

## DN5: Linearna diskriminantna analiza

1. Opis podatkov. Ime raziskave oziroma podatkov. Leto in kraj/država, če je to za podatke smiselno. Enota analize. Velikost vzorca. Uporabljene spremenljivke (ime v datoteki in vsebinski pomen). Ostale posebnosti, za katere menite, da so vredne omembe (obravnavajo manjkajočih vrednosti, rekodirane spremenljivke ...).
2. Skupine in povezanost. Kolikšna je velikost skupin na vzorcu? Na podlagi grafičnega prikaza analizirajte povezanost med spremenljivkami in skupinami. Ali spremenljivke dobro ločijo med skupinami?
3. Predpostavke. Preverite predpostavki o multikolinearnosti spremenljivk (korelacije, rang, lastne vrednosti, občutljivost) in homogenosti kovariančnih matrik po skupinah.
4. Linearna diskriminantna analiza. Navedite metodo linearne diskriminantne analize, število skupin in predpostavljeno velikost skupin na populaciji. Izračunajte prvi dve DF. Ali obe DF statistično značilno razlikujeta med skupinami? Kolikšen delež variabilnosti med skupinami pojasni posamezna DF? Zapišite in interpretirajte povprečja (centroide) DF po skupinah. Dodatno si pomagajte s standardiziranimi koeficienti DF in opišite pomen posamezne DF. Grafično prikažite porazdelitev vrednosti (scores) DF po skupinah.
5. Kvaliteta modela. Kolikšen bi bil delež pravilno razvrščenih enot, če bi vse enote razvrstili v največjo skupino? Izpišite kontingenčno tabelo, zapišite delež pravilno razvrščenih enot in izračunajte popravljen Randov indeks. Kolikšna bi bila nepristranska ocena za delež pravilno razvrščenih enot (CV)?
6. Grafični prikaz. Narišite razsewni grafikon, kjer sta osi določeni s prvima dvema DF. Enote obravnavajte kot pripadnike skupin. Skupine v razsewnem grafikonu naj se med seboj ločijo po barvi. Zapišite glavne ugotovitve (prekrivanje skupin in uspešnost diskriminantne analize), ki so razvidne iz grafičnega prikaza.

---

MASS: lda, predict.lda  
mclust: adjustedRandIndex

## **DODATNO (NI ZA DOMAČO NALOGO)**

Večrazmernostno lestvičenje (MDS).

- Klasično metrično, ne-metrično in Sammon.
- Lastne vrednosti in/ali stress.
- Grafični prikaz 2D rešitve.

Korespondenčna analiza (CA).

- Kontingenčna tabela.
- Primerjava vrstic in stolpcev: strukturni odstotki (profile) in vztrajnost (inertia).
- Hi-kvadrat test, skupna vztrajnost (total inertia) in sled (trace).
- Lastne vrednosti in delež pojasnjene variabilnosti.
- Grafični prikaz 2D rešitve.
- Kakovost predstavitve podatkov ( $\cos^2$ ).

---

MASS: corresp, isoMDS, sammon