

Cvičenie 4.

- Pre daný súbor `farmaka.txt` spravte analýzu hlavných komponent: Datový súbor obsahuje $100 \times R_f$ pre 20 zlúčenín a 18 elučných činidel F (sloupce). Prvý stĺpec určuje poradie a zvyšné hodnoty pre: Toluen : aceton : ethanol : 30-percentný amoniak = 45 : 45 : 7 : 3; Ethylacetát: benzen : methanol : 30-percentný amoniak = 60 : 35 : 6.5 : 2.5; Benzen : dioxan : ethanol : 30-percentný amoniak = 50 : 40 : 7.5 : 2.5; Methanol : 30-percentný amoniak = 100 : 1.5; Benzen : 2-propanol : methanol : 30-percentný amoniak = 70 : 30 : 20 : 5; Ethylacetat : methanol : 30-percentný amoniak = 85 : 10 : 5; Cyklohexan : toluen : diethylamin = 65 : 25 : 10; Cyklohexan : toluen : diethylamin = 75 : 15 : 10; Cyklohexan : benzen : metanol : diethylamin = 70 : 20 : 10 : 5; Chloroform : aceton : diethylamin = 50 : 40 : 10; Cyklohexan : chloroform : diethylamin = 50 : 40 : 10; Benzen : ethylacetat : diethylamin = 50 : 40 : 10; Xylen : methylethylketon : methanol : diethylamin = 40 : 40 : 6 : 2; Diethylether : diethylamin = 95 : 5; Ethylacetat : chloroform = 50 : 50; Ethylacetat :chloroform [A] = 50 : 50; Butanol : methanol = 40 : 60; Butanol : methanol [A] = 40 : 60, kde [A] značí, že byl užit 0.1M methanolát draselný. Testované zlúčeniny boli: atropine, biperid, caffeine, cocaine, codein, cyclizin, diazepam, ketamine, lignocaine, lorazepam, mebeveri, methadon, morphine, naloxone, papaverine, pentazoc, phenacet, phenazon, prazepam, procaine.

1. určte hlavné komponenty a príslušné koeficienty. Koľko hlavných komponent je potrebné použiť?
2. zobrazte Cattelov indexový graf vlastných čísel a ich kumulativnu cetnost.
3. zobrazte grafy komponentných váh pre hlavné komponenty.
4. zobrazte rozptylový diagram komponentného skóre.
5. zobrazte dvojný graf (biplot) zdrojovej matice.
6. popíšte informácie, ktoré sme schopní z daných grafov odčítať.

Opakovanie

Stiahnite si súbor s dátami `Klas vin.txt` a spravte všetky kroky potrebné pre úspešnú klasifikáciu. Súbor obsahuje údaje o troch odrodách vín 1.Barolo, 2.Grigolino a 3.Barbela. Každý objekt je popísaný ôsmimi znakmi: obsah alkoholu, necukorný extrakt, fosfáty, celkové fenoly, flavanoidy, pomer absorbančí pri 280 a 315 nm pre nariedené víno, pomer absorbančí pri 280 a 315 nm pro flavanoidy, obsah prolinu. Súbor má 10 stĺpcov, prvý určuje poradie vzorky, druhý určuje odrodu vína a zvyšné obsahujú hodnoty príslušných znakov.

- učte základné štatistické vlastnosti jednotlivých tried (vektor stredných hodnôt, variančná matica) a zobrazte si jednotlivé znaky(2D a 3D grafy).
- odhadnite a overte rozloženie znakov v každej skupine.

- otestujte zhodnosť variančných matíc a vektorov stredných hodnôt.
- pomocou PCA určte, či sú pre diskriminácia významné všetky znaky, alebo niektoré môžeme vylúčiť.
- na základe grafického zobrazenie vyberte aspoň tri vhodné spôsoby klasifikácie. Určte úspešnosti jednotlivých typov. Vyberte najúspešnejšiu klasifikáciu.
- pomocou najlepšej klasifikácie zaraďte nové dáta v súbore Klas vin zaraďit.txt.

Podrobne popísanú dávku s oddôvodneniami, prečo ste si vybrali dané klasifikácie a ktorá a prečo bola najlepšia odovzdajte do odovzávare predmetu. .