1. Министерство науки и высшего образования Российской Федерации
2. Санкт-Петербургский Политехнический Университет Петра Великого
3. —
4. Институт компьютерных наук и технологий
5. Высшая школа искусственного интеллекта

**ЛАБОРАТОРНАЯ РАБОТА №3**

**«Метод опорных векторов»**

по дисциплине «Машинное обучение»

1. Выполнил: студент группы
2. 3540201/20302 А.И. Обидина

*<подпись>*

1. Проверил: Л.В. Уткин
2. д.т.н., профессор
3. *<подпись>*

Санкт-Петербург

2022

**Задание**

Данные для обучения и тестирования SVM-моделей, которые необходимо построить в приведенных ниже заданиях, хранятся в файлах с именами svmdataI.txt и svmdataItest.txt, где I номер задания.

1. Постройте алгоритм метода опорных векторов типа "C-classification" с параметром C = 1, используя ядро "linear". Визуализируйте разбиение пространства признаков на области с помощью полученной модели. Выведите количество полученных опорных векторов, а также ошибки классификации на обучающей и тестовой выборках.

2. Используя алгоритм метода опорных векторов типа "C-classification" с линейным ядром, добейтесь нулевой ошибки сначала на обучающей выборке, а затем на тестовой, путем изменения параметра C. Выберите оптимальное значение данного параметра и объясните свой выбор. Всегда ли нужно добиваться минимизации ошибки на обучающей выборке?

3. Среди ядер "polynomial", "radial" и "sigmoid" выберите оптимальное в плане количества ошибок на тестовой выборке. Попробуйте различные значения параметра degree для полиномиального ядра.

4. Среди ядер "polynomial", "radial" и "sigmoid" выберите оптимальное в плане количества ошибок на тестовой выборке.

5. Среди ядер "polynomial", "radial" и "sigmoid" выберите оптимальное в плане количества ошибок на тестовой выборке. Изменяя значение параметра gamma, продемонстрируйте эффект переобучения, выполните при этом визуализацию разбиения пространства признаков на области.

6. Постройте алгоритм метода опорных векторов типа "eps-regression" с параметром C = 1, используя ядро "radial". Отобразите на графике зависимость среднеквадратичной ошибки на обучающей выборке от значения параметра ε. Прокомментируйте полученный результат.

# Ход работы

## **Пункт 1**

Был разработал классификатор SVM с линейным ядром.

*svm <- svm(factor(Color)~.,data=train,type = "C-classification",cost = 1,kernel = "linear")*

Результаты кластеризации представлены на рисунке 1. Точность работы получившегося классификатора – 1, так как данные были сильно кластеризованы.

Количество опорных векторов – 2.

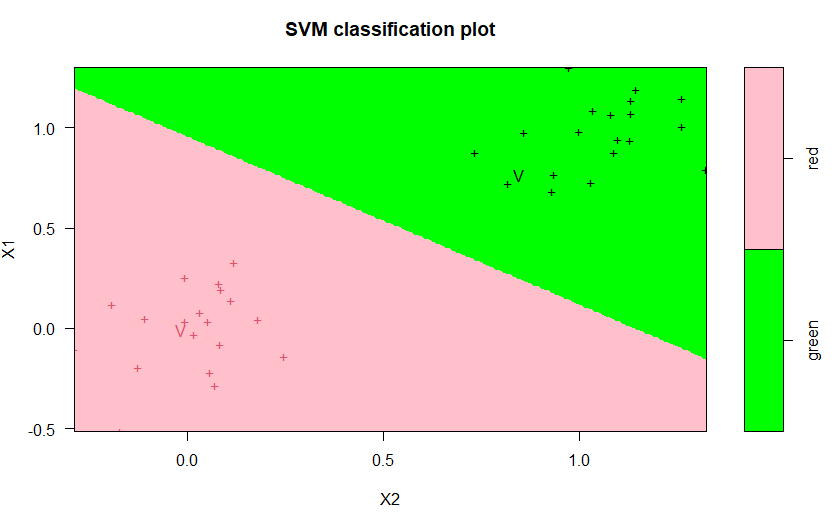


Рисунок 1 – SVM-классификация с линейным ядром

Код:

library(e1071)

train<-read.table("D:\\svmdata1.txt",stringsAsFactors = TRUE)

test<-read.table("D:\\svmdata1test.txt",stringsAsFactors = TRUE)

features<-data.frame(X1=test$X1,X2=test$X2)

svm <- svm(factor(Color)~.,data=train,type = "C-classification",cost = 1,kernel = "linear")

pr <- predict(svm, features)

res<-table(test$Color, pr)

res

acc<-(res[1,1]+res[2,2])/sum(res)

acc

plot(svm,test,grid = 250,col=c("green","pink"), dataSymbol="+", svSymbol="V")

## **Пункт 2**

Был построен алгоритм метода опорных векторов данных svmdata2.txt и svmdata2test.txt с линейным ядром. Была изучена зависимость качества классификации от значения cost, результаты представлены на рисунках 2-10.

