

# INTELIGENTNI SISTEMI

as. ms Vladimir Jocović  
as. ms Adrian Milaković



# NENADGLEDANO MAŠINSKO UČENJE

# 08

---

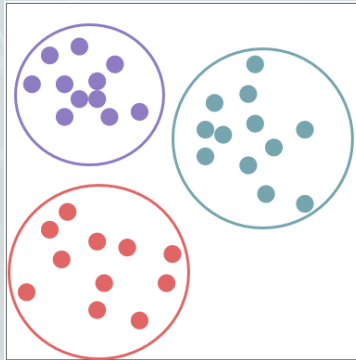
*„Where data points make new friends based  
on common interests.“  
- Unknown author*

# NENADGLEDANO UČENJE

## Tipovi izlazne vrednosti

**Klasterovanje** je tip nenadgledanog mašinskog učenja čiji je cilj da razdvoji podatke na određen broj grupa (klastera) na osnovu sličnosti i razlika.

Za razliku od klasifikacije, kod problema klasterovanja nemamo labelirane podatke.



# ALGORITAM K-MEANS

## Kako funkcioniše algoritam *k-means*?

Algoritam  $k$  srednjih vrednosti (*k-means*) je tip nenadgledanih algoritama mašinskog učenja. Koristi se kod problema klasterovanja.

Algoritam polazi od pretpostavke da će se slične instance nalaziti bliže u prostoru.

Algoritam započinje biranjem vrednosti za  $k$  (broj klastera), potom kreira  $k$  centroida (težišta klastera) na slučajan način, nakon čega iterativno dodeljuje svakom podatku jedan od uspostavljenih klastera na osnovu najbližeg centroida i ažurira sve centroide dokle god se ne uspostavi stabilno stanje (nema promena u klasterima) ili se ne dostigne maksimalan broj iteracija.

# ALGORITAM K-MEANS

## Kako funkcionise algoritam k-means?

Valjanost dobijenih klastera može da se izračuna na osnovu funkcije greške:

$$J(c, t) = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - t_c^{(i)})^2$$

$c$  – klaster

$t$  – težište klastera (centroid)

$x$  – primerci

Da bi se izbeglo zaglavljivanje u lokalnom minimumu, algoritam se često pušta nekoliko stotina puta i bira rešenje za minimalnom vrednošću funkcije greške.

Vrednost  $k$  se često bira na osnovu domenskog znanja, ili upoređivanjem funkcije greške za različite vrednosti  $k$ .

# Zadatak 1 - Oblici



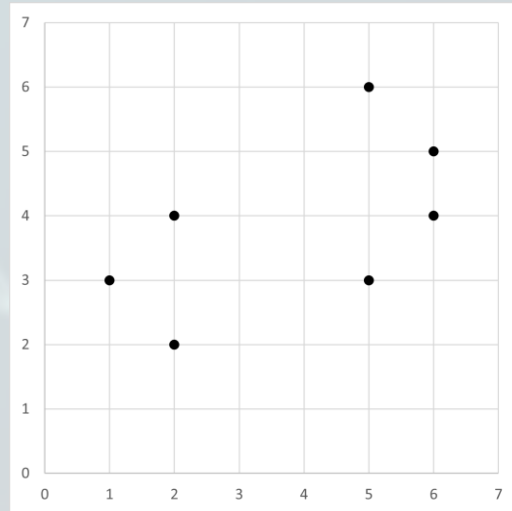
Na osnovu posmatranog skupa o dimenzijama oblika, podeliti skup na dve grupe koristeći *k-means* algoritam sa Euklidskim rastojanjem i izračunati grešku.

Nezavisni atributi		Atribut odluke
Širina (cm)	Visina (cm)	Grupa
1	3	?
5	6	?
5	3	?
2	2	?
6	4	?
6	5	?
2	4	?

# Zadatak 1 - Rešenje

Na grafiku su prikazani posmatrani primerci.

