

Mašinsko učenje

Primenjeni algoritmi

Šta je mašinsko učenje?

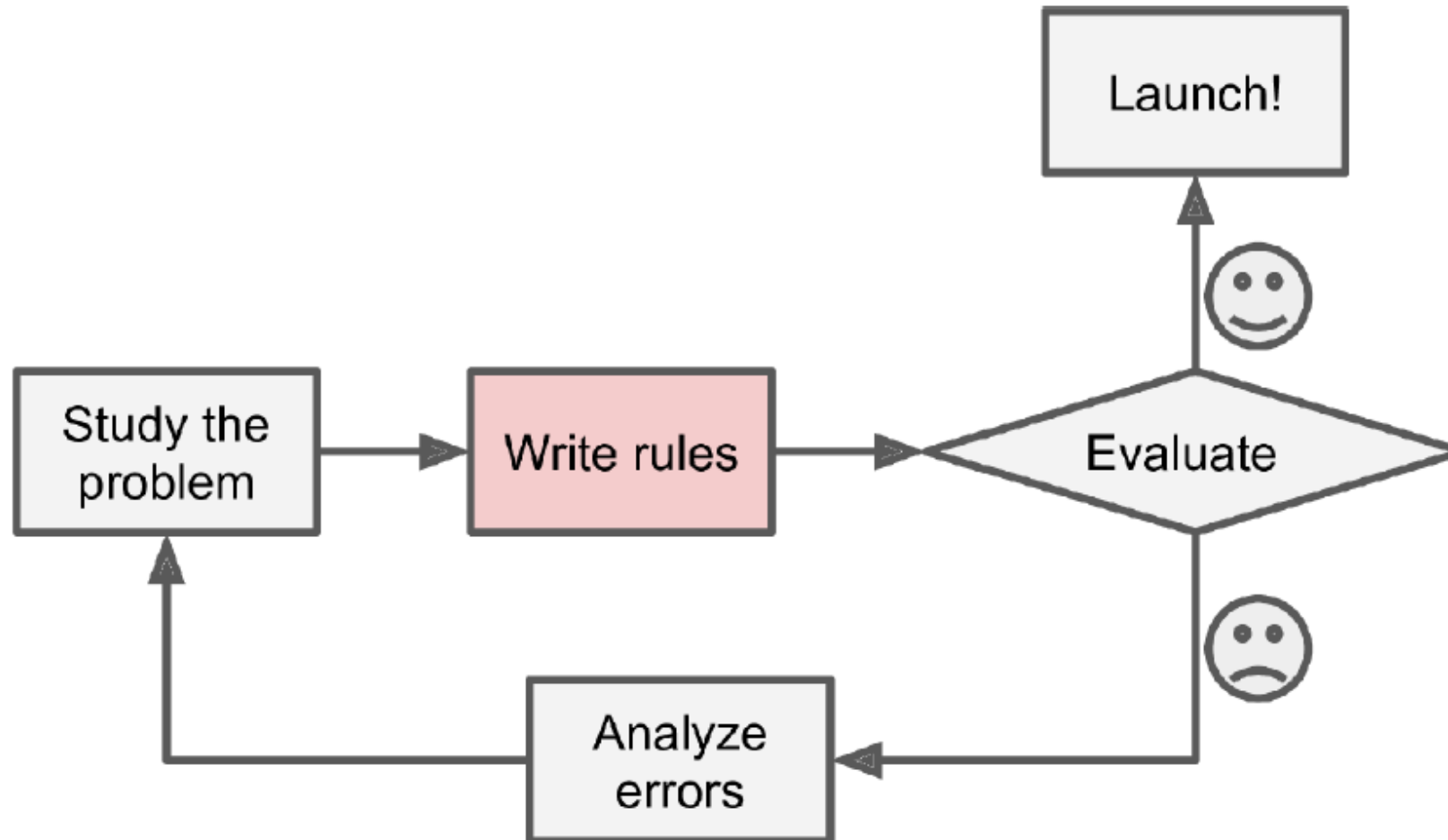
The field of study that gives computers the ability to learn without being explicitly programmed.

—Arthur Samuel, 1959

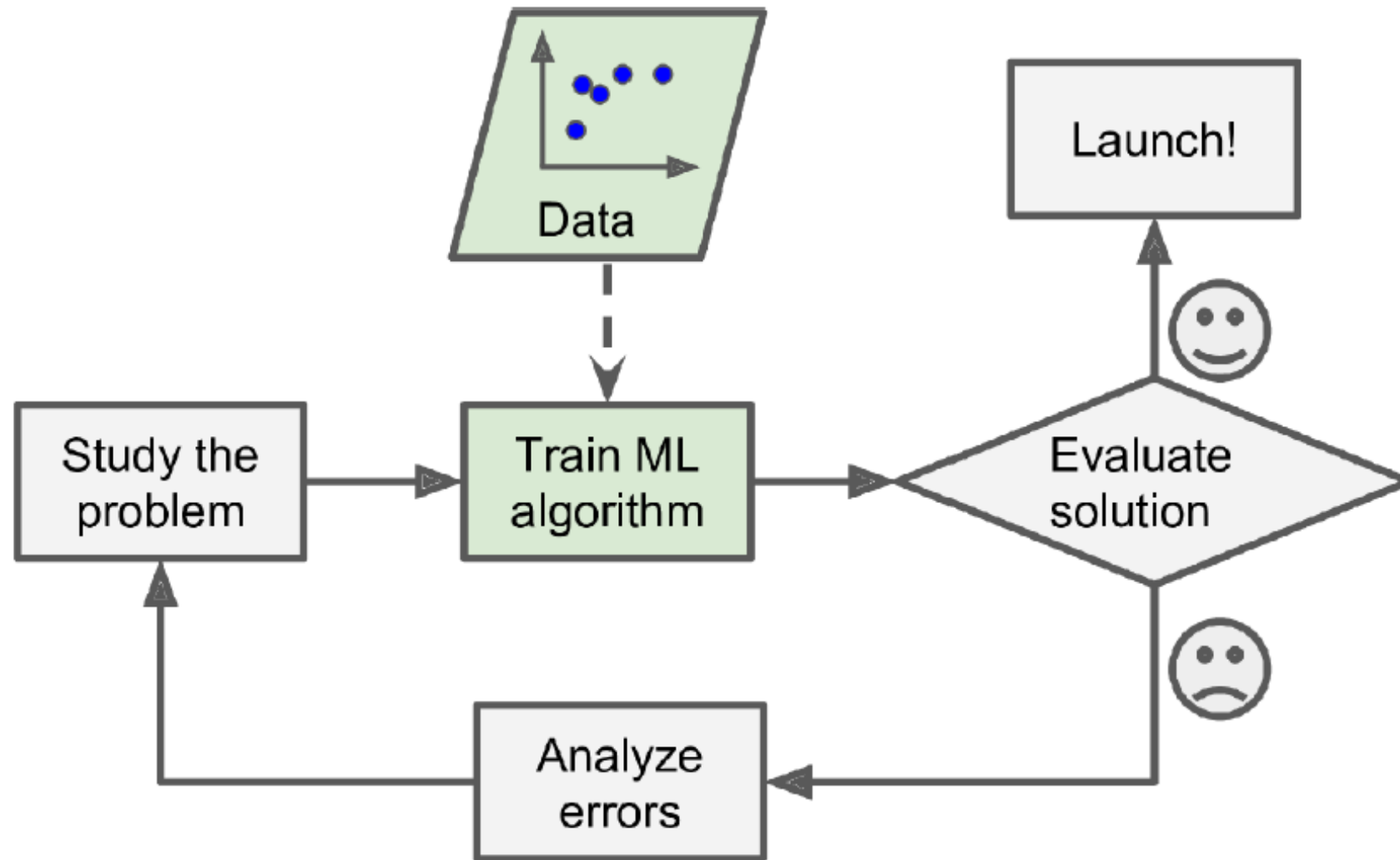
A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E .

—Tom Mitchell, 1997

Primer Filter za SPAM email - Tradicionalna tehnika



Primer Filter za SPAM email - ML tehnika



Komponente mašinskog učenja

- Model
- Funkcija cilja za procenu dobrog modela
- Metoda optimizacije za učenje modela koji optimizuje funkciju cilja

Podela mašinskog učenja

- Nadgledano učenje (Supervised Learning)
- Nenadgledano učenje (Unsupervised Learning)
- Polunadgledano učenje (Semisupervised Learning)
- Učenje sa podsticajem (Reinforcement Learning)

Nadgledano učenje

$$(x, y)$$
$$\begin{array}{c|c} x^{(1)} & y^{(1)} \\ \hline x^{(2)} & y^{(2)} \end{array}$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 185 \\ 42 \\ 53 \\ 2 \end{bmatrix}$$

