

Министерство образования и науки Российской Федерации  
Калужский филиал  
федерального государственного бюджетного образовательного  
учреждения высшего образования  
**«Московский государственный технический университет  
имени Н.Э. Баумана  
(национальный исследовательский университет)»**  
(КФ МГТУ им. Н.Э. Баумана)

И.И. Кручинин  
(к.т.н. доцент)

**ЛАБОРАТОРНАЯ РАБОТА № 2  
по курсу «Методы машинного обучения»**

**Метрические методы классификации  
многомерных объектов пересекающихся  
классов**

**Приложение**

Калуга

## Теоретические основы.

### Метрические методы классификации

Во многих прикладных задачах измерять степень сходства объектов существенно проще, чем формировать признаковые описания. Если мера сходства объектов введена достаточно удачно, то, как правило, оказывается, что схожим объектам очень часто соответствуют схожие ответы. В задачах классификации это означает, что классы образуют компактно локализованные подмножества. Для формализации понятия «сходства» вводится функция расстояния в пространстве объектов  $X$ . Методы обучения, основанные на анализе сходства объектов, будем называть метрическими.

Метрический алгоритм классификации с обучающей выборкой  $X^\ell$  относит объект  $u$  к тому классу  $y \in Y$ , для которого суммарный вес ближайших обучающих объектов  $\Gamma_y(u, X^\ell)$  максимален:

$$a(u; X^\ell) = \arg \max_{y \in Y} \Gamma_y(u, X^\ell);$$

$$\Gamma_y(u, X^\ell) = \sum_{i=1}^{\ell} [y_u^{(i)} = y] w(i, u);$$

где весовая функция  $w(i, u)$  оценивает степень важности  $i$ -го соседа для классификации объекта  $u$ . Функция  $\Gamma_y(u, X^\ell)$  называется оценкой близости объекта  $u$  к классу  $y$ .

Обучающая выборка  $X^\ell$  играет роль параметра алгоритма  $a$ . Настройка сводится к запоминанию выборки, и, возможно, оптимизации каких-то параметров весовой функции, однако сами объекты не подвергаются обработке и сохраняются «как есть». Алгоритм  $a(u; X^\ell)$  строит локальную аппроксимацию выборки  $X^\ell$ , причём вычисления откладываются до момента, пока не станет известен объект  $u$ . По этой причине метрические алгоритмы относятся к методам ленивого обуче-

ния (lazy learning), в отличие от усердного обучения (eager learning), когда на этапе обучения строится функция, аппроксимирующая выборку.

**Алгоритм  $k$  ближайших соседей** ( $k$  nearest neighbors,  $k$ NN). Чтобы сгладить влияние выбросов (объектов, находящийся в окружении объектов чужого класса), будем относить объект  $u$  к тому классу, элементов которого окажется больше среди ближайших соседей.

$$w(i, u) = [i \leq k]; \quad a(u; X^\ell, k) = \arg \max_{y \in Y} \sum_{i=1}^k [y_u^{(i)} = y].$$

При  $k = 1$  этот алгоритм совпадает с предыдущим, следовательно, неустойчив к шуму. При  $k = \ell$ , наоборот, он чрезмерно устойчив и вырождается в константу.

Таким образом, крайние значения  $k$  нежелательны. На практике оптимальное значение параметра  $k$  определяют по критерию скользящего контроля с исключением объектов по одному (leave-one-out, LOO). Для каждого объекта  $x_i \in X^\ell$  проверяется, правильно ли он классифицируется по своим  $k$  ближайшим соседям.

$$\text{LOO} \quad (k, X^\ell) = \sum_{i=1}^{\ell} [a(x_i; X^\ell \setminus \{x_i\}, k) \neq y_i] \rightarrow \min_k.$$

Заметим, что если классифицируемый объект  $x_i$  не исключать из обучающей выборки, то ближайшим соседом  $x_i$  всегда будет сам  $x_i$ , и минимальное (нулевое) значение функционала  $\text{LOO}(k)$  будет достигаться при  $k = 1$ .

