

Учреждение образования

«Белорусский государственный университет
информатики и радиоэлектроники»

Кафедра информационных технологий автоматизированных систем

Экономико-математические методы в системном анализе

Отчёт по лабораторной №1
«ВВЕДЕНИЕ В R»

Проверил:

Гуринович А.Б.

Выполнили ст. гр. 025941:

Колесников В.Г.
Денисик П.Д.

Минск 2021

Цель работы:

Изучение и выработка навыков работы с R.

Задания:

1. Сгенерируйте вектор длины $N = 1000$, элементами которого являются реализации нормально распределенной случайной величины с математическим ожиданием равным 1, и стандартным отклонением 0.3. Подсчитайте статистические оценки мат. ожидания, стандартного отклонения, квантилей уровней 0.95 и 0.99 с и без использования встроенных функций. Сравните результат. Исследуйте отклонение статистического мат. ожидания от 1 при росте N ($N = 1000, 2000, 4000, 8000, 16000, 32000$).
2. Создайте фрейм данных из $N = 20$ записей со следующими полями: `Nrow` – номер записи, `Name` – имя сотрудника, `BirthYear` – год рождения, `EmploYear` – год приема на работу, `Salary` – зарплата. Заполните данный фрейм данными так, что `Nrow` изменяется от 1 до N , `Name` задается произвольно, `BithYear` распределен равномерно (случайно) на отрезке $[1960, 1985]$, `EmployYear` распределен равномерно на отрезке $[\text{BirthYear} + 18, 2006]$, `Salary` для работников младше 1975 г.р. определяется по формуле $\text{Salary} = (\ln(2007 - \text{EmployYear}) + 1) * 8000$, для остальных $\text{Salary} = (\log_2(2007 - \text{EmployYear}) + 1) * 8000$. Подсчитайте число сотрудников с зарплатой, большей 15000. Добавьте в таблицу поле, соответствующее суммарному подоходному налогу (ставка 13%), выплаченному сотрудником за время работы в организации, если его зарплата за каждый год начислялась согласно формулам для `Salary`, где вместо 2007 следует последовательно подставить каждый год работы сотрудника в организации.
3. Напишите функцию, которая принимает на вход числовой вектор x и число разбиений интервала k (по умолчанию равное числу элементов вектора, разделенному на 10) и выполняет следующее: находит минимальное и максимальное значение элементов вектора x , разделяет полученный отрезок $[x_{\min}; x_{\max}]$ на k равных интервалов и подсчитывает число элементов вектора, принадлежащих каждому интервалу. Постройте график, где по оси абсцисс откладываются середины интервалов, по оси ординат – число элементов вектора, принадлежащих интервалу, разделенное на общее число точек. Проведите эксперимент на данной функции, где x – вектор длины 5000, выборка реализаций экспоненциально распределенной случайной величины с параметром $\lambda =$

500. Приближение какого графика мы получаем в итоге при большом числе точек и числе разбиений?

4. Пусть дан набор точек в $x_i \in \mathbb{R}^d$, $i = 1, 2, \dots, n$, каждой из которых поставлено в соответствие некоторое вещественное число y . Реализуйте метод наименьших квадратов, находящий гиперплоскость, обеспечивающую наименьшую невязку по y . Для случая $d = 1$ нарисуйте график, содержащий исходные точки и найденную прямую.

Ход работы:

Задание 1. Листинг кода (В папке lab1/1.R)

```
> n = 1000
> s = rnorm(n, mean = 1, sd = 0.3)
> c(sum(s) / n, mean(s))
[1] 0.9910271 0.9910271
> p = 0
> for(i in seq(n))
+ {
+   p = p + (s[i] - mean(s))^2 / (n - 1)
+ }
> c(sqrt(p), sd(s))
[1] 0.2945941 0.2945941
> c(sort(s)[round(n * 0.95 + 1)], quantile(s, probs = seq(0, 1, 0.95))["95%"])
      95%
1.463989 1.462310
> c(sort(s)[round(n * 0.99 + 1)], quantile(s, probs = seq(0, 1, 0.99))["99%"])
      99%
1.661720 1.659084
```

Рисунок 1.1

Задание 2. Листинг кода (В папке lab1/2.R)

```
Employees with salary > 15000: 20
```

Рисунок 1.2

```

> newDF <- data.frame(
+ names=names,
+ birthYears=birthYears,
+ employYears=employYears,
+ salaries=salaries,
+ taxes=taxes
+ )
> print(newDF)

```

	names	birthYears	employYears	salaries	taxes
1	Magomed1	1974	2005	40000.00	88400.00
2	Magomed2	1971	1991	47255.12	190438.15
3	Magomed3	1960	1986	49034.26	229480.36
4	Magomed4	1980	2000	32356.18	92538.67
5	Magomed5	1983	2006	29664.40	61701.96
6	Magomed6	1972	2000	43138.54	123376.22
7	Magomed7	1985	2004	30665.71	71757.75
8	Magomed8	1961	1991	47255.12	190438.15
9	Magomed9	1968	1990	47633.57	198155.65
10	Magomed10	1975	1999	43675.45	130589.60
11	Magomed11	1964	1998	44188.50	137868.11
12	Magomed12	1963	2006	39255.12	81650.66
13	Magomed13	1967	1985	49359.40	237418.71
14	Magomed14	1965	1994	46039.10	167582.32
15	Magomed15	1966	2000	43138.54	123376.22
16	Magomed16	1984	2004	30665.71	71757.75
17	Magomed17	1970	1997	44679.70	145209.03
18	Magomed18	1982	2000	32356.18	92538.67
19	Magomed19	1962	1987	48699.70	221583.65
20	Magomed20	1973	2002	41983.42	109156.89

Рисунок 1.3

Задание 3. Листинг кода (В папке lab1/3.R)

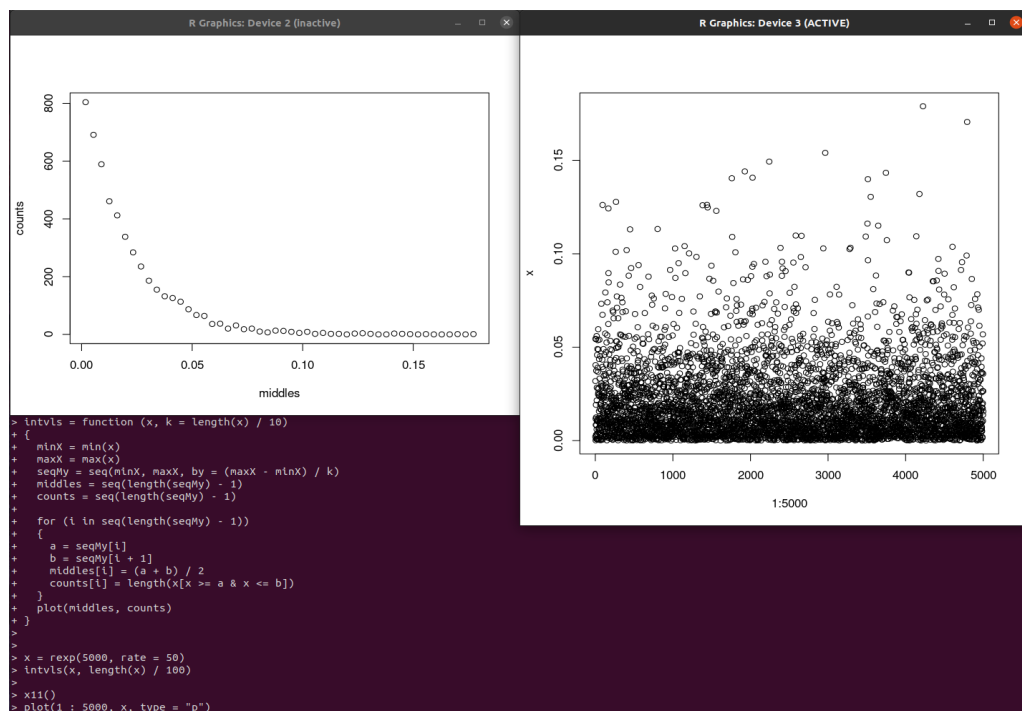


Рисунок 1.4

Задание 4. Листинг кода (В папке lab1/4.R)