VISITKA

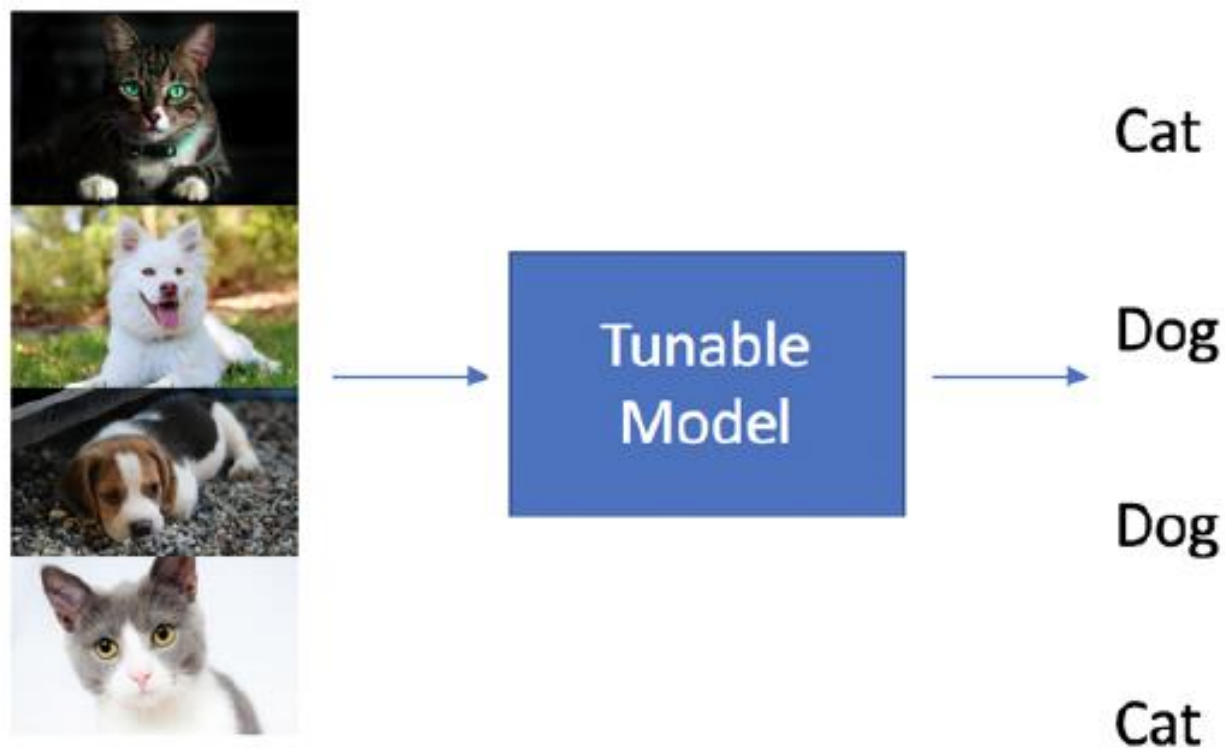
VISITKA = 185

- Polazi od skupa parova (vektor **svojstava (obeležja)**, atribut=vrednost)
- Cilj: na osnovu ulaznih podataka-parova izvesti pravilo koje predviđa vrednost asociranu sa novim vektorom svojstava.

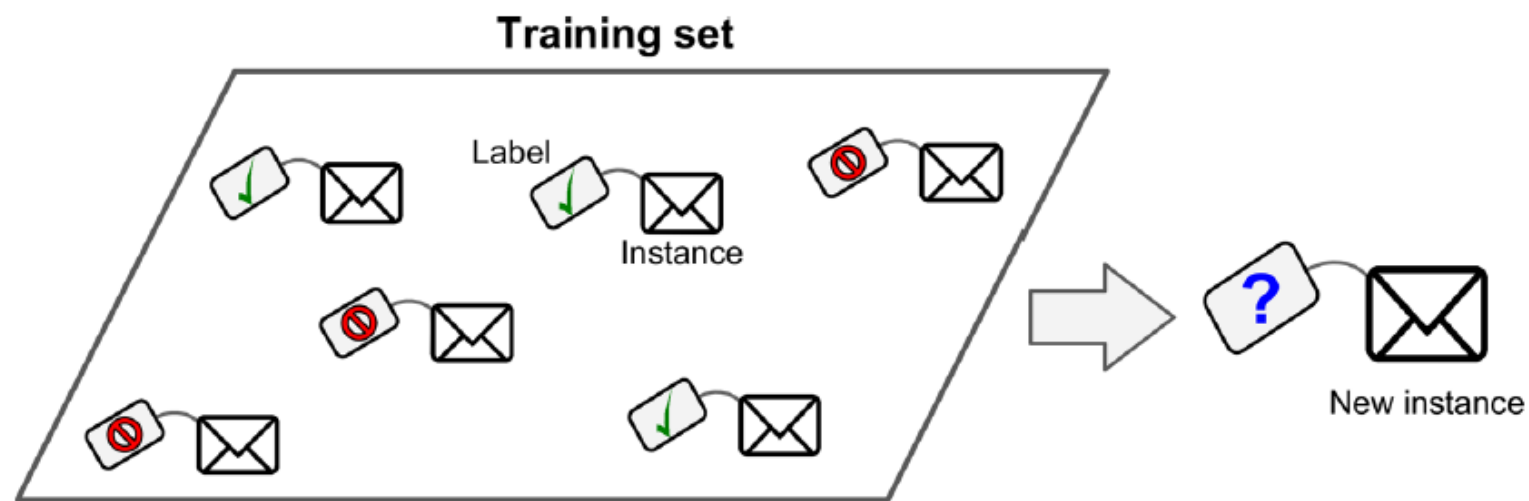
$$\underline{x} \rightarrow \underline{y}$$

- **Regresivni** modeli asociraju realan broj sa svakim vektorom svojstava
- **Klasifikacioni** modeli asociraju jedan simbol iz konačnog skupa sa svakim vektorom svojstava

