

MNOHORozMĚRNÁ STATISTIKA — POKRAČ.

4.1.2013

ÚVODNÍ NASTAVENÍ.

- Otevřete si R Studio. Z internetu si stáhněte data `ccor.csv`, budeme s nimi pracovat v druhé části hodiny.

OPAKOVÁNÍ

- mnohorozměrná data,
- metoda hlavních komponent,
- faktorová analýza,
- shluková analýza.

DISKRIMINAČNÍ ANALÝZA

1. Budeme uvažovat data `iris`, která obsahují informace o 150 kosatcích ze tří tříd (`setosa`, `virginica`, `versicolor`). Na každé květině byla provedena čtyři měření: šířka a délka kališních lístků a okvětních lístků. Otázkou je, zda je možné na základě měření těchto čtyř veličin zařadit květinu do příslušné třídy.

- (a) Nejprve si data načteme a prohlédneme

```
data(iris)

summary(iris)
pairs(iris[,1:4], col=iris[,5])

par(mfrow=c(2,2))
for(i in 1:4)
  boxplot(iris[,i]~iris[,5],main=names(iris)[i])
```

- (b) Na základě všech dat vytvoříme diskriminační pravidlo (použijeme metodu lineární diskriminace):

```
library(MASS)

da= lda(Species ~Sepal.Length + Sepal.Width + Petal.Length+
        Petal.Width, data=iris)
# nebo zkracene: da= lda(Species ~., data=iris)
print(da)
```

- (c) Zhodnocení přesnosti klasifikace (hodnoceno na „trénovacích“ datech):

```
pred=predict(da)

#podrobny vypis:
pred
```

```

table(iris$Species, pred$class)
prop.table(table(iris$Species, pred$class),1)

#celkove procento spravne zarazenych kvetin:
sum(diag(prop.table(table(iris$Species, pred$class))))

```

(d) Obrázek:

```

par(mfrow=c(1,1))
plot(da,col=as.numeric(iris$Species))

plot(da,dimen=1)

```

(e) Nyní na základě modelu můžeme klasifikovat nové pozorování - např. květinu s hodnotami 6,2,1.7,0.8

```

new=data.frame(Sepal.Length=6, Sepal.Width=2,Petal.Length=1.7,Petal.Width=0.8)
predLDA <- predict(da,newdata=new)
predLDA

```

SAMOSTATNÁ PRÁCE

8. Uvažujte data `crabs`. Cílem je vytvořit diskriminační pravidlo, pomocí něhož lze určit pohlaví kraba pouze na základě morfologických měření (FL, RW, CL, CW). Zjistěte, jak je taková diskriminace úspěšná.

```

data(crabs)
?crabs

```

9. Určete (odhadněte) pohlaví kraba, na kterém jsme naměřili FL=10.2, RW=7.5, CL=19.3, CW=22.1 (vše v mm).

KANONICKÉ KORELACE

1. Budeme uvažovat data `ccor.csv`. Data obsahují informace o 600 studentech prvního ročníku VŠ. Pro každého jsou zaznamenány tři psychologické proměnné (těžiště kontroly, self-koncept, motivace) a čtyři akademické proměnné (čtení, psaní, matematika, společenské vědy) — vše jsou standardizované skóry příslušných testů. Navíc známe pro každého pohlaví (nebudeme uvažovat v analýze). Cílem analýzy je popsat (kvantifikovat) vztah mezi psychologickými a akademickými proměnnými.

Načtěte si data a prohlédněte si jednotlivé proměnné a vztah mezi nimi.

2. Prohlédneme si podrobněji zvlášť psychologické a akademické proměnné

```

psych = ccor[, 1:3]
acad = ccor[, 4:7]

pairs(psych)
pairs(acad)

```

```
cor(psych)
cor(acad)
cor(psych,acad)
```

3. Spočítáme kanonické korelace:

```
install.packages("CCA")
library(CCA)

cc1=cc(psych,acad)

cc1$cor

# koeficienty kanonických promenných:
cc1[3:4]

# interpretace kanon. promenných - korelace s původními
cs=cc1$scores
cs[3:6]

#grafické zobrazení
plt.indiv(cc1,1,2)
plt.var(cc1,1,2,var.label=T)
```