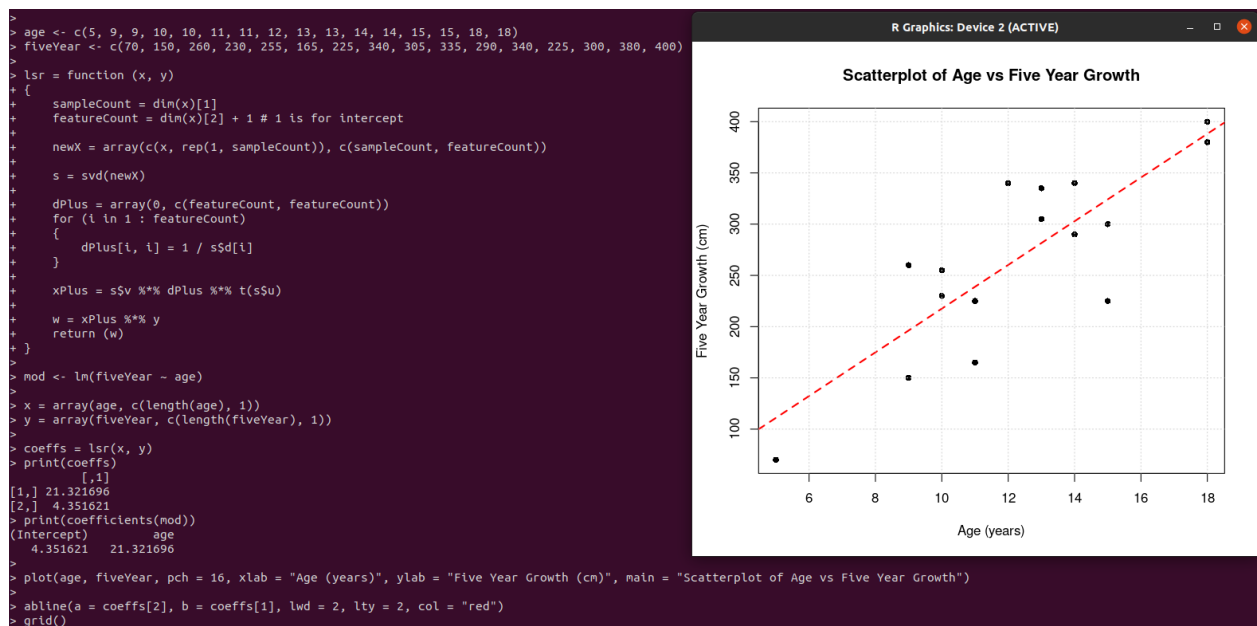


Рисунок 1.5

Вывод:

В ходе выполненной работы был сгенерирован вектор длины $N = 1000, 2000$ выполнены статистические оценки мат. ожидания, стандартного отклонения, квантилей уровней 0.95 и 0.99 с и без использования встроенных функций, создан фрейм данных из $N = 20$ записей с полями Nrow, Name, BirthYear, EmplYear, Salary, фрейм заполнен данными согласно заданию, написана функция, которая принимает на вход числовой вектор x и число разбиений интервала k и находит минимальное и максимальное значение элементов вектора x , разделяет полученный отрезок $[x_{\min}; x_{\max}]$ на k равных интервалов и подсчитывает число элементов вектора, принадлежащих каждому интервалу, реализован метод наименьших квадратов, находящий гиперплоскость, обеспечивающую наименьшую невязку по y .

Учреждение образования

«Белорусский государственный университет
информатики и радиоэлектроники»

Кафедра информационных технологий автоматизированных систем

Экономико-математические методы в системном анализе

Отчёт по лабораторной №2
«ПРОВЕРКА СТАТИЧЕСКИХ ГИПОТЕЗ»

Проверил:

Гуринович А.Б.

Выполнили ст. гр. 025941:

Колесников В.Г.
Денисик П.Д.

Минск 2021

Цель работы:

Изучение и выработка навыков работы с машиной опорных векторов.

Задания:

1. Используя тест Шапиро-Уилка, проверьте, являются ли нормально распределенными характеристики цветов ириса (фрейм данных iris). Уровень значимости $\alpha = 0.05$).
2. Для $k = 10, 15, 20, 25, 30$ сгенерируйте 200 реализаций нормальной распределенной случайной величины с мат. ожиданием, равным k , и стандартным отклонением, равным \sqrt{k} , и 200 реализаций случайной величины, распределенной по закону χ^2 с k степенями свободы. Используя тест Колмогорова-Смирнова, проверьте гипотезу о том, что данные выборки относятся к одному непрерывному распределению. Уровень значимости $\alpha = 0.05$.
3. Загрузите таблицу из файла allcountries.txt, содержащую информацию о населении, площади и ряде других характеристик современных государств. Выберите из таблицы те страны, для которых доступна информация о населении и площади (нет отсутствующих значений NA) и площадь больше 10. Пусть $\text{area_log} = \log_{10}(\log_{10}(\text{area}))$, $\text{population_log} = \log_{10}(\log_{10}(\text{population}))$. Методом наименьших квадратов постройте функцию $f(\cdot)$, моделирующую зависимость population_log от area_log с помощью линейной функции $\text{population_log} = f(\text{area_log}) = \beta_0 + \beta_1 \text{area_log}$, т.е. подберите коэффициенты β_0 и β_1 . Используя тест Колмогорова-Смирнова, проверьте гипотезу о том, что population_log и $f(\text{area_log})$ относятся к одному непрерывному распределению. Уровень значимости $\alpha = 0.05$.
4. Используя критерий χ^2 проверьте гипотезу, состоящую в том, что цвет глаз женщин не зависит от цвета волос (на фрейме данных HairEyeColor).
5. Загрузите таблицу из файла readingspeed.txt, которая содержит информацию о скорости чтения у детей в зависимости от применяемой методики обучения (DRA – direct reading activities, SC – standart curriculum). Используя t-тест, проверьте гипотезу о том, что среднее время чтения для обеих методик совпадает (используйте разные альтернативные гипотезы). Объясните полученные результаты.

Ход работы:

Задание 1. Листинг кода (lab2/1.R)

```
> data = iris[, "Petal.Length"]
> set.seed(0)
> ts = shapiro.test(data)
> p_value = ts["p_value"]
> print(ts)

      Shapiro-Wilk normality test

data:  data
W = 0.87627, p-value = 7.412e-10
```

Рисунок 2.1

Задание 2. Листинг кода (lab2/2.R)

```
> kss = seq(10, 30, 5)
> n = 200
> alph = 0.05
> norm = c()
> chiss = c()
> for (k in kss)
+ {
+   ns = rnorm(n, mean = k, sd = sqrt(k))
+   chis = rchisq(n, k)
+   norm = c(norm, ns)
+   chiss = c(chiss, chis)
+   tst = ks.test(ns, chis)
+   print(c(k, tst["p.value"] > alph))
+ }

      p.value
10         1
      p.value
15         0
      p.value
20         0
      p.value
25         1
      p.value
30         1
```

Рисунок 2.2

Задание 3. Листинг кода (lab2/3.R)

```
> ts = read.table(file.choose(), header = TRUE)
Enter file name: ./allcountries.txt
> ts = na.omit(ts)
> ts["area_log"] = log10(log10(ts$area))
> ts["population_log"] = log10(log10(ts$population))
> b0 = 0.2
> b1 = 0.9
> func = approx(ts$area, b0 + b1 * (ts$area_log), ts$area, method = "constant")$y
Warning message:
In regularize.values(x, y, ties, missing(ties)) :
  collapsing to unique 'x' values
> ks.test(ts$population_log, func)

      Two-sample Kolmogorov-Smirnov test

data:  ts$population_log and func
D = 0.11173, p-value = 0.2138
alternative hypothesis: two-sided
```

