

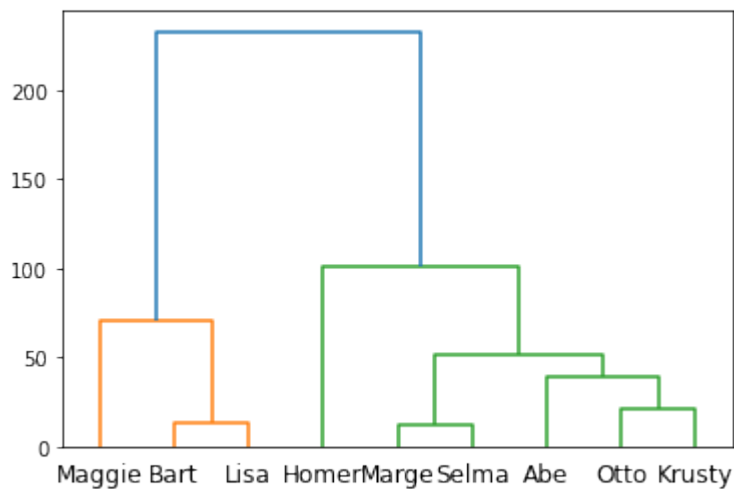
Clusteranalyse Oplossingen

1 De Simpsons revisited

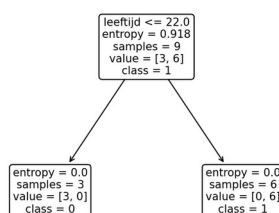
- 1.
2. wis eerst de kolommen "naam" en "geslacht" want die zijn nominaal!

	Homer	Marge	Bart	Lisa	Maggie	Abe	Selma	Otto	Krusty
Homer	0	100,5	162,1	174,4	232,7	86,9	90,5	70,7	51,2
Marge		0	65,1	76,7	134,3	42,2	12,4	30,3	51,4
Bart			0	12,8	70,6	100,0	76,8	94,6	115,5
Lisa				0	58,5	111,1	88,4	106,4	127,5
Maggie					0	165,1	145,7	164,3	185,3
Abe						0	31,5	34,7	39,4
Selma							0	20,3	40,3
Otto								0	21,6
Krusty									0