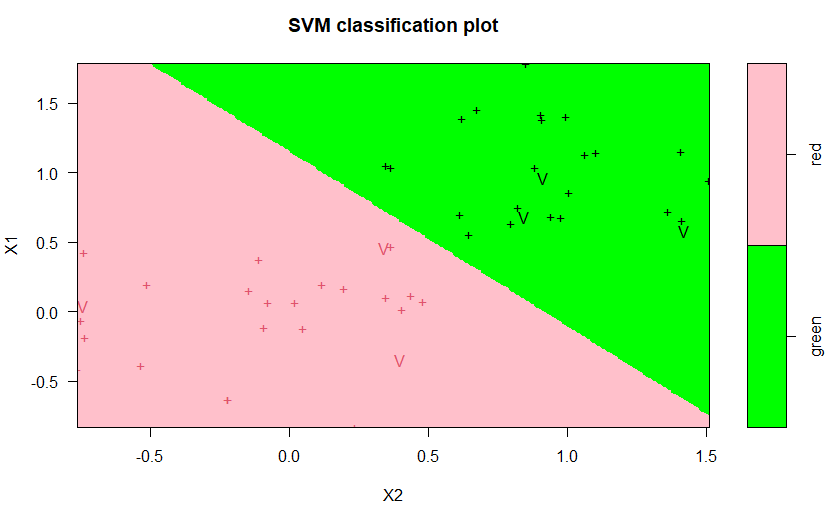


Рисунок 2 – SVM-классификация с cost=1

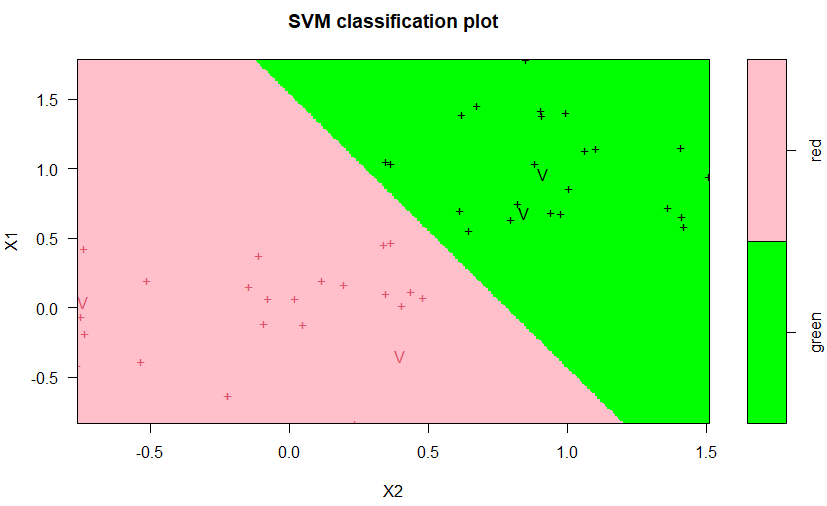


Рисунок 3 – SVM-классификация с cost=5



Рисунок 4 – SVM-классификация с cost=10

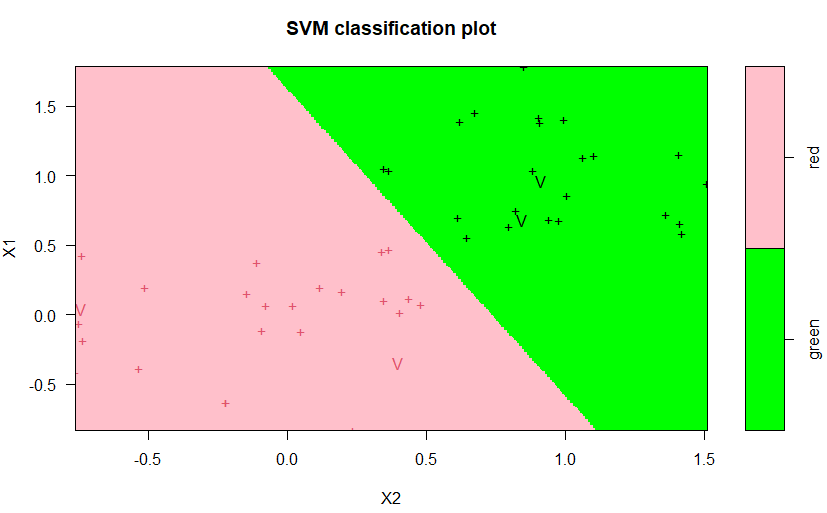


Рисунок 5 – SVM-классификация с cost=30



Рисунок 6 – SVM-классификация с cost=50

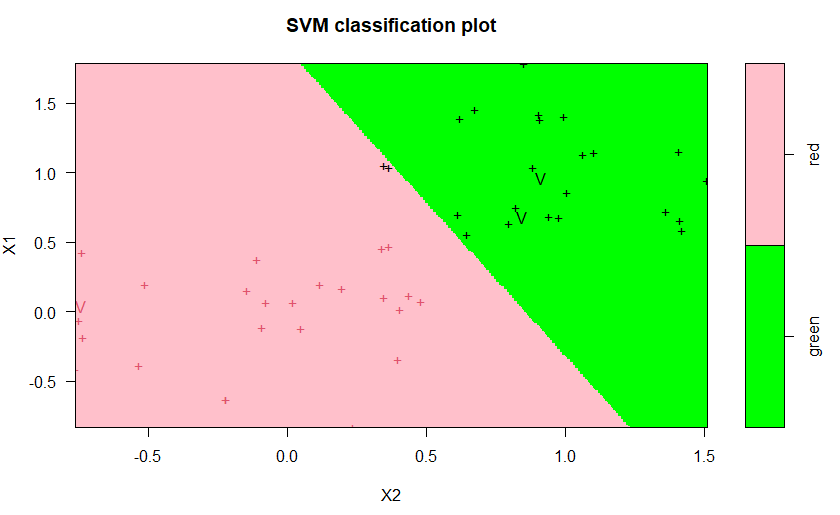


Рисунок 7 – SVM-классификация с cost=100

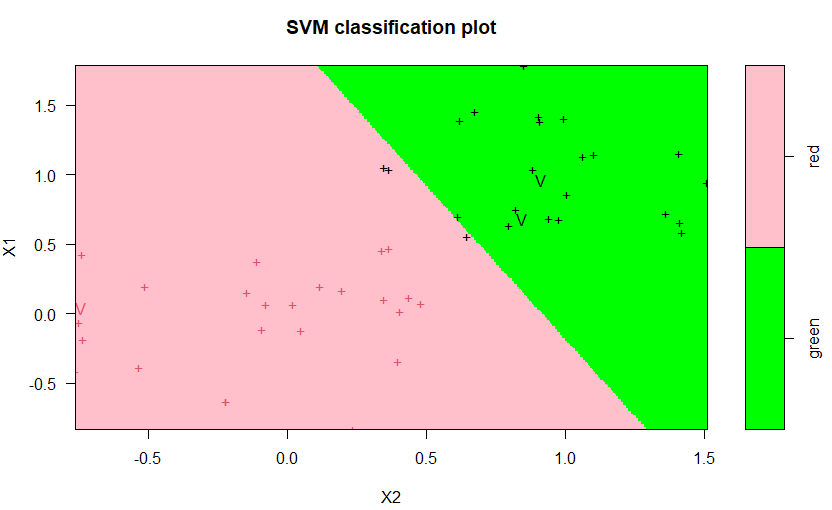


Рисунок 8 – SVM-классификация с cost=300



Рисунок 9 – SVM-классификация с cost=500



Рисунок 10 – SVM-классификация с cost=1000

Таблица 1. Результаты SVM-классификации

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cost** | 1 | 5 | 10 | 30 | 50 | 100 | 300 | 500 | 1000 |
| **Accuracy** | 1 | 1 | 1 | 1 | 1 | 0,96 | 0,94 | 0,94 | 0,94 |
| **Количество опорных векторов** | 6 | 4 | 4 | 3 | 3 | 3 | 3 | 3 | 3 |

Согласно результатам, представленным в таблице 1, наилучшие значения штрафа cost для классификации – 1, 5, 10, 30.

При классификации тестовых данных с Accuracy = 1, качество классификации обучающих данных равно 0,98. Следовательно, значение штрафа cost нужно подбирать для наилучшего качества классификации и тестовой, и обучающей выборок.

Код:

library(e1071)

train<-read.table("D:\\svmdata2.txt",stringsAsFactors = TRUE)

test<-read.table("D:\\svmdata2test.txt",stringsAsFactors = TRUE)

features<-data.frame(X1=test$X1,X2=test$X2)

costs<-c(1,5,10,30,50,100,300,500,1000)

for (c in costs) {

print(c)

svm <- svm(factor(Colors)~.,data=train,type = "C-classification",cost = 1,kernel = "linear")

pr <- predict(svm, features)

res<-table(test$Colors, pr)

print(res)

acc<-(res[1,1]+res[2,2])/sum(res)

acc

plot(svm,test,grid = 250,col=c("green","pink"), dataSymbol="+", svSymbol="V")

}

plot(X1 ~ X2,train, col = c("green","pink"))

plot(X1 ~ X2,test, col = c("green","pink"))

## **Пункт 3**

Были построены алгоритмы метода опорных векторов данных svmdata3.txt с разными типами ядер – polynomial, radial и sigmoid. Была изучена зависимость качества классификации от значения degree для полиномиального ядра, результаты представлены на рисунках 11-16.