Рисунок 2.3

Задание 4. Листинг кода (lab2/4.R)

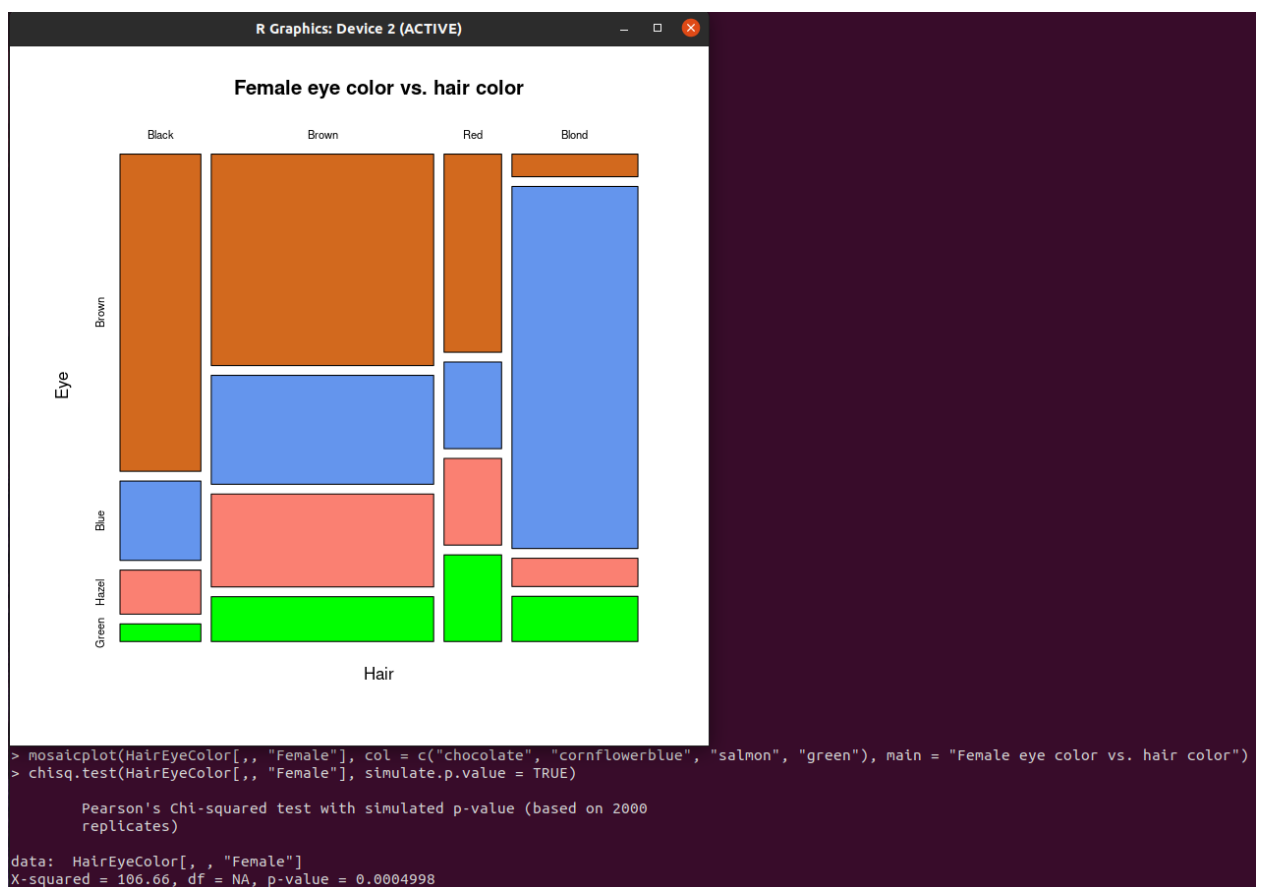


Рисунок 2.4

Задание 5. Листинг кода (lab2/5.R)

```
> t = read.table(file.choose(), header = TRUE)
Enter file name: ./readingspeed.txt
> group1 = t[t$LearningType == "DRA",]
> group1 = group1$Score
>
> group2 = t[t$LearningType == "SC",]
> group2 = group2$Score
>
> t.test(group1, group2)

        Welch Two Sample t-test

data:  group1 and group2
t = 2.3109, df = 37.855, p-value = 0.02638
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.23302 18.67588
sample estimates:
mean of x mean of y
 51.47619  41.52174
```

Рисунок 2.5

Выводы:

Проверено, являются ли нормально распределенными характеристики цвета зрачка, сгенерированно 200 реализаций нормальной распределенной случайной величины, проверены гипотезы: о том, что данные выборки относятся к одному непрерывному распределению, что `population_log` и `f(area_log)` относятся к одному непрерывному распределению, что цвет глаз женщин не зависит от цвета волос (на фрейме данных `HairEyeColor`), и что среднее время чтения для обеих методик совпадает.

Учреждение образования

«Белорусский государственный университет
информатики и радиоэлектроники»

Кафедра информационных технологий автоматизированных систем

Экономико-математические методы в системном анализе

Отчёт по лабораторной №3
«ЛИНЕЙНАЯ РЕГРЕССИЯ»

Проверил:

Гуринович А.Б.

Выполнили ст. гр. 025941:

Колесников В.Г.
Денисик П.Д.

Минск 2021

Цель работы:

Изучение и выработка навыков работы с линейной регрессией в R.

Задания:

1. Загрузите данные из файла reglab1.txt. Используя функцию lm, постройте регрессию (используйте разные модели). Выберите наиболее подходящую модель.
2. Реализуйте следующий алгоритм для уменьшения количества признаков, используемых для построения регрессии: для каждого $k \in 0, 1, \dots, d$ выбрать подмножество признаков мощности $k + 1$, минимизирующее остаточную сумму квадратов RSS. Используя полученный алгоритм, выберите оптимальное подмножество признаков для данных из файла reglab2.txt. Объясните свой выбор. Дайте интерпретацию вычисленным значениям t-статистики и p-value для коэффициентов β .
3. Загрузите данные из файла cugage.txt. Постройте регрессию, выражающую зависимость возраста исследуемых отложений от глубины залегания, используя веса наблюдений. Оцените качество построенной модели.
4. Загрузите данные из файла alligators.txt. Выберите лучшую регрессионную модель (возможно нелинейную), отражающую зависимость веса аллигатора от его длины.
5. Исключите из набора данных longley переменную "Population". Разделите данные на тестовую и обучающую выборки равных размеров случайным образом. Постройте гребневую регрессию для значений $\lambda = 10 - 3 + 0.2 \cdot i$, $i = 0, 25$, подсчитайте ошибку на тестовой и обучающей выборке для данных значений λ , постройте графики. Объясните полученные результаты.