3. Selma en Marge (12,4)
4. 188
5. 2,887089
- 6.

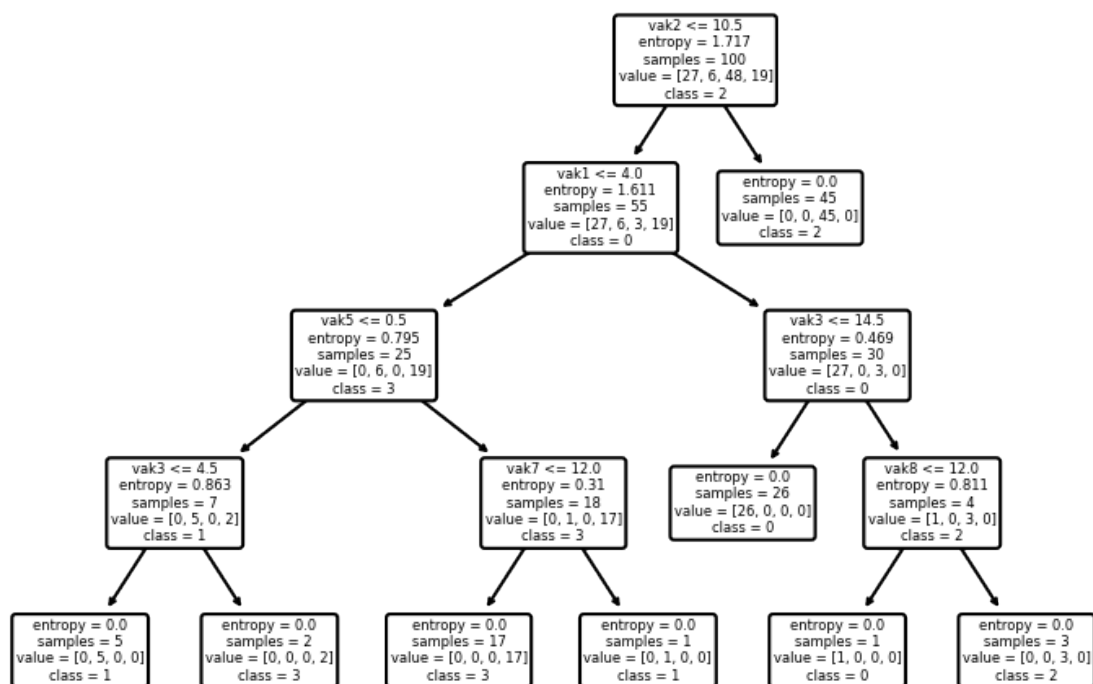


7. de kinderen en de volwassenen (de lichte, jonge simpsons en de oude, zwaardere simpsons)
8. deze clusters zijn op basis van de leeftijd en gewicht
9. nu zit Homer in een aparte cluster. Hij heeft het grootste gewicht. Dat maakt waarschijnlijk dat hij apart wordt gezet.
10. centroids (kan afwijken): (4, 62,667, 6,333) en (5.833, 185, 44)
de centroids verschillen het meest in gewicht en leeftijd
- 11.



2 Studenten

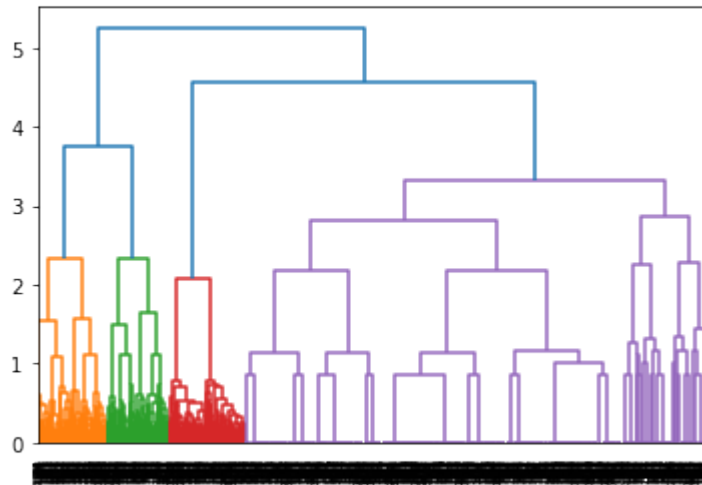
- 1.
2. afstanden:
 - euclidisch: 8,3666
 - manhattan: 20
 - gestandaardiseerd euclidisch: 1.513
 - gestandaardiseerd manhattan: 3.694
3. afstanden:
 - student1: 8.124038, 24.41311, 19.339080, 21.424285
 - student10: 13.92839, 19.39072, 17.14643, 16.82260
 - student20: 19.57039, 26.32489, 12.60952, 27.38613
 - student30: 24.33105, 24.28992, 14.96663, 22.51666
4. zie onderliggende stukken in vorig antwoord
5. er zijn 25 studenten in iedere cluster
6. als je met de gegeven centroids start bevatten de clusters 27, 6, 48 en 19 studenten
- 7.



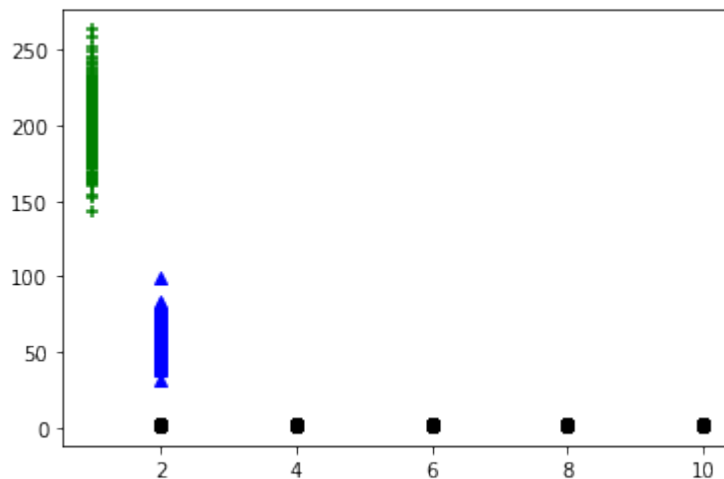
8. cluster nummer 2 (op basis van score van vak2)

3 Extraterrestrial life...

- 1.
- 2.

