

Domača naloga 4

Neža Kržan, Tom Rupnik

1 Opis podatkov

Za analizo sva si izbrala podatkovni okvir z naslovom *Wine*, ki vsebuje 178 vrstic. Enoto oz. vrstico podatkov predstavlja posamezno vino. Podatki vsebujejo vrednosti kemijske analize treh različnih sort vina iz enake regije v Italiji. Analiza vsebuje vrednosti 13 različnih komponent (lastnosti) najdenih v vsaki izmed treh sort vina. Iz teh komponent želimo določiti kateri sorti pripada vino (stolpec *Class*). Komponente oz. stolpci v podatkovni okvirju so:

Tabela 1: Spremenljivke v podatkovnem okvirju.

Ime	Pomen
Class	razred (uporabimo za analizo)
Alcohol	% alkohola
Malic_acid	jabolčna kislina
Ash	vsebnost pepela
Alcalinity_of_ash	alkalnost pepela
Magnesium	Magnezij
Total_phenols	skupaj fenolov
Flavanoids	
Nonflavanoid_phenols	
Proanthocyanins	
Color_intensity	intenziteta barve
Hue	
OD280_OD315	
Proline	

2 Skupine in povezanost

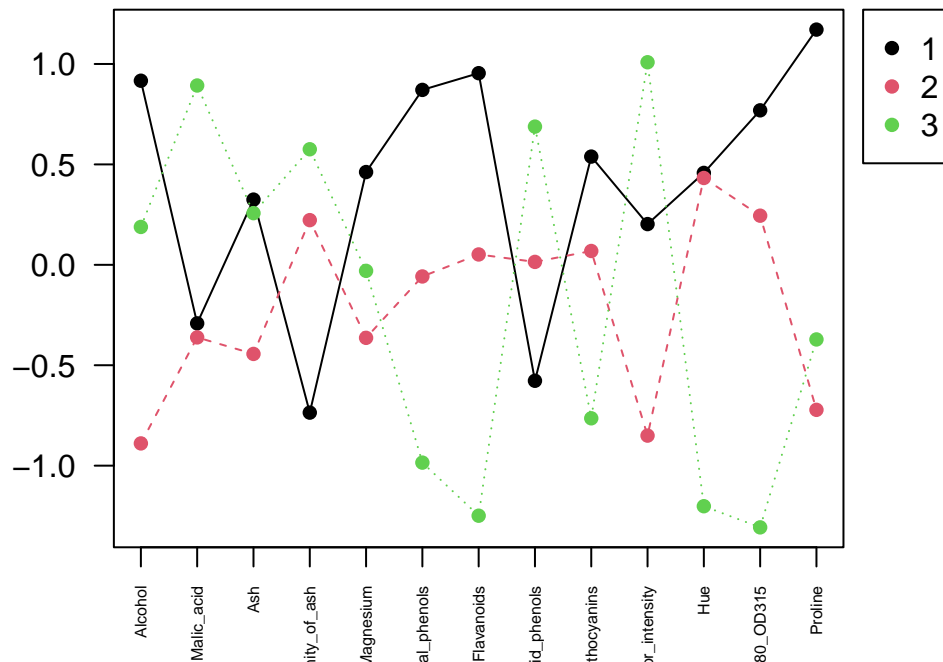
Iz tabele 3 vidimo, da je največ enot prisotnih v skupini **2** (71) in najmanj v skupini **3** (48). Preostanek enot je v skupini **1** in sicer 59.

Tabela 2: Opisna statistika spremenljivk.

Ime	min	max	povprecje
Alcohol	11.03	14.83	13.00
Malic_acid	0.74	5.80	2.34
Ash	1.36	3.23	2.37
Alcalinity_of_ash	10.60	30.00	19.49
Magnesium	70.00	162.00	99.74
Total_phenols	0.98	3.88	2.30
Flavanoids	0.34	5.08	2.03
Nonflavanoid_phenols	0.13	0.66	0.36
Proanthocyanins	0.41	3.58	1.59
Color_intensity	1.28	13.00	5.06
Hue	0.48	1.71	0.96
OD280_OD315	1.27	4.00	2.61
Proline	278.00	1680.00	746.89

Tabela 3: Velikost skupin.

skupina	koliko	delež
1	59	33.15
2	71	39.89
3	48	26.97



Slika 1: Povprecja neodvisnih spremenljivk po skupinah (standardizirane vrednosti)

Iz grafičnega prikaza lahko vidimo, da večina spremenljivk dobro loči med skupinami, je pa nekaj takih kjer to ne velja. Primer take spremenljivke je na primer *Hue* ali *Ash*.

3 Predpostavke