### Метод парzenовского окна (Parzen Window)

Ещё один способ задать веса соседям — определить  $w_i$  как функцию от расстояния  $\rho(u, x_u^{(i)})$ , а не от ранга соседа  $i$ . Введём функцию ядра  $K(z)$ , невозрастающую на  $[0, \infty)$ . Положив  $w(i, u) = K\left(\frac{1}{h} \rho(u, x_u^{(i)})\right)$  в общей формуле

$$a(u; X^\ell, h) = \arg \max_{y \in Y} \sum_{i=1}^{\ell} [y_u^{(i)} = y] K\left(\frac{\rho(u, x_u^{(i)})}{h}\right).$$

## Пример программного кода

Алгоритм К ближайших соседей

```
#data(iris)
library(kohonen)
library(RSNNS)
library(class)
library(gmodel)
library(modeest)
library(nnet)

ramFo = data.frame(read.table("cleverK.txt", header = TRUE, sep = ""))
#print("Исходные данные")
#print(ramFo)

alg <- matrix(0:0, nrow=150, ncol=4)
for (i in 1:50) {
  for (j in 1:4) {
    alg[i,j] = sample(5:50,1)
  }
}

for (i in 51:100) {
  for (j in 1:4) {
    alg[i,j] = sample(51:95,1)
  }
}

for (i in 101:150) {
  for (j in 1:4) {
    alg[i,j] = sample(1:4,1)
  }
}
```

```

}

#alg
write.table(alg, file="GMB1.txt")
ramFo2 = data.frame(read.table("GMB1.txt", header = TRUE, sep = ""))

#C1 <- c("Выгодная")
ramFoT = data.frame(read.table("org.txt", header = TRUE, sep = ""))
ramFo2 <- cbind(ramFo2, ramFoT$V1)
#ramFo2

table(ramFoT$V1)
ramFoT$V1 <- factor(ramFoT$V1, levels = c("S", "M"),
labels = c("Super", "Medium"))
round(prop.table(table(ramFoT$V1))*100, digits = 1)
#ramFo2 <- ramFo2[-1]
summary(ramFo2[c("V1", "V2", "V3", "V4")])
normalize <- function(x) {return ((x - min(x)) / (max(x) - min(x)))}
rm_n <- as.data.frame(lapply(ramFo2[1:4], normalize))
rm_train <- rm_n[1:89, ]
rm_test <- rm_n[90:149, ]
rm_train_labels <- ramFo2[1:89, 1]
rm_test_labels <- ramFo2[90:149, 1]
#library("class")
rm_test_pred <- knn(train = rm_train, test = rm_test, cl =
rm_train_labels, k=17)

library("gmodels")
CrossTable(x = rm_test_labels, y = rm_test_pred, prop.chisq=FALSE)

```

## Алгоритм Парзена

```

#data(iris)
library(kohonen)
library(RSNNS)
library(class)
library(gmodel)
library(modeest)

```

```

library(nnet)

ramFo = data.frame(read.table("cleverK.txt", header = TRUE, sep = ""))
#print("Исходные данные")
#print(ramFo)

alg <- matrix(0:0, nrow=150, ncol=4)
for (i in 1:50) {
  for (j in 1:4) {
    alg[i,j] = sample(5:50,1)
  }
}

for (i in 51:100) {
  for (j in 1:4) {
    alg[i,j] = sample(51:95,1)
  }
}

for (i in 101:150) {
  for (j in 1:4) {
    alg[i,j] = sample(1:4,1)
  }
}

#alg
write.table(alg, file="GMB1.txt")
ramFo2 = data.frame(read.table("GMB1.txt", header = TRUE, sep = ""))

#C1 <- c("Выгодная")
ramFoT = data.frame(read.table("org.txt", header = TRUE, sep = ""))
ramFo2 <- cbind(ramFo2, ramFoT$V1)
#ramFo2

#x <- rbeta(1000,23,4)
#x

stano <- as.data.frame(scale(ramFo2[,1:4]))
x <- rbeta(stano,23,4)

```

```
M1 <- mlv(x, method = "parzen", kernel = "gaussian")  
M2 <- mlv(x, method = "tsybakov", kernel = "gaussian")
```

```
plot(M1)
```

```
j=3  
for (i in 1:50000000) {  
  j=j+1  
}
```

```
plot(M2)
```