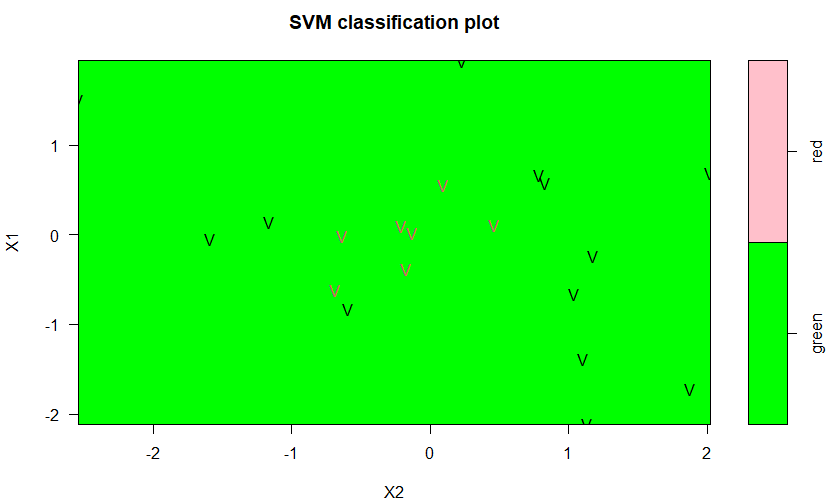


Рисунок 11 – SVM-классификация с degree=1 и полиномиальным ядром

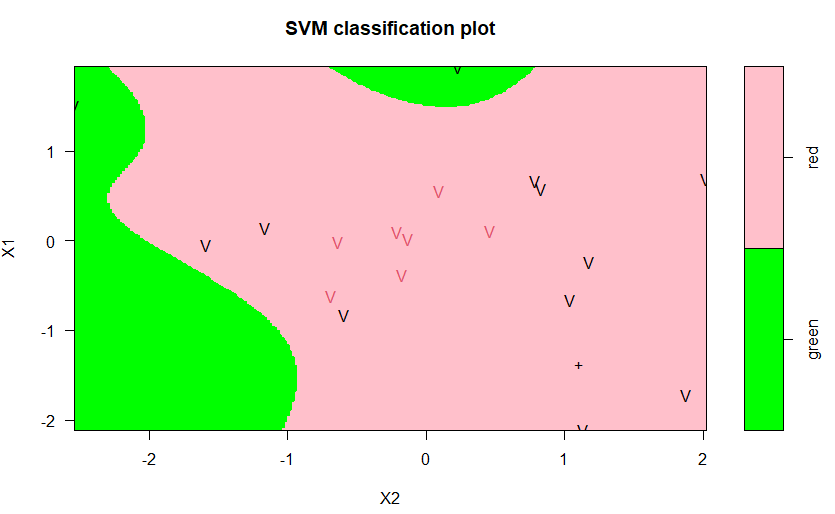


Рисунок 12 – SVM-классификация с degree=5 и полиномиальным ядром

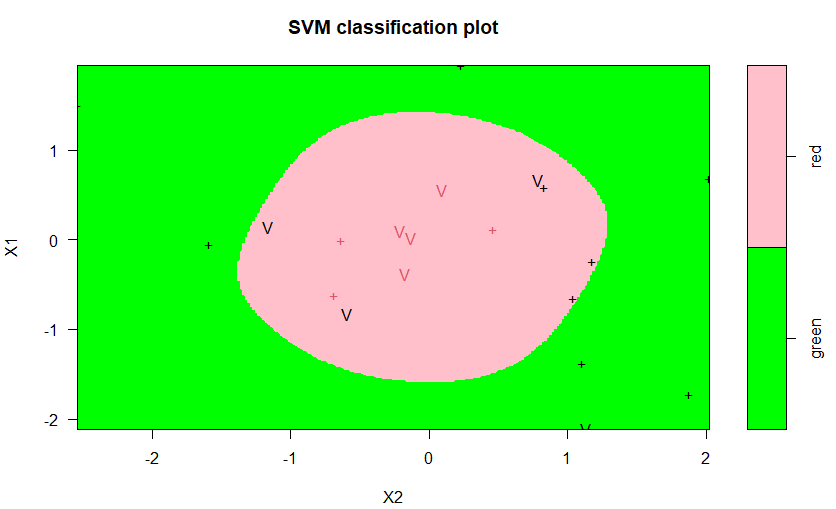


Рисунок 13 – SVM-классификация с degree=10 и полиномиальным ядром

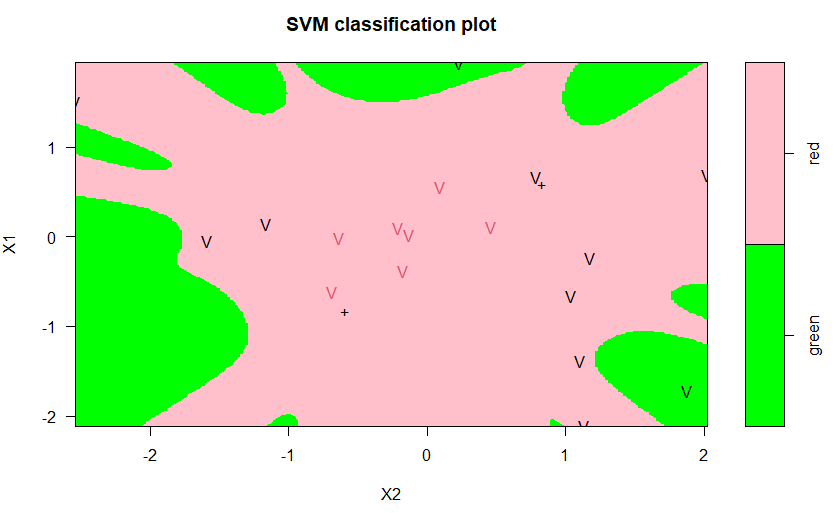


Рисунок 14 – SVM-классификация с degree=15 и полиномиальным ядром

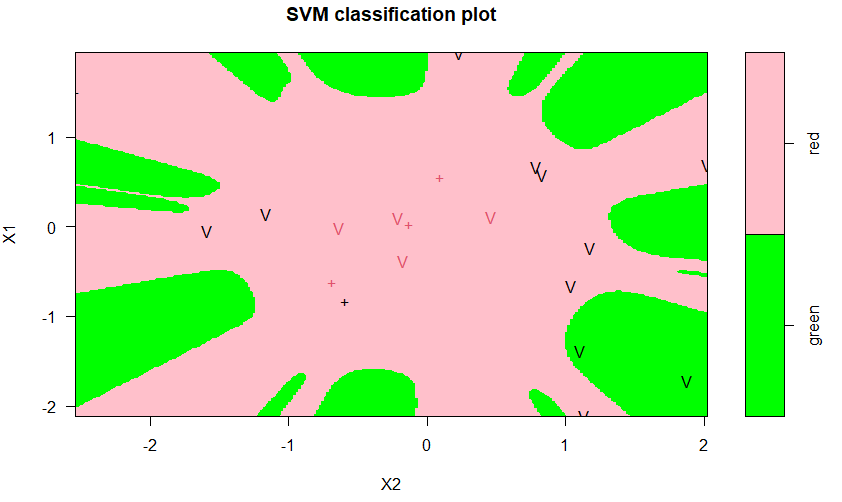


Рисунок 15 – SVM-классификация с degree=25 и полиномиальным ядром

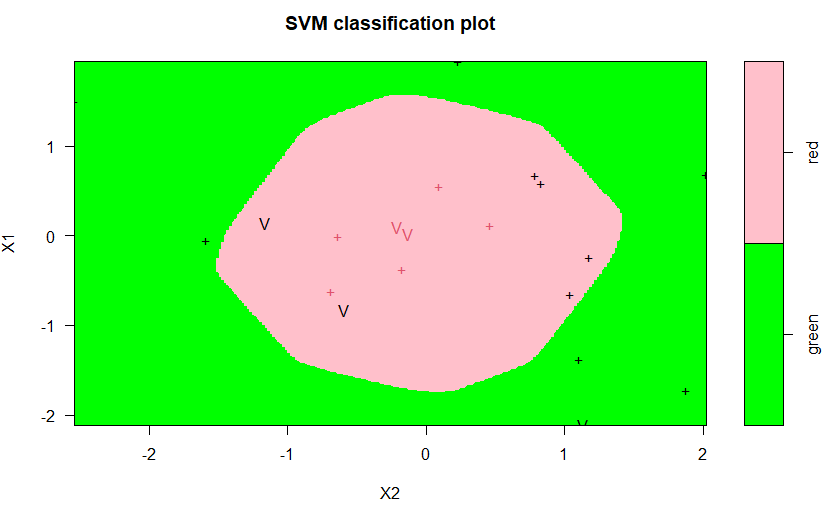


Рисунок 16 – SVM-классификация с degree=50 и полиномиальным ядром

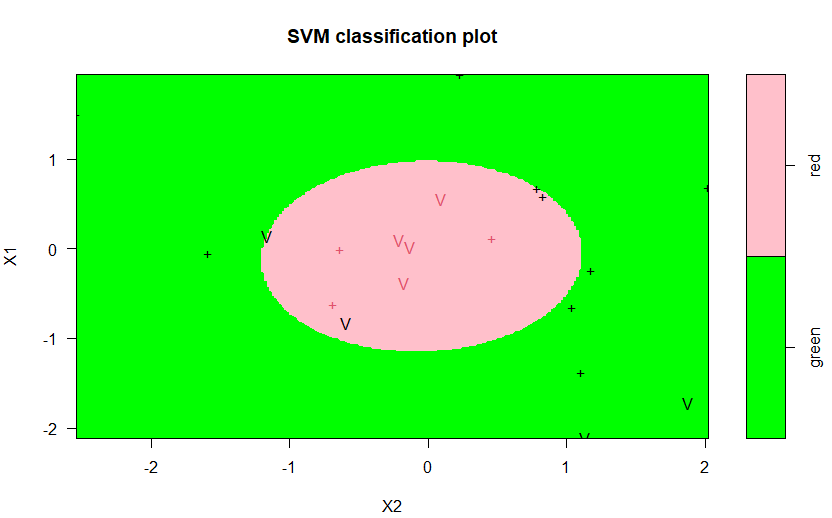


Рисунок 16 – SVM-классификация с degree=2 и полиномиальным ядром

Таблица 2. Результаты SVM-классификации с полиномиальным ядром

