

# Mašinsko učenje

## Vežbe 04

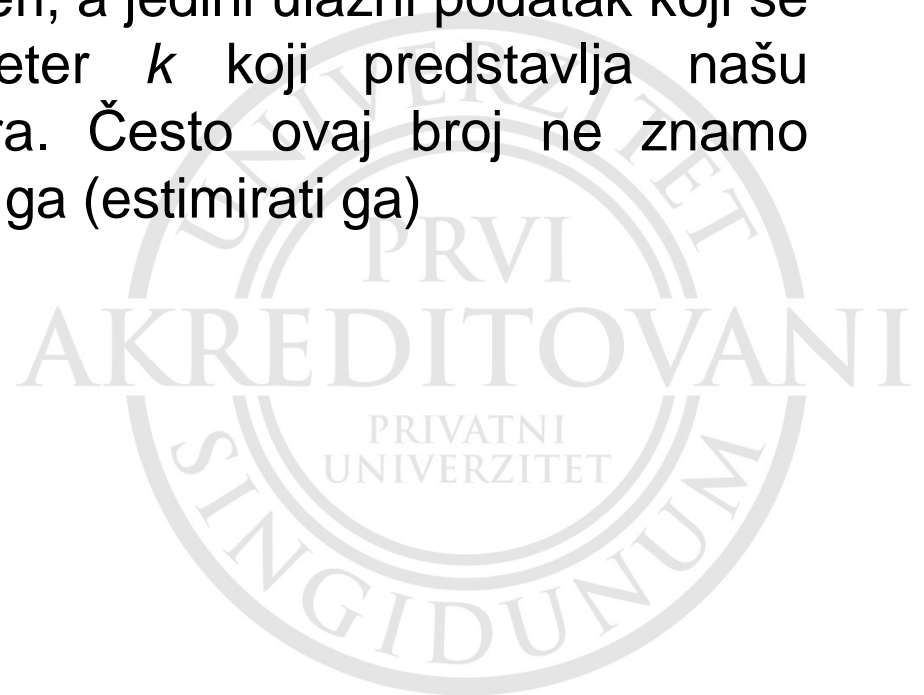
profesor: Prof. Dr. Milan Milosavljević

asistent: doc. dr. Vladimir Matic

email: [vmatic@singidunum.ac.rs](mailto:vmatic@singidunum.ac.rs)

# Klasterovanje

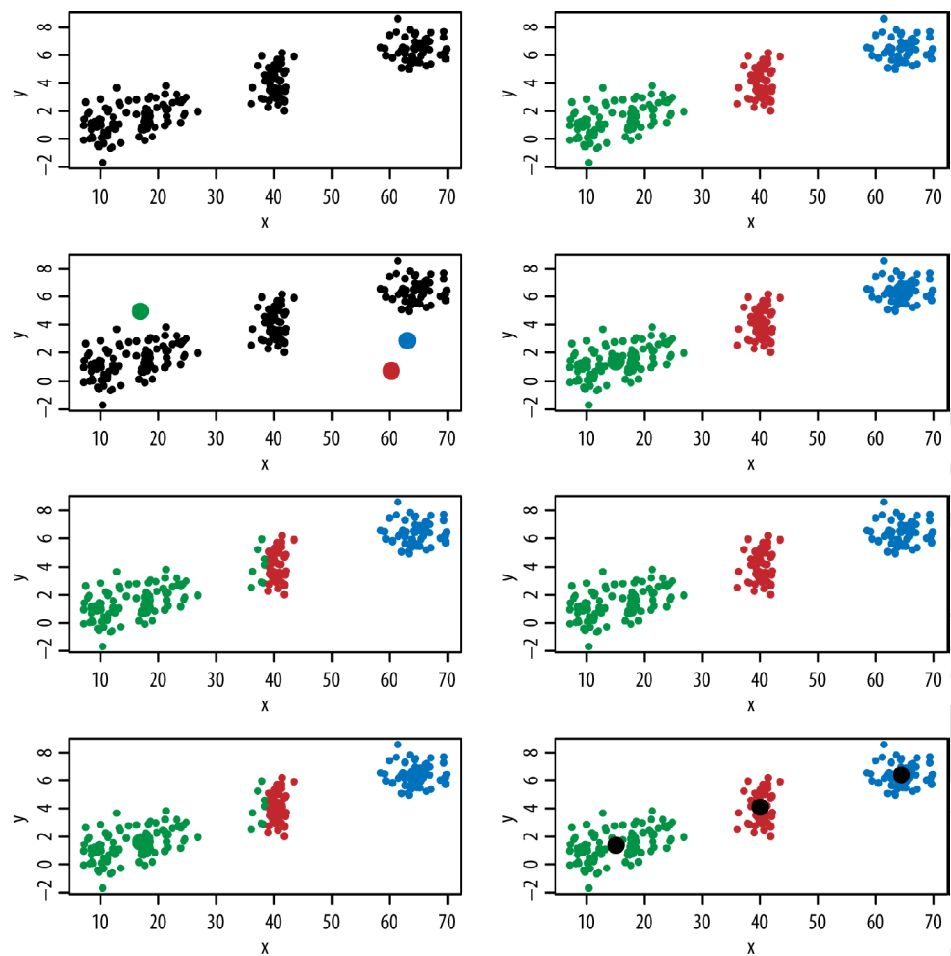
- Klasterovanje predstavlja *unsupervised learning* problem mašinskog učenja jer klase elemenata nisu unapred poznate.
- *k-means* je jedan od algoritama za klasterovanje.
  - Zadatak je da se kreiraju klasteri, a jedini ulazni podatak koji se zadaje algoritmu je parameter  $k$  koji predstavlja našu pretpostavku o broju klastera. Često ovaj broj ne znamo unapred i potrebno je proceniti ga (estimirati ga)



# Primer klasterovanja

- Dobili ste veliki broj tekstualnih podataka (ili e-knjiga). Potrebno je razvrstati ih automatski po oblastima (folderima), tako da se slične knjige nalaze u blizini.
- Pratite grupu pacijenata i potrebno je da odredite grupe kojima ćete prepisati odgovarajuće terapije, na osnovu vremenske serije bio-markera koje pratite.
- Potrebno je da kreirate profile ljudi – grupe, kojima ćete slati personalizovane oglase. Npr. Kojim korisnicima ponuditi oglase o putovanjima, a kojima o online naručivanju hrane.

# k-means algoritam :: ilustracija



# k-means algoritam

1. Inicijalizacija. Izabrati na slučajan način  $k$  centroida. Centroidi su tačke koje predstavljaju centre klastera
2. Dodeliti svim podacima pripadnost klasteru čiji je centroid njemu najbliži
3. *Update*-ovati, centre klaster izračunavanjem srednje vrednosti koordinata svih elemenata iz datog klastera
4. Ponavljati korake 2. i 3. sve do trenutka dok primena koraka 2. više ne dovodi do promena pripadnosti klasterima ili dok se ne dostigne unapred definisan maksimalni broj iteracija (npr.  $n\_iter = 1000$ )