## Задания к лабораторной работе

### Вариант 1

1. Определим уровень финансовой устойчивости предприятия, как составной части общей устойчивости предприятия (при этом соблюдаются сбалансированность финансовых потоков, наличие средств, позволяющих организации поддерживать свою деятельность в течение определенного периода времени, в том числе обслуживая полученные кредиты и производя продукцию). При этом необходимо использовать показатели финансовой устойчивости: коэффициент автономии, Коэффициент финансового левериджа, Коэффициент обеспеченности собственными оборотными средствами, Коэффициент покрытия инвестиций, Коэффициент маневренности собственного капитала, Коэффициент мобильности имущества, Коэффициент мобильности оборотных средств, Коэффициент обеспеченности запасов, Коэффициент краткосрочной задолженности, коэффициент текущей ликвидности и коэффициент быстрой ликвидности, коэффициент капитализации, коэффициент покрытия активов, коэффициент покрытия инвестиций.

Исходные данные следует организовать в виде таблицы:

RES	K1	K2	K3	K4	K5	...	KN
T	0.01						
F	0.022						
...	...	...	...	...	...	...	...
T	0.451						

В первом столбце заносится значение бинарной классификации – финансовый состояние предприятия - устойчиво или нет (True, False). В данном варианте признаками финансовой устойчивости будут: коэффициент автономии, Коэффициент финансового левериджа, коэффициент текущей ликвидности, коэффициент капитализации. Строк в таблице должно быть 110 (каждая строка - сведения по проверенному предприятию).

- Используем метод К-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода К соседей параметр K =18, для метода Парзена тип ядра выбрать "triangular", "uniform", а параметр optim. method ="Nelder-Mead", "BFGS". Проверить точность прогнозов.

## Вариант 2

- Определим уровень финансовой устойчивости предприятия, как составной части общей устойчивости предприятия (при этом соблюдаются сбалансированность финансовых потоков, наличие средств, позволяющих организации поддерживать свою деятельность в течение определенного периода времени, в том числе обслуживая полученные кредиты и производя продукцию).

Исходные данные следует организовать в виде таблицы:

RES	K1	K2	K3	K4	K5	...	KN
T	0.05						0.509
F	0.027			0.611			
...	...	...	...	...	...	...	...
T	0.458						

В первом столбце заносится значение бинарной классификации – финансовый состояние предприятия - устойчиво или нет (True,



False). В данном варианте признаками финансовой устойчивости будут: Коэффициент обеспеченности собственными оборотными средствами, Коэффициент покрытия инвестиций, Коэффициент маневренности собственного капитала, Коэффициент мобильности имущества. Строк в таблице должно быть 140 (каждая строка - сведения по проверенному предприятию).

- Используем метод К-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода К соседей параметр  $K=20$ , для метода Парзена тип ядра выбрать "epanechnikov", "uniform", а параметр `optim. method` ="SANN", "BFGS". Проверить точность прогнозов.

### Вариант 3

- Определим уровень финансовой устойчивости предприятия, как составной части общей устойчивости предприятия (при этом соблюдаются сбалансированность финансовых потоков, наличие средств, позволяющих организации поддерживать свою деятельность в течение определенного периода времени, в том числе обслуживая полученные кредиты и производя продукцию).

Исходные данные следует организовать в виде таблицы:

RES	K1	K2	K3	K4	K5	...	KN
T	0.05						0.539
F	0.027	0.240		0.671			
...	...	...	...	...	...	...	...
T	0.458			0.683			

В первом столбце заносится значение бинарной классификации – финансовой состояние предприятия - устойчиво или нет (True, False). В данном варианте признаками финансовой устойчивости будут: Коэффициент краткосрочной задолженности, коэффициент текущей ликвидности и коэффициент быстрой ликвидности, коэффициент капитализации, коэффициент покрытия активов, коэффициент покрытия инвестиций. Строк в таблице должно быть 130 (каждая строка - сведения по проверенному предприятию).

- Используем метод К-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода К соседей параметр  $K = 16$ , для метода Парзена тип ядра выбрать "gaussian", "optcosine" а параметр `optim. method = "CG", "L-BFGS-B"`. Проверить точность прогнозов.

## Вариант 4

- Определим уровень финансовой устойчивости предприятия, как составной части общей устойчивости предприятия (при этом соблюдаются сбалансированность финансовых потоков, наличие средств, позволяющих организации поддерживать свою деятельность в течение определенного периода времени, в том числе обслуживая полученные кредиты и производя продукцию).

