

Cvičenie 1.

Načítajte si data troch skupín kosatcov `iris.dat`. Tento súbor je v knižnici `matlabu`, preto ho netreba stahovať alebo ukladať. Svoju prácu s komentárom ukladajte do dávkky a túto davku na konci hodiny nahrajte do odovzdávarene predmetu.

1. Súbor rozdeľte podľa jednotlivých skupín a pre každú z nich určte základné charakteristiky (vektor stredných hodnôt a variančnú maticu).

- načítanie dát do matice A:

```
A=load('iris.dat');
```

- načítanie skupiny 1 do matice:

```
k1=A(1:50,1:4);
```

- určenie vektora stredných hodnôt 1 skupiny:

```
mi1=1/length(k1).*(sum(k1));
```

- vyrátanie variačnej matice pre skupinu 1 :

```
m=0;
d=0;
for i=1:50
m=k1(i,1:4)-mi1;
d=d+m'*m;
end
sigma1=1/(length(k1)-1)*d
```

2. Pre každú skupinu zvolte apriórne pravdepodobnosti podľa četnosti prípadov jednotlivých skupín v celom súbore.
3. Dáta si zobrazte a porovnajte na 2D a 3D úrovni. Spravte obrázok, na ktorom bude graf pre každú dvojicu znakov a ďalší pre každú trojicu znakov. V každom grafe zobrazte tiež etalony jednotlivých skupín. Pomocou obrázkov rozhodnite, či sa dá pre každý prípad súboru jednoznačne určiť, do ktorej skupiny patrí. Ďalej rozhodnite, či nestačí rozdelenie priamkami u jednej dvojice znakov a či sú všetky znaky potrebné na zaradenie do skupiny.
 - zobrazenie viac grafov na jednom obrázku: `subplot(klm)`, kde `k`-počet riadkov, `l`-počet stĺpcikov, `m`-poradie grafu;
 - zobrazenie 3D grafu: `plot3`;
4. Pre každú skupinu určte najvhodnejšie viacrozmerne rozloženie pravdepodobnosti.

- pri rozhodovaní je užitočné zobrazíť histogram: histfit;
 - tiež skúsiť pre každý znak pravdepodobnostný graf: probplot;
5. Pomocou vhodných testov potvrdíte svoju hypotézu o danom rozložení. Prípadné odchýlky by sa mohli ukázať už na jednorozmernej úrovni.
 - príkladom testu, ktorý testuje, či dáta pochádzajú z nejakej rodiny rozložení je Lillieforsov test: lillietest;
 6. Použitím vhodného testu porovnajete variančné matice a v prípade zhodnosti aj vektory stredných hodnôt. Prečo je takýto test potrebný?
 7. Určte funkcie kritéria maximálnej aposteriórnej pravdepodobnosti pre zaradenie do každej skupiny a následne túto klasifikáciu použite na tento súbor, čím dostanete nové zaradenie prípadov. Stratová funkcia bude v tomto prípade nulová.
 8. Určte úspešnosť klasifikácie. Napríklad ako pomer správne zaradených prvkov podľa novej klasifikácie ku počtu všetkých prvkov.
 9. Vhodne zmeňte apriórne pravdepodobnosti. Určte novú klasifikáciu a jej úspešnosť.
 10. Vyberte si vhodné hodnoty stratovej funkcie pre každú skupinu a určte funkcie kritéria minimálnej strednej straty. Túto klasifikáciu použite opäť na pôvodný súbor.
 11. Určte úspešnosť novej klasifikácie.
 12. Zmeňte počiatočný súbor napr. o 10 vzoriek menej v každej skupine. Určte nové vektory stredných hodnôt, variančné matice a funkcie Bayesovho klasifikátora. Následne skúste tieto odstránené prípady zaradiť. Určte úspešnosť.
 13. Určte úspešnosť zaradenia. Porovnajte jednotlivé úspešnosti a rozhodnite, ktorá klasifikácia bola najúspešnejšia.