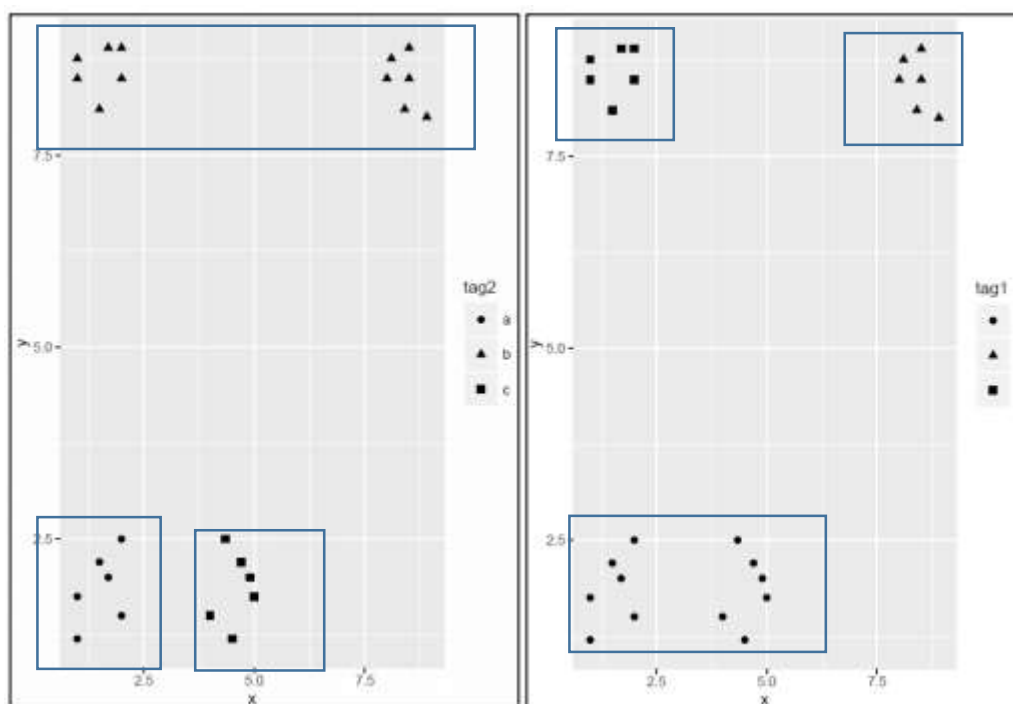


פתוחה 3: (4 נקודות, מוכפל ב1.25 כדי שיהיה 5 כמו כל השאר)



- א. סמנו על כל שרטוט את גבולות ה clusters-הצפויים להתקבל לאחר התכנסות האלגוריתם. (נקי לכל שרטוט)
- ב. האם יתכן שהאלגוריתם יתכנס לשני מצבים שונים? מדוע: (חצי נקי)
מאחר ואלגוריתם k means מתכנס לנקודת מינימום מקומי (ולא גלובלי), תנאי התחלה שונים עשויים להוביל להתכנסות למצב אחר.
- ג. עבור אילו ערכים של k לא יתכן כי האלגוריתם יתכנס לשני מצבים שונים?
 א. עבור $k=1$: יתקבל cluster אחד המכיל את כל הנקודות (חצי נקי)
 ב. עבור $k=24$ (כמספר הנקודות): כל cluster יכיל נקודה אחת בלבד (מאחר והclusters זרים זה לזה). (חצי נקי)
 אם אין פתרונות לא נכונים – חצי נקי ($k=2$) לא נחשב כפתרון לא נכון).

פתוחה 4: 5 נקודות.

- באם נבצע PCA על דוגמאות בגרף זה – לאיזה משני הממדים יהיה משקל (loading) גדול יותר ב component Principle הראשון? (1.5 נקי)
מאחר והשונויות של הנקודות גדולה יותר בציר ה y בהשוואה לציר x, למימד זה יהיה משקל גדול יותר ב PC הראשון.
- כיצד ישתנה ה loading-ז' באם לא נתחשב בשלוש הנקודות הקופות בעיגול? (1.5 נקי)
שלוש הנקודות המוקפות בעיגול מגדילות את השונויות בציר ה x (מאחר ונראה לעין כי הן רחוקות מהממוצע של הנקודות על ציר זה). לעומת זאת, על ציר ה y שלוש הנקודות לא משפיעות על השונויות מאחר והן יחסית קרובות לממוצע הנקודות על ציר זה. לכן, ללא שלוש הנקודות הללו השונויות של הנתונים על ציר x קטנה בעוד ששונויות הנתונים על ציר y לא משתנה הרבה. לכן, בעקבות הורדת הנקודות השונות היחסית של הנתונים על ציר y גדלה. על כן loading של ציר y ב PC הראשון יגדל.
- סעיף מבוטל. כולם מקבלים 2 נקודות