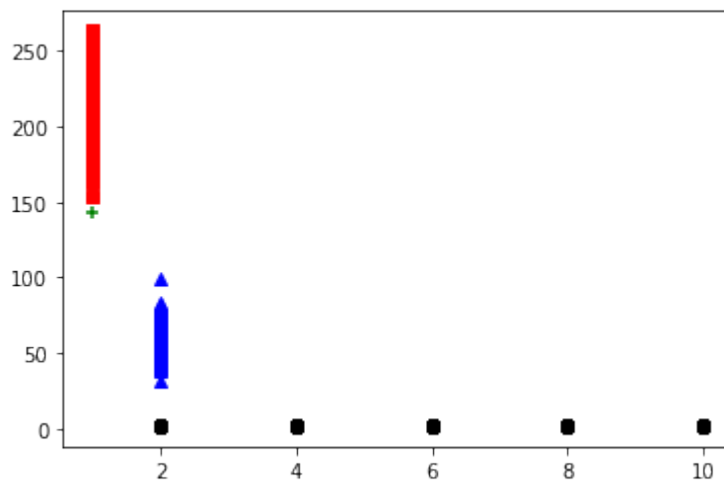


3. 1230, 350 en 200
- 4.

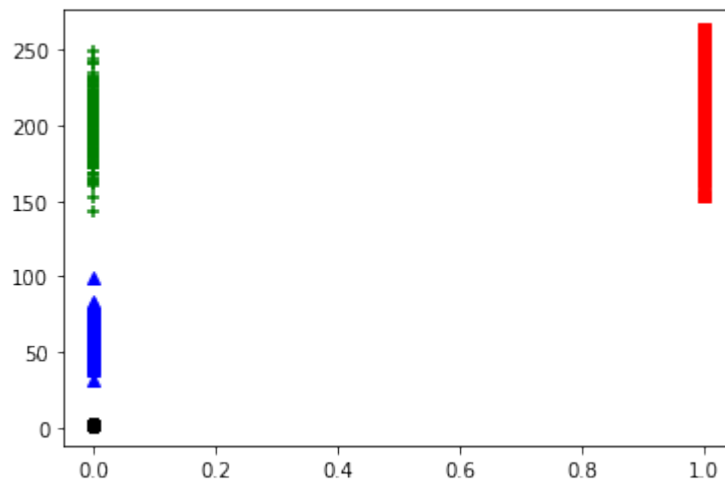


De categorie lijkt bepaald te zijn aan de hand van de lengte (het zou natuurlijk kunnen dat er overlappen zijn die je niet kan zien in deze grafiek omdat je maar 2 variabelen kan zien)

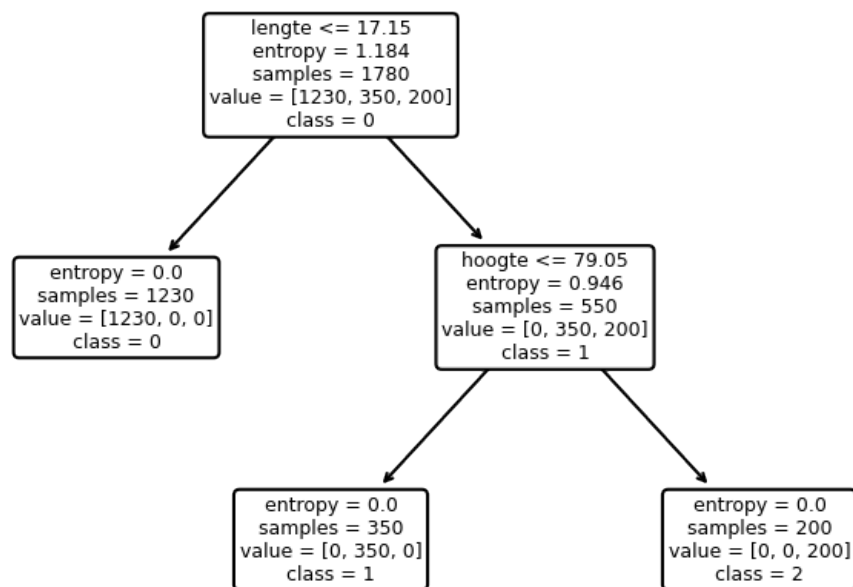
- 5.



De cluster van de grootste aliens werd opgesplitst. Er is maar 1 alien te zien die is afgesplitst. Dit zou echter kunnen komen door de gemaakte projectie. Als je "staart" en "lengte" in de grafiek zet, dan zie je dat "staart" gebruikt wordt voor de 4e cluster:



6. Als je k-means met een andere seed doet, krijg je soms licht verschillende indelingen.
 Met seed gelijk aan 42 krijg je volgende categorieën:
1. 6-7 poten, 3-4 ogen, 2 lengte, 0.7 breedte, 0.7 hoogte, geen staart, soms vleugels
 2. 10 poten, 1 oog, 201 lengte, 101 breedte, 30 hoogte, soms staart, geen vleugels
 3. 4 poten, 2 ogen, 60 lengte, 70 breedte, 178 hoogte, geen staart, geen vleugels
- 7.



lengte is het belangrijkste.