Ход работы:

Задание 1. Листинг кода задания (lab3/1.R)

```
> d = read.table("./reglab1.txt", header = TRUE)
> f = lm(z ~ ., data = d)
> summary(f)

Call:
lm(formula = z ~ ., data = d)

Residuals:
    Min       1Q   Median       3Q      Max
-0.97246 -0.16759  0.01308  0.20537  0.81127

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.02163     0.06384  -0.339   0.735
x             4.10248     0.08698  47.168 <2e-16 ***
y             4.94308     0.08035  61.517 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3376 on 197 degrees of freedom
Multiple R-squared:  0.9686,    Adjusted R-squared:  0.9683
F-statistic: 3041 on 2 and 197 DF,  p-value: < 2.2e-16
```

Рисунок 3.1

```
> f = lm(z ~ x + y, data = d)
> summary(f)

Call:
lm(formula = z ~ x + y, data = d)

Residuals:
    Min       1Q   Median       3Q      Max
-0.97246 -0.16759  0.01308  0.20537  0.81127

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.02163     0.06384  -0.339   0.735
x             4.10248     0.08698  47.168 <2e-16 ***
y             4.94308     0.08035  61.517 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3376 on 197 degrees of freedom
Multiple R-squared:  0.9686,    Adjusted R-squared:  0.9683
F-statistic: 3041 on 2 and 197 DF,  p-value: < 2.2e-16
```

Рисунок 3.2

```

> f = lm(z ~ x * y, data = d)
> summary(f)

Call:
lm(formula = z ~ x * y, data = d)

Residuals:
    Min       1Q   Median       3Q      Max
-0.086847 -0.022270  0.001217  0.020802  0.092255

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.003994    0.009804   102.4  <2e-16 ***
x             1.975466    0.017672   111.8  <2e-16 ***
y             3.001298    0.016180   185.5  <2e-16 ***
x:y           4.020776    0.029009   138.6  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03402 on 196 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 2.061e+05 on 3 and 196 DF,  p-value: < 2.2e-16

```

Рисунок 3.3

Задание 2. Листинг кода задания (lab3/2.R)

```

> t = read.table("./reglab2.txt", header = TRUE)
> g = colnames(t)[2 : length(colnames(t))]
>
> for(i in seq(length(g)))
+ {
+   combs = combn(g, i)
+   ii = dim(combs)[1]
+   jj = dim(combs)[2]
+   for(j in seq(jj))
+   {
+     names = c()
+     for(i in seq(ii)){
+       names = c(names, combs[i, j])
+     }
+     names = c("y", names)
+     it = t[names]
+     cat(c(names, ";", lm(y ~ ., data = it)$coefficients, "\n"))
+   }
+ }
y x1 ; 1.66792978553119 3.75485986294696
y x2 ; 2.28031373380239 2.69752481088658
y x3 ; 3.45371578163341 0.26897014607118
y x4 ; 3.55117440502059 0.0732954794623247
y x1 x2 ; 0.0842462458504668 3.99970178676916 3.01262259481575
y x1 x3 ; 1.55575779533087 3.75253132179838 0.229893903452898
y x1 x4 ; 1.66526249582932 3.75476394234834 0.00566057897951267
y x2 x3 ; 2.20140180316801 2.69153511421278 0.165911948637967
y x2 x4 ; 2.19779377997981 2.70277535093225 0.166668430802927
y x3 x4 ; 3.43441498327733 0.264311513747633 0.0450087481232859
y x1 x2 x3 ; 0.0317180166733256 3.99822745252188 3.00845792302292 0.112142303929776
y x1 x2 x4 ; 0.0328678956647152 3.99817490600051 3.01582472555088 0.105463883331735
y x1 x3 x4 ; 1.5638029266332 3.75283529597068 0.23186970009174 -0.0191194746948446
y x2 x3 x4 ; 2.13443951635192 2.69684231850616 0.150142076536048 0.150395192656927
y x1 x2 x3 x4 ; -0.00964100302917558 3.99699068850118 3.01169105847903 0.102260673660886 0.0943984276221464
>
> sm = sum(f$residuals)

```

Рисунок 3.4

Задание 3. Листинг кода задания (lab3/3.R)

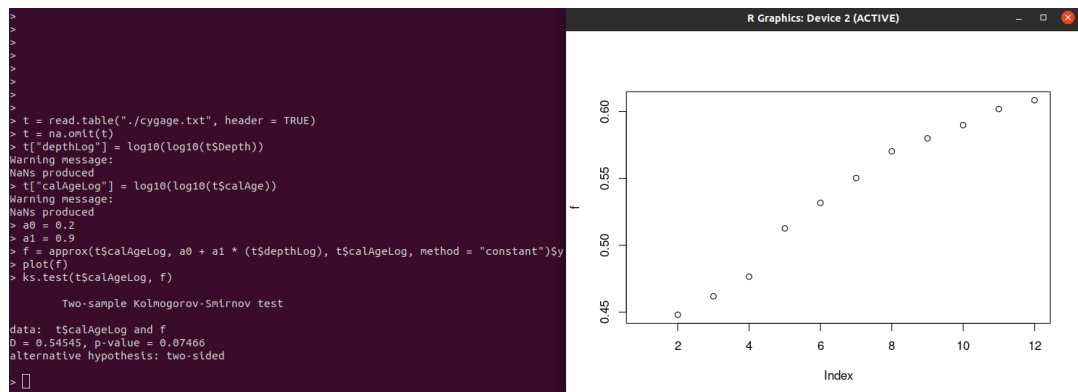


Рисунок 3.5

Задание 4. Листинг кода задания (lab3/4.R)

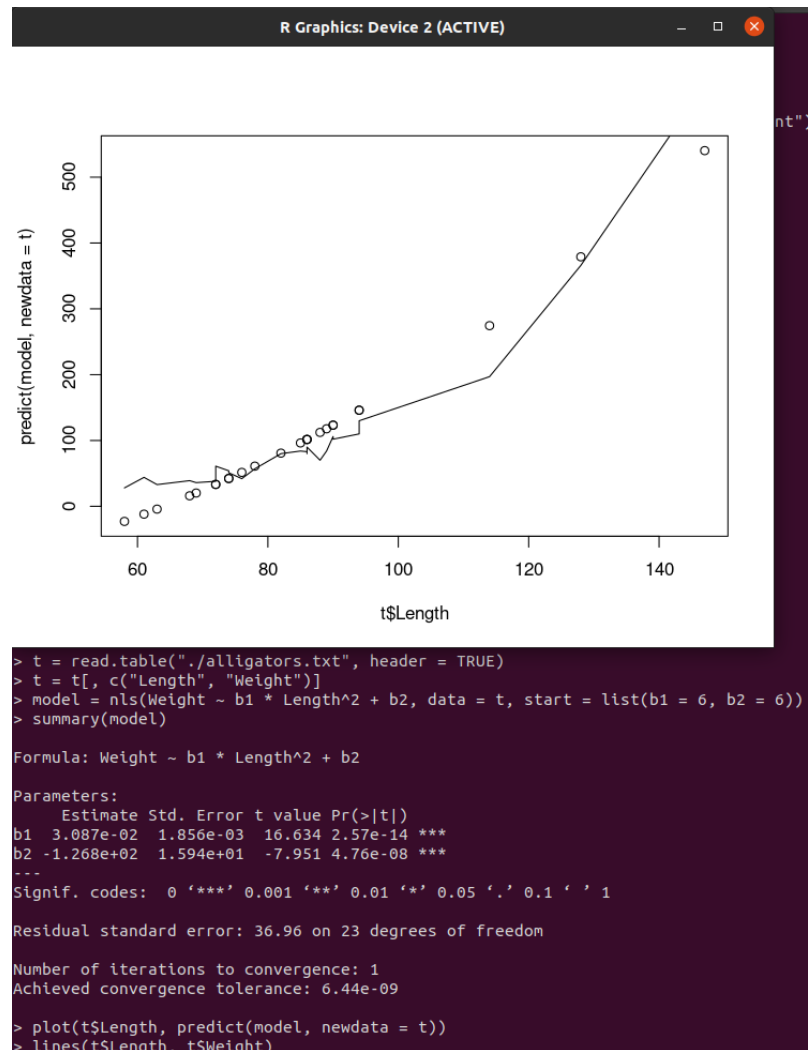


Рисунок 3.6

Задание 5. Листинг кода задания (lab3/5.R)