Nadgledano učenje – primer Klasifikacije

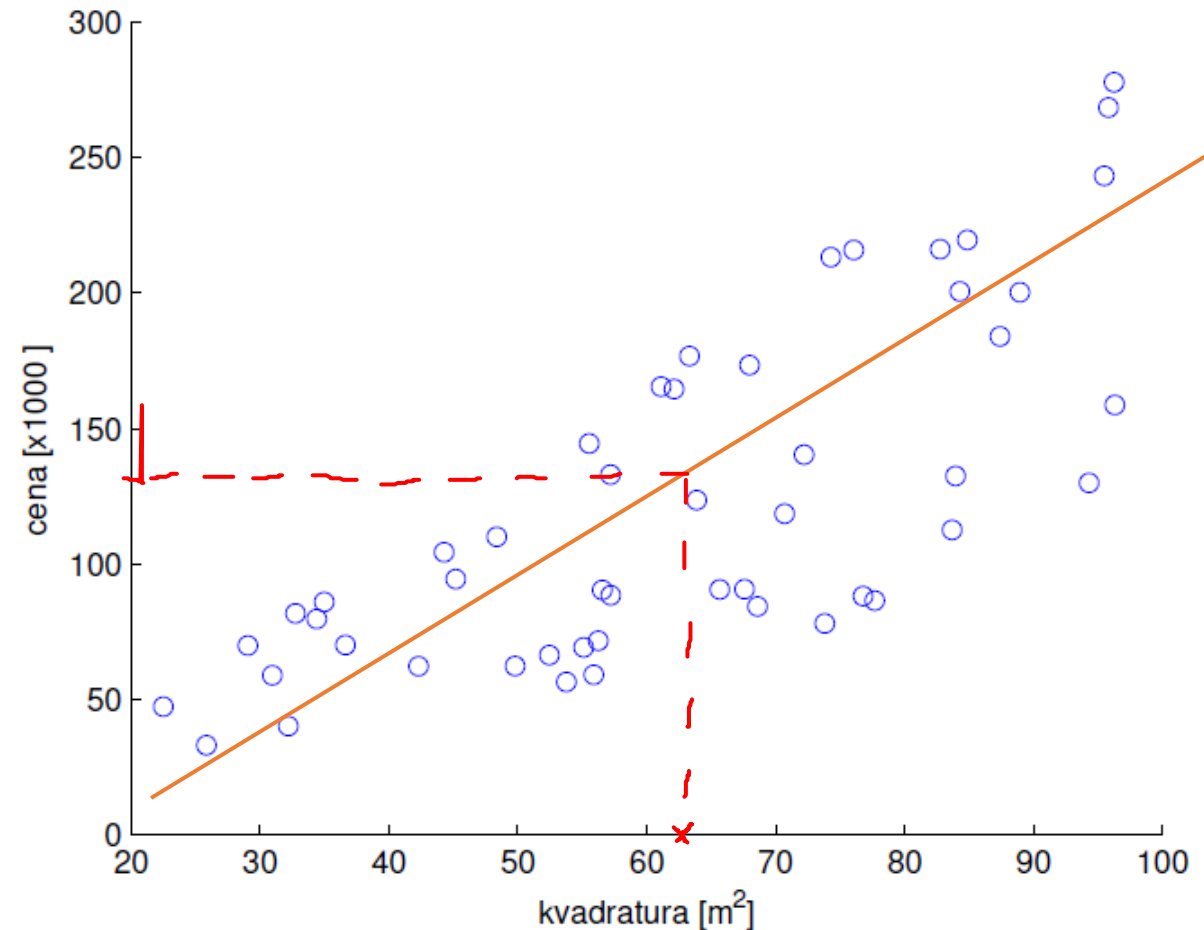


Nadgledano učenje – primer Klasifikacije



Nadgledano učenje – primer Regresije

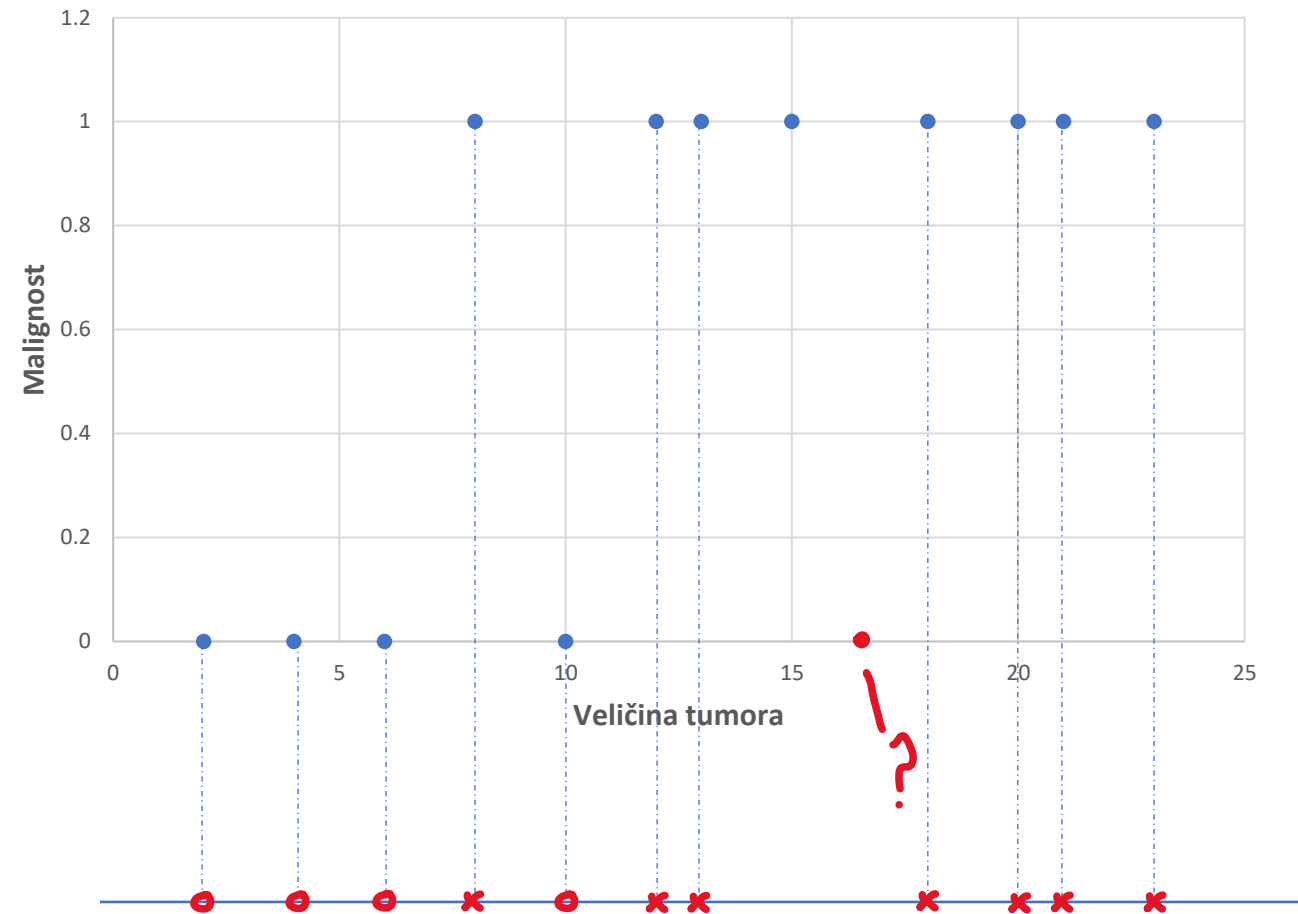
- Cena kuće u zavisnosti od kvadrature
 - Dato je (x,y) i treba odrediti kako se mapira $x \rightarrow y$
 - Problem Regresije



Problem klasifikacije

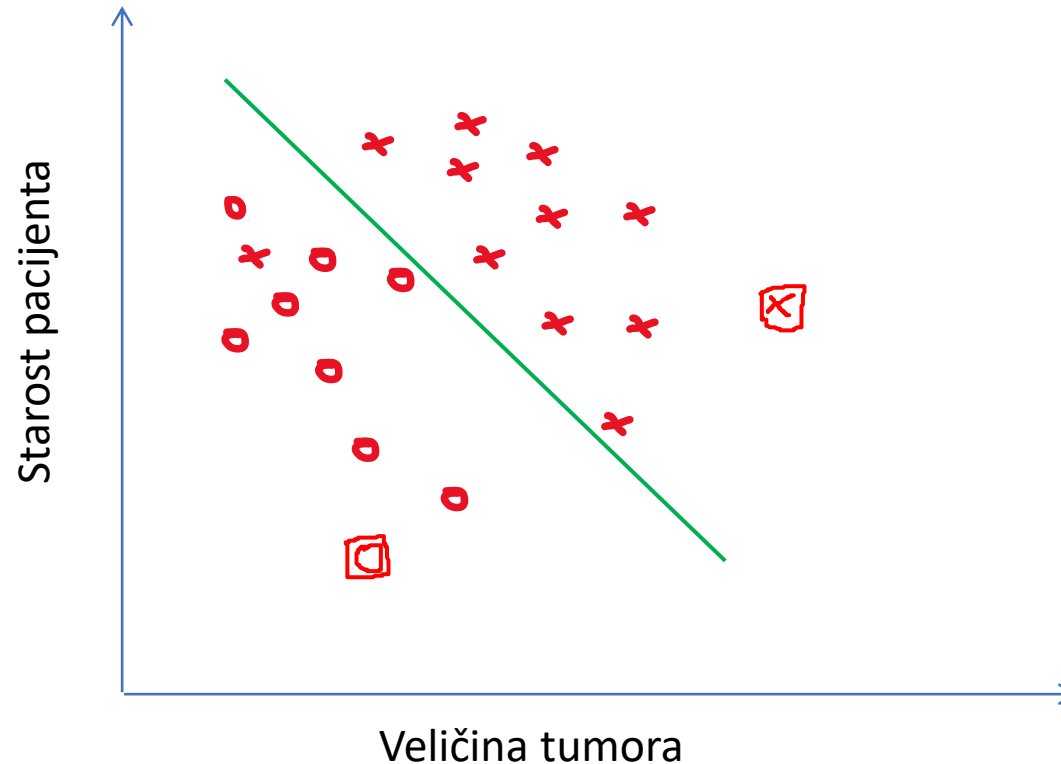
- Primer

- Drugi način prikaza



Dimenzija problema

- Kod oba primera je ulaz X jednodimenzionalan, ali je u praksi najčešće višedimenzionalan, npr.

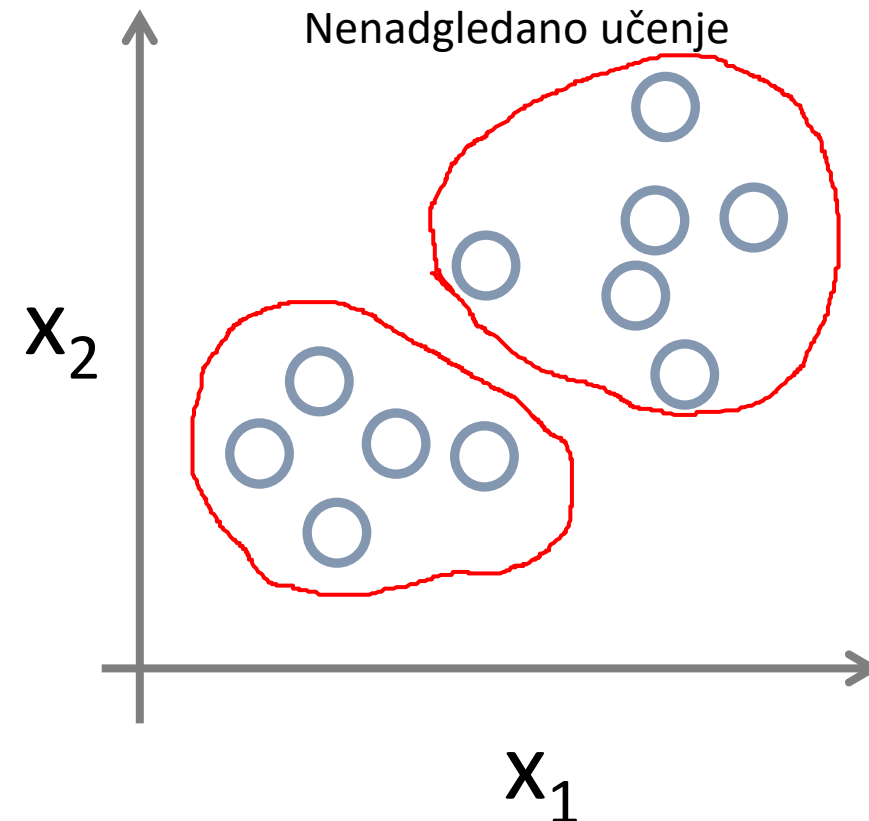
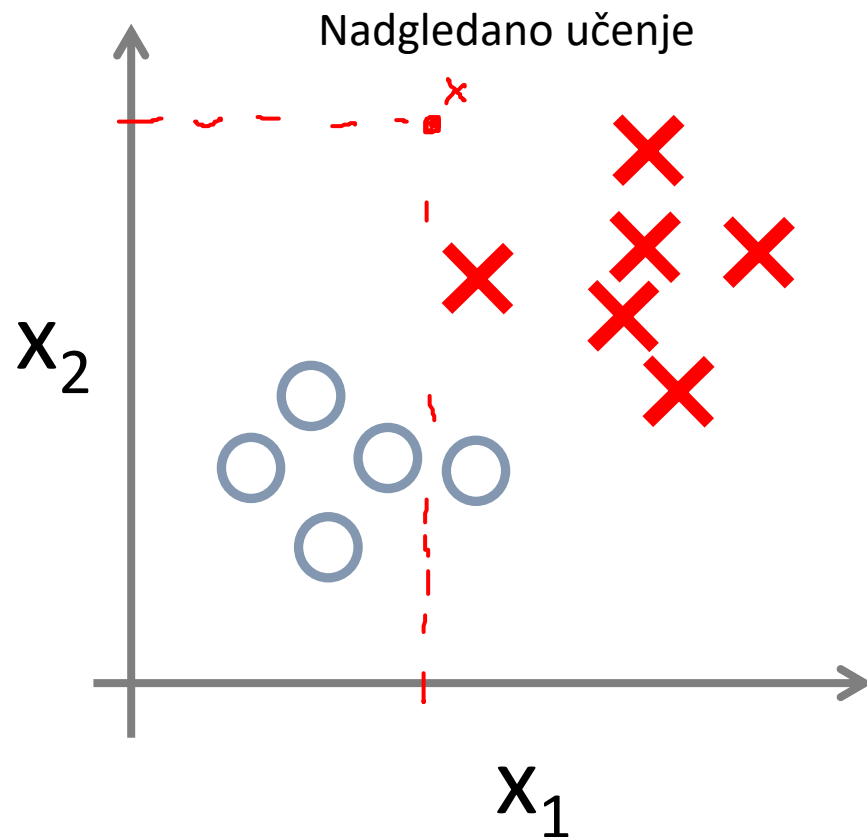