Исходные данные следует организовать в виде таблицы:

RES	K1	K2	K3	K4	K5	...	KN
T	0.252	0.301					0.539
F	0.327	0.240		0.671			
...	...	...	...	...	...	...	...
T	0.458			0.683			

В первом столбце заносится значение бинарной классификации – финансовой состояние предприятия - устойчиво или нет (True, False). В данном варианте признаками финансовой устойчивости будут: Коэффициент мобильности имущества, Коэффициент мобильности оборотных средств, Коэффициент обеспеченности запасов, Коэффициент краткосрочной задолженности. Строк в таблице должно быть 120 (каждая строка - сведения по проверенному предприятию).

- Используем метод К-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода К соседей параметр  $K = 22$ , для метода Парзена тип ядра выбрать "rectangular", "triangular" а параметр `optim. method = "Brent", "Nelder-Mead"`. Проверить точность прогнозов.

## Вариант 5

1. Определим уровень финансовой устойчивости предприятия, как составной части общей устойчивости предприятия (при этом соблюдаются сбалансированность финансовых потоков, наличие средств, позволяющих организации поддерживать свою деятельность в течение определенного периода времени, в том числе обслуживая полученные кредиты и производя продукцию).

Исходные данные следует организовать в виде таблицы:

RES	K1	K2	K3	K4	K5	...	KN
T	0.252	0.301					0.539
F	0.327	0.240		0.695	0.671		
...	...	...	...	...	...	...	...
T	0.458			0.683			

В первом столбце заносится значение бинарной классификации – финансовый состояние предприятия - устойчиво или нет (True, False). В данном варианте признаками финансовой устойчивости будут: коэффициент автономии, коэффициент покрытия инвестиций, коэффициент быстрой ликвидности, Коэффициент мобильности имущества. Строк в таблице должно быть 127 (каждая строка - сведения по проверенному предприятию).

2. Используем метод K-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода K соседей параметр K = 24, для метода Парзена тип ядра выбрать "biweight", "cosine", "eddy" а параметр optim. method = "CG". Проверить точность прогнозов.

## Вариант 6

Характеристики грибов	K1	K2	K3	K4
Калории	34	9	29	7
Белки	3.7	0.24	3.1	0.12
Углеводы	1.1	0.32	0.97	0.04
Диаметр шляпки	25	21	14	13
Диаметр ножки	10	11	8	9
Длина ножки	20	18	13	12
Цвет шляпки	коричневый	Светло-коричневый	Рыжий	Грязно-рыжий
Цвет ножки	Белый	Грязно-желтый	Рыжий	Рыжий
Вид ножки	Трубчатая сплошная	Трубчатая	Трубчатая	Трубчатый
Содержание железа	5.2	0.88	4.4	0.53
Содержание кобальта	6	0.7	5.1	0.72
Содержание калия	4.6	0.5	3.8	0.84
Тип ножки	Гладкая	Неровная	Гладкая	Шероховатая
Слой под шляпкой	трубчатый	трубчатый	Трубчатый	Трубчатый
Жиры	1.7	0.7	1.2	0.2
Тип среза шляпки	Светлый	Грязно-розовый	Светло-желтый	Грязно-желтый
Тип среза ножки	Светлый	Грязно-розовый	Светло-желтый	Грязно-желтый
Название	Белый боровик	Желчный гриб (ложный белый)	Лисичка	Ложная лисичка

Используем классификацию грибов по вкусовым качествам:

1 категория: белый гриб, польский гриб, рыжик, груздь

2 категория: подосиновик, подберезовик, масленок, шампиньон, волнушка, белая, лисичка

3 категория: моховик, козляк, сморчок

4 категория: горькушка, скрипица, рядовка

По строению шляпки можно выделить грибы: трубчатые, пластинчатые, сумчатые. К трубчатым можно отнести грибы из 1 и 2 вкусовой категории. К пластинчатым можно отнести грибы из 3 и 4 вкусовой категории. К сумчатым грибам относят сморчки, строчки и трюфельные грибы (т.е. из 3 вкусовой категории)

1. Необходимо разработать метрический классификатор, который функционирует на предметной области исследования лесной растительности. Различные виды и категории грибов часто употребляются в пищу в столовых, ресторанах, в домашних условиях. При этом необходимо использовать характеристические показатели качества грибов (см. сводную таблицу). В данном варианте надо использовать признаки грибов: содержание калорий, жиров, углеводов, белков, тип слоя под шляпкой, длина ножки и шляпки, цвет шляпки.