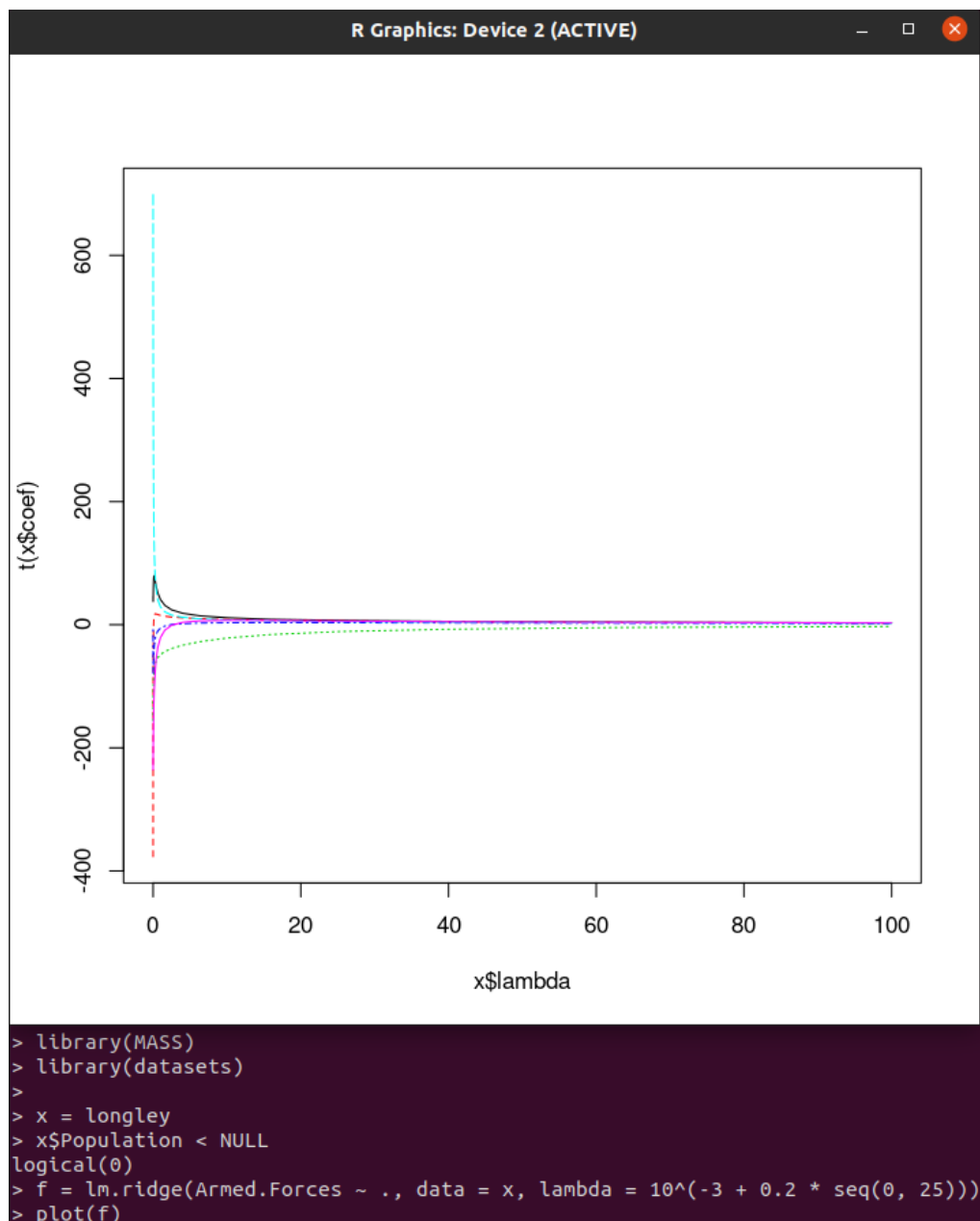


Рисунок 3.7

Вывод:

В ходе выполненной работы была построена регрессия, реализован алгоритм для уменьшения количества признаков, используемых для построения регрессии, построена регрессия, выражающая зависимость возраста исследуемых отложений от глубины залегания, используя веса наблюдений; построена гребневая регрессия.

Учреждение образования

«Белорусский государственный университет
информатики и радиоэлектроники»

Кафедра информационных технологий автоматизированных систем

Экономико-математические методы в системном анализе

Отчёт по лабораторной №4
«МАШИНА ОПОРНЫХ ВЕКТОРОВ»

Проверил:

Гуринович А.Б.

Выполнил ст. гр. 025941:

Колесников В.Г.
Денисик П.Д.

Минск 2021

Цель работы:

Изучение и выработка навыков работы с машиной опорных векторов.

Задания:

1. Постройте машину 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 машину 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 машину 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:

Задание 1. Листинг код построения модели (lab4/1.R)

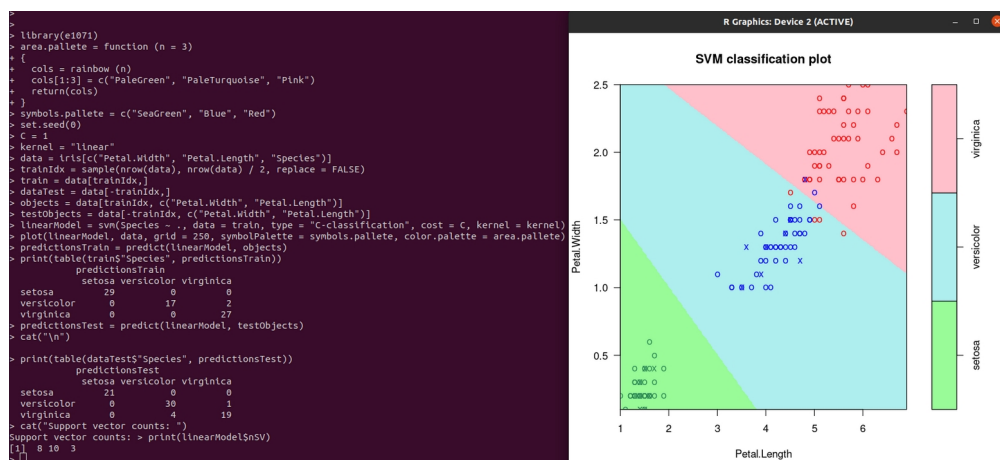


Рисунок 4.1 - Разбиение пространства признаков на области

Задание 2. Нулевой ошибки для тренировочного набора данных получилось достичь при $C = 225$. Однако при таком значении параметра C модель сделала 5 ошибок на тестовой выборке. Минимальной ошибки на тестовом наборе данных получилось достичь при $C = 0.1$. При данном значении параметра модель сделала 2 ошибки на тренировочном наборе данных и 2 ошибки на тестовом. Параметр $C = 0.1$ является оптимальным, т. к. при нем достигается наименьшая ошибка на тестовой выборке, соответственно достигнута наилучшая обобщающая способность модели.

Листинг программы (lab4/2.R)

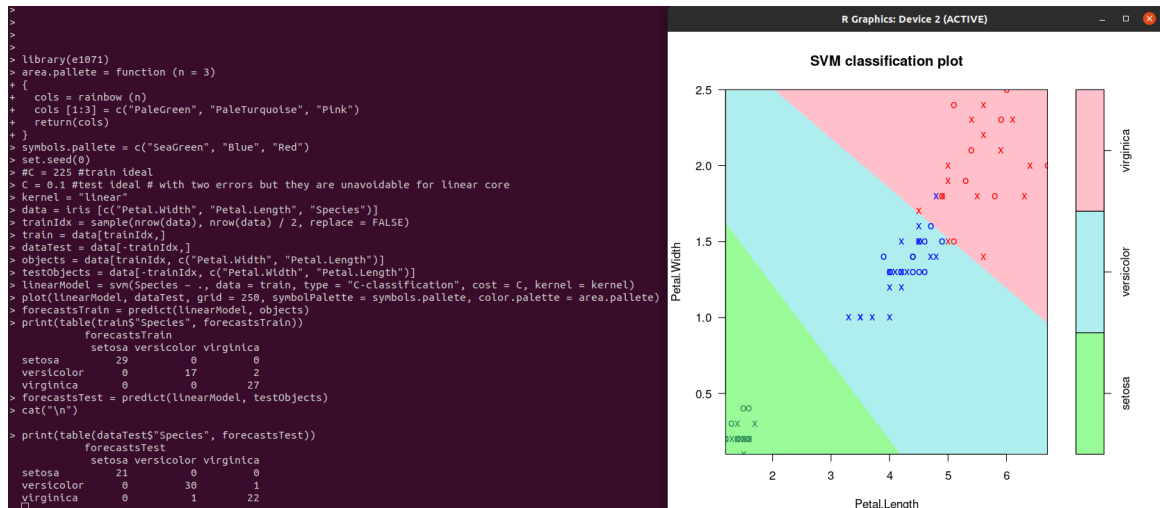


Рисунок 4.2

Задания 3-4. Листинг программы (lab4/3&4.R)

