

## 2.1

Zbiór danych *wine.data* zawiera dane dotyczące wyników chemicznej analizy win pochodzących z tego samego regionu Włoch, ale od trzech różnych plantatorów (Barolo, Grignolino, Barbera). W zbiorze mamy m.in zmienne:

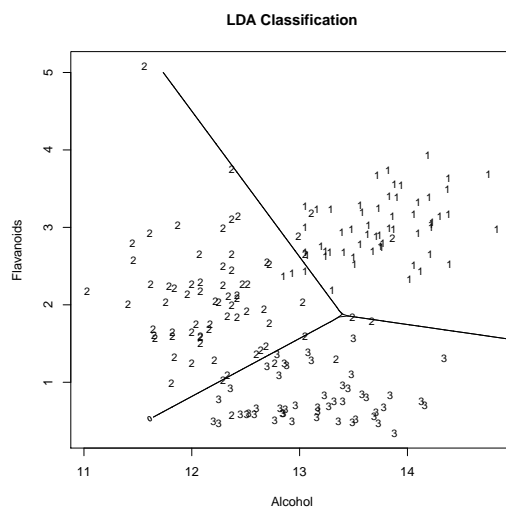
- **V1**- zmienna grupująca, przyjmuje wartości 1, 2, 3 (numer plantatora),
- **V2**- zawartość alkoholu,
- **V8**- zawartość flawanoidów,
- **V14**- protolina (aminokwas), i inne (dokładny opis na <http://archive.ics.uci.edu/ml/datasets/Wine>).

a) Skorzystać z funkcji `lda(MASS)` i utworzyć regułę dyskryminacyjną przy wykorzystaniu dwóch atrybutów: **V2** i **V8**. Dla każdego obiektu wypisać prawdopodobieństwo zaklasyfikowania do jednej z trzech klas.

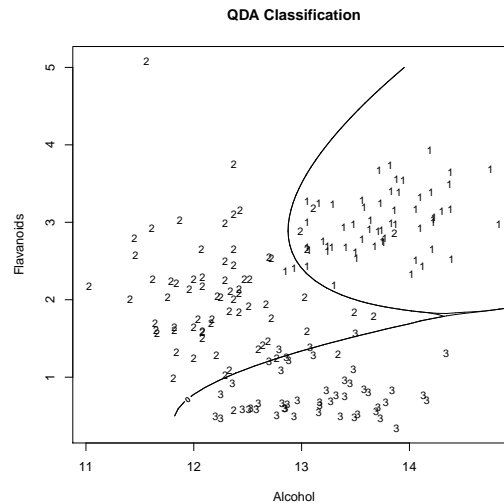
b) Zaklasyfikować wszystkie obiekty stosując wyznaczoną regułę, sporządzić tabelę klasyfikacji (zastosuj podział na próbę testową i treningową). Obliczyć procent poprawnych klasyfikacji.

c) Zaprezentować graficznie sposób w jaki dokonujemy klasyfikacji obiektów za pomocą funkcji `lda` (patrz Rysunek 1), wykorzystaj funkcje `contour` i `expand.grid`.

d) Wykonać podpunkty (a), (b), (c) używając kwadratowej funkcji dyskryminacyjnej (funkcja `qda(MASS)`). Porównać wyniki z otrzymanymi dla liniowej funkcji dyskryminacyjnej.



Rysunek 1: Klasyfikacja za pomocą liniowej funkcji dyskryminacyjnej.



Rysunek 2: Klasyfikacja za pomocą kwadratowej funkcji dyskryminacyjnej.

## 2.2

Zbiór danych skoringowych *kredit.asc* zawiera informacje o klientach banku. Zmienna **kredit** zawiera informację czy klient spłacił kredyt (wartość 1) czy nie spłacił kredytu (wartość 0). Każdy z klientów jest opisany w terminach punktów skoringowych. Dokładny opis danych znajduje się na stronie <http://www.ipipan.eu/~teisseyrep/DM/DANE/kredit.htm>. W analizie będziemy rozważać jedynie pierwsze 9 atrybutów.

- a) Zbudować klasyfikator na podstawie wszystkich atrybutów używając funkcji `lda(MASS)` i `qda(MASS)`. Wypisać tabelę klasyfikacji i podać procent poprawnych klasyfikacji (zastosuj podział na próbę testową i treningową).
- b) Zbudować klasyfikator na podstawie pierwszych 9 zmiennych używając funkcji `lda(MASS)` i `qda(MASS)`. Wypisać tabelę klasyfikacji i podać procent poprawnych klasyfikacji (zastosuj podział na próbę testową i treningową).
- b) Używając procedury `stepclass(klar)` (z opcją `dierction="backward"`) dokonać eliminacji zmiennych. Dla otrzymanego modelu wypisać tabelę klasyfikacji i podać procent poprawnych klasyfikacji.
- d) W jakim celu dokonuje się selekcji zmiennych?

## 2.3

Dane *mtcars* w pakiecie MASS zawierają informacje dotyczące parametrów wybranych marek samochodów. Niech zmienna **cyl** (liczba cylindrów) będzie zmienną grupującą. Przedstawić graficznie obszary klasyfikacji na podstawie par atrybutów (spośród zmiennych **mpg**, **disp**, **hp**) dla metod `lda` oraz `qda`. Skorzystać z funkcji `partimat(klar)`.