ANALIZA DISCRIMINANTĂ LINIARĂ – ALGORITM DE CLASIFICARE –

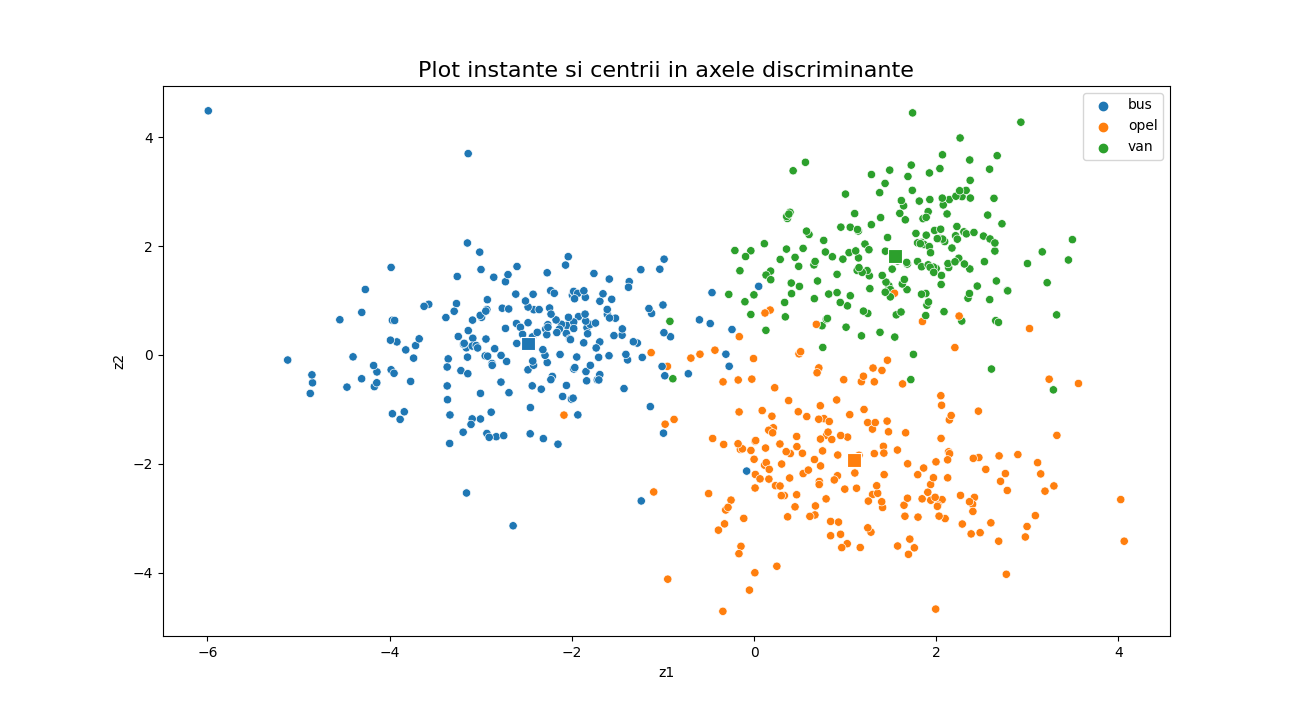
Analiza discriminantă liniară este un model de clasificare dar şi un model de reducere a dimensionalităţii.

Obiectiv: Clasificarea unei structuri date ca aparţinând uneia dintre cele trei tipologii de autovehicule (bus, opel, van), folosind un set de caracteristici specifice structurii. Autovehiculul poate fi văzut din diferite unghiuri.

Setul de date conţine 629 de observaţii şi 18 caracteristici corespunzătoare celor trei tipologii de autovehicule (setul original conţinea 846 de observaţii şi 4 clase).