Данный код перебирает различные всевозможные сочетания значений гиперпараметров из следующих возможных значений:

kernel: "polynomial", "radial", "sigmoid"

C: 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50

degree: 1-15

gamma: 0.1, 0.25, 0.5, 0.75, 1, 2, 5, 10, 25, 50

По результатам работы кода наилучшими комбинациями значений гиперпараметров с точки зрения ошибки на тестовой выборке являются:

```

> print(dataFrame[dataFrame$test_errors == min_error_count,])

```

	kernel	C	degree	gamma	errors	train_errors	test_errors
107	polynomial	0.01	1	10.00	4	2	2
227	polynomial	0.05	1	2.00	4	2	2
362	polynomial	0.10	1	1.00	4	2	2
467	polynomial	0.50	1	0.25	4	2	2
602	polynomial	1.00	1	0.10	4	2	2
1228	radial	0.10	1	5.00	5	3	2
1232	radial	0.50	1	0.10	4	2	2
1322	sigmoid	1.00	1	0.10	4	2	2

Рисунок 4.3

Задание 5. Для демонстрации эффекта переобучения было выбрано радиальное ядро. При значении $\gamma = 50$ модель делает 0 ошибок на тренировочной выборке и 3 ошибки на тестовой. При значении $\gamma = 500$ модель делает 0 ошибок на тренировочной выборке и 15 ошибок на тестовой. Листинг программы (lab4/5.R)

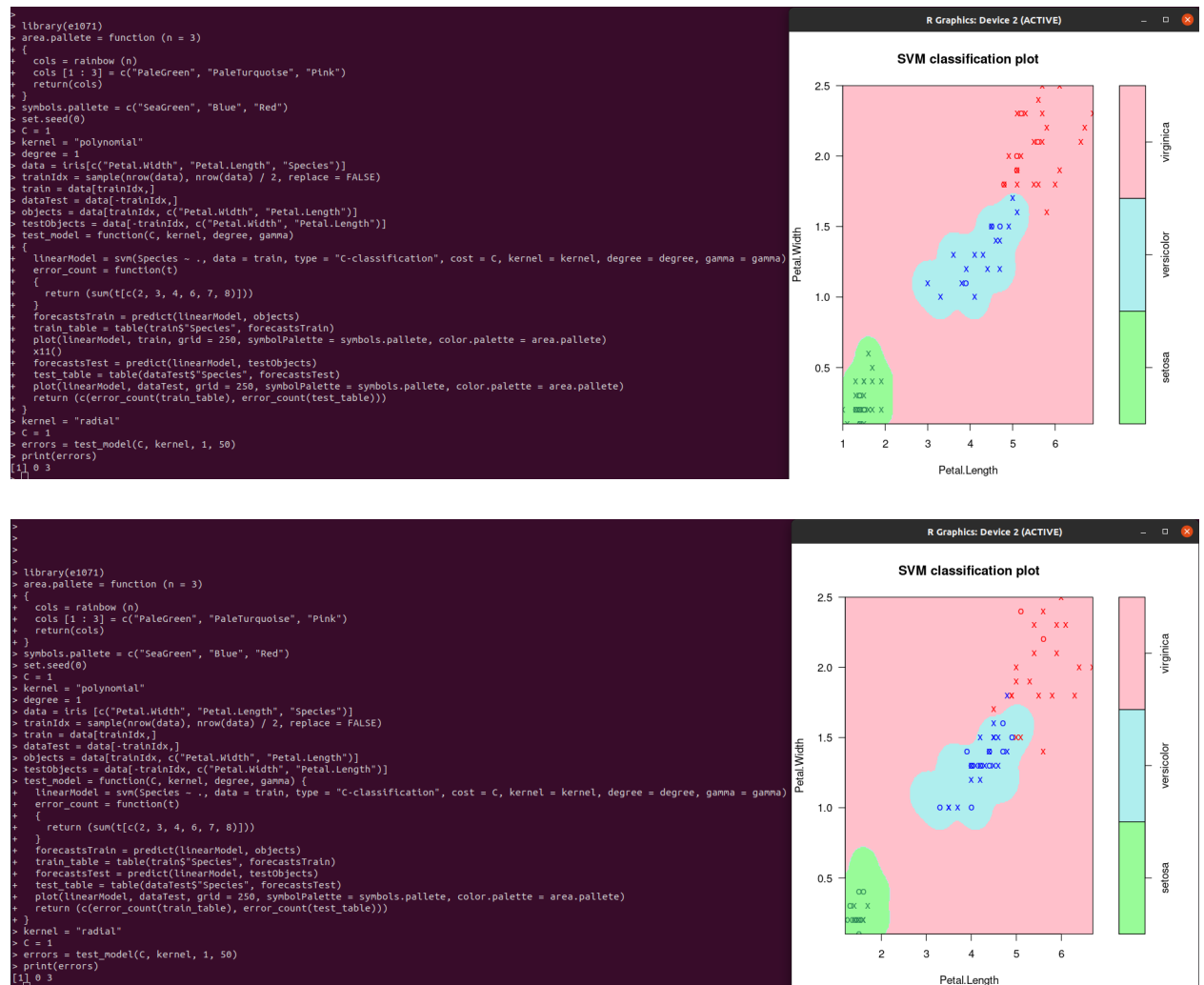
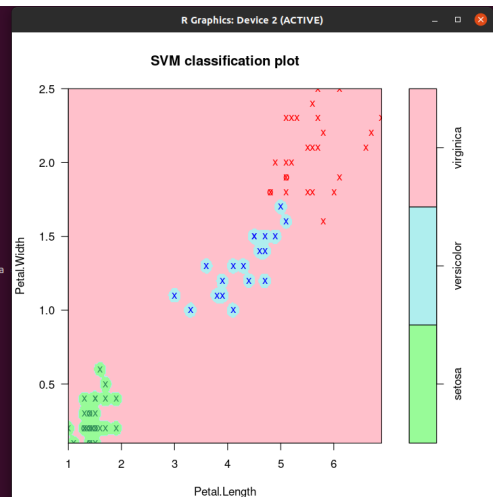


Рисунок 4.4 - Разбиение пространства признаков при параметре γ равном 50 (тренировочная выборка сверху, тестовая снизу)

```

library(e1071)
area.pallete = function (n = 3)
{
  cols = rainbow (n)
  cols [1 : 3] = c("PaleGreen", "PaleTurquoise", "Pink")
  return(cols)
}
symbols.pallete = c("SeaGreen", "Blue", "Red")
set.seed(0)
C = 1
kernel = "polynomial"
degree = 1
data = iris[c("Petal.Width", "Petal.Length", "Species")]
trainIdx = sample(nrow(data), nrow(data) / 2, replace = FALSE)
train = data[trainIdx,]
dataTest = data[-trainIdx,]
objects = data[trainIdx, c("Petal.Width", "Petal.Length")]
testObjects = data[-trainIdx, c("Petal.Width", "Petal.Length")]
test_model = function(C, kernel, degree, gamma)
{
  linearModel = svm(Species ~ ., data = train, type = "C-classification", cost = C, kernel = kernel, degree = degree, gamma = gamma)
  error_count = function(t)
  {
    return (sum(t[c(2, 3, 4, 6, 7, 8)]))
  }
  forecastsTrain = predict(linearModel, objects)
  train_table = table(train$Species, forecastsTrain)
  plot(linearModel, train, grid = 250, symbolPalette = symbols.pallete, color.palette = area.pallete)
  x11()
  forecastsTest = predict(linearModel, testObjects)
  test_table = table(dataTest$Species, forecastsTest)
  plot(linearModel, dataTest, grid = 250, symbolPalette = symbols.pallete, color.palette = area.pallete)
  return (c(error_count(train_table), error_count(test_table)))
}
kernel = "radial"
C = 1
errors = test_model(C, kernel, 1, 500)
print(errors)
[1] 0 15