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Degree** | 1 | 5 | 10 | 15 | 25 | 50 | 2 |
| **Accuracy** | 0,65 | 0,45 | 0,7 | 0,45 | 0,45 | 0,7 | 0,8 |

Согласно результатам из таблицы 2 для полиномиального ядра наибольшую точность имеет модель с degree=2 (0,8). С увеличением degree точность классификации уменьшается.

Модель с радиальным ядром имеет точность 0,95, результат классификации представлен на рисунке 17.

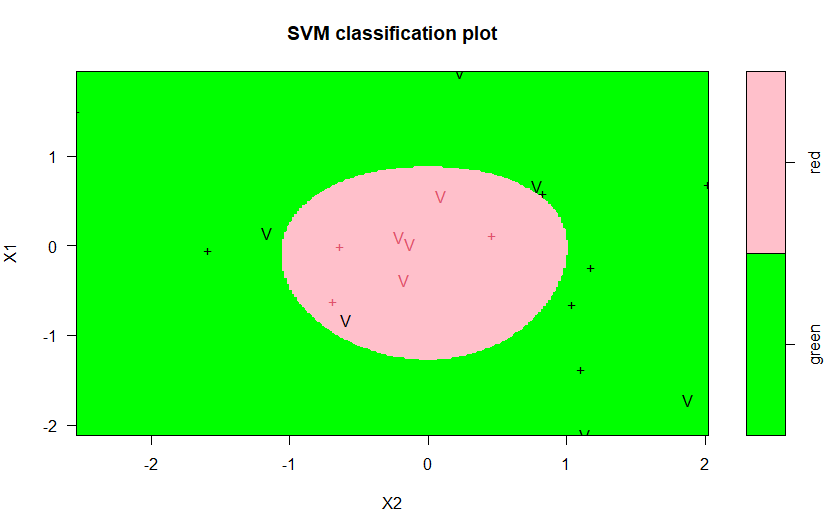


Рисунок 17 – SVM-классификация с радиальным ядром

Модель с сигмоидальным ядром имеет точность 0,6, результат классификации представлен на рисунке 18.

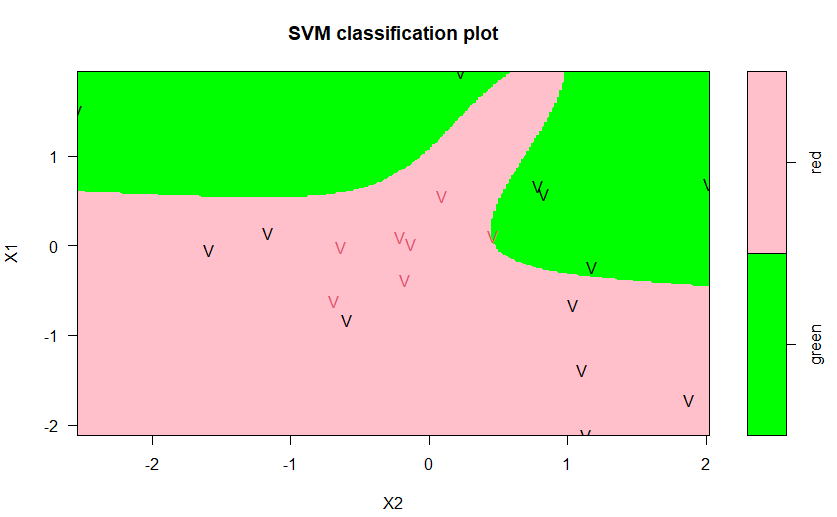


Рисунок 18 – SVM-классификация с сигмоидальным ядром

Таким образом, наиболее достоверная классификация данных наблюдается у модели с радиальным ядром.