Исходные данные следует организовать в виде таблицы:

RES	K1	K2	K3	K4	K5	...	KN
T	0.01			11			
F	0.028		5.1		4.3		
...	...	...	...	...	...	...	...
T	0.451			12			

В первом столбце заносится значение бинарной классификации – съедобный или не съедобный гриб (True, False). Строк в таблице должно быть 80 (каждая строка - сведения по выбранному грибу).

2. Используем метод K-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода K соседей параметр K = 38, для метода Парзена тип ядра выбрать "eddy", "epanechnikov", а параметр optim. method = "Nelder-Mead", "BFGS". Проверить точность прогнозов.

## Вариант 7

1. Необходимо разработать метрический классификатор, который функционирует на предметной области исследования лесной растительности. Различные виды и категории грибов часто употребляются в пищу в столовых, ресторанах, в домашних условиях. При этом необходимо использовать характеристические показатели качества грибов (см. сводную таблицу). В данном варианте надо использовать признаки грибов: содержание кобальта, железа, калия, диаметр ножки и шляпки и типы срезов ножки и шляпки.

Исходные данные следует организовать в виде таблицы:

RES	K1	K2	K3	K4	K5	...	KN
T	0.01						
F	0.028		5.1		4.3		
...	...	...	...	...	...	...	...
T	0.451						

В первом столбце заносится значение бинарной классификации – съедобный или не съедобный гриб (True, False). Строк в таблице должно быть 70 (каждая строка - сведения по выбранному грибу).

2. Используем метод K-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода K соседей параметр  $K = 36$ , для метода Парзена тип ядра выбрать "biweight", "cosine", а параметр optim. method ="CG", "L-BFGS-B". Проверить точность прогнозов.

## Вариант 8

1. Необходимо разработать метрический классификатор, который функционирует на предметной области исследования лесной растительности. Различные виды и категории грибов часто употребляются в пищу в столовых, ресторанах, в домашних условиях. При этом необходимо использовать характеристические показатели качества грибов

(см. сводную таблицу). В данном варианте надо использовать признаки грибов: диаметр ножки и шляпки, длина ножки и шляпки, тип слоя под шляпкой, содержание кобальта.

Исходные данные следует организовать в виде таблицы:

RES	K1	K2	K3	K4	K5	...	KN
T	0.01						
F	0.028		5.1		4.3		
...	...	...	...	...	...	...	...
T	0.451						

В первом столбце заносится значение бинарной классификации – гриб первой или второй категории (True, False). Строк в таблице должно быть 75 (каждая строка - сведения по выбранному грибу).

- Используем метод K-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода K соседей параметр  $K = 34$ , для метода Парзена тип ядра выбрать "gaussian", "optcosine", а параметр optim. method ="SANN", "Brent". Проверить точность прогнозов.

## Вариант 9

- Необходимо разработать метрический классификатор, который функционирует на предметной области исследования лесной растительности. Различные виды и категории грибов часто употребляются в пищу в столовых, ресторанах, в домашних условиях. При этом необходимо использовать характеристические показатели качества грибов (см. сводную таблицу). В данном варианте надо использовать признаки грибов: диаметр ножки и шляпки, цвет ножки и шляпки, тип слоя под шляпкой, содержание железа, жиров и углеводов.

Исходные данные следует организовать в виде таблицы:

RES	K1	K2	K3	K4	K5	...	KN
T	0.01						

F	0.028		5.1		4.3		
...	...	...	...	...	...	...	...
T	0.451						

В первом столбце заносится значение бинарной классификации – гриб сумчатый или пластинчатый (True, False). Строк в таблице должно быть 85 (каждая строка - сведения по выбранному грибу).

- Используем метод К-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода К соседей параметр  $K = 29$ , для метода Парзена тип ядра выбрать "rectangular", "triangular", "uniform", а параметр `optim. method` ="BFGS". Проверить точность прогнозов.

