K-means pitanja

- Da li konvergira? Da li u lokalni ili u globalni minimum?
- Kako da procenimo kvalitet dobijenih klastera?
- Kako da odredimo broj klastera *K*?
- Kako definisati metriku udaljenosti/sličnosti?

K-means kao *Coordinate Descent*

- *Distortion function*: $J(c,\mu) = \sum_{i=1}^{N} \left\| x^{(i)} - \mu_{z^{(i)}} \right\|_2^2$

- K-means alternira između dva koraka:
  1. Dodeli opservacije najbližem centroidu
  $$z^{(i)} \longleftarrow \arg\min_{j} \left\| \mu_j - x^{(i)} \right\|_2^2$$
  2. Ažuriraj centroide da budu srednja vrednost dodeljenih opservacija
  $$\mu_j = \frac{1}{n_j} \sum_{i:z^{(i)}=j} x^{(i)} : \ \mu_j \leftarrow \arg\min_{\mu} \sum_{i:z^{(i)}=j} \left\| \mu - x^{(i)} \right\|_2^2$$

- Alteriranje minimizacije
  1. Minimizacija *J* promenom z za fiksno μ
  2. Minimizacija *J* promenom μ za fiksno z

  → *coordinate descent* → Vrednost *J* će se monotono smanjivati do konvergencije

K-means konvergencija

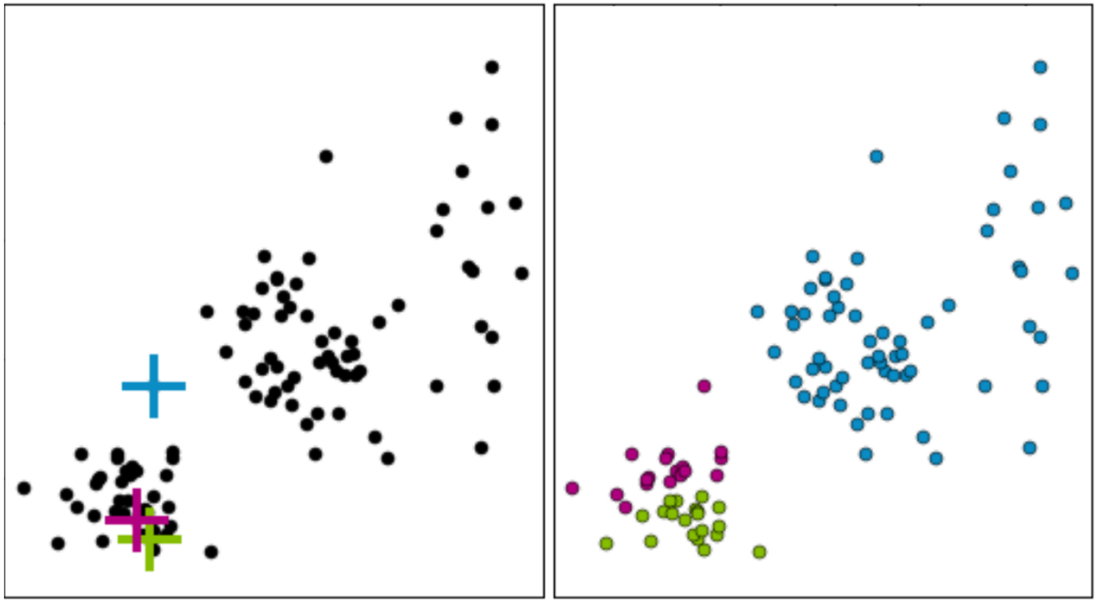
- Dakle, sigurno konvergira

- Ali, da li će završiti u lokalnom ili globalnom optimumu?

K-means konvergencija

- Ciljna funkcija *J* nije konveksna – ne možemo da garantujemo da ćemo pronaći globalni optimum

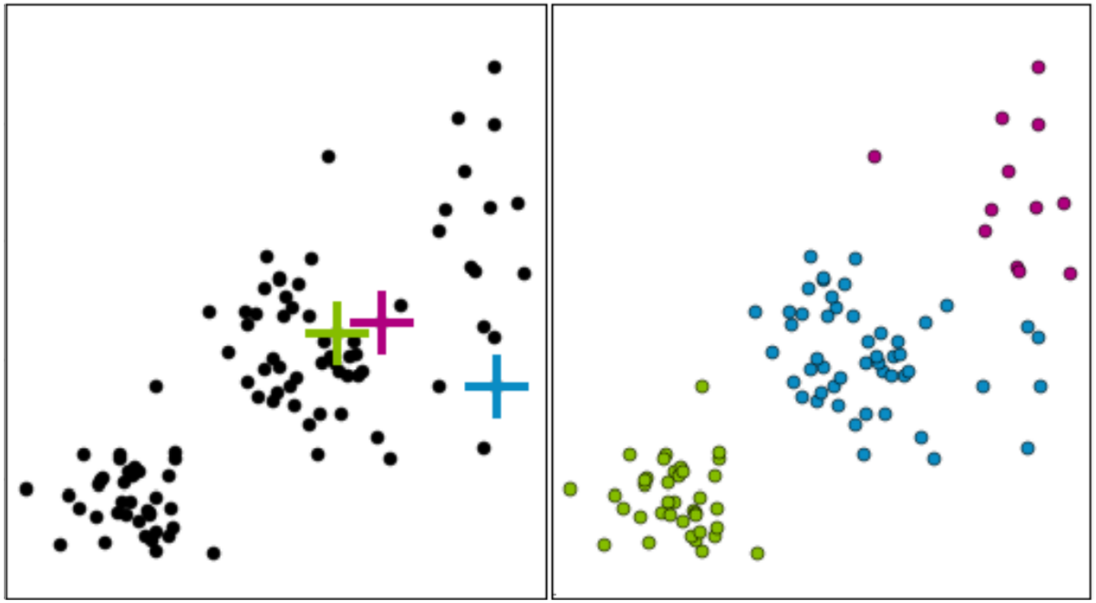
- Rezultati *K-means* veoma zavise od inicijalizacije centroida



