

# Классификация с обучением

Alexeyeva N.P.

24 04 2020

## Подготовка данных

```
library("knitr")  
library(klaR)
```

```
### Loading required package: MASS
```

```
data <- read.csv("~/Documents/share/data/data_big.csv")  
df <- na.omit(data.frame(weak=as.factor(ifelse(data$weakness.1<=1,0,1)),  
                        scale(data[,c(23:30)])))
```

Исследуем зависимость между фактором “Слабость” и кардиологическими данными.

## Дискриминантный анализ

```
LDA <- lda(weak ~ ., df)  
kable(data.frame(LDA$scaling), digits=3,  
       caption = "веса дискриминантной функции")
```

Таблица 1: веса дискриминантной функции

	LD1
HR.1	-0.550
SBP.1	-0.333
DBP.1	0.963
MBP.1	-0.511
SV.1	0.591
CO.1	-1.701
SI.1	-2.901
CI.1	3.607

```
print("матрица классификации")
```

```
## [1] "матрица классификации"
```

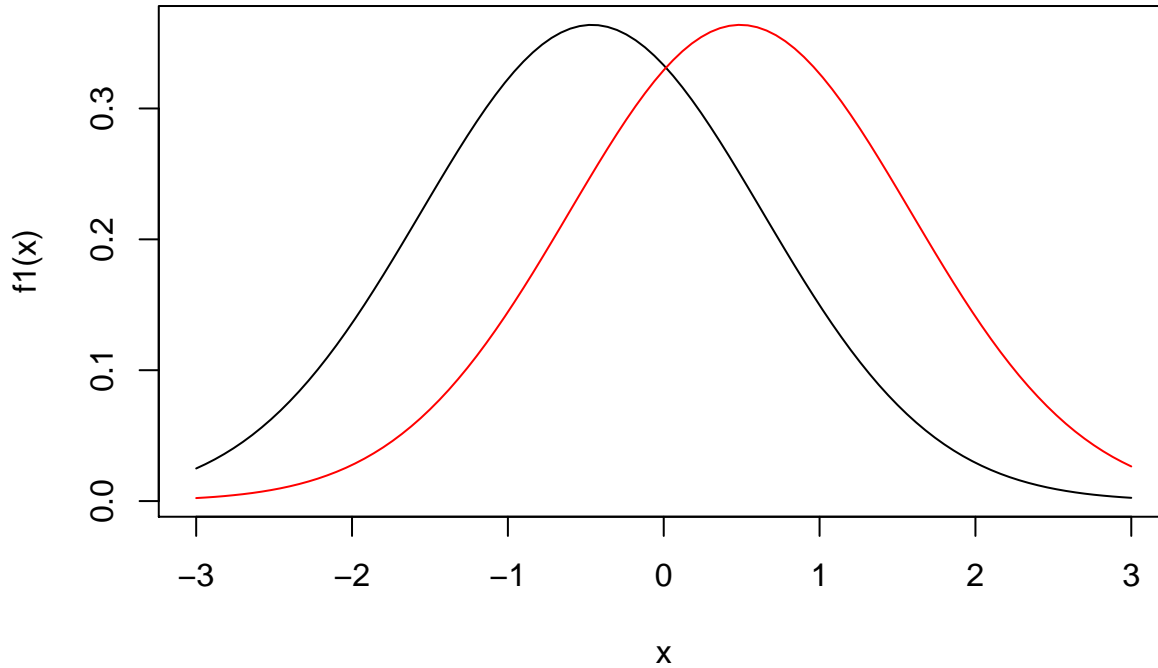
```
table(weak=df$weak, Predict=predict(LDA,df)$class)
```

```
## Predict  
## weak 0 1  
## 0 12 5  
## 1 5 11
```

```
mm <- tapply(predict(LDA,df)$x, df$weak, function(x) mean(x, na.rm=TRUE))  
SD <- sd(predict(LDA,df)$x)
```

```
f1<-function(x)dnorm(x,mean = mm[[1]],sd=SD)
f2<-function(x)dnorm(x,mean = mm[[2]],sd=SD)

curve(f1,-3,3)
curve(f2,-3,3,col=2,add=TRUE)
```



## Дерево решений

```
library("tree")
i.tree<-tree(weak~.,df,split = "gini")
P<-predict(i.tree,df)
Class<-apply(P,1,function(x)ifelse(x[1]>x[2],0,1))
table(df$weak,Class)
```

```
## Class
## 0 1
## 0 17 0
## 1 6 10
```

```
plot(i.tree)
text(i.tree)
```



## Случайный лес

```
library('randomForest')
```

```
## randomForest 4.6-12
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
i.rf<-randomForest(weak~.,df)
```

```
i.rfp<-predict(i.rf,df)
```

```
table(weak=df$weak,Predict=i.rfp)
```

```
## Predict
```

```
## weak 0 1
```

```
## 0 17 0
```

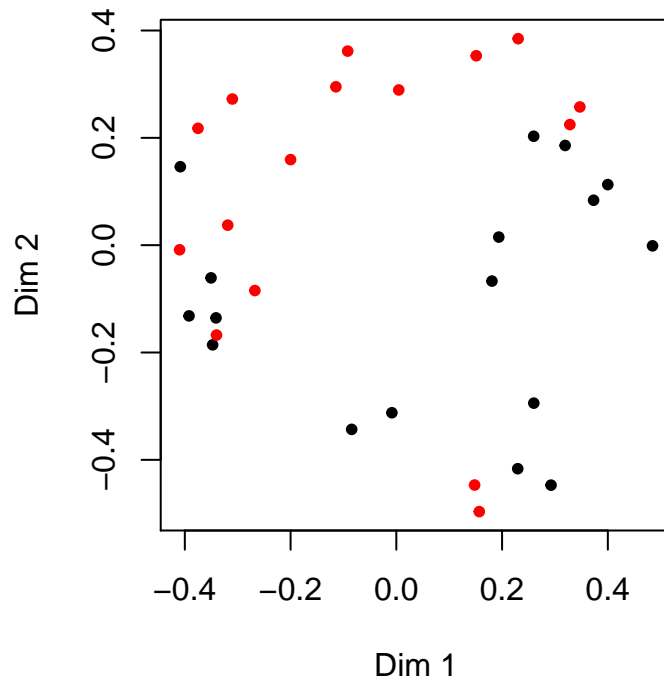
```
## 1 0 16
```

```
i.urf<-randomForest(df)
```

```
MDSplot(i.urf,df$weak,
```

```
palette=c(1,2)
```

```
)
```



#Метод опорных векторов

```
library(e1071)
i.svm<-svm(weak~.,df)
i.svmp<-predict(i.svm,df)
table(weak=df$weak,Predict=i.svmp)
```

```
##   Predict
## weak 0 1
##    0 15 2
##    1  4 12
```

## Задание для самостоятельной работы

В качестве переменной типа класс выбрать другой фактор и проверить возможность классификации по кардиологическим переменным при помощи дискриминантного анализа, дерева решений, случайный лес и метод опорных векторов.