|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **Министерство науки и высшего образования 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)*** |

**ФАКУЛЬТЕТ** ***ИУ-КФ «Информатика и управление»***

**КАФЕДРА** ***ИУ4-КФ «Программное обеспечение ЭВМ, информационные технологии»***

**ЛАБОРАТОРНАЯ РАБОТА №2**

**«Метрические методы классификации многомерных объектов пересекающихся классов»**

**ДИСЦИПЛИНА: «Методы машинного обучения»**

|  |  |  |
| --- | --- | --- |
| Выполнил: студент гр. ИТД.Б-62 | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (Чулин К.В.)  (Подпись) |
| Проверил: | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (Кручинин И.И.)  (Подпись) |
| Дата сдачи (защиты):  Результаты сдачи (защиты): | | |
|  | - Балльная оценка:  - Оценка: | |

Калуга , 2020

Цель работы: сформировать практические навыки разработки программ с использованием метрических методов классификации многомерных объектов пересекающихся классов на языке R.

Вариант 2

Задание

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 | 0.05 |  |  |  |  |  | 0.509 |
| F | 0.027 |  |  | 0.611 |  |  |  |
| … | … | … | … | … | … | … | … |
| T | 0.458 |  |  |  |  |  |  |

В первом столбце заносится значение бинарной классификации – финансовой состояние предприятия - устойчиво или нет (True, False). В данном варианте признаками финансовой устойчивости будут: [Коэффициент обеспеченности собственными оборотными средствами](https://www.audit-it.ru/finanaliz/terms/liquidity/working_capital_ratio_means.html), [Коэффициент покрытия инвестиций](https://www.audit-it.ru/finanaliz/terms/solvency/koeffitsient_pokrytiya_investitsiy.html), Коэффициент маневренности собственного капитала, Коэффициент мобильности имущества. Строк в таблице должно быть 140 (каждая строка - сведения по проверенному предприятию).

1. 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 К =20, для метода Парзена тип ядра выбрать "epanechnikov", "uniform", а параметр optim. method ="SANN", "BFGS". Проверить точность прогнозов.

Листинг программы:

library(class)

level <- sample(0:1, 140, replace=T)

security <- runif(140, 0, 5)

coating <- runif(140, 0, 5)

maneuver <- runif(140, 0, 5)

mobility <- runif(140, 0, 5)

company <- data.frame(Level = level, Security = security, Coating = coating,

Maneuver = maneuver, Mobility = mobility)

set.seed(4948493)

comp\_sample<-sample(1:nrow(company),size=nrow(company)\*.7)

comp\_train<-company[comp\_sample,]

comp\_test<-company[-comp\_sample,]

comp\_acc<- sample(0:0, 50, replace=T)

for(i in 1:50){

predict<-knn(comp\_train[,-5],comp\_test[,-5],comp\_train$Level,k=i)

comp\_acc<-c(comp\_acc,mean(predict==comp\_test$Level))

}

plot(1-comp\_acc,type="l",ylab="Error Rate",xlab="K",main="Company")

trial\_sum<-numeric(20)

trial\_n<-numeric(20)

set.seed(6033850)

for(i in 1:100){

comp\_sample<-sample(1:nrow(company),size=nrow(company)\*.7)

comp\_train<-company[comp\_sample,]

comp\_test<-company[-comp\_sample,]

test\_size<-nrow(comp\_test)

for(j in 1:20){

predict<-knn(comp\_train[,-5],comp\_test[,-5], comp\_train$Level,k=j)

trial\_sum[j]<-trial\_sum[j]+sum(predict==comp\_test$Level)

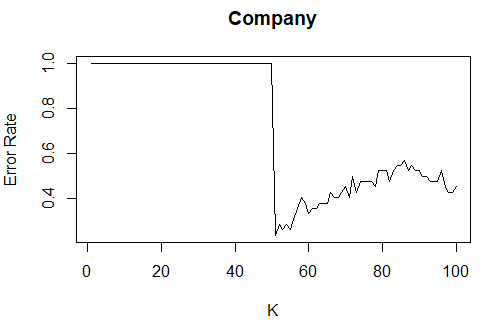
trial\_n[j]<-trial\_n[j]+test\_size

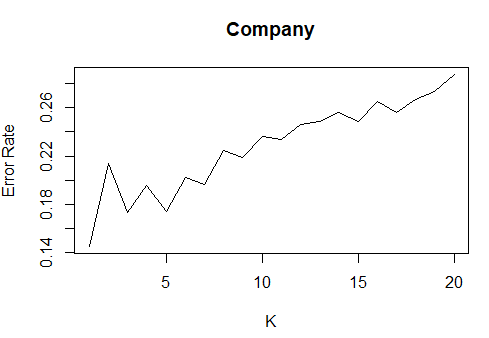
}

}

plot(1-trial\_sum / trial\_n,type="l",ylab="Error Rate", xlab="K",main="Company")

