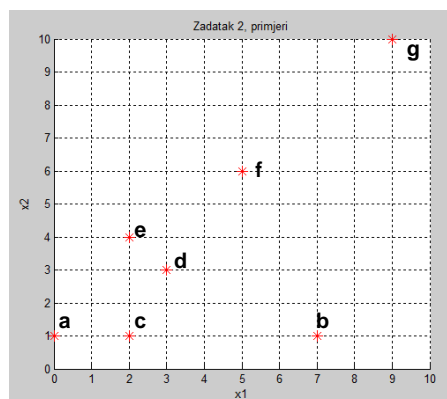


Zadatak 2: Hijerarhijsko aglomerativno grupiranje

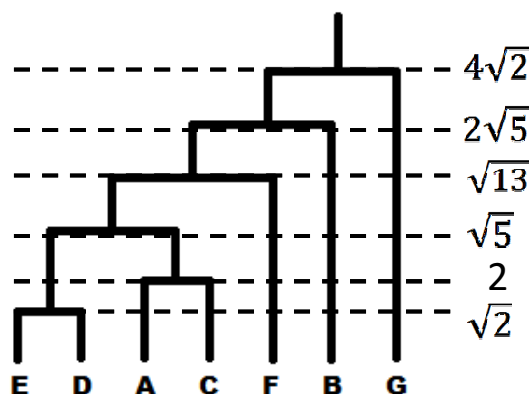
Zadan je skup primjera

$$\mathcal{D} = \{a = (0, 1), b = (7, 1), c = (2, 1), d = (3, 3), e = (2, 4), f = (5, 6), g = (9, 10)\}.$$

- (a) Skicirajte primjere. Nacrtajte dendrogram koji se dobiva hijerarhijskim aglomerativnim grupiranjem ovih primjera temeljem jednostruke povezanosti. Na dendrogramu naznačite udaljenost na kojoj je provedeno stapanje grupa.

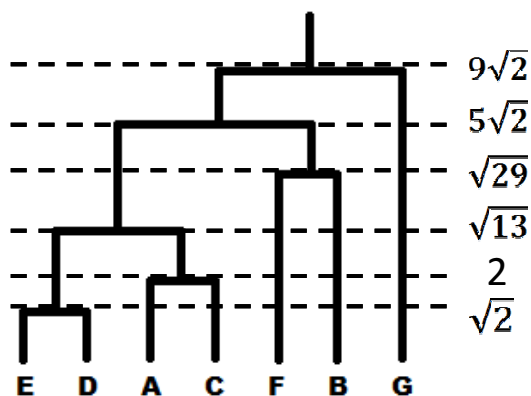


Slika 1. Primjeri



Slika 2. Dendrogram, jednostruke povezanosti

- (b) Nacrtajte dendrogram koji se dobiva grupiranjem temeljem potpune povezanosti.



Slika 3. Dendrogram, potpune povezanosti

- (c) Kako biste odredili optimalan broj grupa K temeljem dendrograma dobivenog u prethodnom zadatku? Obrazložite odgovor. Kojem pragu udaljenosti odgovara odabrani broj grupa? Presijecite dendogram na odabranom pragu udaljenosti i napišite particiju koja se time dobiva.

Pogledao bih gdje je dendrogram izdužen, i tu ga presjekao. Pošto su kod mene jednako nacrtane sve razine, gledao bih gdje je velika razlika u vrijednostima udaljenosti stapanja. U prethodnom primjeru, odokativno bih odabrao $K = 3$, na temelju toga što je velika razlika između $\sqrt{29}$ i $5\sqrt{2}$, to bi bio prag $\sqrt{29}$.

Particije: $C_1=[a,c,d,e]$, $C_2=[b,f]$, $C_3=[g]$

- (d) Napišite matricu sličnosti za primjere iz \mathcal{D} . Pretpostavite da raspolazete samo tom matricom. Koje algoritme grupiranja (od onih koje smo upoznali na predavanjima) možemo primijeniti na takvu matricu?

	a	b	c	d	e	f	g
a	1	0.125	0.333	0.217	0.217	0.124	0.073
b	0.125	1	0.167	0.183	0.146	0.157	0.098
c	0.333	0.167	1	0.309	0.250	0.146	0.081
d	0.217	0.183	0.309	1	0.414	0.217	0.098
e	0.217	0.146	0.250	0.414	1	0.217	0.098
f	0.124	0.157	0.146	0.217	0.217	1	0.150
g	0.073	0.098	0.081	0.098	0.098	0.150	1

Tablica 1. Matrica sličnosti

Ako imamo samo matricu sličnosti, koristili bi algoritam k-medoida.

Na predavanju su spomenute sljedeće implementacije:

- PAM (*partitioning around medoids*)
- CLARA (*clustering in large applications*)
- CLARANS (*CLARA based upon randomized search*)