

R instructions for the 7th seminar

In a data set *Holiday.RData* there are information about 50 families and their attitude to holiday destinations:

<i>ID</i>	yes/no indicator showing whether a family visited particular holiday destination during last 2 years
<i>X</i> ₁	states an annual family income in thousands of dollars
<i>X</i> ₂	states an attitude to travelling (on a scale 1 to 9, where 1 = absolutely negative, 9 = absolutely positive)
<i>X</i> ₃	states an importance attributed to a family holiday (on a scale 1 to 9, where 1 = the lowest, 9 = the highest)
<i>X</i> ₄	states the number of family members
<i>X</i> ₅	states an age of the oldest member of a family
<i>V</i>	states planed holiday expenditures of a family: little (1), medium (2), much (3)

All following tasks do first when:

a) The *ID* is a grouping variable (this must be of class factor). Explanatory variables are X_1, \dots, X_5 . (Data are considered to be training data set.)

At home when:

b) The *V* is a grouping variable. Explanatory variables are X_1, \dots, X_5 the same.

R Instructions for the problem 1:

Get familiar with data

- Plot 2D graphs with pairs of explanatory variables and distinguish the points with regard to grouping variable.

```
plot(Holiday[,2:6],col=Holiday[,1])
```

- Calculate means of all explanatory variables separated by grouping variable; what do the results suggest?

```
for(i in 2:6)cat("X",i-1," : ",tapply(Holiday[,i],Holiday[,1],mean), ' \n')
```

CANONICAL DISCRIMINANT ANALYSIS:

R Instructions for the problem 2:

- How many canonical discriminant variables are in the model?

(Answer: for grouping variable = *ID* it is $l = \min\{p = 5, k - 1 = 1\} = 1$. Thus we are looking for just one discriminant variable.)

R Instructions for the problem 3:

- Create the object in R bearing all essential results of Canonical discriminant analysis.

```
library(candisc)
```

```
LinModel<-lm(cbind(X1,X2,X3,X4,X5) ID,data=Holiday)
```

```
CanResults1<-candisc(LinModel,term="ID")
```

or for groupin variable = *V*:

```
LinModel2<-lm(cbind(X1,X2,X3,X4,X5) V,data=Holiday)
```

```
CanResults2<-candisc(LinModel2,term="V")
```

R Instructions for the problem 4:

- Find out raw and standardized discriminant coefficients. (Standardized allows better interpretation.) Which variable bears the largest peace of information for discrimination?

```
CanResults1$coeffs.raw
```

```
CanResults1$coeffs.std
```

(for standardized: $Y = -0.848X_1 + \dots + -0.469X_5$ Thus the direction of discriminant variable *Y* is mostly influenced by X_1 which is an income...)

R Instructions for the problem 5:

- Find out correlations between original variables and canonical variable(s). Compare its signs with signs of discriminant coefficients. Interpret it.

`CanResults1$structure`

(e.g.: $R(X_1, Y) = -0.929$).

R Instructions for the problem 6:

- Find out canonical scores. In other words express all cases by the mean of the new canonical variable(s).

`CanResults1$scores`

R Instructions for the problem 7:

- Find out means of the new canonical variable(s) for both groups.

`CanResults1$means`

("No" group mean of Y: 1.059855; "Yes" group mean of Y: -1.463609)

R Instructions for the problem 8:

- Express graphically the results of canonical discriminant analysis.

```
plot(CanResults1, fill=TRUE)
```

```
NewData<-data.frame(CanResults1$scores,rep(0,times=50))
```

```
head(NewData)
```

```
plot(NewData[,2:3],col=NewData[,1])
```

for volunteers TODO: LINEAR DISCRIMINANT ANALYSIS LDA:

```
library(MASS)
```

R Instructions for the problem 9:

- Check assumptions (multivariate normality in all groups (we wish not to reject), equality of Variance matrices (we wish not to reject), equality of mean vectors (we wish to reject))

- Create objects for LDA:

```
LdaResults1<-lda(ID ~ X1+X2+X3+X4+X5,data=Holiday,prior=c(0.5,0.5),CV=F)
```

```
LdaResults1
```

To get the classification matrix:

```
LdaPredict1<-predict(LdaResults1,newdata=Holiday[,2:6])$class
```

```
LdaPredict1
```

```
table(LdaPredict1,Holiday[,1])
```

```
LdaResults1.1<-lda(ID ~ X1+X2+X3+X4+X5,data=Holiday,prior=c(0.5,0.5),CV=T)
```

```
LdaResults1.1
```

Nefachci pro jiné prior, kořeny i tabulka úspěšnosti stejné, jako pro 0,5

```
LdaResults2<-lda(ID ~ X1+X2+X3+X4+X5,data=Holiday,prior=c(0.58,0.42),CV=F)
```

```
LdaResults2
```

```
LdaResults2.1<-lda(ID ~ X1+X2+X3+X4+X5,data=Holiday,prior=c(0.58,0.42),CV=T)
```

```
LdaResults2.1
```