Код:

library(e1071)

data<-read.table("D:\\svmdata3.txt",stringsAsFactors = TRUE)

num<-nrow(data)

num

nt<-as.integer(num\*0.8)

nt

rdata<-data[order(runif(num)),]

rdata

train=rdata[1:nt, ]

train

test=rdata[(nt+1):100, ]

test

features<-data.frame(X1=test$X1,X2=test$X2)

degr<-c(1,5,10,15,25,50,2)

for (d in degr) {

print(d)

svm <- svm(factor(Colors)~.,data=train,type = "C-classification",cost = 1,kernel = "polynomial", degree=d)

pr <- predict(svm, features)

res<-table(test$Colors, pr)

print(res)

acc<-(res[1,1]+res[2,2])/sum(res)

print(acc)

summary(svm)

plot(svm,test,grid = 250,col=c("green","pink"), dataSymbol="+", svSymbol="V")

}

svm <- svm(factor(Colors)~.,data=train,type = "C-classification",cost = 1,kernel = "radial")

pr <- predict(svm, features)

res<-table(test$Colors, pr)

print(res)

acc<-(res[1,1]+res[2,2])/sum(res)

acc

summary(svm)

plot(svm,test,grid = 250,col=c("green","pink"), dataSymbol="+", svSymbol="V")

svm <- svm(factor(Colors)~.,data=train,type = "C-classification",cost = 1,kernel = "sigmoid")

pr <- predict(svm, features)

res<-table(test$Colors, pr)

print(res)

acc<-(res[1,1]+res[2,2])/sum(res)

acc

summary(svm)

plot(svm,test,grid = 250,col=c("green","pink"), dataSymbol="+", svSymbol="V")

## **Пункт 4**

Были построены аналогичные модели задания 3 для датасета svmdata4.txt и svmdata4test.txt, результаты представлены на рисунках 19-21.

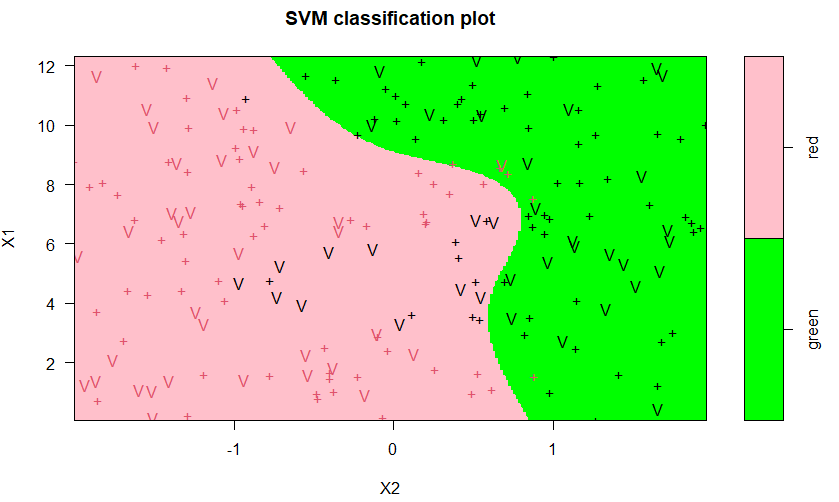


Рисунок 19 – SVM-классификация с полиномиальным ядром

Точность равна 0,87.

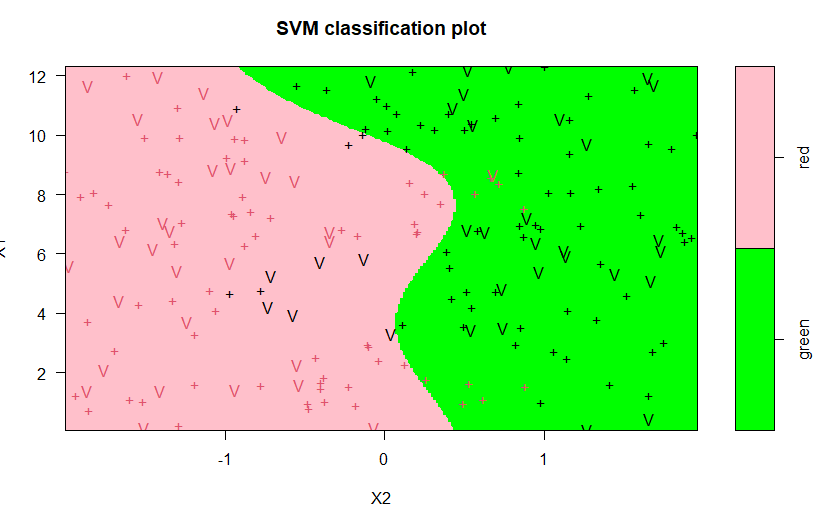


Рисунок 20 - SVM-классификация с радиальным ядром

Точность равна 0,89.

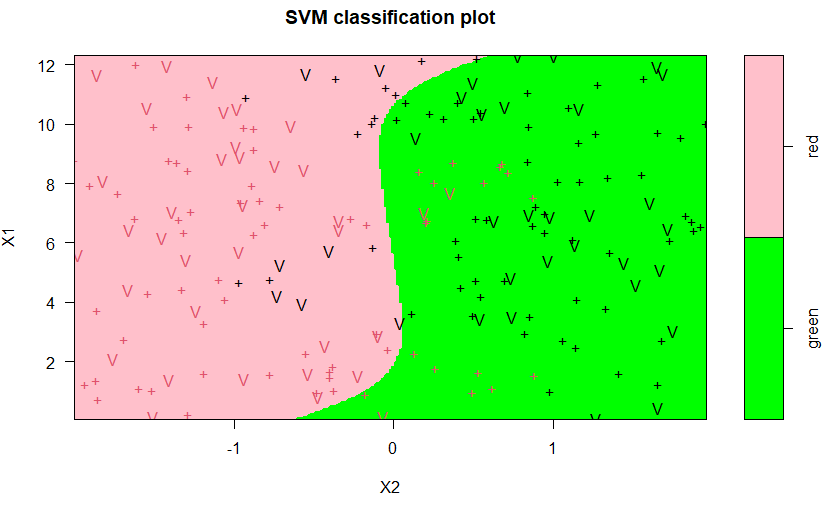


Рисунок 21 - SVM-классификация с сигмоидальным ядром

Точность равна 0,805.

Таким образом, наибольшая точность классификации у модели с радиальным ядром.

Код:

library(e1071)

train<-read.table("D:\\svmdata4.txt",stringsAsFactors = TRUE)

test<-read.table("D:\\svmdata4test.txt",stringsAsFactors = TRUE)

features<-data.frame(X1=test$X1,X2=test$X2)

svm <- svm(factor(Colors)~.,data=train,type = "C-classification",cost = 1,kernel = "polynomial")

pr <- predict(svm, features)

res<-table(test$Colors, pr)

print(res)

acc<-(res[1,1]+res[2,2])/sum(res)

acc

summary(svm)

plot(svm,test,grid = 250,col=c("green","pink"), dataSymbol="+", svSymbol="V")

svm <- svm(factor(Colors)~.,data=train,type = "C-classification",cost = 1,kernel = "radial")

pr <- predict(svm, features)

res<-table(test$Colors, pr)

print(res)

acc<-(res[1,1]+res[2,2])/sum(res)

acc

summary(svm)

plot(svm,test,grid = 250,col=c("green","pink"), dataSymbol="+", svSymbol="V")

svm <- svm(factor(Colors)~.,data=train,type = "C-classification",cost = 1,kernel = "sigmoid")

pr <- predict(svm, features)

res<-table(test$Colors, pr)

print(res)

acc<-(res[1,1]+res[2,2])/sum(res)

acc

summary(svm)

plot(svm,test,grid = 250,col=c("green","pink"), dataSymbol="+", svSymbol="V")

## **Пункт 5**

Были построены аналогичные модели задания 3 для датасета svmdata5.txt и svmdata5test.txt, изменялся параметр gamma (1 или 50), результаты представлены на рисунках 22-27.