Algoritmi nadgledanog učenje

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests
- Neural networks

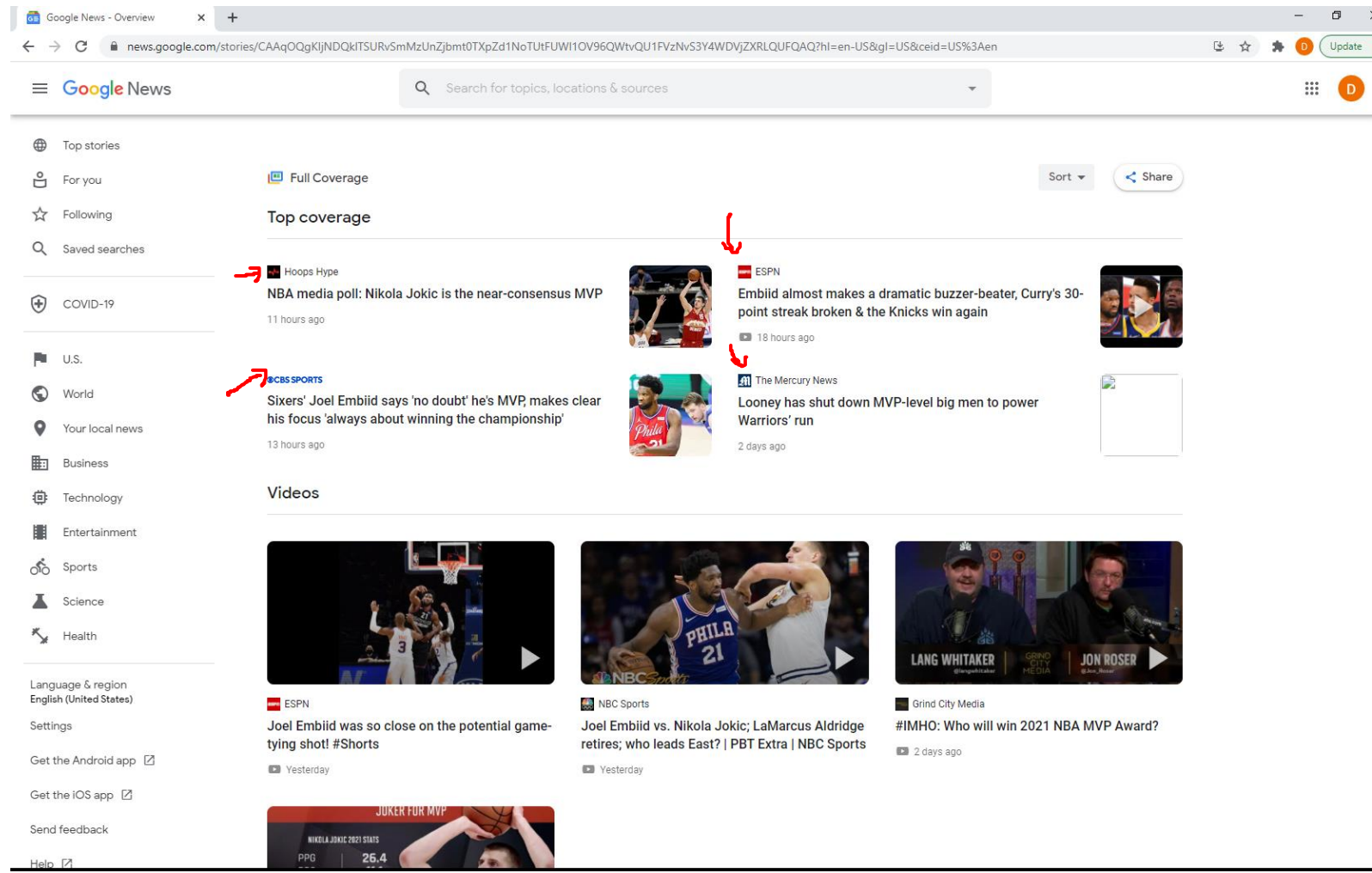
Nenadgledano učenje

- Nenadgledano učenje je koncept koji koristi neoznačene podatke samo X i nalaze interesantne stvari o njima



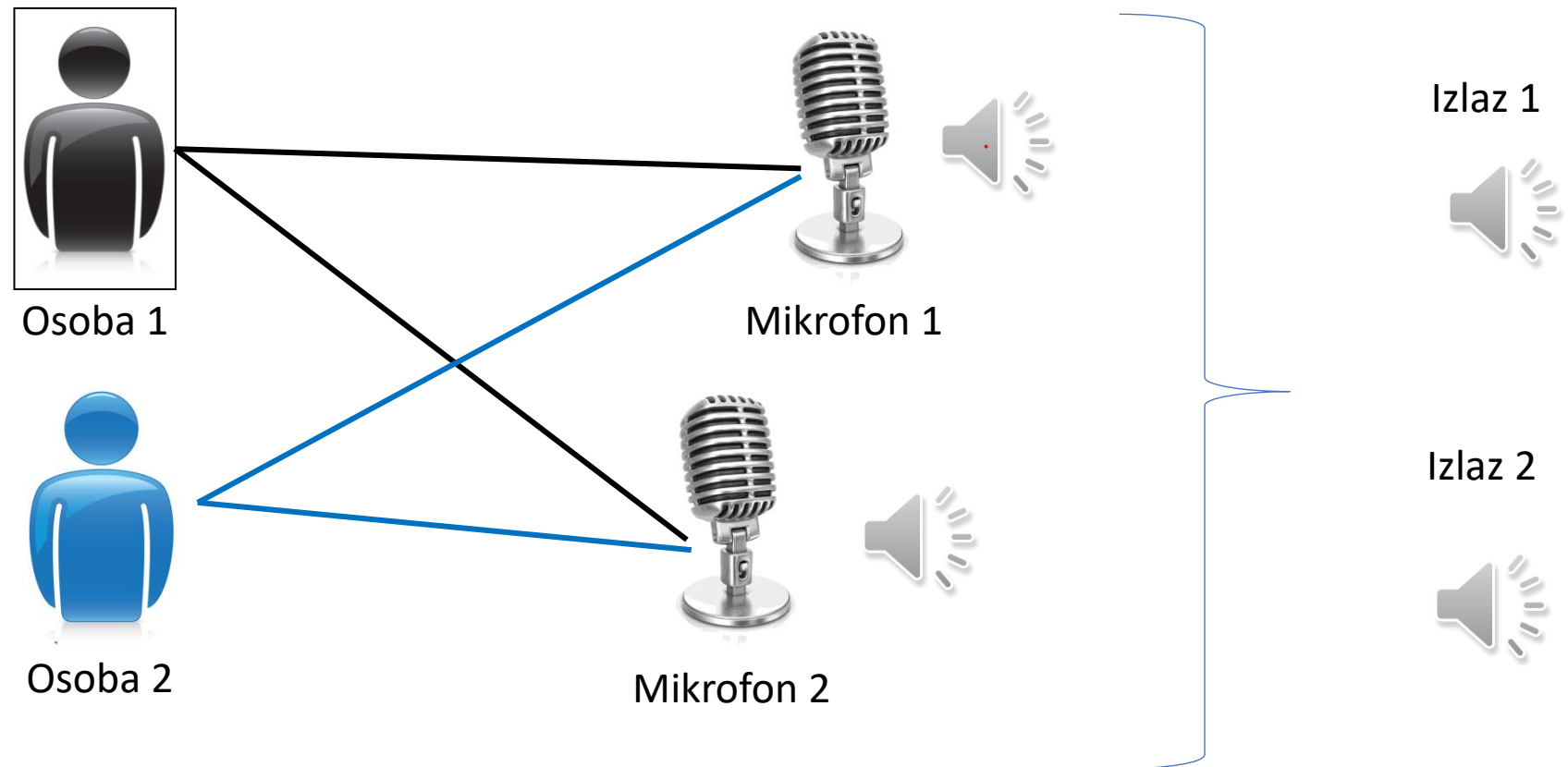
Nenadgledano učenje

- Najčešće se primenjuje za grupisanje (klastering) podataka
- Primer: Google new web site - izdvaja članke na istu temu od različitih autora



