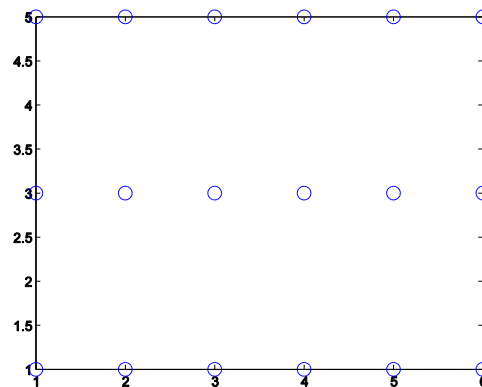
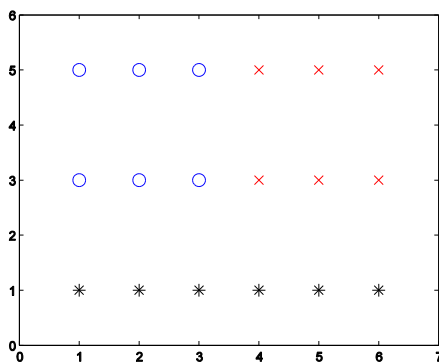


Vzorový test pro první část předmětu M33SAD

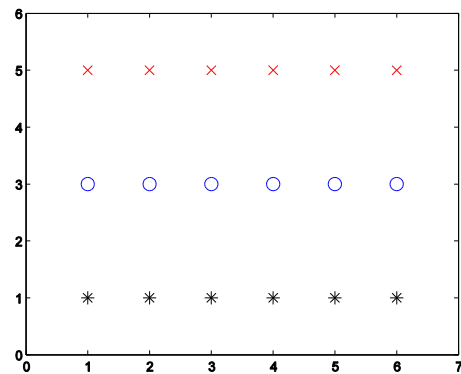
1. Na obrázku 1 jsou zobrazena vstupní data pro shlukování. Obrázky 2 a 3 odpovídají shlukování pomocí algoritmu k-means (s euklidovskou vzdáleností) a pomocí hierarchického shlukování (single linkage, euklidovská vzdálenost). Rozhodněte, který z obrázků 2, 3 odpovídá algoritmu k-means a který odpovídá hierarchickému shlukování.



