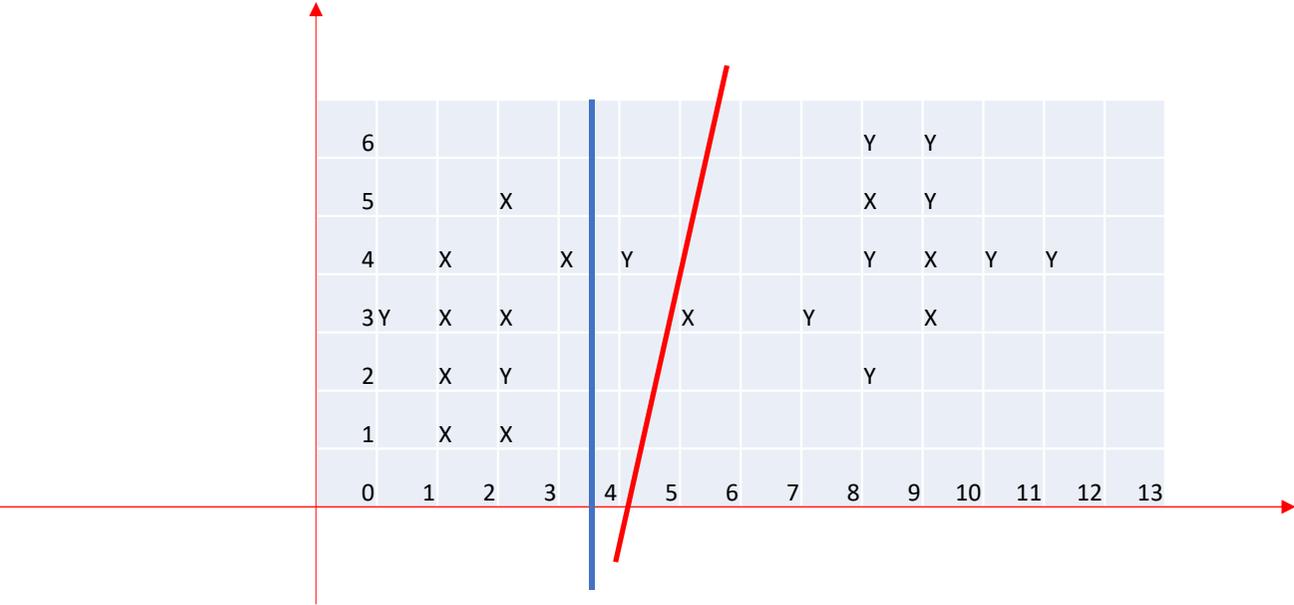


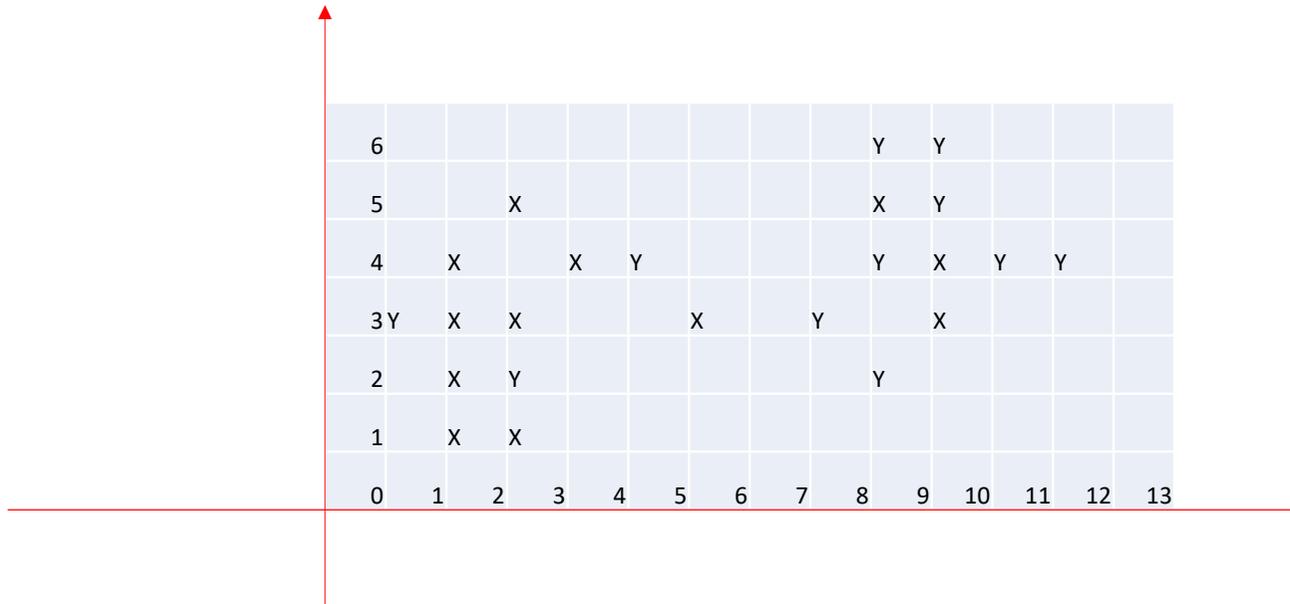
Exercise 020620

Exercise 1



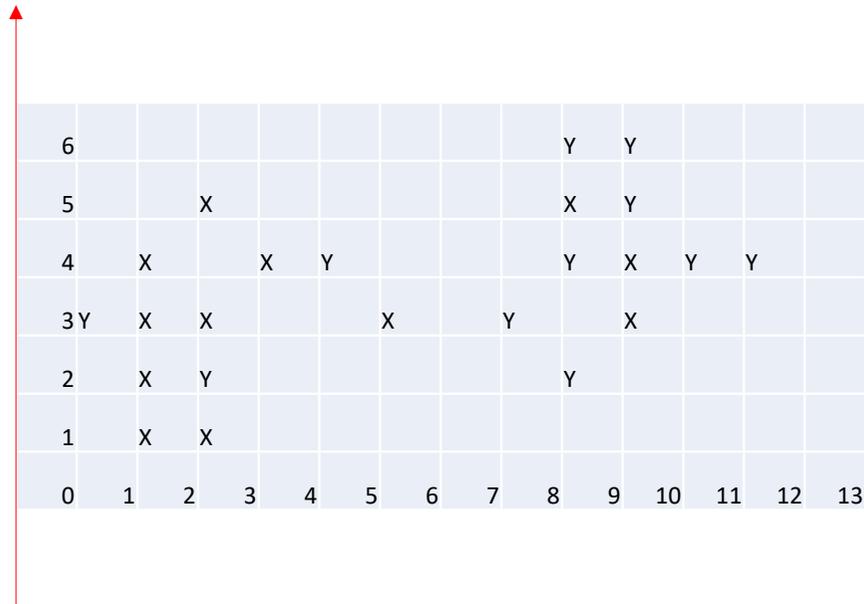
Klassifikation: Welche ist die Beste Splitting Linie? Blau oder rot? Beweis!  
Note: X, Y sind als Klasse betrachtet.

## Exercise 2



Verwenden Sie eine Data Mining Technik um diese Daten zu klassifizieren?