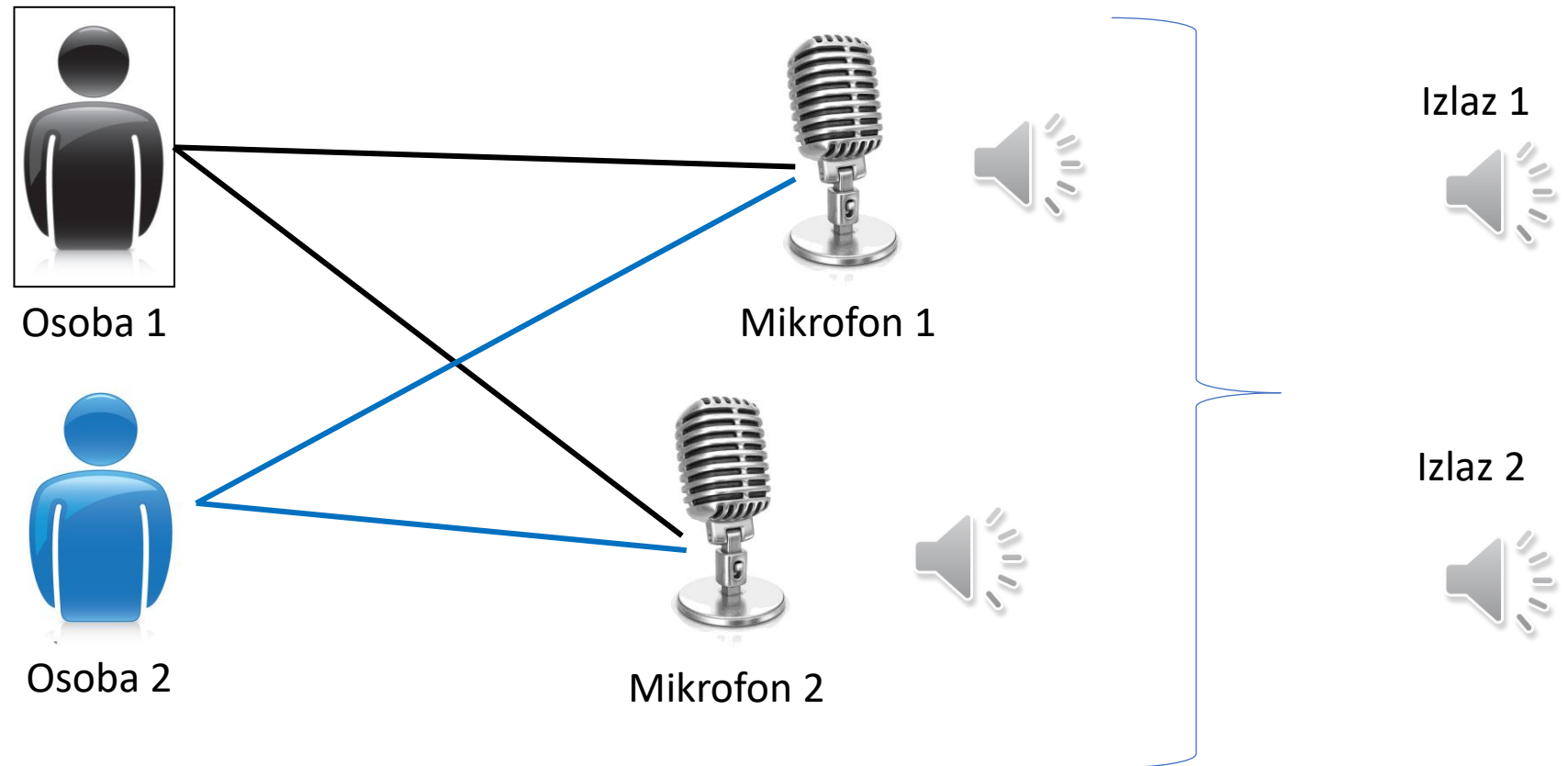
Primer: Koktel parti problem

- Kako razlikovati glasove različitih ljudi na žurci sa više mikrofona (120cm od govornika)?



Primer: Koktel parti problem

- Kako razlikovati glasove različitih ljudi u Sali za sastanke sa više mikrofona (60cm od govornika)?



Primer: Pronaći na internetu sajtove koji sadrže FTN

The screenshot shows a Google Scholar search interface. The search bar contains the text "fakultet tehnickih nauka". The results are displayed in a list format. On the left side, there are filters for "Чланци" (Articles) and "Око 16.900 резултата (0,07 сек)". Below the filters, there are options to sort by "Сортирај према значају" (Sort by relevance) and "Сортирај према датуму" (Sort by date). There are also checkboxes for "укључи патенте" (include patents) and "укључи цитате" (include citations), and a button for "Направи обавештење" (Create alert).

Било када
Од 2021.
Од 2020.
Од 2017.
Прилагођени опсер...

Сортирај према значају
Сортирај према датуму

☐ укључи патенте
☒ укључи цитате

☒ Направи обавештење

[PDF] Međunarodni projekti doprinos internacionalizaciji Fakulteta tehničkih nauka [PDF] uns.ac.rs
I Čosić, J Hodolić, V Katić, I Kovačević, S Kolaković... - 2012 - trend.uns.ac.rs
Fakulteta tehničkih nauka u Novom Sadu u skladu sa Planom naučnoistraživačkog rada kao jedno od svojih osnovnih načela navodi upravo internacionalizaciju, odnosno uspostavljanje i jačanje saradnje sa drugim visoko obrazovnim i naučnoistraživačkim ...
☆ 3 пута наведен Сродни чланци

[PDF] Elektronsko učenje: Fakultet tehničkih nauka [PDF] uns.ac.rs
B Lalić, U Marjanović, B Bogojević, I Žunić... - ... I nauke u Srbiji" (130 ..., 2017 - trend.uns.ac.rs
This paper presents the highlights of Moodle eLLab platform for distance learning at the Faculty of Technical Sciences, University of Novi Sad for the 2016/17 year. Comparative analysis between school year 2015/16 and winter semester 2016/17 has been conducted ...
☆ 3 пута наведен Сродни чланци

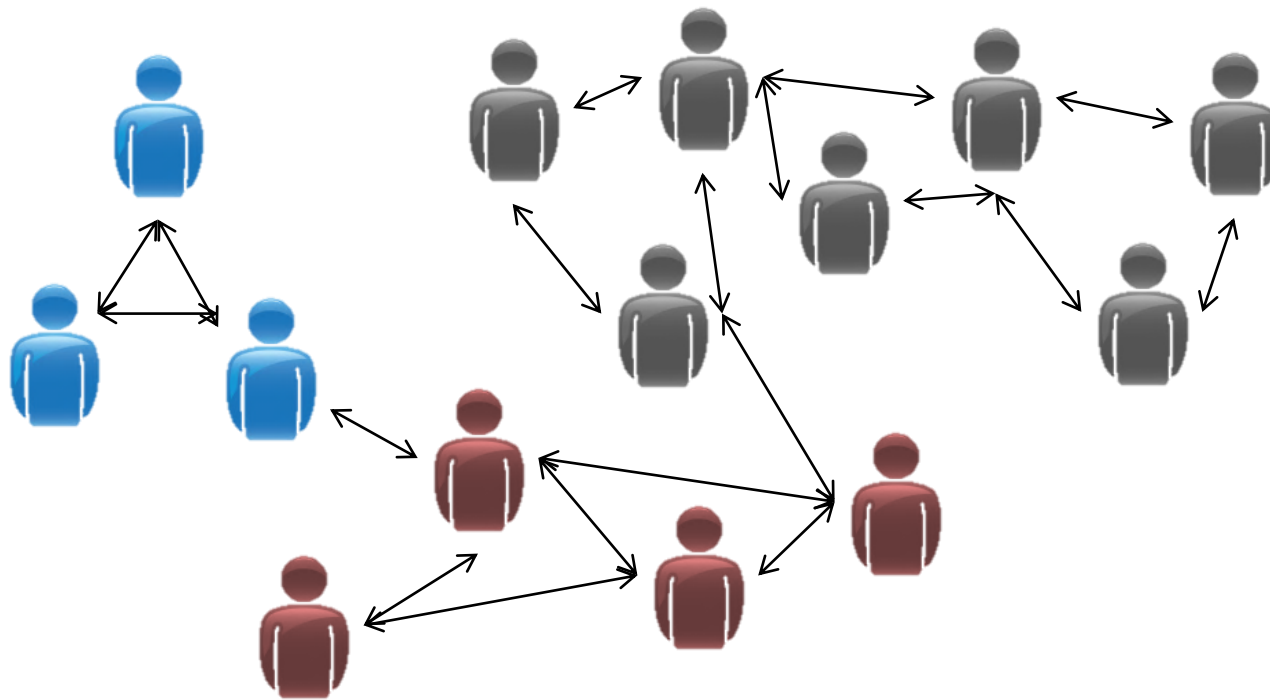
Preduzetničke težnje studenata Fakulteta tehničkih nauka
V Todorović, Ž Tekić, M Pečujlija - Učenje za poduzetništvo, 2012 - k2net.info
• Veoma veliki broj studenata, čak 83, 1% razmišljalo o pokretanju sopstvenog biznisa, međutim oni odustaju od te ideje pre nego što i pokušaju. • Na odluku o pokretanju sopstvenog biznisa, za studente presudan uticaj ima državna politika, odnosno nesigurna ...
☆ 6 пута наведен Сродни чланци Све верзије (3)

[PDF] Primena digitalnih tehnologija učenja: Fakultet tehničkih nauka [PDF] uns.ac.rs
B Lalić, U Marjanović, B Bogojević, I Žunić - 2018 - trend.uns.ac.rs
Technological development has influenced the ways in which learning takes place, and a variety of technological tools now supplement and partly replace paper books. Digital media have been transformed the knowledge dissemination process. This paper presents results of ...
☆ 3 пута наведен Сродни чланци