Širina (cm)	Visina (cm)
1	3
5	6
5	3
2	2
6	4
6	5
2	4



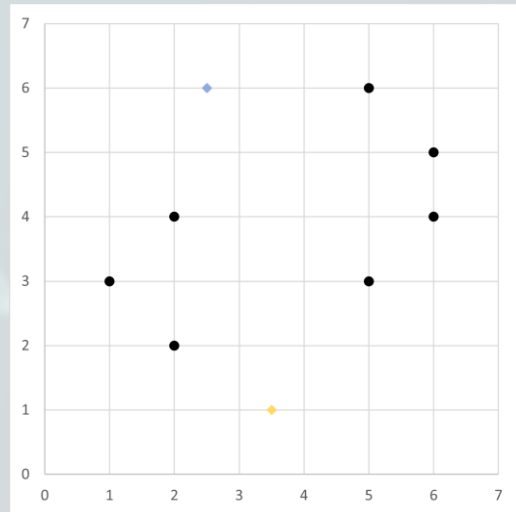
# Zadatak 1 - Rešenje

Algoritam započinje nasumičnim odabirom dve tačke koje služe kao početni centroidi klastera.

Širina (cm)	Visina (cm)
1	3
5	6
5	3
2	2
6	4
6	5
2	4

Širina (cm)	Visina (cm)
2.5	6
3.5	1





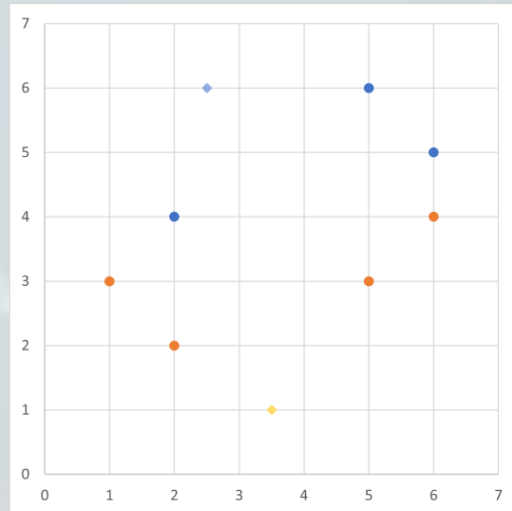
# Zadatak 1 - Rešenje

Za svaki primerak se računa udaljenost od oba centroida i primerak se pridružuje klasteru koji mu je bliži.

Š (cm)	V (cm)	C1	C2
1	3	3.35	3.2
5	6	2.5	5.22
5	3	3.9	2.5
2	2	4.03	1.8
6	4	4.03	3.9
6	5	3.64	4.71
2	4	2.06	3.35

Širina (cm)	Visina (cm)
2.5	6
3.5	1



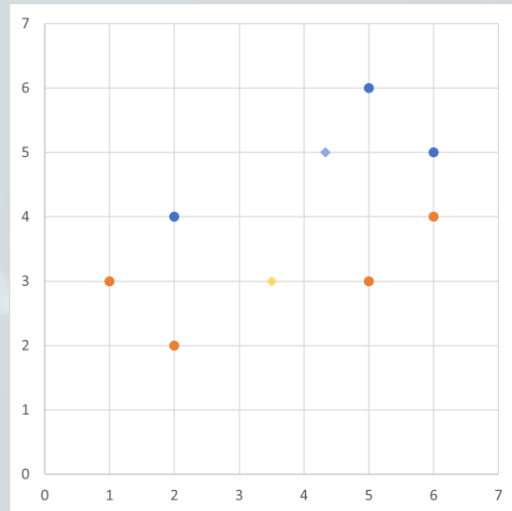
# Zadatak 1 - Rešenje

Potom se ažuriraju centroidi tako što se on pomera u težište svoje grupe.