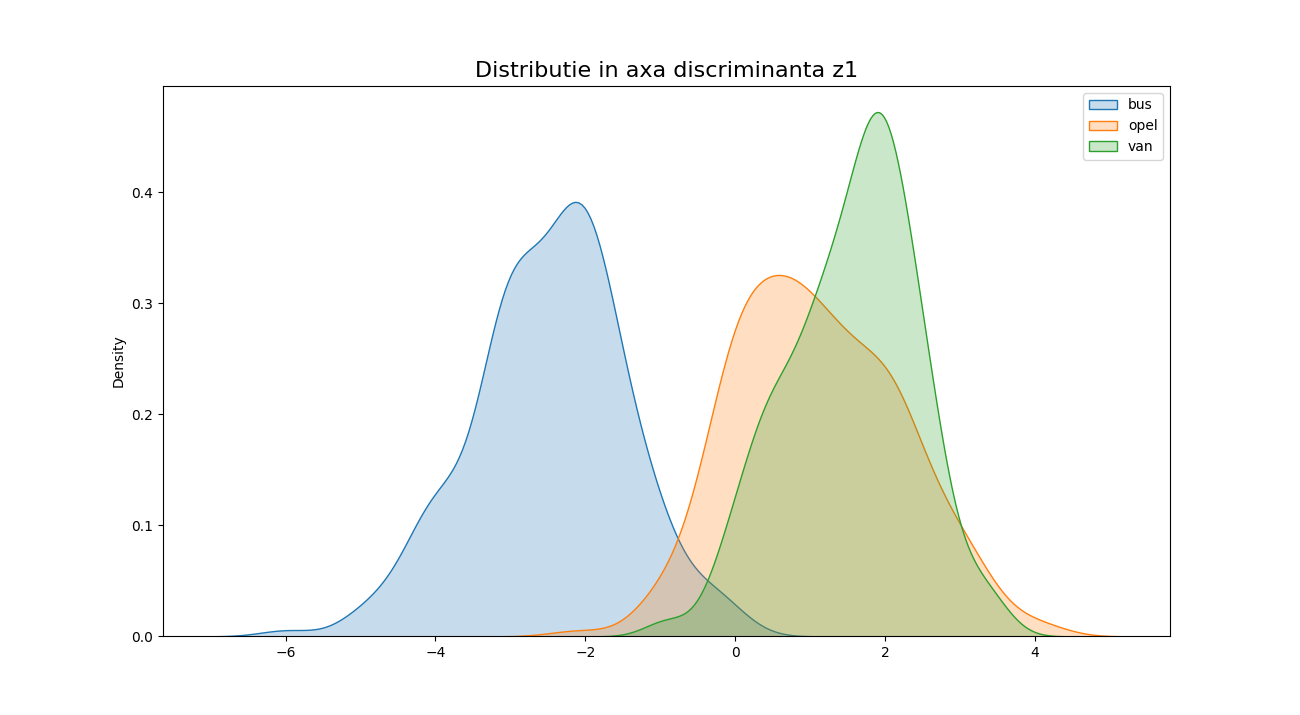
Am calculat scorurile dicriminante (valorile luate de instanţe în axele discriminante). Numărul axelor discriminante este egal cu minimul dintre numărul de predictori (18) şi numărul de clase (3) minus 1.

Pentru a obţine o imagine asupra clasificării au fost realizate urmatoarele grafice :

