

鸢尾花数据SVM与LDA练习

```
#data processing
#加载并查看R内置iris数据集（共150行，包含花萼长度、花萼宽度、花瓣长度、花瓣宽度、
类别5个变量）
data(iris)
iris
#设定种子，随机取样70%行数据（105个数据样本）作为训练数据集，其余剩下30%行数据作
为测试数据集
set.seed(44)
num<-sample(nrow(iris),0.7*nrow(iris))
train_data<-iris[num,]
test_data<-iris[-num,]
train_data
test_data

#SVM
#安装并加载SVM所需R包"e1071"
install.packages("e1071")
library(e1071)
#训练得到模型：用train_data来训练，选用$Species为因变量，其它所有变量为自变量
model <- svm(Species ~., train_data)
#利用test_data测试
#新建y变量记录测试数据集中每行数据所对应的鸢尾花的类别，用于SVM测试较结果的比较
y.test_data<-test_data$Species
#删去测试数据集中的类别变量数据
test_data$Species=NULL
#利用训练所得模型对测试数据集每行数据的y进行预测
y.predict<-predict(model,test_data)
#将测试数据集真实类别y与利用模型预测得到的y进行比较
y.test_data==y.predict
#y.test_data==y.predict
# [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
TRUE  TRUE
#[13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
FALSE  TRUE
#[25]  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
TRUE  TRUE
#[37]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
```

#第23、第26个、第41个样本数据的类别预测错了，其它42个都预测正确，分类器分类的正确率为93.3%。#

#LDA

#安装并加载LDA所需的包

```
install.packages("MASS")
```

```
library(MASS)
```

#利用train_data训练数据、创建模型

```
model <- lda(Species ~., data = train_data)
```

#利用所得模型对test_data进行预测，并查看预测到的class

```
y.predict<-predict(model,test_data)
```

```
y.predict$class
```

#将预测所得类别与测试数据集类别y进行比较

```
y.test_data==y.predict$class
```

```
#y.test_data==y.predict$class
```

```
# [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  
TRUE  TRUE
```

```
#[13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  
TRUE  TRUE
```

```
#[25]  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  
TRUE  TRUE
```

```
#[37]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
```

#第26个、第41个样本数据的类别预测错了，其它43个都预测正确，此处线性判别分析准确率为95.5%。

#此处种子设置为44的情况下LDA分类效果优于SVM，不过其实不设种子的时候LDA和SVM结果基本是差不多的，在测试阶段基本都是错两个左右。