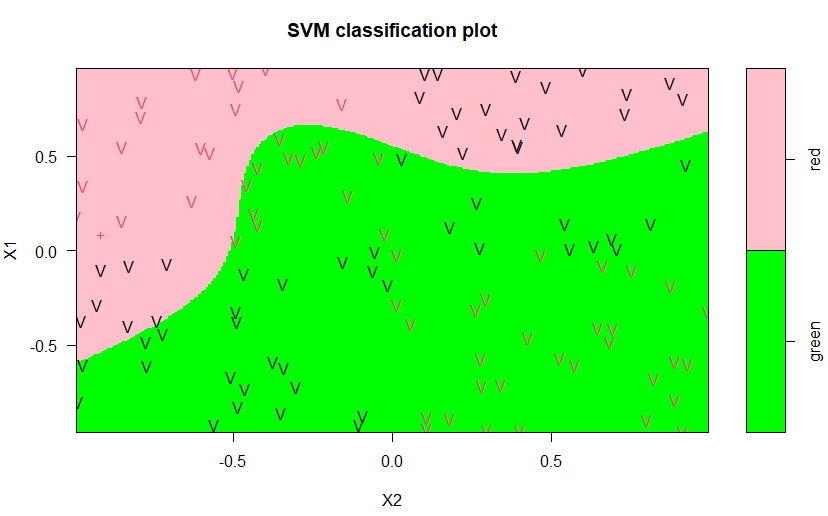


Рисунок 28 - SVM-классификация с полиномиальным ядром, gamma = 1

Точность классификации – 0,425.

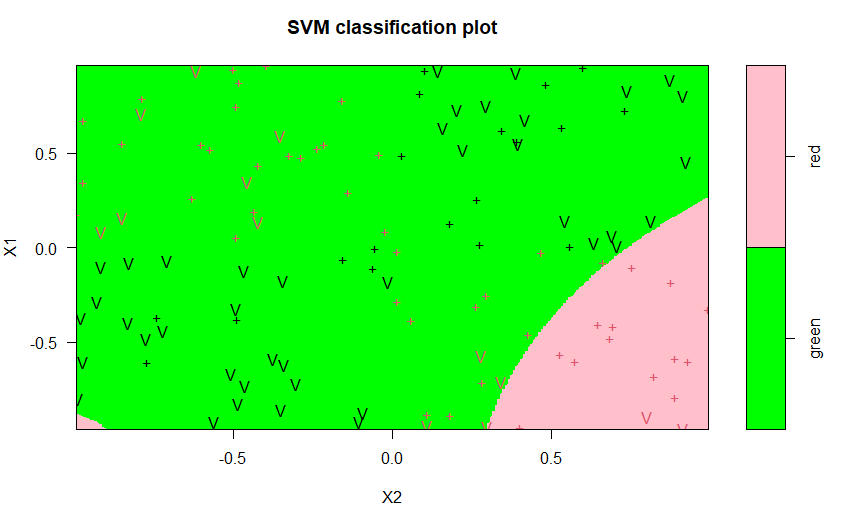


Рисунок 29 - SVM-классификация с полиномиальным ядром, gamma = 50

Точность классификации – 0,633.

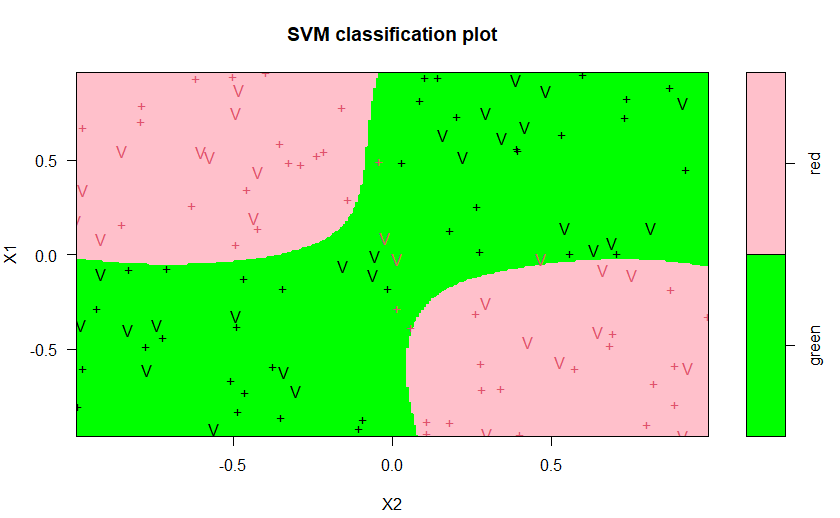


Рисунок 30 - SVM-классификация с радиальным ядром, gamma = 1

Точность классификации – 0,958.

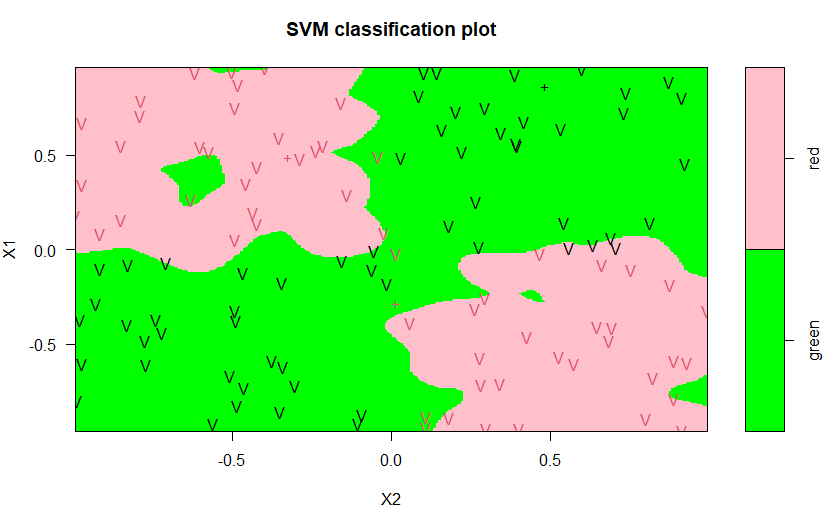


Рисунок 31 - SVM-классификация с радиальным ядром, gamma = 50

Точность классификации – 0,908.

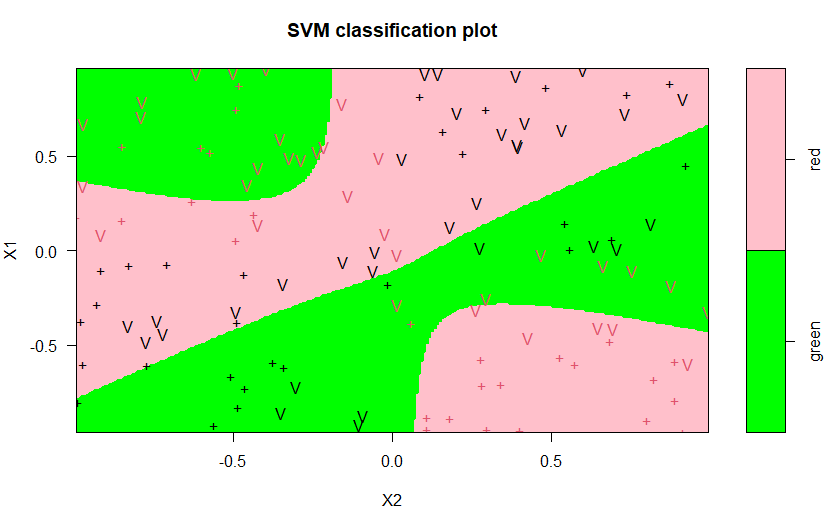


Рисунок 32 - SVM-классификация с сигмоидальным ядром, gamma = 1

Точность классификации – 0,458.