Š (cm)	V (cm)	C1	C2
1	3	3.35	3.2
5	6	2.5	5.22
5	3	3.9	2.5
2	2	4.03	1.8
6	4	4.03	3.9
6	5	3.64	4.71
2	4	2.06	3.35

Širina (cm)	Visina (cm)
4.33	5
3.5	3



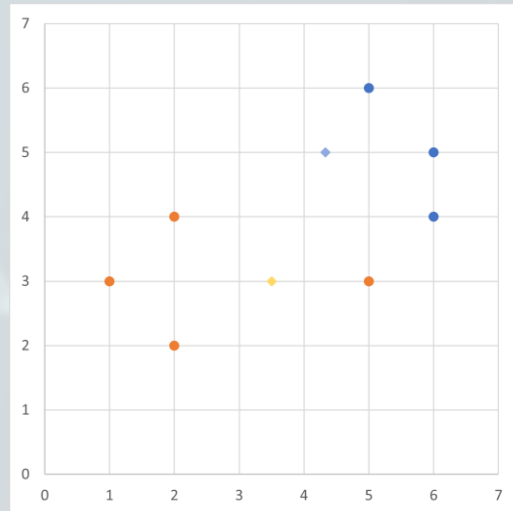
# Zadatak 1 - Rešenje

Algoritam se iterativno ponavlja dokle god primerci menjaju klaster kojem pripadaju.

Š (cm)	V (cm)	C1	C2
1	3	3.88	2.5
5	6	1.2	3.35
5	3	2.1	1.5
2	2	3.79	1.8
6	4	1.94	2.69
6	5	1.67	3.2
2	4	2.53	1.8

Širina (cm)	Visina (cm)
4.33	5
3.5	3



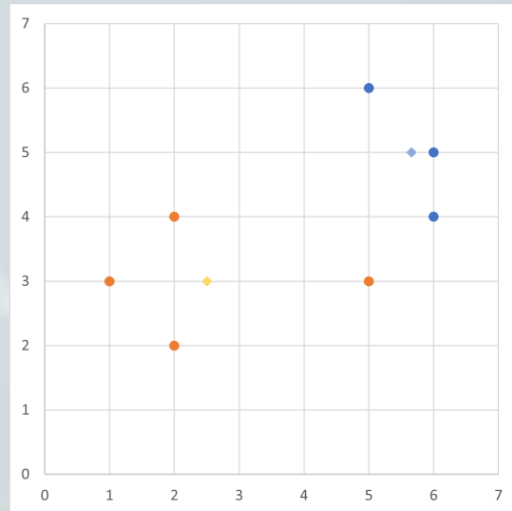
# Zadatak 1 - Rešenje

Algoritam se iterativno ponavlja dokle god primerci menjaju klaster kojem pripadaju.

Š (cm)	V (cm)	C1	C2
1	3	3.88	2.5
5	6	1.2	3.35
5	3	2.1	1.5
2	2	3.79	1.8
6	4	1.94	2.69
6	5	1.67	3.2
2	4	2.53	1.8

Širina (cm)	Visina (cm)
5.67	5
2.5	3



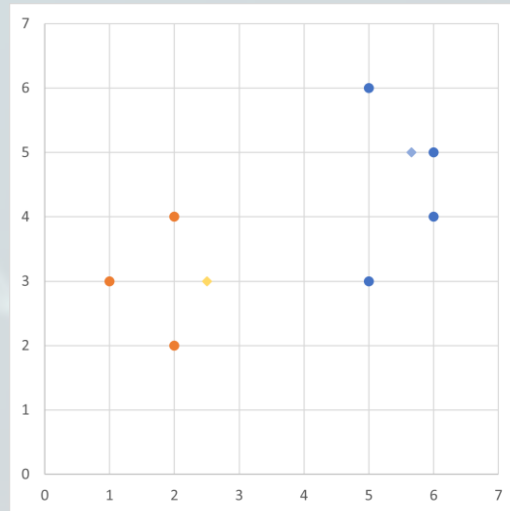
# Zadatak 1 - Rešenje

Algoritam se iterativno ponavlja dokle god primerci menjaju klaster kojem pripadaju.