## Вариант 10

- Необходимо разработать метрический классификатор, который функционирует на предметной области исследования лесной растительности. Различные виды и категории грибов часто употребляются в пищу в столовых, ресторанах, в домашних условиях. При этом необходимо использовать характеристические показатели качества грибов (см. сводную таблицу). В данном варианте надо использовать признаки грибов: содержание белков, жира, кобальта, диаметр шляпки, длина ножки, тип среза шляпки, цвет ножки.

Исходные данные следует организовать в виде таблицы:

RES	K1	K2	K3	K4	K5	...	KN
T	0.01						
F	0.028		5.1		4.3	8.1	
...	...	...	...	...	...	...	...
T	0.451						

В первом столбце заносится значение бинарной классификации – гриб трубчатый или пластинчатый (True, False). Строк в таблице должно быть 72 (каждая строка - сведения по выбранному грибу).



2. Используем метод К-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода К соседей параметр  $K = 33$ , для метода Парзена тип ядра выбрать , "gaussian", "triangular", "uniform", а параметр `optim. method = "BFGS", "SANN"`. Проверить точность прогнозов.

## Вариант 11

Больные увеличением щитовидной железы общим числом 50 человек были разделены на две группы.

*Группа 1.* Лечение оказалось успешным; пациент здоров.

*Группа 2.* Лечение безуспешно, состояние больного осталось без изменения.

По результатам обследования 50 пациентов имеются следующие измерения:

y6 – йод, регистрируемый через 2 часа после принятия испытательной дозы;

y9 – йод, регистрируемый через 16 часов после принятия испытательной дозы;

y10 – содержание в крови белковосвязанного йода ( $PB^{131}J$ ) через 32 часа;

kl – номер группы.

*Таблица*

№	Kl	y6	y9	y10
1	1	14.4	25.1	0.20
2	1	20.1	40.1	0.11
3	1	24.1	32.1	0.17
4	1	11.1	16.9	0.12
5	1	16.3	32.1	0.36
6	1	40.5	64.4	0.21
7	1	52.7	50.0	0.53
8	1	20.8	22.3	0.13
9	1	14.0	3.1	0.18
10	1	27.0	41.7	0.19

11	1	44.3	63.8	0.22
12	1	47.5	50.1	0.29
13	1	54.0	57.0	0.19
14	1	16.1	20.6	0.22
15	1	57.5	74.5	0.49
16	1	37.8	63.0	0.32
17	2	55.8	48.0	2.74
18	2	75.0	60.0	1.37
19	2	72.0	65.0	0.70
20	2	70.6	45.0	1.40
21	2	24.1	45.0	0.22
22	2	33.2	55.0	0.01
23	2	30.4	44.6	0.09
	...	...		
50	2		61.1	

Необходимо разработать метрический классификатор, который функционирует на предметной области исследования медицинских анализов. Три характеристических показателя надо взять из таблицы. Используем метод К-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода К соседей параметр  $K = 11$ , для метода Парзена тип ядра выбрать , "gaussian", , "uniform", а параметр `optim. method = "SANN"`. Проверить точность прогнозов.

## Вариант 12

Больные увеличением щитовидной железы общим числом 55 человек были разделены на две группы.

*Группа 1.* Лечение оказалось успешным; пациент здоров.

*Группа 2.* Лечение безуспешно, состояние больного осталось без изменения.

По результатам обследования 55 пациентов имеются следующие измерения:

УЗ – йод, регистрируемый через 1 час после принятия испытательной дозы;

У5 – йод, регистрируемый через 5 часов после принятия испытательной дозы;

Y7 – содержание в крови белковосвязанного йода ( $PB^{131}J$ ) через 17 часа;

Y8 – йод, регистрируемый через 8 часов после принятия испытательной дозы;

k1 – номер группы.