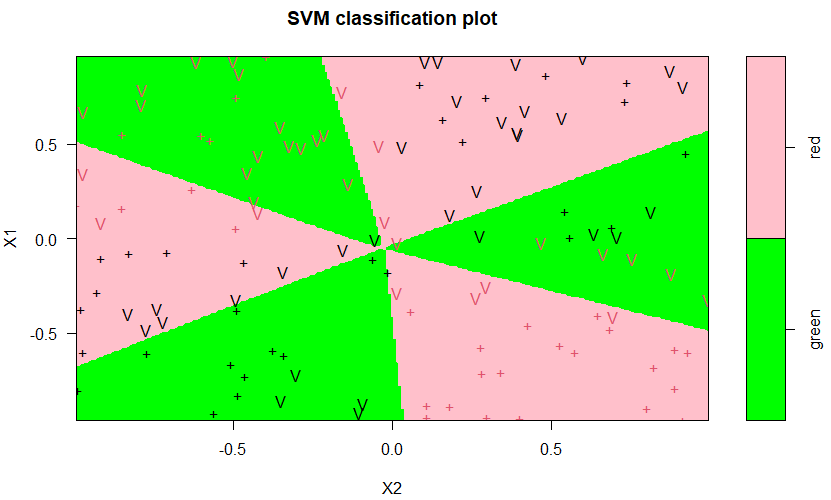


Рисунок 33 - SVM-классификация с сигмоидальным ядром, gamma = 50

Точность классификации – 0,483.

Из полученных результатов видно, что наилучшее качество классификации показывает радиальное ядро при параметре gamma = 1. Эффект переобучения для радиального ядра виден при gamma = 50.

Код:

library(e1071)

train<-read.table("D:\\svmdata5.txt",stringsAsFactors = TRUE)

test<-read.table("D:\\svmdata5test.txt",stringsAsFactors = TRUE)

features<-data.frame(X1=test$X1,X2=test$X2)

gamm<-c(1,50)

for (g in gamm) {

print(g)

svm <- svm(factor(Colors)~.,data=train,type = "C-classification",cost = 1,kernel = "polynomial", gamma=g)

pr <- predict(svm, features)

res<-table(test$Colors, pr)

print(res)

acc<-(res[1,1]+res[2,2])/sum(res)

print(acc)

summary(svm)

plot(svm,test,grid = 250,col=c("green","pink"), dataSymbol="+", svSymbol="V")

}

gamm<-c(1,50)

for (g in gamm) {

print(g)

svm <- svm(factor(Colors)~.,data=train,type = "C-classification",cost = 1,kernel = "radial", gamma=g)

pr <- predict(svm, features)

res<-table(test$Colors, pr)

print(res)

acc<-(res[1,1]+res[2,2])/sum(res)

print(acc)

summary(svm)

plot(svm,test,grid = 250,col=c("green","pink"), dataSymbol="+", svSymbol="V")

}

gamm<-c(1,50)

for (g in gamm) {

print(g)

svm <- svm(factor(Colors)~.,data=train,type = "C-classification",cost = 1,kernel = "sigmoid", gamma=g)

pr <- predict(svm, features)

res<-table(test$Colors, pr)

print(res)

acc<-(res[1,1]+res[2,2])/sum(res)

print(acc)

summary(svm)

plot(svm,test,grid = 250,col=c("green","pink"), dataSymbol="+", svSymbol="V")

}

## **Пункт 6**

Были построены модели для восстановления регрессии для датасета svmdata6.txt, результаты представлены на рисунках 34-38.