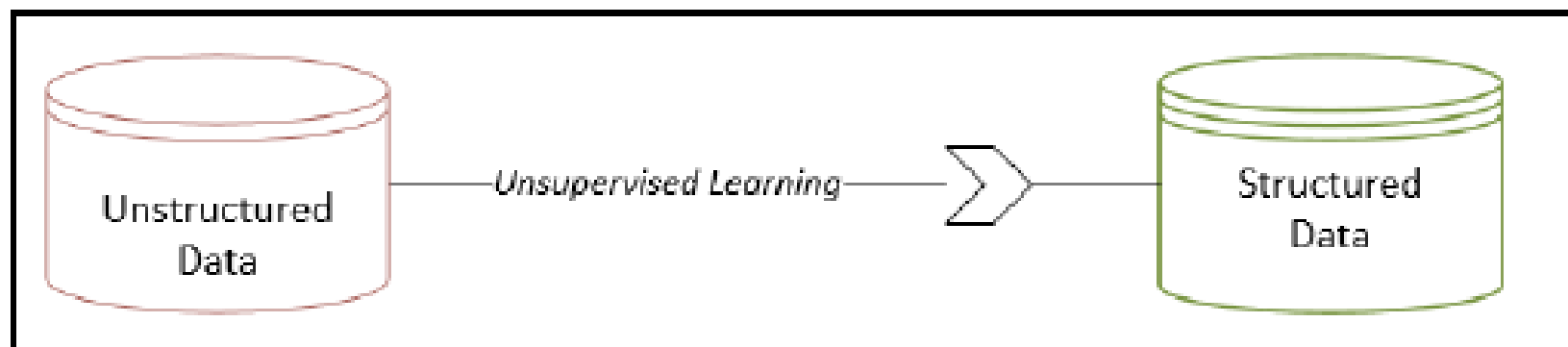
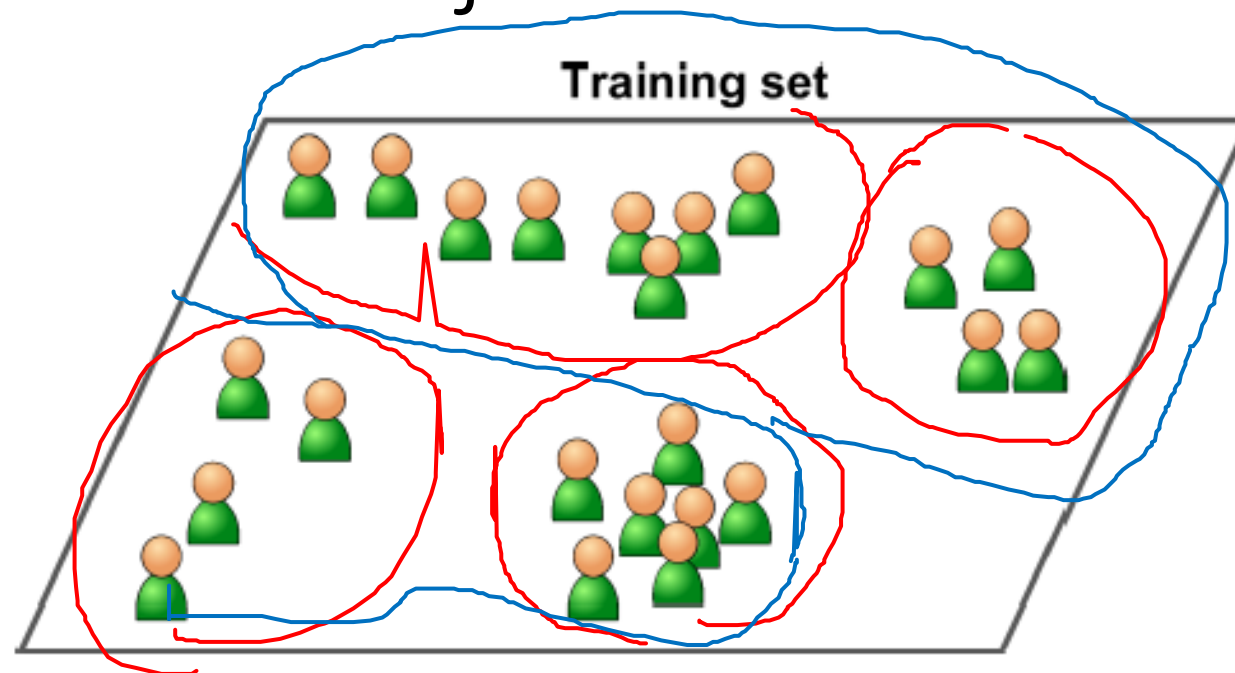
[PDF] Razvoj marketinških aktivnosti na Fakultetu tehničkih nauka u Novom Sadu [PDF] uns.ac.rs
N Simeunović, R Bojanić, D Medvecki, M Vrtunski... - 2018 - trend.uns.ac.rs
Recognizing the importance of permanent monitoring of changes in the knowledge market

Primer: Analiza socijalnih mreža

- Facebook i sl



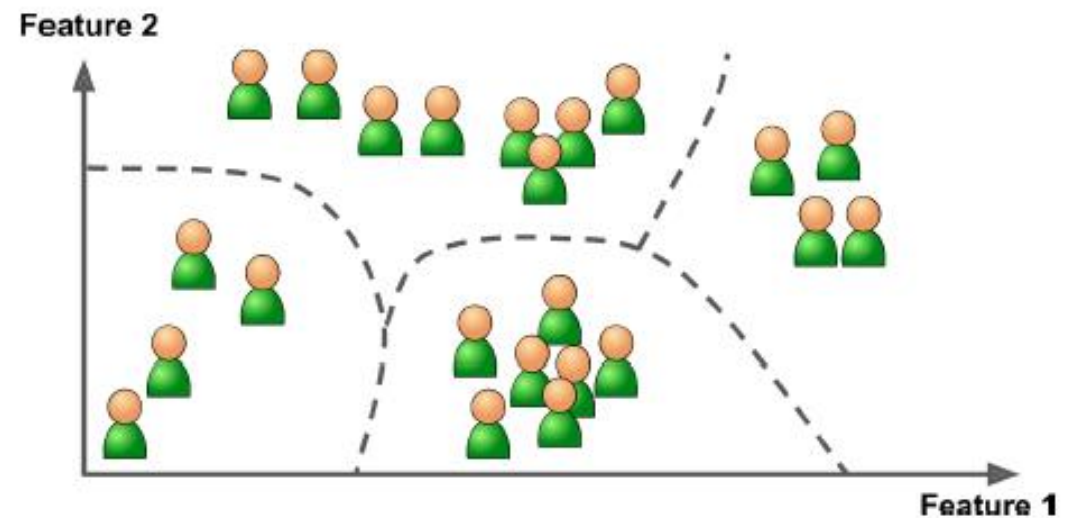
Nendgledano učenje



Algoritmi nenadgledanog učenja

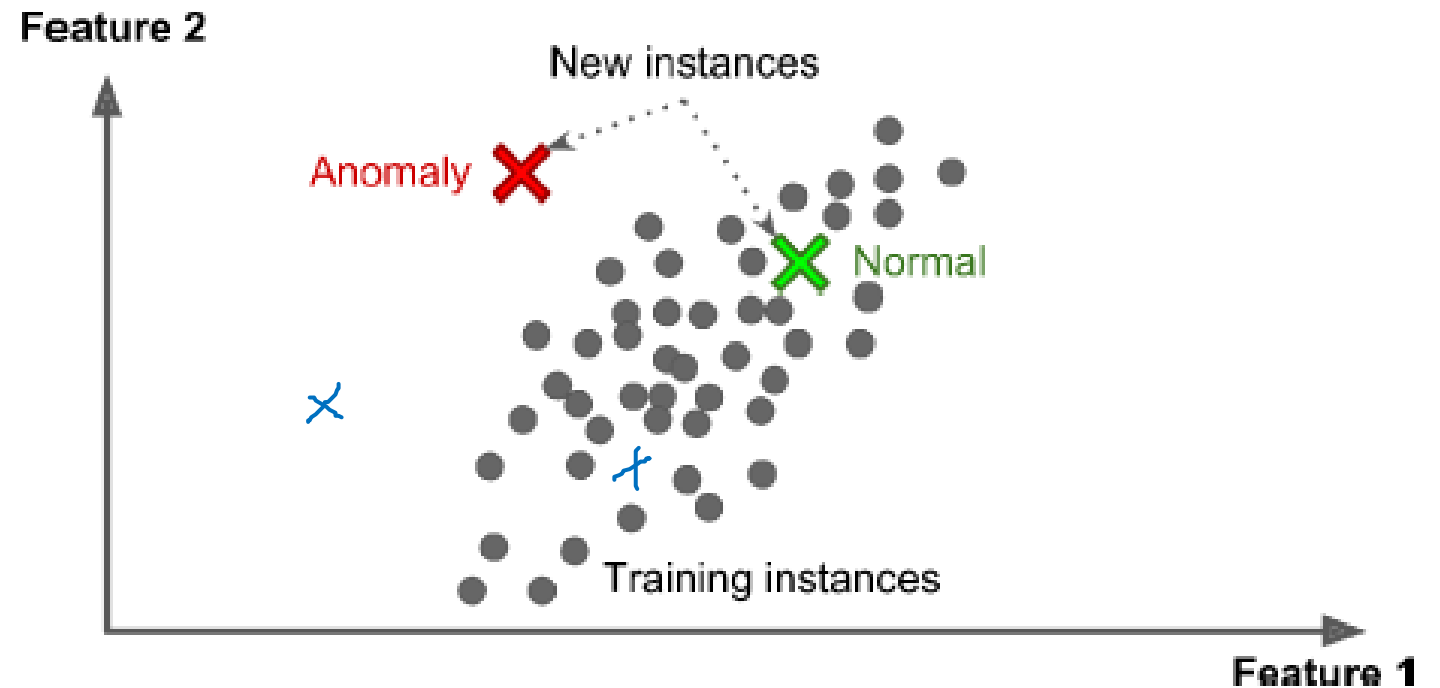
- **Grupisanje (Clustering)**

- K-Means
- DBSCAN (Density-based spatial clustering of applications with noise)
- Hierarchical Cluster Analysis (HCA)
- SVM (Support Vector Machine)



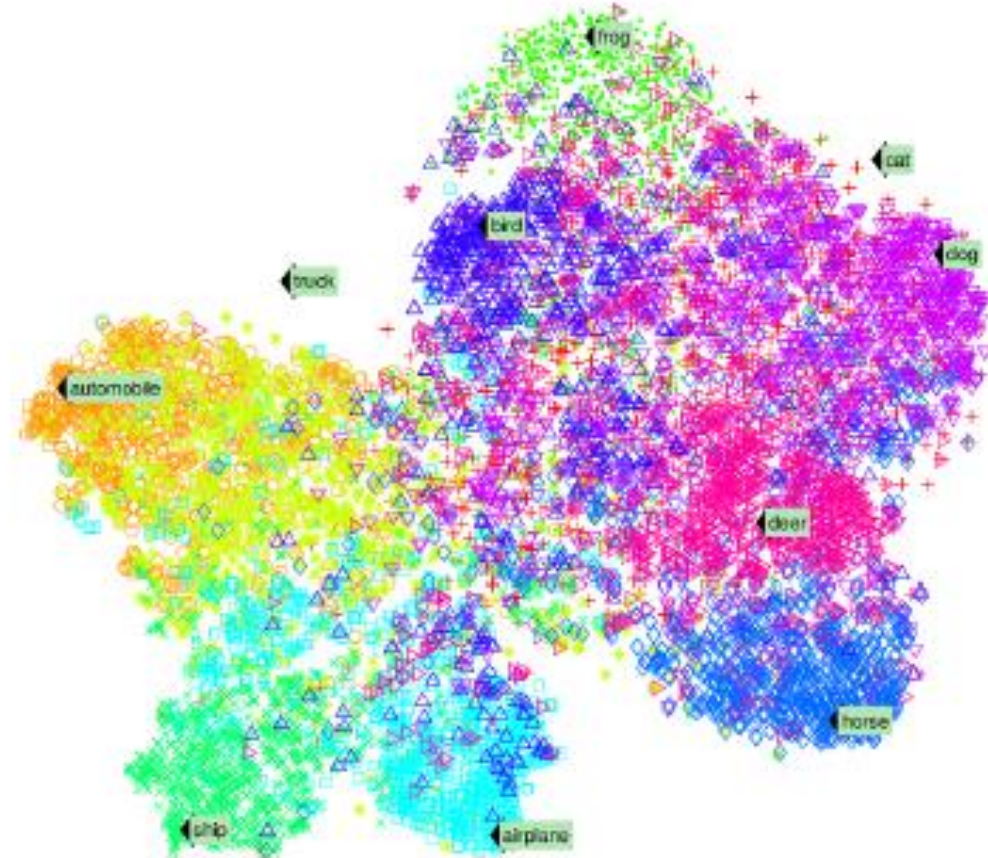
Algoritmi nenadgledanog učenja

- Grupisanje
- Detekcija anomalija
 - One-class SVM
 - Isolation Forest



Algoritmi nenadgledanog učenja

- Grupisanje
- Detekcija anomalija
- **Vizuelizacija i smanjenje dimenzionalnosti**
 - Principal Component Analysis (PCA)
 - Kernel PCA
 - Locally Linear Embedding (LLE)
 - t-Distributed Stochastic Neighbor Embedding (t-SNE)