```



```

library(e1071)
area.pallete = function (n = 3)
{
  cols = rainbow (n)
  cols [1 : 3] = c("PaleGreen", "PaleTurquoise", "Pink")
  return(cols)
}
symbols.pallete = c("SeaGreen", "Blue", "Red")
set.seed(0)
C = 1
kernel = "polynomial"
degree = 1
data = iris[c("Petal.Width", "Petal.Length", "Species")]
trainIdx = sample(nrow(data), nrow(data) / 2, replace = FALSE)
train = data[trainIdx,]
dataTest = data[-trainIdx,]
objects = data[trainIdx, c("Petal.Width", "Petal.Length")]
testObjects = data[-trainIdx, c("Petal.Width", "Petal.Length")]
test_model = function(C, kernel, degree, gamma)
{
  linearModel = svm(Species ~ ., data = train, type = "C-classification", cost = C, kernel = kernel, degree = degree, gamma = gamma)
  error_count = function(t)
  {
    return (sum(t[c(2, 3, 4, 6, 7, 8)]))
  }
  forecastsTrain = predict(linearModel, objects)
  train_table = table(train$Species, forecastsTrain)
  forecastsTest = predict(linearModel, testObjects)
  test_table = table(dataTest$Species, forecastsTest)
  plot(linearModel, dataTest, grid = 250, symbolPalette = symbols.pallete, color.palette = area.pallete)
  return (c(error_count(train_table), error_count(test_table)))
}
kernel = "radial"
C = 1
errors = test_model(C, kernel, 1, 500)
print(errors)
[1] 0 15

```

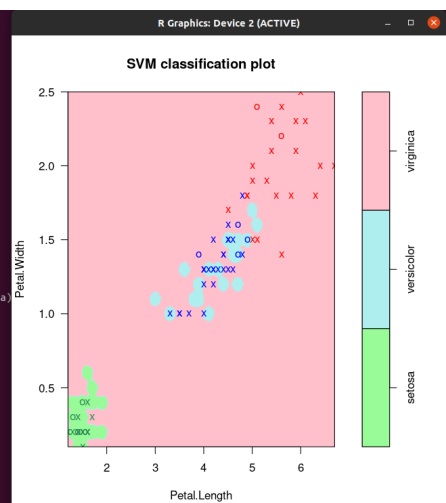


Рисунок 4.5 - Разбиение пространства признаков при параметре гамма равном 500 (тренировочная выборка сверху, тестовая снизу)

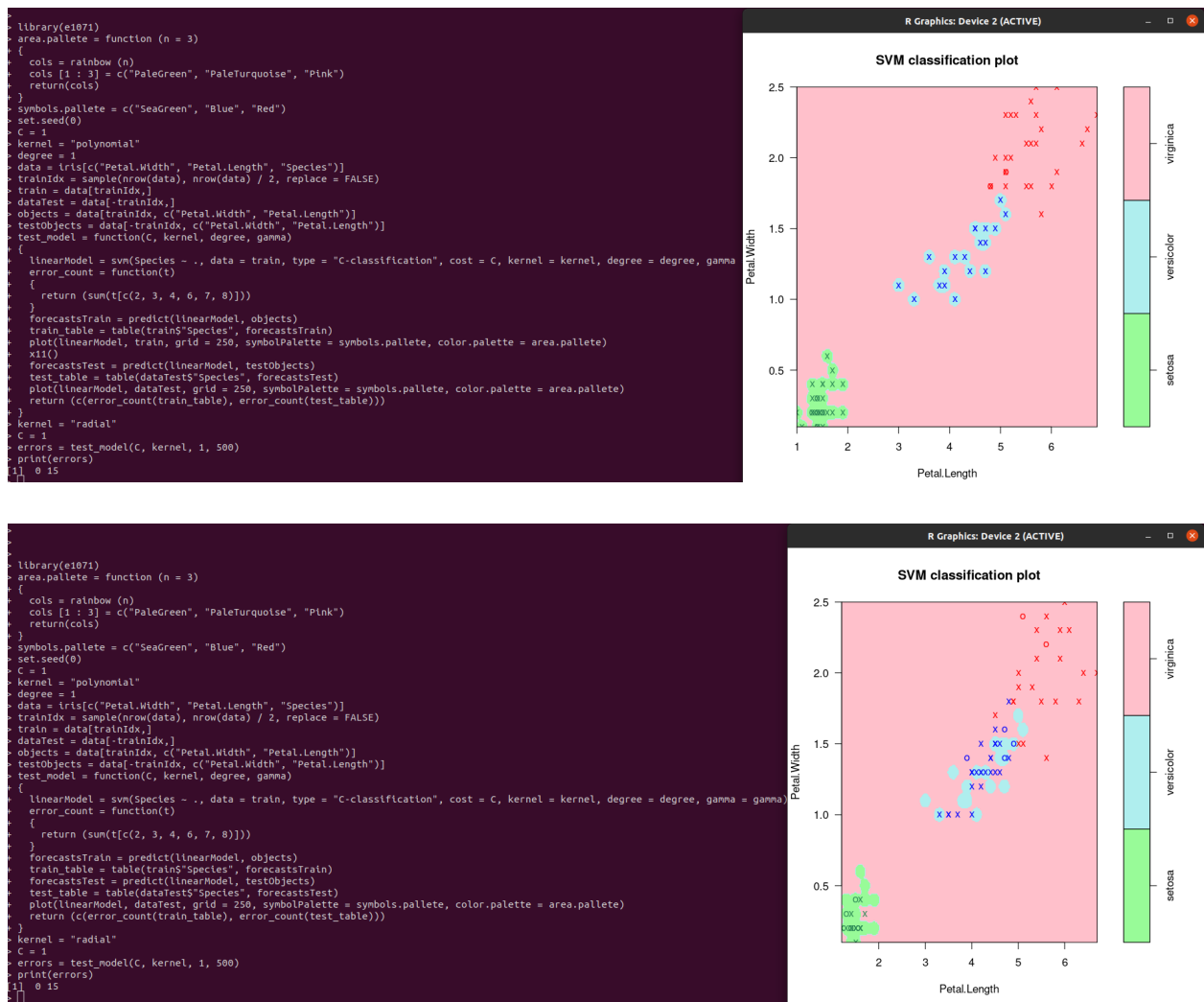


Рисунок 4.6 – Разбиение пространства признаков при параметре гамма равном 500 (тренировочная выборка сверху, тестовая снизу)

Задание 6. Листинг кода (lab4/6.R)

