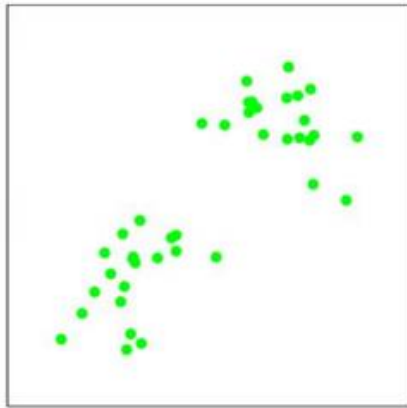


Klasterovanje

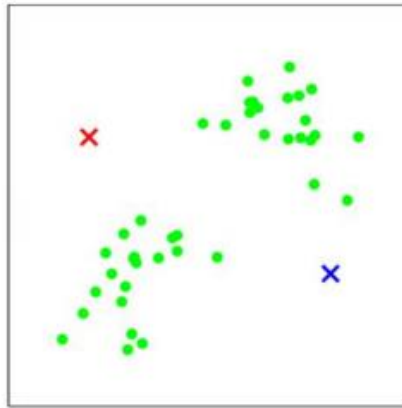
K-means

K-means

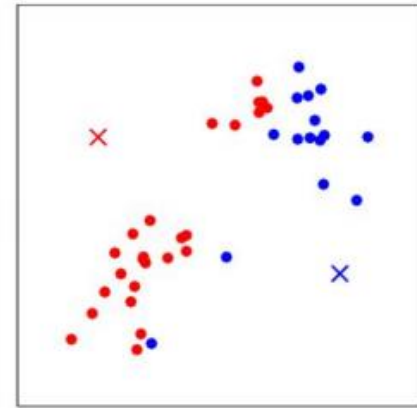
- Algoritam za klasterovanje koji uvodi pretpostavku:
 - *Score* = rastojanje od centra klastera (manje rastojanje je bolje)



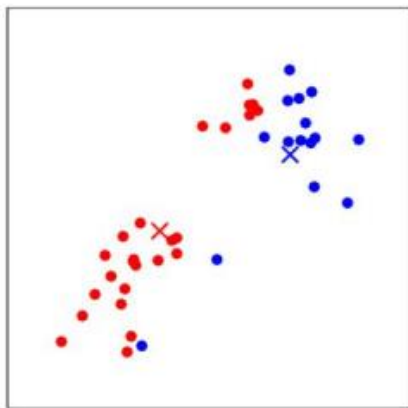
(a)



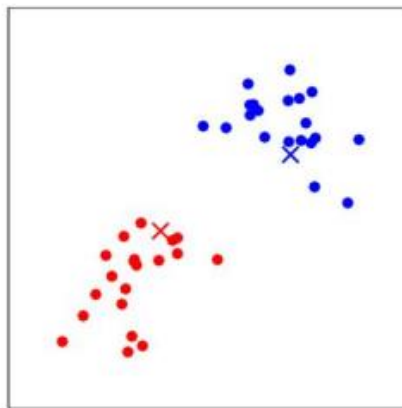
(b)



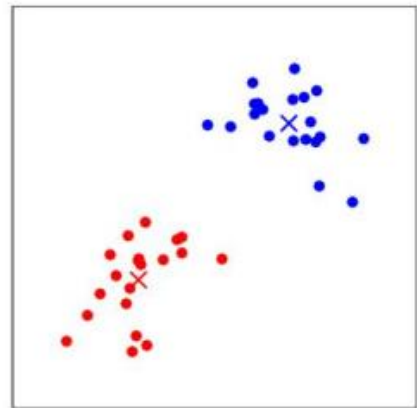
(c)



(d)



(e)



(f)

K-means

| | |
|----------|---|
| Ulaz | <ul style="list-style-type: none">$T = \{x^{(i)}, i = 1, \dots, N\}$ – trening skupK – broj klastera Kako da odredimo K? |
| Postupak | <ul style="list-style-type: none">Na slučajan način inicijalizovati centroide μ_1, \dots, μ_KDok ne dođe do konvergencije: Da li konvergira?<ul style="list-style-type: none">Za svaku opservaciju $x^{(i)}$ pronaći klaster sa najbližim centroidom (i dodeliti opservaciju tom klasteru): Šta znači „blizu“?$z^{(i)} \leftarrow \arg \min_j \ \mu_j - x^{(i)}\ _2^2$Ažurirati centroide klastera prema opservacijama koje su im dodeljene (<i>mean</i>)$\mu_j = \frac{1}{n_j} \sum_{i: z^{(i)}=j} x^{(i)}$$n_j$ - broj opservacija u klasteru j |
| Izlaz | <ul style="list-style-type: none">$\{(x^{(i)}, z^{(i)}), i = 1, \dots, N\}$ Da li smo dobili kvalitetno rešenje? |