Obrázek 1: Vstupní data

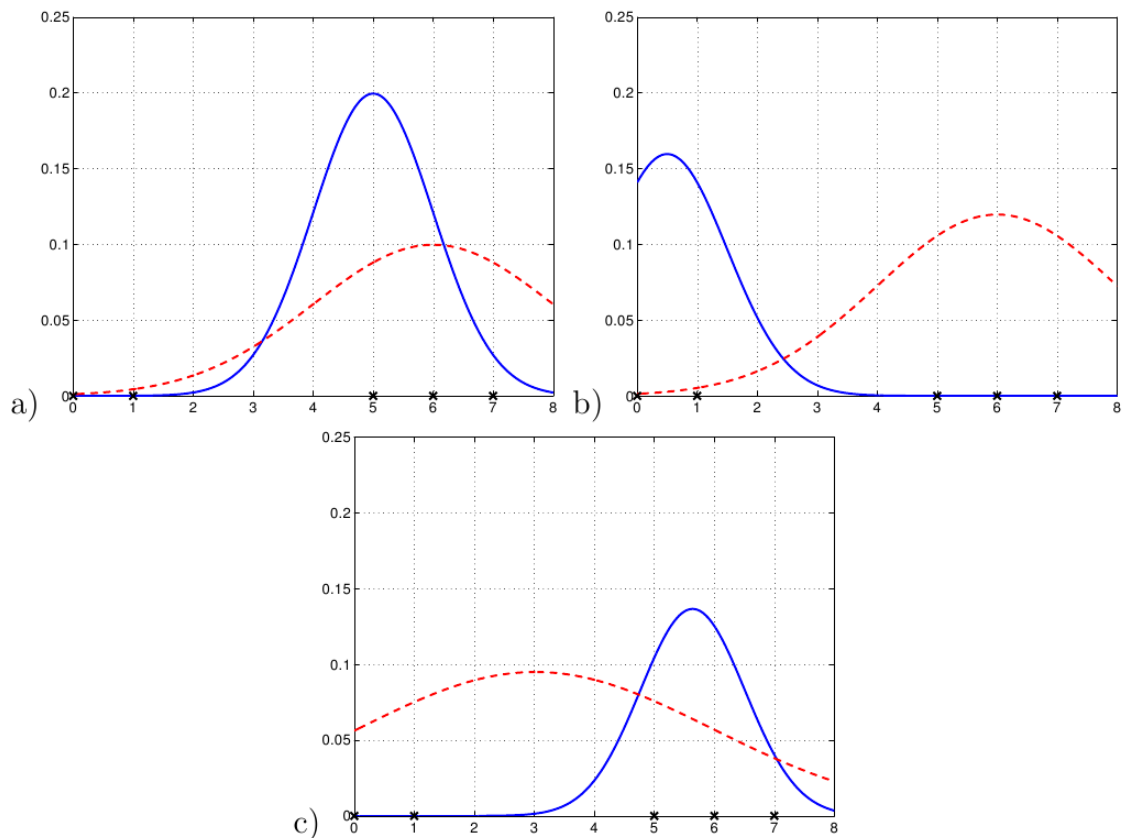


Obrázek 2



Obrázek 3

2. Pomocí EM algoritmu odhadujete parametry směsi 2 normálních rozdělání. Rozdělení směsi podle příznaku x lze zapsat takto: $f(x, \vartheta) = \alpha N(x; \mu_1, \sigma_1^2) + (1 - \alpha) N(x; \mu_2, \sigma_2^2)$. Obrázky uvedené níže ilustrují kroky EM algoritmu (na horizontální ose je parametr x , na vertikální ose je hustota pravděpodobnosti, pozorování jsou značena křížkem). Na jednom z obrázků je uveden náhodný inicializační krok (*init*), na druhém je uveden první optimalizační krok (*step1*). Třetí z obrázků je navíc. Obrázky jsou seřazeny náhodně. Rozhodněte, která dvojice obrázků odpovídá uvedeným krokům *init* a *step1*. Vysvětlete, proč uvedené pořadí dává smysl a jak *step1* vychází z *init*.



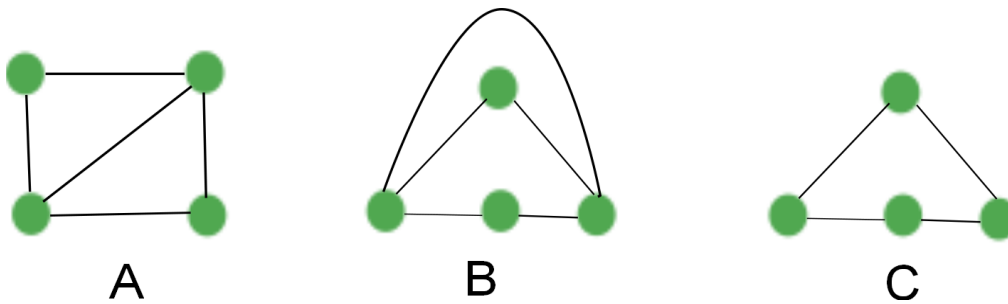
3. Mějme nějakou blíže nspecifikovanou databázi transakcí. Předpokládejme, že jsme zjistili, že jediné časté množiny položek velikosti 3 jsou: $\{a,b,c\}$, $\{a,b,d\}$, $\{b,c,d\}$, $\{a,c,d\}$, $\{b,c,e\}$. O kterých z následujících množin položek můžeme říct, že určitě nejsou časté: $\{a,b,c,d\}$, $\{a,b,c,e\}$, $\{b,c,d,e\}$.

4. Mějme databázi transakcí zobrazenou v tabulce 1. Najděte všechna asociační pravidla s podporou alespoň 50% a spolehlivostí více než 60%.

| Transakce | Položky |
|-----------|------------------------------|
| T1 | pivo, chléb |
| T2 | chléb, burákové máslo |
| T3 | pivo, mléko |
| T4 | chléb, džem, burákové máslo |
| T5 | chléb, mléko, burákové máslo |

Tabulka 1

5. Rozhodněte, které z následujících grafů jsou navzájem izomorfní.



6. Mějme abecedu dvou symbolů $\{a,b\}$. Uvažujme neorientované sekvence. Zodpovězte následující otázky:
- Kolik existuje různých neorientovaných sekvencí délky 3?
 - Naznačte, jak budete generovat různé sekvence délky 4. Ukažte alespoň jednu duplicitní sekvenci délky 4.
 - U sekvencí délky 3 jste ověřili, že časté jsou pouze sekvence $\{aab, bab, bbb\}$. Které sekvence délky 4 ještě stále mohou být časté? Proč?