Š (cm)	V (cm)	C1	C2
1	3	5.08	1.5
5	6	1.2	3.9
5	3	2.11	2.5
2	2	4.74	1.12
6	4	1.05	3.64
6	5	0.33	4.03
2	4	3.8	1.12

Širina (cm)	Visina (cm)
5.67	5
2.5	3



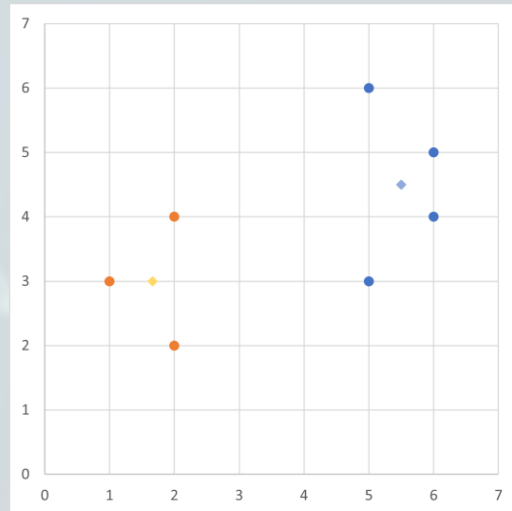
# Zadatak 1 - Rešenje

Algoritam se iterativno ponavlja dokle god primerci menjaju klaster kojem pripadaju.

Š (cm)	V (cm)	C1	C2
1	3	5.08	1.5
5	6	1.2	3.9
5	3	2.11	2.5
2	2	4.74	1.12
6	4	1.05	3.64
6	5	0.33	4.03
2	4	3.8	1.12

Širina (cm)	Visina (cm)
5.5	4.5
1.67	3



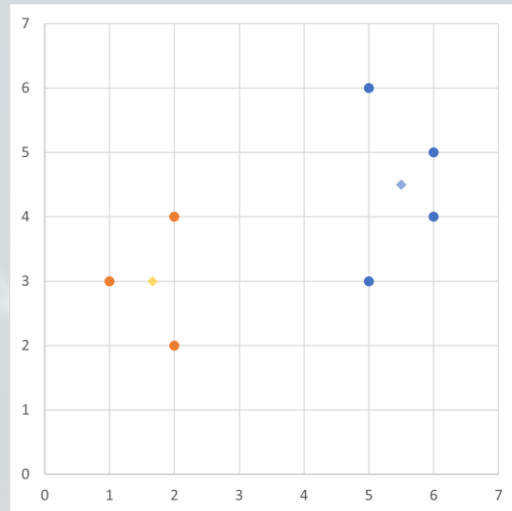
# Zadatak 1 - Rešenje

Kako nema izmena u klasterima nakon izračunavanja udaljenosti od centroida, algoritam se završava i konačan klaster se dodeljuje primercima.

Š (cm)	V (cm)	C1	C2
1	3	4.74	0.67
5	6	1.58	4.48
5	3	1.58	3.33
2	2	4.3	1.05
6	4	0.71	4.44
6	5	0.71	4.77
2	4	3.53	1.05

Širina (cm)	Visina (cm)
5.5	4.5
1.67	3



$$\text{Greška} = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - t_c^{(i)})^2$$

$$= (1/7) * (0.67^2 + 1.58^2 + 1.58^2 + 1.05^2 + 0.71^2 + 0.71^2 + 1.05^2) = 1.23$$

# Zadatak za samostalnu vežbu - Oblici



Na osnovu posmatranog skupa o dimenzijama oblika, podeliti skup na dve grupe koristeći *k-means* algoritam sa Euklidskim rastojanjem i izračunati grešku. Centroidi su se nasumično formirali na tačkama (2, 1) i (6, 6).

Nezavisni atributi		Atribut odluke
Širina (cm)	Visina (cm)	Grupa
6	1	?
4	4	?
2	5	?
3	6	?
5	3	?
1	6	?
5	1	?



# PITANJA?

<http://ri4es.etf.rs/>

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.