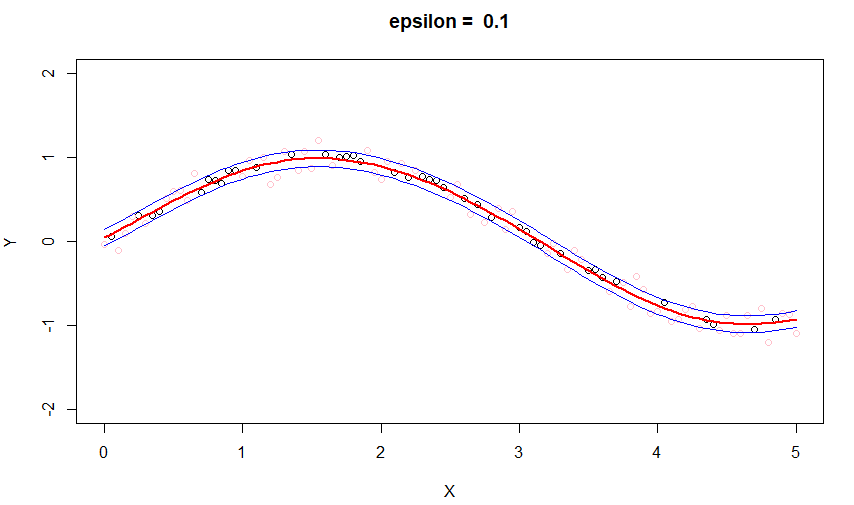


Рисунок 34 - SVM-регрессия с радиальным ядром, epsilon=0.1

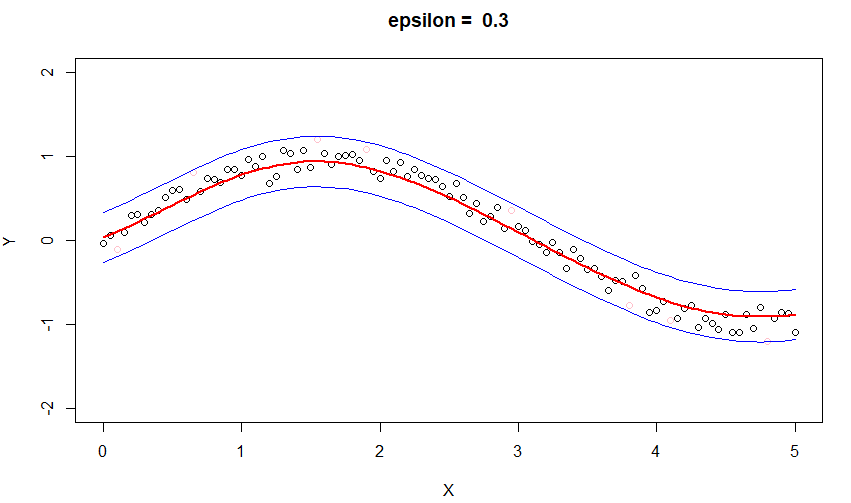


Рисунок 35 - SVM-регрессия с радиальным ядром, epsilon=0.3

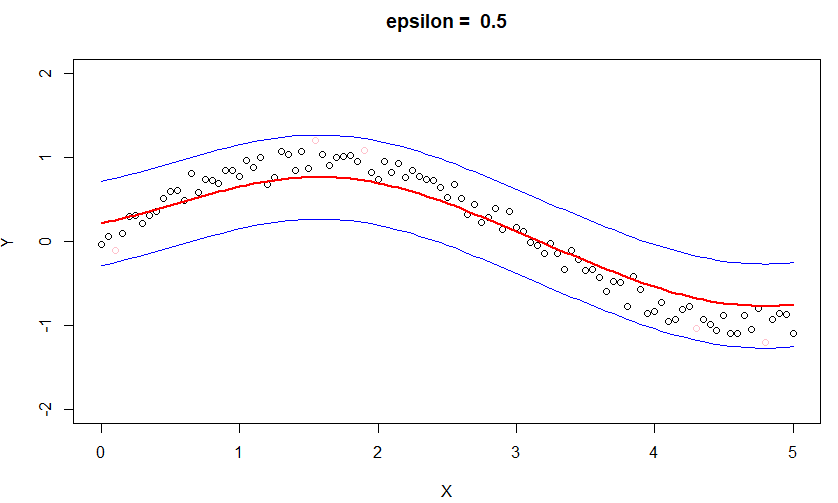


Рисунок 36 - SVM-регрессия с радиальным ядром, epsilon=0.5

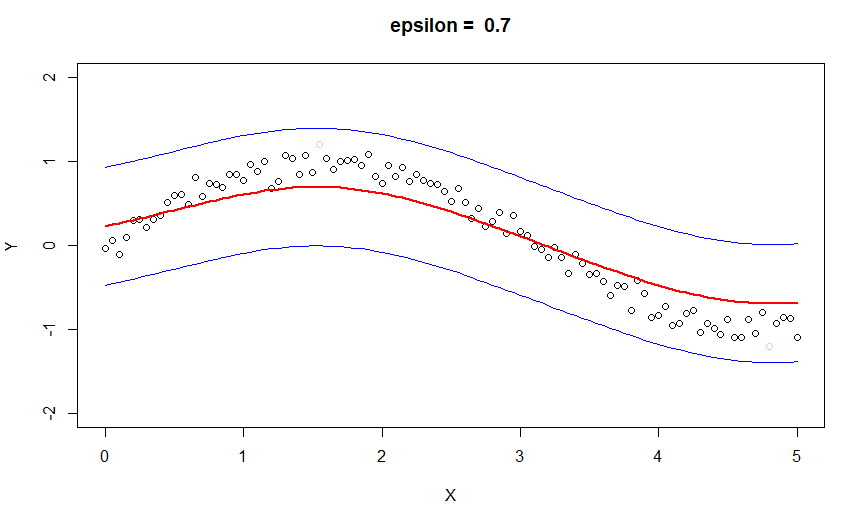


Рисунок 37 - SVM-регрессия с радиальным ядром, epsilon=0.7

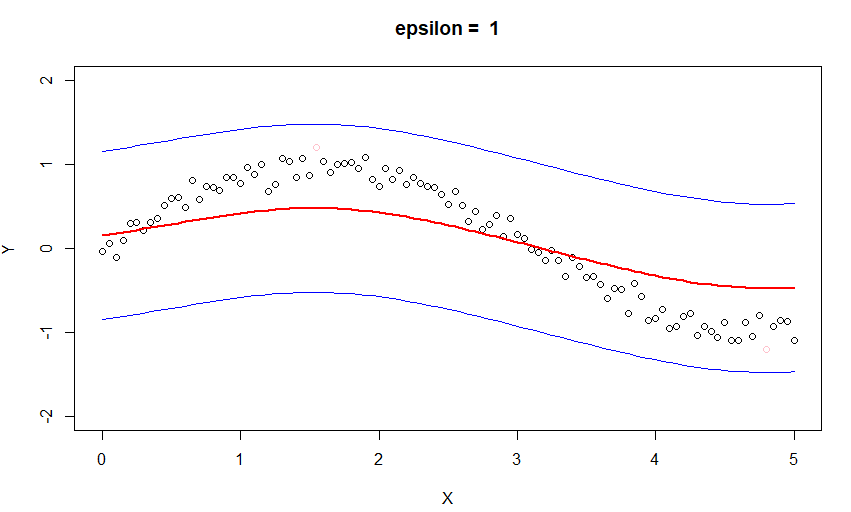


Рисунок 38 - SVM-регрессия с радиальным ядром, epsilon=1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Epsilon** | 0.1 | 0.3 | 0.5 | 0.7 | 1 |
| **MSE** | 0.01064825 | 0.01542817 | 0.04302801 | 0.06486077 | 0.15815918 |

Таблица 3. Результаты SVM-регрессии с радиальным ядром

Согласно графику зависимости MSE от значения параметра функции потерь можно сказать, что MSE экспоненциально возрастает при увеличении epsilon.

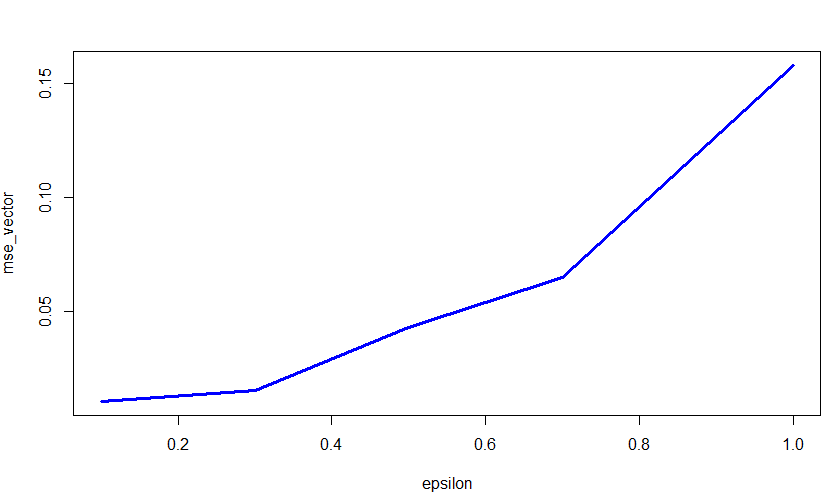


Рисунок 39 – Зависимость MSE от значения epsilon

Код:

library(Metrics)

library(e1071)

train<-read.table("D:\\svmdata6.txt",stringsAsFactors = TRUE)

features<-data.frame(X=train$X)

train

features

mse\_vector<-vector()

epsilon<-c(0.1, 0.3, 0.5, 0.7, 1)

for (e in epsilon) {

print(e)

svm <- svm(train$X, train$Y,type = "eps-regression",cost = 1,kernel = "radial", eps = e)

pr <- predict(svm, features)

res<-table(train$Y, pr)

mse<-mse(pr,train$Y)

mse\_vector<-append(mse\_vector, mse)

print(mse\_vector)

acc<-(res[1,1]+res[2,2])/sum(res)

print(acc)

summary(svm)

x<-train$X

y<-train$Y

plot(x, y, xlab="X", ylab="Y", ylim=c(-2,2), main = paste("epsilon = ", e))

points(x[svm$index], y[svm$index], col = "pink")

lines(x, pr, col = "red", lwd = 2)

lines(x, pr + svm$epsilon, col = "blue")

lines(x, pr - svm$epsilon, col = "blue")

}

mse\_vector

plot(y=mse\_vector, x=epsilon, col="blue", lwd=3, type="l")

## **Вывод**

При выполнении данной лабораторной работы был изучен метод опорных векторов, реализованный в пакете е1071 языка R. Были созданы модели SVM, которые хорошо справились с задачами классификации и восстановления регрессии. Для того, чтобы точность классификации была высока, необходимо подбирать параметры - ядро, значения степени полинома и штрафного параметра.