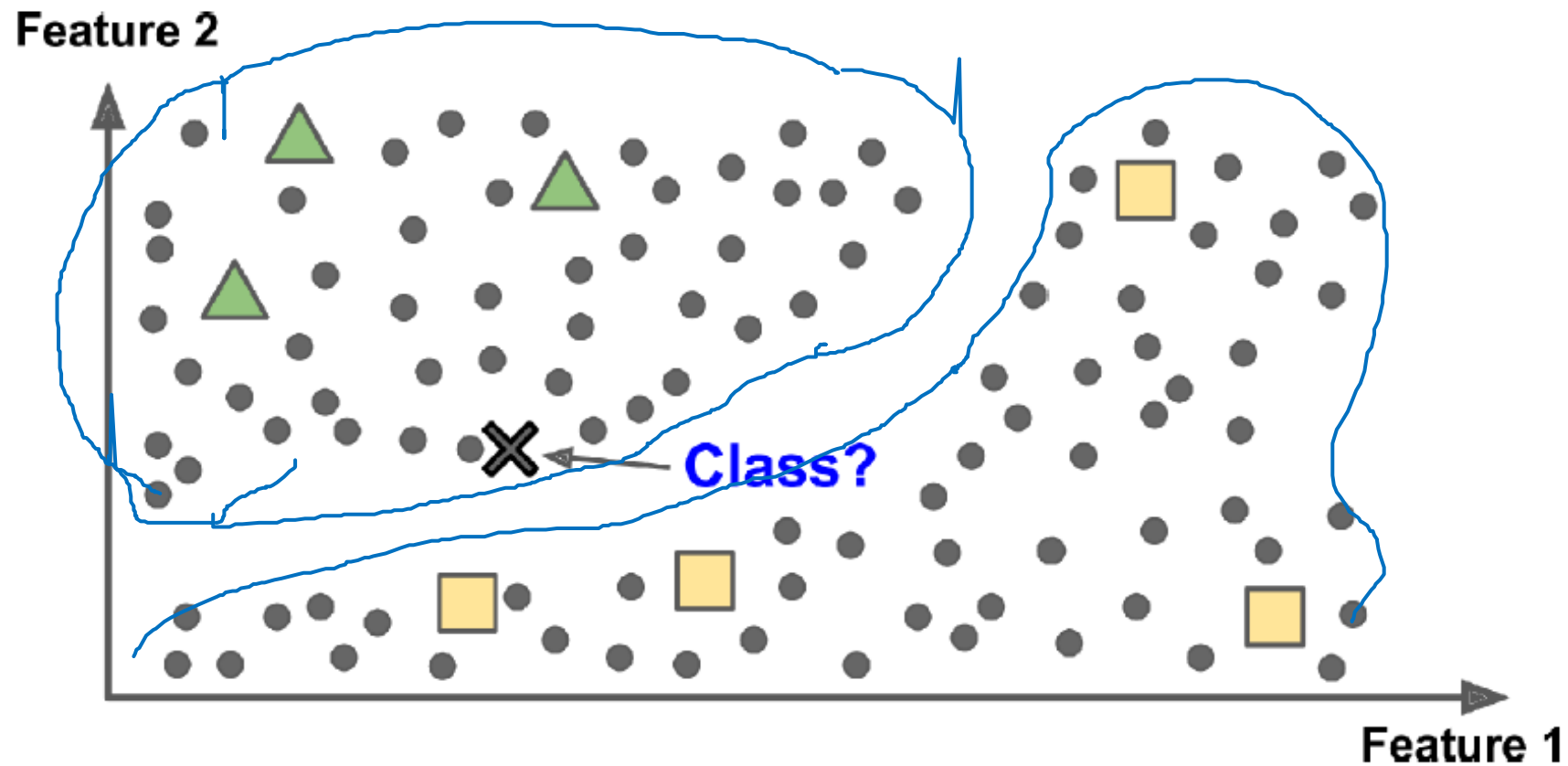


Algoritmi nenadgledanog učenja

- Grupisanje
- Detekcija anomalija
- Vizuelizacija i smanjenje dimenzionalnosti
- **Učenje pravila udruživanja**
 - Apriori
 - Eclat (Equivalence Class Clustering and bottom-up Lattice Traversal)

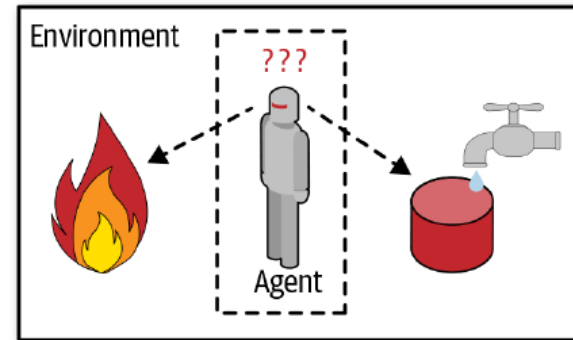
Polunadgledano učenje

- Delimično obeleženi podaci + neobeleženi podaci



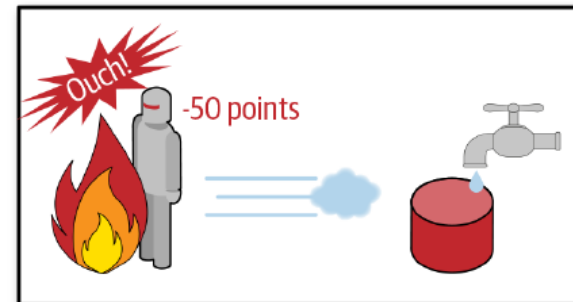
Učenje sa podsticajem (Reinforcement learning)

- Agenti izvode aktivnosti po nekoj strategiji
- Dobije se „nagrada“ ili „kazna“
- Uči koja je najbolja strategija - pravilo sa najviše nagrada
- Primer: Pas robot



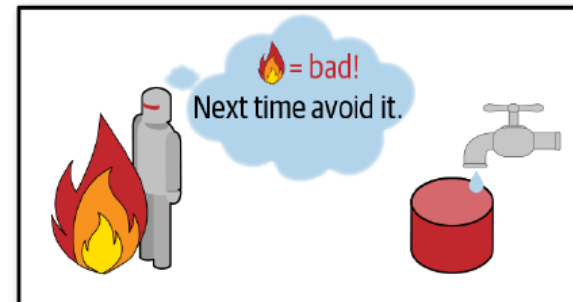
1 Observe

2 Select action using policy



3 Action!

4 Get reward or penalty



5 Update policy (learning step)

6 Iterate until an optimal policy is found

Notacija

- Podaci (tačke, skupovi) – x (poznato m tačaka)

- Svojstvo (vektor) – svojstvo je karakteristika podatka $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

- numeričko (visina čoveka = 1.83 m)
 - opisno (boja očiju je zelena) – prevodi se u numeričko
- $x^{(i)}$ – i -ti ulaz - vektor dimenzije n
- $y^{(i)}$ - izlaz za i – ti ulaz