Таблица

№	K1	Y3	Y5	Y7	Y8
1	1	14.4	25.1	0.25	32
2	1	20.1	40.1	0.17	54
3	1	24.1	32.1	0.19	27
4	1	13.1	16.9	0.14	11
5	1	17.3	32.1	0.39	14
6	1	40.5	64.4	0.22	51
7	1	52.7	50.0	0.56	7
8	1	20.8	22.3	0.17	9
9	1	14.0	3.1	0.19	10
10	1	27.0	41.7	0.13	21
11	1	44.3	63.8	0.26	43
12	1	47.5	50.1	0.21	56
13	1	54.0	57.0	0.17	12
14	1	16.1	20.6	0.29	34
15	1	87.5	74.5	0.48	12
16	1	37.8	63.4	0.31	17
17	2	35.8	48.0	2.74	14
18	2	65.0	60.0	1.37	17
19	2	62.0	65.0	0.70	23
20	2	71.6	45.0	1.40	10
21	2	24.1	45.0	0.22	9
22	2	37.2	55.0	0.01	31
23	2	35.4	44.6	0.09	65
	...	...			28
55	2		61.1		42

Необходимо разработать метрический классификатор, который функционирует на предметной области исследования медицинских анализов. Четыре характеристических показателя надо взять из таблицы. Используем метод К-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода К-соседей параметр  $K = 16$ , для метода Парзена тип ядра выбрать "cosine", "uniform", а параметр `optim. method` = "CG", "L-BFGS-B". Проверить точность прогнозов.

### Вариант 13

Больные увеличением щитовидной железы общим числом 48 человек были разделены на две группы.

*Группа 1.* Лечение оказалось успешным; пациент здоров.

*Группа 2.* Лечение успешно на 85 %, возможен рецидив.

По результатам обследования 48 пациентов имеются следующие измерения:

Y3 – йод, регистрируемый через 2 часа после принятия испытательной дозы;

Y5 – йод, регистрируемый через 10 часов после принятия испытательной дозы;

Y7 – содержание в крови белковосвязанного йода ( $PB^{131}J$ ) через 19 часов;

Y8 – йод, регистрируемый через 21 час после принятия испытательной дозы;

Y9 – содержание в крови белковосвязанного йода ( $PB^{131}J$ ) через 26 часов;

kl – номер группы.

Таблица

№	Kl	Y3	Y5	Y7	Y8	Y9
1	1	14.4	25.1	0.25	32	1.77
2	1	20.1	40.1	0.17	54	2.86
3	1	24.1	32.1	0.19	27	3.11
4	1	13.1	16.9	0.14	11	1.48
5	1	17.3	32.1	0.39	14	12
6	1	40.5	64.4	0.22	51	40.3
7	1	52.7	50.0	0.56	7	1.97
8	1	20.8	22.3	0.17	9	13.86
9	1	14.0	3.1	0.19	10	15.23
10	1	27.0	41.7	0.13	21	24.08
11	1	44.3	63.8	0.26	43	...
12	1	47.5	50.1	0.21	56	
13	1	54.0	57.0	0.17	12	...
14	1	16.1	20.6	0.29	34	
15	1	87.5	74.5	0.48	12	
16	1	37.8	63.4	0.31	17	
17	2	35.8	48.0	2.74	14	
18	2	65.0	60.0	1.37	17	
19	2	62.0	65.0	0.70	23	

20	2	71.6	45.0	1.40	10	
21	2	24.1	45.0	0.22	9	
22	2	37.2	55.0	0.01	31	
23	2	35.4	44.6	0.09	65	
	...	...			28	
48	2		61.1		42	0.98

Необходимо разработать метрический классификатор, который функционирует на предметной области исследования медицинских анализов. Пять характеристических показателей надо взять из таблицы. Используем метод К-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода К соседей параметр  $K = 44$ , для метода Парзена тип ядра выбрать , "erapachnikov", , "uniform", а параметр `optim. method = "Brent"`. Проверить точность прогнозов.

## Вариант 14

Больные увеличением щитовидной железы общим числом 70 человек были разделены на две группы.

*Группа 1.* Лечение оказалось успешным; пациент здоров.

*Группа 2.* Лечение успешно на 75 %, возможен рецидив.

По результатам обследования 70 пациентов имеются следующие измерения:

Y3 – йод, регистрируемый через 3 часа после принятия испытательной дозы;

Y5 – йод, регистрируемый через 12 часов после принятия испытательной дозы;

Y7 – содержание в крови белковосвязанного йода ( $PB^{131}J$ ) через 20 часов;

Y9 – содержание в крови белковосвязанного йода ( $PB^{131}J$ ) через 33 часа;

kl – номер группы.

