

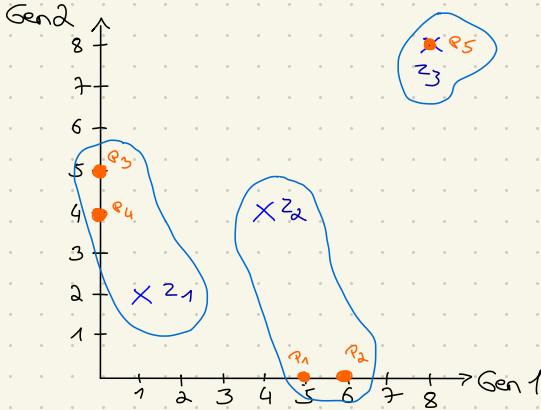
Übung 8

8.1) k-means Clustering

Initiale Zentren:

$(1,2), (4,4), (8,8) \rightarrow k=3$

	P_1	P_2	P_3	P_4	P_5
G_1	5	6	0	0	8
G_2	0	0	5	4	8



Euklidische Distanz:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

$$d_{2,1,P_1} = \sqrt{4^2 + 2^2} = 4,47$$

$$d_{2,1,P_2} = 5,39$$

$$d_{2,1,P_3} = 3,16$$

$$d_{2,1,P_4} = 2,24$$

$$d_{2,1,P_5} = 9,22$$

$$d_{2,2,P_1} = 4,12$$

$$d_{2,2,P_2} = 4,47$$

$$d_{2,2,P_3} = 4,12$$

$$d_{2,2,P_4} = 4$$

$$d_{2,2,P_5} = 5,66$$

$$d_{2,3,P_1} = 8,54$$

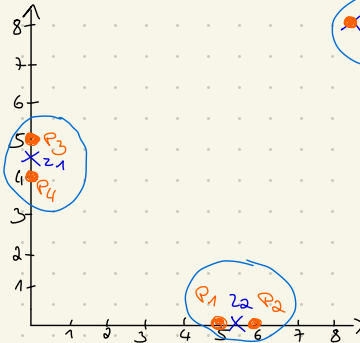
$$d_{2,3,P_2} = 8,25$$

$$d_{2,3,P_3} = 8,54$$

$$d_{2,3,P_4} = 8,94$$

$$d_{2,3,P_5} = 0$$

Anpassung \rightarrow



Within-cluster-sum-of-square:

$$\text{Cluster 1: } 0,5^2 + 0,5^2 = 0,5$$

$$\text{Cluster 2: } 0,5^2 + 0,5^2 = 0,5$$

$$\text{Cluster 3: } 0$$

8.2) charakteristische biologische Prozesse:

- Regulation der Zelldifferenzierung im Innenohr
- Regulation der prä-3-Zellteilung
- Festlegung anteriorer / posteriorer Regionen
- Entwicklung vom proximalen Tubulus
- zelluläre Antwort auf Tumorzellen

False discovery rate:

- wenn viele Hypothesen getestet werden
- erwartetes Verhältnis aus fälschlicherweise zurückgewiesenen Nullhypothesen und zurückgewiesenen Nullhypothesen insgesamt
- Fehler 1. Art nimmt bei Testung multipler Hypothesen zu
 - ↳ Nullhypothese wird abgelehnt, obwohl sie eigentlich richtig ist

↳ α muss strenger (niedriger) sein als bei einzelnen Hypothesentest

8.3)	Nicht Apoptose	Apoptose	Summe	
nicht differenziell	390 $400 \cdot 450 / 500 = 360$	60 $100 \cdot 450 / 500 = 90$	450	erwartete Werte
differenziell	10 $400 \cdot 50 / 500 = 40$	40 $100 \cdot 50 / 500 = 10$	50	
Summe	400	100	500	

H_0 : keine Assoziation, Zufall

H_1 : statistisch signifikant, Abhängigkeit

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

$$E_i = \frac{(r \cdot c)}{N}$$

r : Zeilensumme

c : Spaltensumme

$$df = (r-1) \cdot (c-1)$$

$$df = 1$$

$$\chi^2 = \frac{(390-360)^2}{360} + \frac{(60-90)^2}{90} + \frac{(10-40)^2}{40} + \frac{(40-10)^2}{10} = 125$$

Tabelle: $\chi^2(1; 0,05) = 3,8415 < 125 \rightarrow H_0$ wird abgelehnt
 → Abhängigkeit, in den Genen sind signifikant solche mit Funktion Apoptose vorhanden