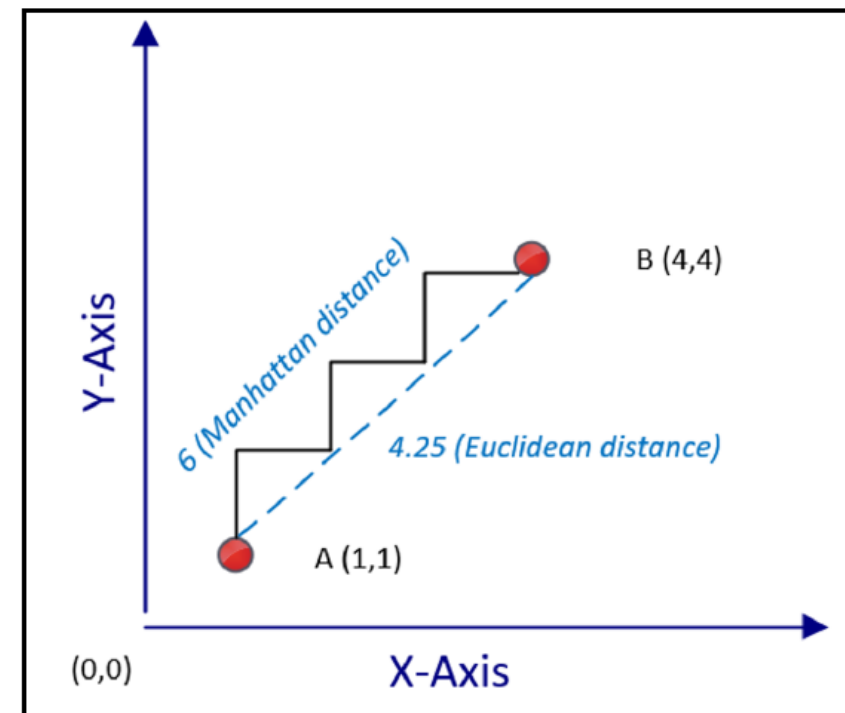
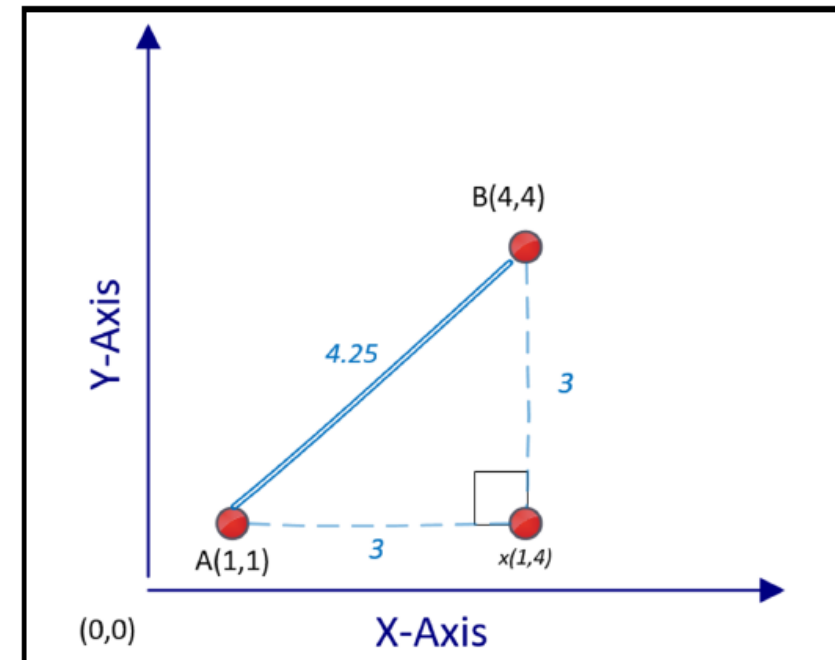
Merenje rastojanja

- Euklidsko rastojanje – Pitagorina teorema

$$d(A, B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

- Menhetn rastojanje – Suma abs vrednosti razlika ulaza u vektorima

$$\text{Manhattan} = \sum_{i=1}^n |x_i - y_i|$$



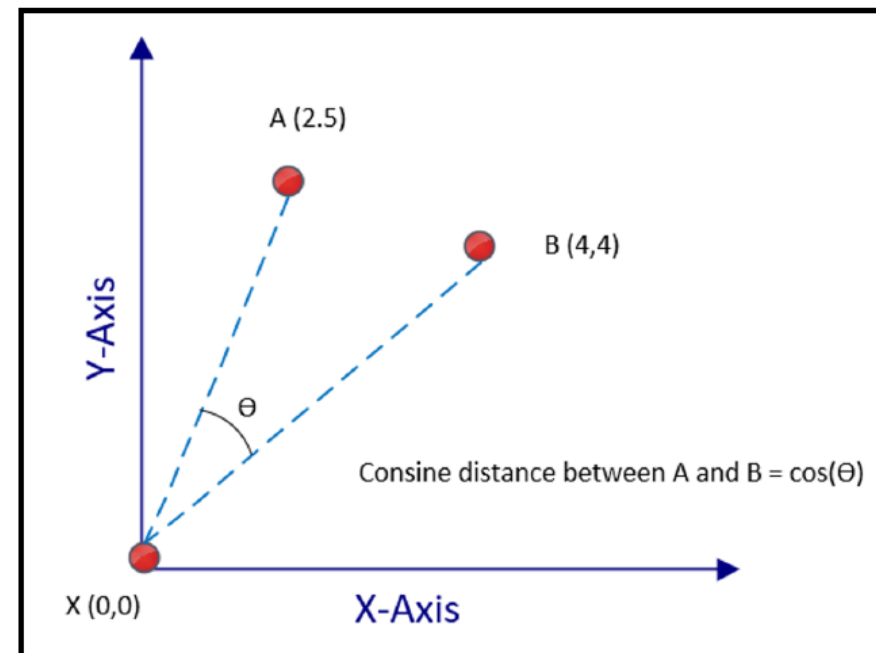
Merenje rastojanja

- Čebiševo rastojanje – maksimum abs vrednosti razlika između svih elemenata vektora

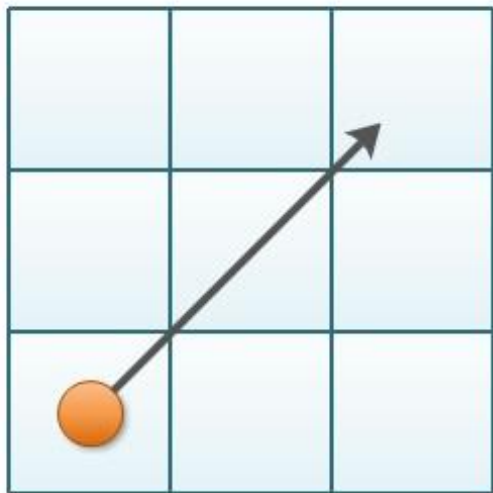
$$Chebyshev = \max(|x_i - y_i|)$$

- Sličnost kosinusa – vektori usmereni u istom smeru su slični (određuje se kosinus između dva vektora, tj. tačkasti proizvod / dve dužine)

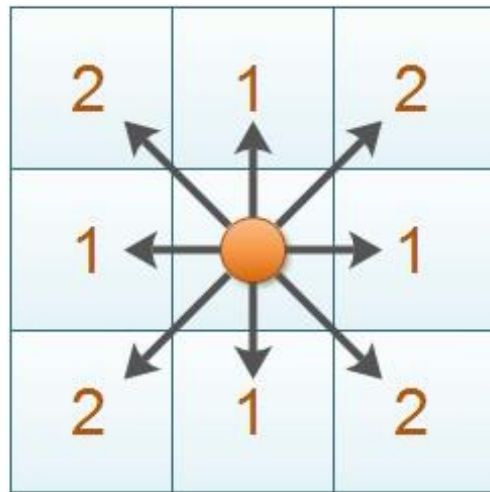
$$similarity(a, b) = \frac{a \cdot b}{\|a\| \cdot \|b\|}$$



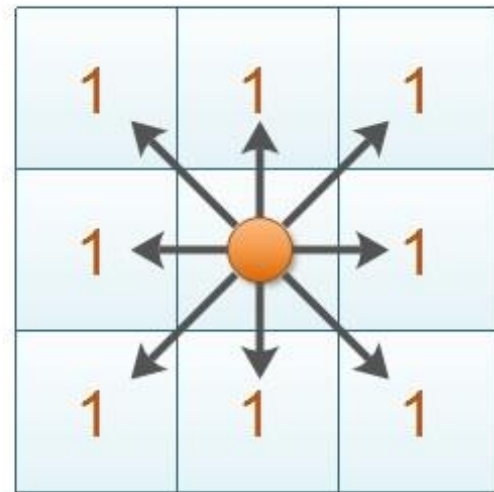
Euclidean Distance



Manhattan Distance



Chebyshev Distance



$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad |x_1 - x_2| + |y_1 - y_2| \quad \max(|x_1 - x_2|, |y_1 - y_2|)$$

Skaliranje podataka

- Prvi korak u cilju pravilnog poređena svojstava kod merenja rastojanja između tačaka

Skaliranje podataka

- Prvi korak u cilju pravilnog poređena svojstava kod merenja rastojanja između tačaka
- Npr. Ako se klasifikuju ljudi prema plati i broju kućnih ljubimaca, vrednosti za platu su preko 30000, a vrednosti za broj kućnih ljubimaca je < 10
- Jedan od načina je da se po svakom svojstvu u svim podacima odrede srednja vrednost i stand. dev. i da se skaliraju vrednosti tog svojstva tako da imaju srednju vrednost 0 i st.dev. 1
- Drugi način je da je minimalna vrednost 0, a najviša 1

Skaliranje obeležja

$$X^{(1)} = \begin{pmatrix} 250 \\ 5 \\ 1 \\ 15 \end{pmatrix} \quad Y^{(1)} = 460$$
$$X^{(2)} = \begin{pmatrix} 145 \\ 3 \\ 2 \\ 20 \end{pmatrix} \quad Y^{(2)} = 232$$

- Sva svojstva postaviti na približno istu skalu
- Preporuka: Opseg vrednosti za svojstva približno $-1 \leq x_i \leq 1$

	x_1 Površina (m ²)	x_2 Broj spavaćih soba	x_3 Spratnost	x_4 Starost kuće (god)	y Cena (€1000)
$x^{(1)}$	250	5	1	15	460
	145	3	2	20	232
	162	4	2	10	315
	93	2	1	32	178

Skaliranje svojstava - normalizacija

$$\begin{aligned} 0 &\leq x_1 \leq 2 \quad \checkmark \\ -2 &\leq x_2 \leq 0.8 \quad \checkmark \\ -150 &\leq x_3 \leq 200 \quad -5 \\ -0.0015 &\leq x_4 \leq 0.0025 \quad -5 \end{aligned}$$

- Svojstvo x_i se zamenjuje sa $x_i - \mu_i$
- Primer: Cena kuće

Površina (m ²)	Broj spavaćih soba	Spratnost	Starost kuće (god)	Cena (€1000)
250	5	1	15	460
145	3	2	20	232
162	4	2	10	315
93	2	1	32	178

$$\bar{x}_1 = \frac{254+145+162+\dots+95}{m} = 164$$

$$x_1 \in [50, 800]$$

$$x_1^{(1)} = \frac{x_1^{(1)} - \bar{x}_1}{\max - \min} = \frac{x_1^{(1)} - 164}{800 - 50} = \frac{250 - 164}{750} = 0.1446$$

$$x_1 = \frac{x_1}{\max(x_1)} \quad x_1 \in [0, 1]$$

$$x_1 = \frac{x_1 - \text{mean}}{\text{st. dev.}}$$