Biometry. Lecture 28

Alexey Shipunov

Minot State University

May 4, 2012

Outline

Outline

Starting...

```
> setwd("<working folder>")
```

or

“Change dir”

in menu!

Previous final question: the answer

Whet is a difference between euclidean and manhattan distances?

Previous final question: the answer

Whet is a difference between euclidean and manhattan distances?

- Euclidean uses hypotenuse whereas manhattan—sum of legs

Multivariate statistics, or Data Mining

Cluster analysis

Another hierarchical clustering example (very simple)

```
> fences <- read.table(
+ "http://ashipunov.info/data/fences.txt", h=T)
> library(cluster)
> str(fences)
> fences.d <- daisy(fences)
> summary(fences.d)
> plot(hclust(fences.d))
```

Fuzzy clustering

```
> iris.f <- fanny(iris[,1:4], 3)
> plot(iris.f, which=1, main="")
> head(data.frame(sp=iris[,5], iris.f$membership))
```

Multivariate statistics, or Data Mining

Classification (machine learning)

Classification (machine learning)

- Machine learning, or classification is alway based on the example where objects already distributed into groups
- These methods are trying to find a best classification algorithm

Multivariate statistics, or Data Mining

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA)

- Linear discriminant analysis is based on the idea that classification could be made on a bases of linear equations
- This is a parametric method

LDA example

```
> library(MASS)
> iris.train <- iris[seq(1,nrow(iris),5),]
> iris.unknown <- iris[-seq(1,nrow(iris),5),]
> iris.lda <- lda(Species ~ . , data=iris.train)
> iris.ldap <- predict(iris.lda, iris.unknown[,1:4])$class
> table(iris.ldap, iris.unknown[,5])
```

LDA testing

```
> ldam <- manova(as.matrix(iris.unknown[,1:4]) ~
+ iris.ldap)
> summary(ldam, test="Wilks")
```

"Wilks" value is not only a statistic, it is also a likelihood ratio: for better classifications, Wilks is closer to 0

LDA visualization

```
> iris.lda2 <- lda(scale(iris[,1:4]), iris[,5])
> iris.ldap2 <- predict(iris.lda2, dimen=2)$x
> plot(iris.ldap2, type="n", xlab="LD1", ylab="LD2")
> text(iris.ldap2, labels=abbreviate(iris[,5], 1,
+ method="both.sides"))
```

Multivariate statistics, or Data Mining

Regression trees (recursive partitioning)

Regression trees (recursive partitioning)

- Regression trees, or recursive partitioning are based on the same idea as biological descriptive keys
- On each step, methods searches for the best separation between members of group

Regression tree example I

```
> library(tree)
> iris.tree <- tree(Species ~ ., data=iris)
> plot(iris.tree); text(iris.tree)
```

Regression tree example II

```
> eq <- read.table("http://ashipunov.info/data/eq.txt", h=TRUE
> eq.tree <- tree(SPECIES ~ ., data=eq)
> plot(eq.tree); text(eq.tree)
```

Multivariate statistics, or Data Mining

Advanced methods of classification

Advanced methods of classification

- “Random Forest” is based on the construction of multiple regression trees
- “Support Vector Machines” try to find a hyperplane which separates objects best

Random Forest example

```
> library(randomForest)
> set.seed(17)
> iris.rf <- randomForest(Species ~ ., data=iris.train)
> iris.rfp <- predict(iris.rf, iris.unknown[,1:4])
> table(iris.rfp, iris.unknown[,5])
```

Random Forest visualization

```
> set.seed(17)
> iris.urf <- randomForest(iris[,1:4])
> MDSplot(iris.urf, iris[,5])
```

SVM example

```
> library(e1071)
> iris.svm <- svm(Species ~ ., data=iris.train)
> iris.svmp <- predict(iris.svm, iris.unknown[,1:4])
> table(iris.svmp, iris.unknown[,5])
```

Finishing...

```
> savehistory("20120504.r")
```

Reference materials

Short anonymous absolutely voluntary survey

1. What do you **like** most in biometrics course?
2. What do you **dislike** most in biometrics course?
3. **Which lab** do you remember most of all?
4. Please grade (1—bad, 5—excellent):
    1. Lectures
    2. Labs
    3. Final questions
    4. Exams

Summary: most important commands

- `lda()`—preforms linear discriminate analysis (LDA)
- `tree()`—performs recursive partitioning

For Further Reading

A. Shipunov.
Biometry [Electronic resource].
2012—onwards.
Mode of access: `http://ashipunov.info/shipunov/school/biol_299`