Exercise 2: X und Y sind unabhängig von Exercise 1



6									Y	Y				
5		X							X	Y				
4	X		X	Y					Y	X	Y	Y		
3	Y	X	X			X		Y		X				
2	X	Y							Y					
1	X	X												
0	1	2	3	4	5	6	7	8	9	10	11	12	13	

Verwenden Sie Assoziation Rules Technik um diese Daten zu klassifizieren?  
Beweis!

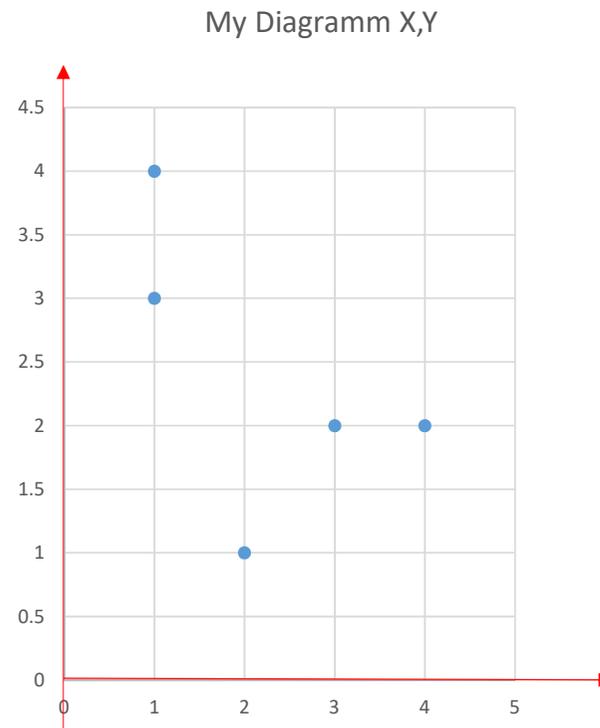
### Bemerkung zu Clusteranalyseverfahren:

1. Clusteranalyse ist **deskriptiv**. Es werden **keine Tests durchgeführt**.
2. Die Effektivität der verschiedenen Clustertechniken hängt von der Lage und Form der 'wahren' Cluster ab, die nicht bekannt sind.
3. Nützlich: **Resultate verschiedener Verfahren vergleichen**. **Führen verschiedene Ansätze zu ähnlichen Ergebnissen**, die auch inhaltlich Sinnmachen, so kann man auf die gefundenen Cluster vertrauen.
4. Clusteranalysemethoden können Ausreisser anfällig sein, insbesondere, wenn sie auf der **euklidischen Distanz basieren**.
5. Ohne die Existenz deutlicher Trennungen keine guten Ergebnisse erzielbar. Nichtlineare Strukturen werden von wenigen Clusteranalyse-verfahren getrennt, z.B. kreisförmige (Gruppe 1) innerhalb U-förmiger Punktwolke (Gruppe 2)
6. Variablen clustern: Alternative zur Faktorenanalyse, extrahiert allerdings keine verborgenen Faktoren, sondern findet nur ähnliche Variablengruppen

### Exercise 3

Verwenden Sie eine Descriptive Data Mining Technik-Clustering- um diese Data zu klassifizieren!  
Stellen Sie diese Data auf 2D Kartesische Diagramm dar!(bereits gemacht).  
Berechnungsmatrix, Darstellung-Dendogramm sind wichtig!

	X	Y	Z	X1	Y1	Z1	usw
A		1	3				
B		3	2				
C		2	1				
D		4	2				
E		1	4				



## Exercise 4

Verwenden Sie eine Data Mining Technik-Clustering- um diese Data zu klassifizieren!

Stellen Sie diese Data auf 2D Kartesische Diagramm dar!(bereits gemacht).

Berechnungsmatrix, Darstellung-Dendogramm sind wichtig!

	X	Y	Z	X1	Y1	Z1	usw
A		1	3	1			
B		3	2	2			
C		2	1	3			
D		4	2	2			
E		1	4	1			

## Exercise 5

### DNA Sequence

CTG GAT ATT ACG GCC TTT TTA 480 AAG ACC GTA AAG AAA AAT  
AAG CAC AAG TTT TAT CCG GCC TTT ATT CAC ATT CTT GCC CGC  
540 CTG ATG AAT GCT CAT CCG GAA TTT CGT ATG GCA ATG AAA  
GAC GGT GAG CTG GTG ATA

ACG

GCC

TTT

TTA

AAG

ACC

GTA

AAG

AAA

AAT

AAG

AAA

Berechnen Sie die Support und  
Confidence von diese Rule(zwei AA  
verfolget mit G:

R1: AA  G

Berechnen Sie die Support und  
Confidence von diese Rule(zwei AA  
verfolget mit G:

R2: A  A

## Ref. Clustering

[https://www-m9.ma.tum.de/material/felix-klein/clustering/Methoden/Hierarchisches\\_Clustern\\_Beispiel.php](https://www-m9.ma.tum.de/material/felix-klein/clustering/Methoden/Hierarchisches_Clustern_Beispiel.php)

[https://www.uni-kassel.de/fb07/fileadmin/datas/fb07/5-Institute/IVWL/Kosfeld/lehre/multivariate/Multivariate12\\_Clusteranalyse2\\_.pdf](https://www.uni-kassel.de/fb07/fileadmin/datas/fb07/5-Institute/IVWL/Kosfeld/lehre/multivariate/Multivariate12_Clusteranalyse2_.pdf)

<https://swl.htwsaar.de/lehre/ws17/ds/slides/2017-vl-ds-kap4-2-clustering.pdf>

<https://www.statistik.tu-dortmund.de/~dvoel/Multivariate/Skript/08-Clustersanalyse.pdf>