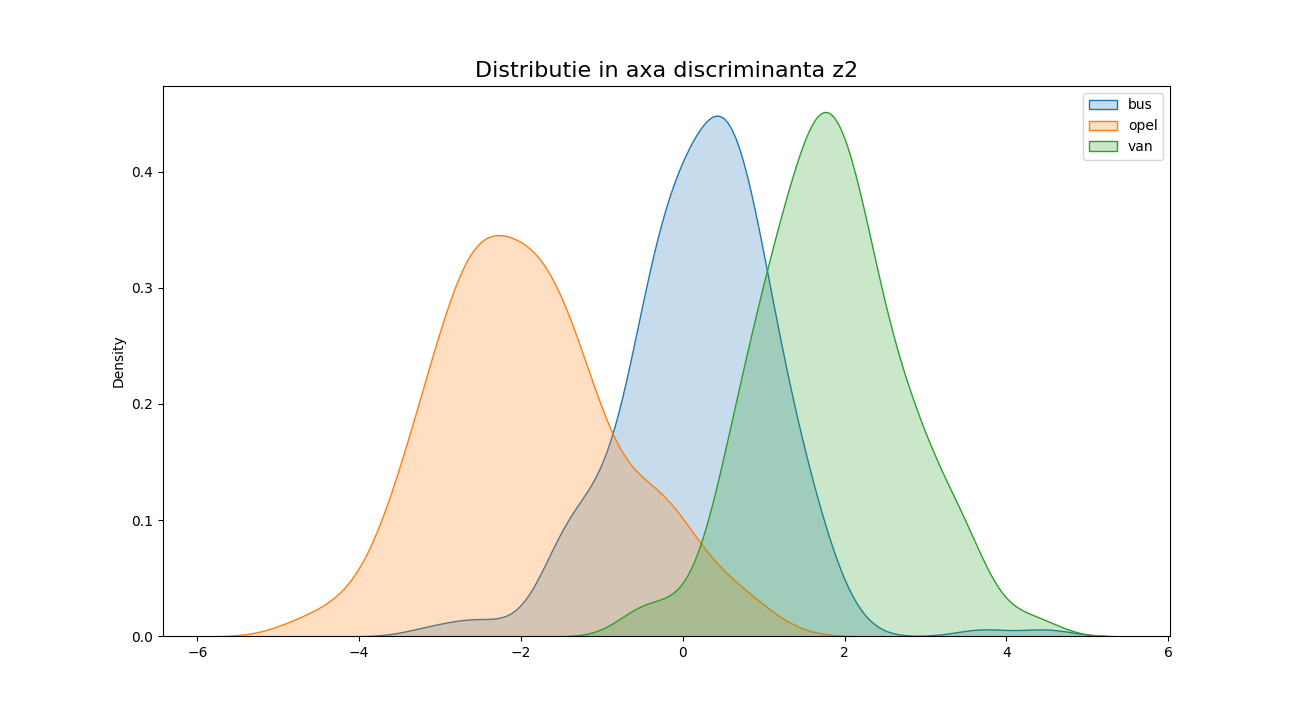


Observăm în graficul de mai sus că în general sunt grupe omogene, în special grupa a 3-a, reprezentând autovehiculele de tip van.

Am realizat apoi un plot de distribuţie în axa discriminantă z1 şi în axa z2.

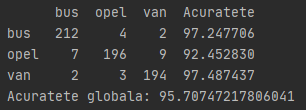


Prima funcţie separă foarte bine grupele bus şi van, mai puţin grupele opel şi van.



A doua funcţie separă grupele opel şi van foarte bine şi mai puţin bine grupele opel şi bus, bus şi van.

Rezultate matrice mal-clasificări :

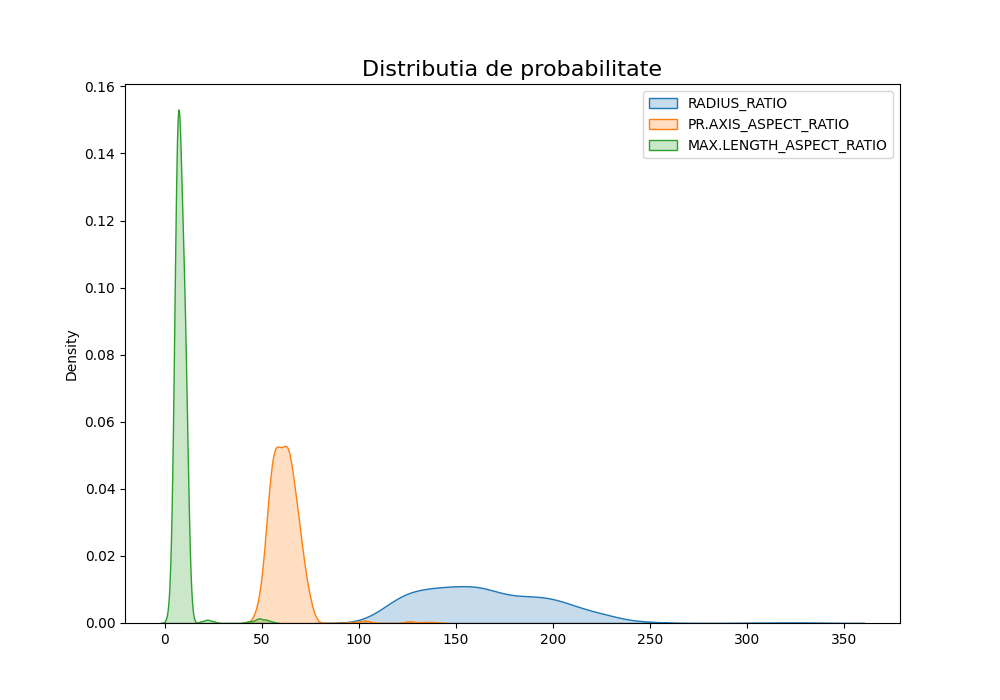


212 instanţe din clasa bus au fost clasificate corect (97%), 6 instanţe au fost clasificate greşit (4 – clasa opel, 2 – clasa van)

196 instanţe din clasa opel au fost clasificate corect (92%), 16 instanţe au fost clasificate greşit (7 – clasa bus, 9 – clasa van)

194 instanţe din clasa van au fost clasificate corect (97%), 5 instanţe au fost clasificate greşit (2 – clasa bus, 3 – clasa opel)

Per total, instanţele au fost clasificare corect în proporţie de 96%.

Am contruit pe tabelul de intrare un grafic de distribuţie pentru a observa dacă datele urmează o distribuţie normală.

Sursa datelor : <https://www.openml.org/d/54>