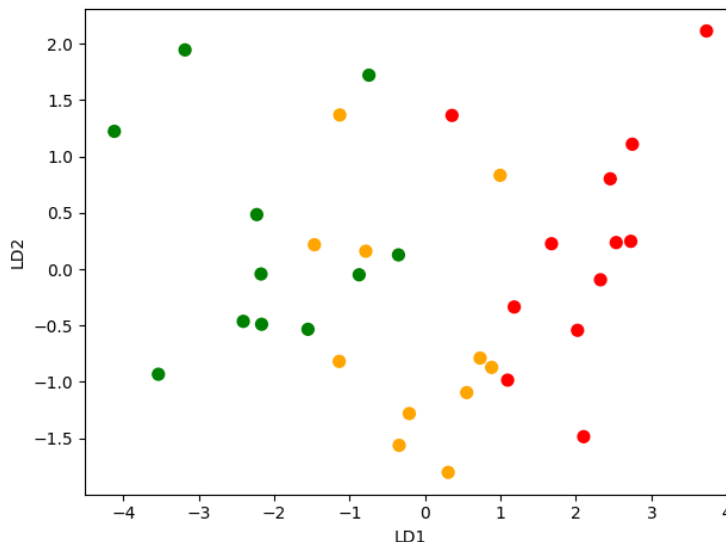


Discriminantanalyse – Oplossingen

Exercise 1:

-
- 2 dimensies: de afhankelijke variable heeft 3 mogelijke waarden en er zijn 4 onafhankelijke variabelen. Dus $N = \min(3-1, 4) = 2$
- Hint: gebruik `plt.scatter`



Een 2D plot waar elke instantie is geplot met als coördinaten de waarde bekomen op de discriminantfuncties LD1 en LD2 en waar verschillende kleuren werden gebruikt om de quality te visualizeren (bad=rood, medium=orangje and good=groen).

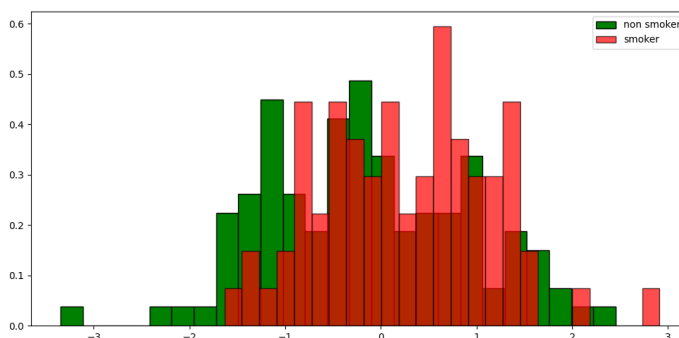
-
-
-
- 0.7941 (= accuracy)

Exercise 2:

- Hint: pas de pandas functie `describe` toe op je dataframe.
- gelijk welke categorische/kwalitatieve variabele kan gebruikt worden als afhankelijke variabele: *low*, *smoke*, *race*, ...

In de oplossing van deze oefening hebben we voor de variabele *smoke* gekozen

- in principe, mogen alleen continue, normal verdeelde lineair, niet-gecorrleerde variabelen gebruikt worden. Alleen de variabelen *age*, *lwt*, *bwt* lijken geschikt
-
- 1 dimensie: de afhankelijke variabele heeft 2 mogelijke waarden en er zijn 3 onafhankelijke variabelen. Dus $N = \min(2-1, 3) = 1$
-



Exercise 3:

- a.
- b.
- c. gelijk welke categorische/kwalitatieve variabele: *Manufacturer, Model, Type, Airbags, DriveTrain, Man.trans.avail, origin, make* mar ook *Cylinders*.
Cylinders is in principe een kwantitatieve variabele, but er is 1 car met de omschrijving 'rotary' in de plaats van een getal
- d. alle kwantitatieve variabelen: *Min.Price, Price, Max.Price, MPG.city, MPG.highway, EngineSize, Horsepower, RPM, Rev.per.mile, Fuel.tank.capacity, Passengers, Length, Wheelbase, Width, Turn.circle, Rear.seat.room, Luggage.room, Weight*
- e.
- f. 5 dimensies: $N = \text{Min}(6-1, 18)$
- g. Neen, enkel 2 op de 3 (accuracy = 0.6667)