**Задание ЛОГ2 - Построение и применение множественной логистической регрессионной модели**

Выполнил студент 2 курса

группы 09-715(1)

Санамян Артак Размикович

**Текст задания:**

1. Найдите коэффициенты логистической регрессионной модели.
2. Выполните прогноз переменной отклика для нескольких новых объектов и нескольких объектов обучающей выборки.
3. Сделайте выводы о качестве логистической регрессионной модели.
4. Результаты оформите в виде Word-файла и прикрепите его здесь (Весь R-код поместите в конец файла как приложение).

Источник числовых данных, использованный для выполнения задания: http://archive.ics.uci.edu/ml/datasets/banknote+authentication

**Описание выполненной работы:**

Данные были извлечены из изображений, которые были взяты для оценки процедуры аутентификации банкнот. Для оцифровки использовалась промышленная камера, обычно используемая для проверки печати. Изображения имеют 400 x 400 пикселей. Для извлечения признаков из изображений использовалось вейвлет-преобразование.

Необходимо предложить инструментарий для поддержки принятия решения о подлинности банкноты (или о подделке банкноты).

Переменная отклика – подлинность образца банкноты:

* 1 – подлинная банкнота;
* 0 – поддельная банкнота.

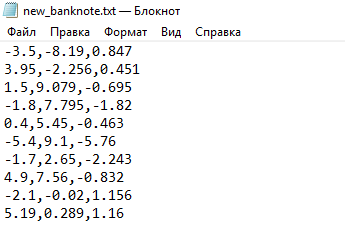
Предикторами являются следующие признаки:

* дисперсия Вейвлет-Преобразованного изображения;
* коэффициент асимметрии Вейвлет-Преобразованного изображения;
* энтропия изображения.

Коэффициенты логистической регрессионной модели:



Выполним прогноз переменной отклика для нескольких новых объектов. Новые данные хранятся в файле new\_banknote.txt:

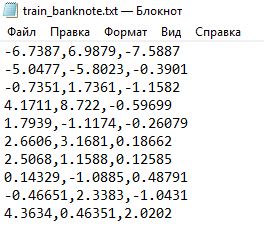


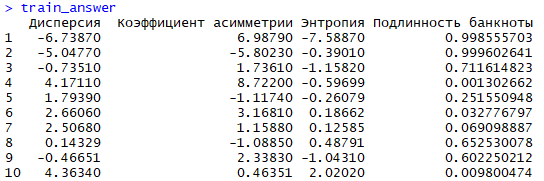


Пусть «пороговое» значение вероятности того, что банкнота окажется подлинной, равно 0.6.

Таким образом, мы получили, что образцы банкнот 1, 6, 7, 9 логистическая модель отнесет к классу подлинных, а образцы банкнот 2, 3, 4, 5, 8, 10 – к классу поддельных.

Выполним прогноз переменной отклика для нескольких объектов обучающей выборки. Данные хранятся в файле train\_banknote.txt:





Как видно, образцы банкнот 1, 2, 3, 8, 9 логистическая модель отнесет к классу подлинных, а образцы банкнот 4, 5, 6, 7, 10 – к классу поддельных. Все 10 предсказаний оказались верными.

**Код программы:**

setwd("C://Users/pc/Documents/Магистратура/2 семестр/R/task7 Log2")

getwd()

#очищаем рабочее пространство

rm (list=ls())

# Прочтём данные о банкнотах

bankdata <- read.csv("banknote.txt", sep=",", head=FALSE)

bankdata = bankdata[-3]

bankdata

# Озаглавим столбцы bankdata

colnames(bankdata)<-c("Variance", "Skewness", "Entropy", "Authenticity")

bankdata

# Представим озаглавленные данные bankdata как таблицу

B<-as.data.frame(bankdata)

B

# Вызовем метод максимизации функции правдоподобия glm

mymodel<-glm(formula = Authenticity ~ Variance + Skewness + Entropy,

family = binomial, data = B)

# Запишем найденные коэффициенты в вектор Teta

Teta=coef(mymodel)

Teta

# Определим подлинность новых банкнот

newBanknote <- read.csv("new\_banknote.txt", sep=",", head=FALSE)

# Озаглавим столбцы newBanknote

colnames(newBanknote)<-c("Variance", "Skewness", "Entropy")

newBanknote

newPredict<-predict(mymodel,

newdata = newBanknote,

type="response")

newPredict

# Число новых банкнот

nnew = length(newBanknote[,1])

nnew

solutions = matrix(1:(nnew\*4),ncol=4)

solutions[,1] = newBanknote[,1]

solutions[,2] = newBanknote[,2]

solutions[,3] = newBanknote[,3]

solutions[,4] = newPredict

# Сохраним матрицу как фрейм данных

answer = as.data.frame(solutions)

# Добавим заголовки

colnames(answer) = c("Дисперсия", " Коэффициент асимметрии", "Энтропия", "Подлинность банкноты")

answer

# Определим подлинность банкнот из обучающей выборки

trainBanknote <- read.csv("train\_banknote.txt", sep=",", head=FALSE)

# Озаглавим столбцы trainBanknote

colnames(trainBanknote)<-c("Variance", "Skewness", "Entropy")

trainBanknote

trainPredict<-predict(mymodel,

newdata = trainBanknote,

type="response")

trainPredict

# Число новых банкнот

nlen = length(trainBanknote[,1])

nlen

#аналогично для банкнот из обучающей выборки

train\_solutions = matrix(1:(nlen\*4),ncol=4)

train\_solutions[,1] = trainBanknote[,1]

train\_solutions[,2] = trainBanknote[,2]

train\_solutions[,3] = trainBanknote[,3]

train\_solutions[,4] = trainPredict

# Сохраним матрицу как фрейм данных

train\_answer = as.data.frame(train\_solutions)

# Добавим заголовки

colnames(train\_answer) = c("Дисперсия", " Коэффициент асимметрии", "Энтропия", "Подлинность банкноты")

train\_answer