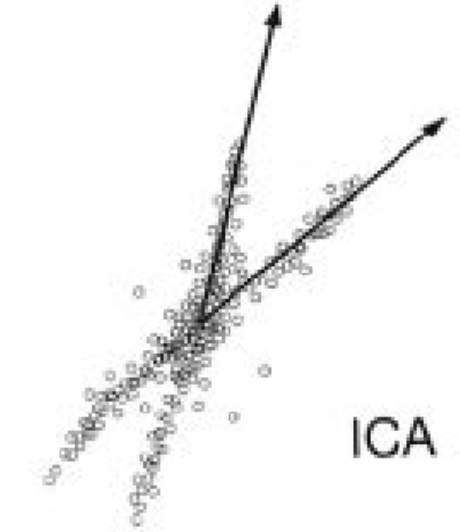
$\Rightarrow$ Koristi činjenicu da Gauss. ima maksim. entropiju (za dane μ, σ) !

- **Najčešće se koriste neke aproksimativne funkcije *G(z)* i na njima temelji određivanje *z***

$$J(z) = (E[G(z)] - E[G(z_{gaus})])^2$$

PCA vs ICA





- PCA: dekoreliranje varijabli

  (statistika drugog reda)

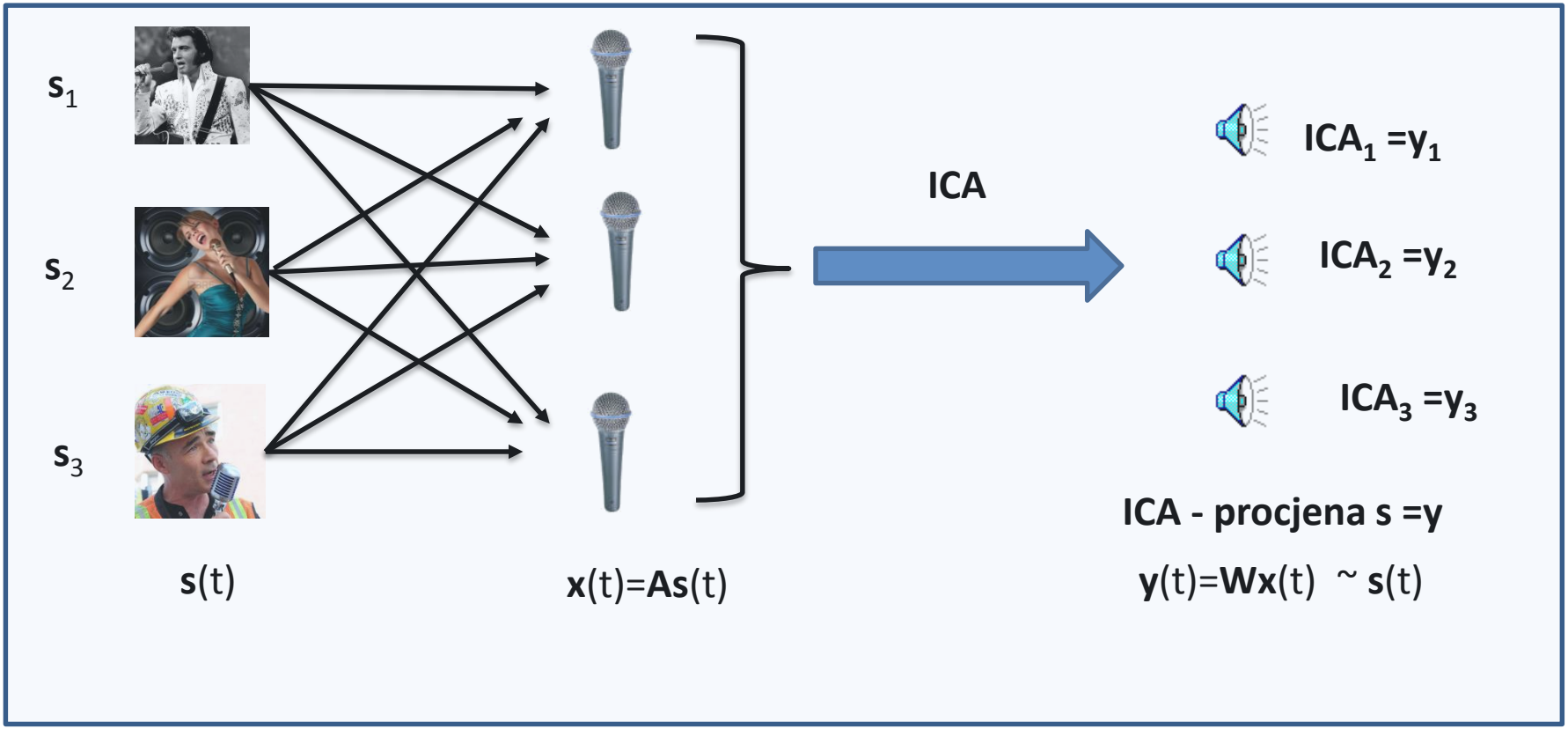
- ICA: nezavisnost

  (uključivanje momenata distribucije višeg reda)

Redukcija dimenzionalnosti: ICA

ICA = metoda kojom se rješava problem slijepe separacije signala
(**Blind Source Separation - BSS**)

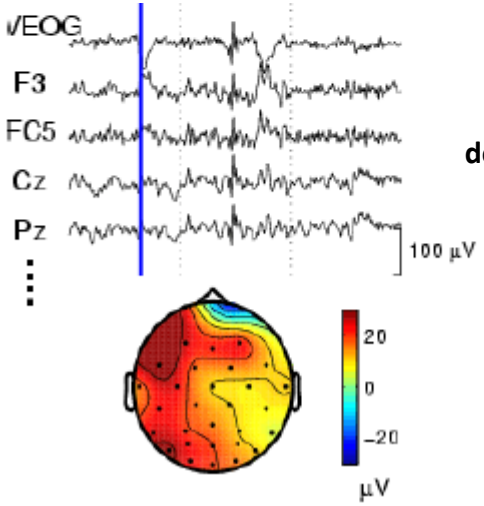
"cocktail party problem"



ICA – Slijepo razdvajanje signala

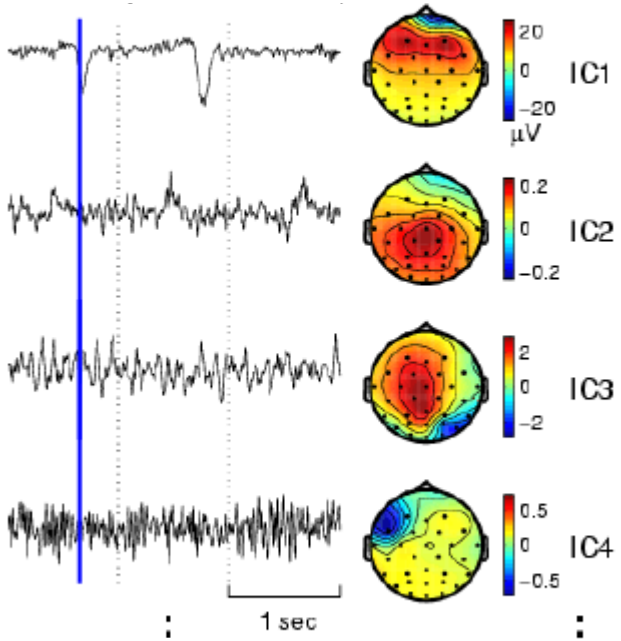
Primjer rješenja

EEG kanali (točke snimanja)



Signali x

ICA dekompozicija



Nezavisni izvori - aktivacije

y=Wx

Matrica miješanja

W^{-1}

The Elements of Statistical Learning

Hastie, Tibshirani, Friedman (ch. 14)

PCA

- I. T. Jolliffe, *Principal Component Analysis*, Springer, 2nd, 2002.

ICA

- A. J. Bell and T. J. Sejnowski, *An information-maximisation approach to blind separation and blind deconvolution*, Neural Computation, 7(6), 1995.
- A. Hyvärinen and E. Oja, *Independent component analysis: a tutorial*, Neural Computation, 13(4-5), pp. 411-420, 2000

NMF

- D. D. Lee and H. S. Seung. *Learning the parts of objects by non-negative matrix factorization*. Nature 401, 788-791 (1999).
- D. D. Lee and H. S. Seung, *Algorithms for non-negative matrix factorization*, NIPS 13, pp. 556-562, 2001.