K-means konvergencija

- Ciljna funkcija *J* nije konveksna – ne možemo da garantujemo da ćemo pronaći globalni optimum

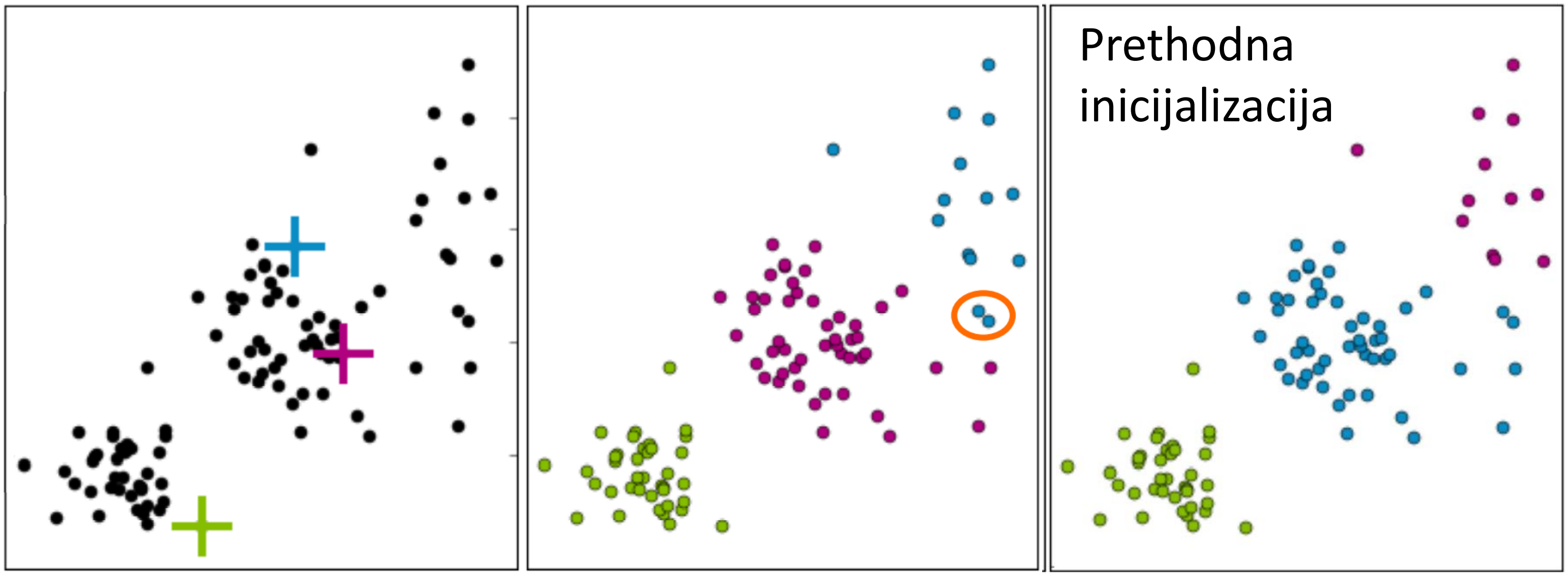
- Rezultati *K-means* veoma zavise od inicijalizacije centroida



K-means konvergencija

- Ciljna funkcija *J* nije konveksna – ne možemo da garantujemo da ćemo pronaći globalni optimum

- Rezultati *K-means* veoma zavise od inicijalizacije centroida



K-means konvergencija

- Dakle, *K-means* je veoma osetljiv na izbor inicijalnih centroida. Kako da rešimo ovaj problem?

1. Više slučajnih inicijalizacija (vratiti podelu sa najnižom vrednošću *J*)

2. Pametna inicijalizacija *K*-means++

Pametna inicijalizacija *K-means++*

1. Na slučajan način odabrati prvi centroid tako da bude jedna od opservacija skupa podataka
2. Za svaku od opservacija $x^{(i)}$ izračunati rastojanje $d\left(x^{(i)}\right)$ od najbližeg centroida
3. Odabrati novi centroid kao jednu od tačaka skupa podataka, s tim da je verovatnoća selekcije opservacije $x^{(i)}$ proporcionalna $d\left(x^{(i)}\right)^2$
4. Ponavljati korake 2 i 3 dok ne odaberemo *K* centroida

Pametna inicijalizacija *K*-means++

- Računarski zahtevno u poređenju sa slučajnom inicijalizacijom
- Ali, sa *K-means++* inicijalizacijom *K*-means često brže konvergira
- Obično popravlja kvalitet lokalnog optimuma i ubrzava vreme izvršavanja