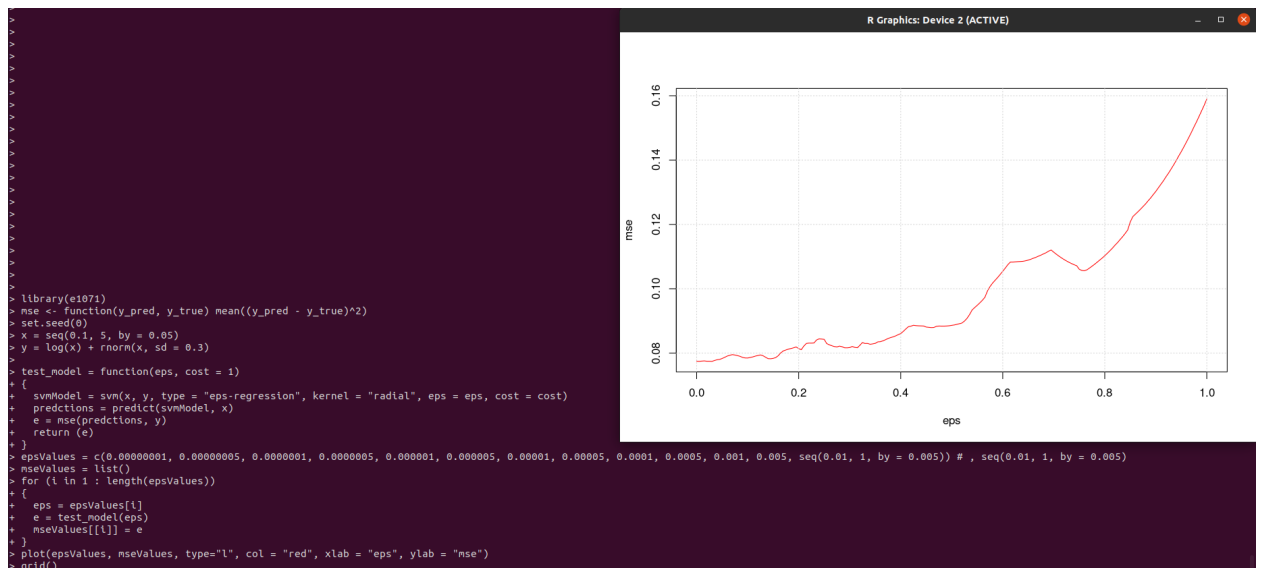


Рисунок 4.7 – Зависимость среднеквадратичной ошибки от параметра ϵ

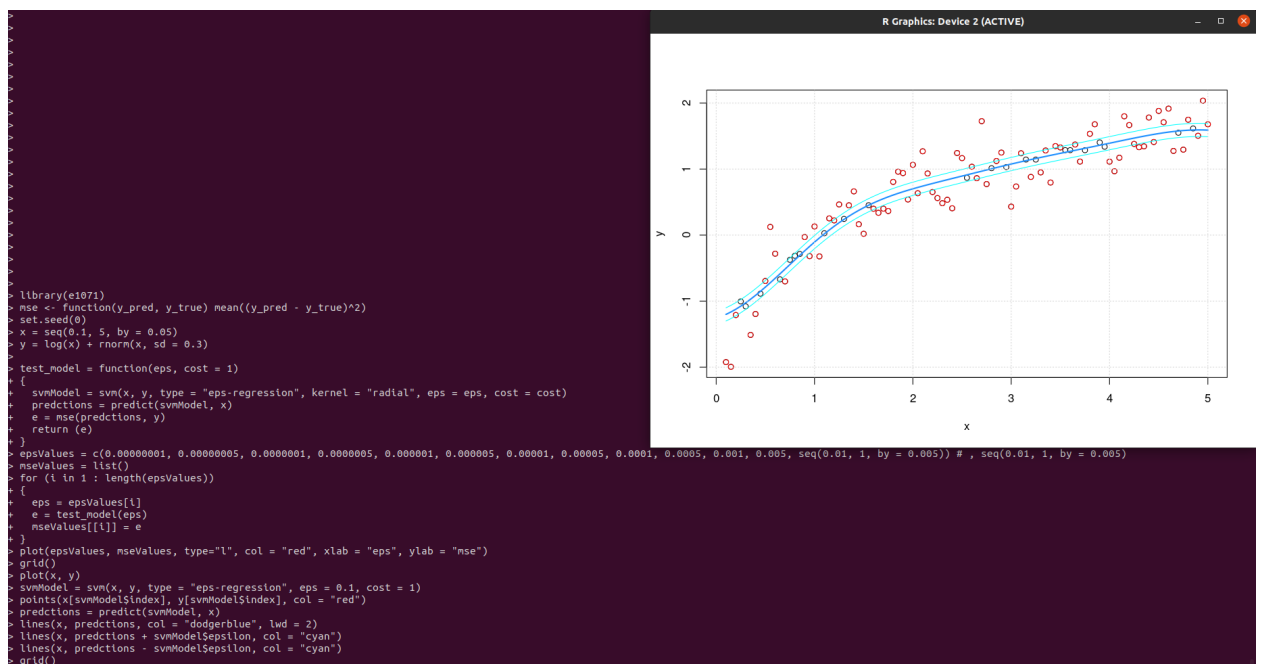


Рисунок 4.8 – Регрессия и ее ϵ -окрестность при значении параметра ϵ равном 0.1

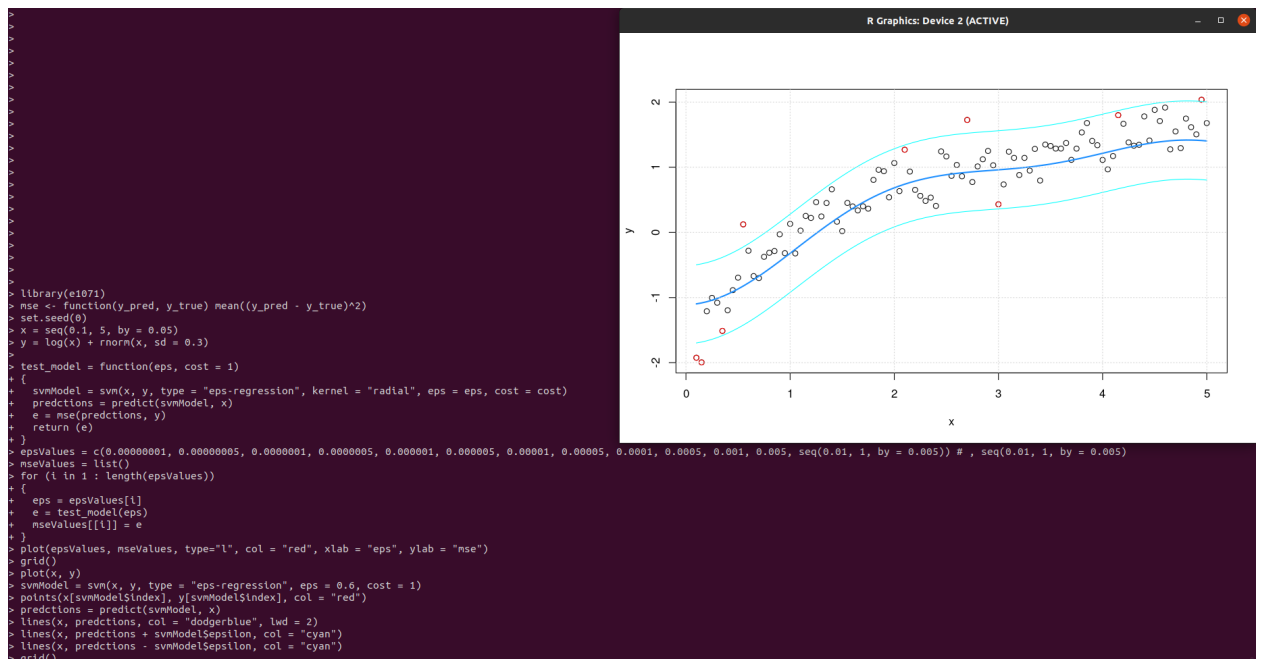


Рисунок 4.9 – Регрессия и ее ϵ -окрестность при значении параметра ϵ равном 0.6

Выводы

В ходе выполнения лабораторных работ были сформированы практические навыки работы с языком R. Получены знания о проверке статистических гипотез согласно условиям и требованиям. Изучено понятие линейной регрессии и выработаны навыки построения классификатора с использованием машины опорных векторов.