Таблица

№	KI	Y3	Y5	Y7	Y9
1	1	14.4	25.1	0.25	1.77
2	1	20.1	40.1	0.17	2.86
3	1	24.1	32.1	0.19	3.11
4	1	13.1	16.9	0.14	1.48
5	1	17.3	32.1	0.39	12
6	1	40.5	64.4	0.22	40.3
7	1	52.7	50.0	0.56	1.97
8	1	20.8	22.3	0.17	13.86
9	1	14.0	3.1	0.19	15.23
10	1	27.0	41.7	0.13	24.08
11	1	44.3	63.8	0.26	...
12	1	47.5	50.1	0.21	
13	1	54.0	57.0	0.17	...
14	1	16.1	20.6	0.29	
15	1	87.5	74.5	0.48	
16	1	37.8	63.4	0.31	
17	2	35.8	48.0	2.74	
18	2	65.0	60.0	1.37	
19	2	62.0	65.0	0.70	
20	2	71.6	45.0	1.40	
21	2	24.1	45.0	0.22	
22	2	37.2	55.0	0.01	
23	2	35.4	44.6	0.09	
	...	...			
70	2		61.1		0.98

Необходимо разработать метрический классификатор, который функционирует на предметной области исследования медицинских анализов. Четыре характеристических показателя надо взять из таблицы. Используем метод К-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода К соседей параметр  $K = 37$ , для метода Парзена тип ядра выбрать , "epanechnikov", "eddy" , а параметр `optim. method = "Nelder-Mead"`. Проверить точность прогнозов.

## Вариант 15

Больные увеличением щитовидной железы общим числом 63 человека были разделены на две группы.

*Группа 1.* Лечение оказалось успешным; пациент здоров.

*Группа 2.* Лечение успешно на 90 %, возможен рецидив.

По результатам обследования 63 пациентов имеются следующие измерения:

Y3 – йод, регистрируемый через 4 часа после принятия испытательной дозы;

Y5 – йод, регистрируемый через 16 часов после принятия испытательной дозы;

Y7 – содержание в крови белковосвязанного йода ( $PB^{131}J$ ) через 19 часов;

Y8 – йод, регистрируемый через 31 час после принятия испытательной дозы;

Y9 – содержание в крови белковосвязанного йода ( $PB^{131}J$ ) через 46 часов;

kl – номер группы.

Таблица

№	Kl	Y3	Y5	Y7	Y8	Y9
1	1	14.4	25.1	0.25	32	1.77
2	1	20.1	40.1	0.17	54	2.86
3	1	24.1	32.1	0.19	27	3.11
4	1	13.1	16.9	0.14	11	1.48
5	1	17.3	32.1	0.39	14	12
6	1	40.5	64.4	0.22	51	40.3
7	1	52.7	50.0	0.56	7	1.97
8	1	20.8	22.3	0.17	9	13.86
9	1	14.0	3.1	0.19	10	15.23
10	1	27.0	41.7	0.13	21	24.08
11	1	44.3	63.8	0.26	43	...
12	1	47.5	50.1	0.21	56	
13	1	54.0	57.0	0.17	12	...
14	1	16.1	20.6	0.29	34	
15	1	87.5	74.5	0.48	12	32.08
16	1	37.8	63.4	0.31	17	
17	2	35.8	48.0	2.74	14	

18	2	65.0	60.0	1.37	17	
19	2	62.0	65.0	0.70	23	
20	2	71.6	45.0	1.40	10	21.75
21	2	24.1	45.0	0.22	9	
22	2	37.2	55.0	0.01	31	22.56
23	2	35.4	44.6	0.09	65	
	...	...			28	
63	2		61.1		42	0.98

Необходимо разработать метрический классификатор, который функционирует на предметной области исследования медицинских анализов. Пять характеристических показателей надо взять из таблицы. Используем метод К-ближайших соседей и метод Парзена. Сформировать обучающие и тестовые выборки. Полученные результаты визуализировать и сравнить. Представить значения параметров с минимальным уровнем ошибки. Для метода К соседей параметр  $K = 8$ , для метода Парзена тип ядра выбрать "gaussian", "optcosine", а параметр `optim. method = "Brent", "CG"`. Проверить точность прогнозов.