Результаты работы:





Листинг программы:

library(class)

s <- sample(0:5, 140, replace=T)

s <- replace(s, s == 0, "a")

s <- replace(s, s == 1, "b")

s <- replace(s, s == 2, "c")

s <- replace(s, s == 3, "d")

s <- replace(s, s == 4, "e")

s <- replace(s, s == 5, "f")

level <- sample(0:1, 140, replace=T)

security <- runif(140, 0, 5)

coating <- runif(140, 0, 5)

maneuver <- runif(140, 0, 5)

mobility <- runif(140, 0, 5)

company <- data.frame(Name = s, Level = level, Security = security, Coating = coating,

Maneuver = maneuver, Mobility = mobility)

colors <- c("a" = "red", "b" = "yellow", "c" = "blue", "d" = "white", "e" = "black")

plot(NULL, NULL, type = "l", xlim = c(min(company[, 3]), max(company[, 3])),

ylim = c(min(company[, 5]), max(company[, 5])), xlab = 'y', ylab = 'x')

eDist <- function(u, v) {

sqrt(sum((u - v)^2))

}

# functions of cores:

CoreEpanechnikova <- function(r) {

if(abs(r) <= 1) {

return((3/4)\*(1 - (r^2)))

} else {

return(0)

}

}

k <- 6

xl <- company[, 3:5]

l <- dim(xl)[1]

n <- dim(xl)[2] - 1

col3 <- seq(from = min(company[, 3]), to = max(company[, 3]), by = 0.101)

col4 <- seq(from = min(company[, 5]), to = max(company[, 5]), by = 0.101)

for(i in col3) {

for(j in col4) {

point <- c(i, j)

distances <- c()

for(p in 1:l) {

distances[p] <- eDist(xl[p, 1:n], point)

}

orderedxl <- xl[order(distances), ]

sortedDistances <- sort(distances)

weights <- c()

for(t in 1:k) {

r <- sortedDistances[t] / sortedDistances[k + 1]

cores <- c(CoreEpanechnikova(r))

weights[t] <- cores[1]

}

orderedxl\_weighted <- cbind(orderedxl[1:k, ], weights)

classes <- orderedxl\_weighted[ , (n + 1):(n + 2)]

sumA <- sum(classes[classes$Name == "A", 2])

sumB <- sum(classes[classes$Name == "B", 2])

sumC <- sum(classes[classes$Name == "C", 2])

sumD <- sum(classes[classes$Name == "D", 2])

sumE <- sum(classes[classes$Name == "E", 2])

answer <- matrix(c(sumA, sumB, sumC, sumD, sumE), nrow = 1, ncol = 4, byrow = T, list(c(1), c('A', 'B', 'C', 'D', 'E')))

points(point[1], point[2], pch = 1, bg = "white", col = colors[which.max(answer)])

}

}

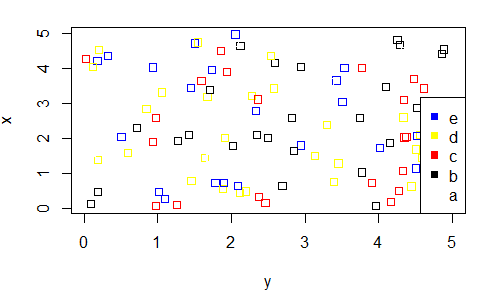
for (i in 1:l) {

points(company[i, 3], company[i, 5], pch = 0.000001, bg = colors[company$Name[i]], col = colors[company$Name[i]])

}

legend("bottomright", c("e", "d", "c", "b", "a"), pch = c(15,15,15),

col = c("blue", "yellow", "red", "black", "white"))



Вывод: были сформированы практические навыки разработки программ с использованием метрических методов классификации многомерных объектов пересекающихся классов на языке R.