Министерство образования и науки Российской Федерации

Севастопольский государственный университет

Кафедра ИС

Отчет

По дисциплине: “Интеллектуальный анализ данных”

Лабораторная работа №5

“Линейный дискриминантный анализ. Построение

канонических и классификационных функций”

Выполнил:

ст.гр. ИС/б-32

Долженко И.А.

Проверила:

Сырых О.А.

Севастополь

2020

1 ЦЕЛЬ РАБОТЫ

Закрепить теоретические знания и приобрести практические навыки в проведении дискриминантного анализа по экспериментальным данным. Исследовать возможности языка R для проведения дискриминантного анализа.

2 ХОД РАБОТЫ

**1. Подготовка данных для дискриминантного анализа.**

С помощью команды проведем кластерный анализ методом k-средних на 3 кластера:

.cluster <- KMeans(model.matrix(~-1 + Безопасность + Здоровье + Индекс.качества.жизни + Климат + Окружающая.среда + Покупательная.способность + Стоимость.жизни + Стоимость.жилья.к.доходу + Транспорт, Dataset), centers = 3, iter.max = 10, num.seeds = 10)

Результаты разбиения на кластеры добавим к данным.

**2. Создание тренировочной выборки.**

Создадим выборку строк от 1 до последней с шагом 5:

Dataset.train <- Dataset [seq (1, nrow(Dataset),5),]

Оставшиеся данные сформируем в выборку для последующей проверки полученной классификации:

Dataset.unknow <- Dataset [-seq (1, nrow(Dataset),5),]

**3. Проведение дискриминантного анализа.**

В качестве переменных используем столбцы 2:10, классификация проводится по столбцу 11:

dataset.lda <- lda (Dataset.train[, 2:10], Dataset.train[,11])

**4. Построение дискриминантных функций.**

Для построения дискриминантных функций из проведенного анализ используются Коэффициенты линейных дискриминантов:

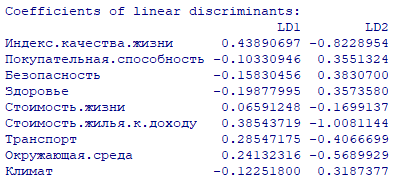


Рисунок 1 – Коэффициенты линейных дискриминантов

Исходя из результатов получаем две дискриминантные функции:

1)

2)

**5. Проведем классификацию и проверку оставшихся данных:**

dataset.ldap <- predict(dataset.lda, Dataset.unknow [, 2:10])$class

Результат полученной классификации:



Рисунок 2 – Результат полученной классификации

Для проверки созданной модели построим матрицу неточностей:

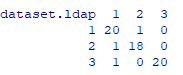


Рисунок 3 – Матрица неточностей

По данной матрице видно, что тренировочная выборка привела к построению гипотезы, по которой 2 объекта первого класса и 1 объект из второго класса попали не в «свою» группу.

**6. Проведем шаговую процедуру выбора переменных для построения дискриминантной модели.**

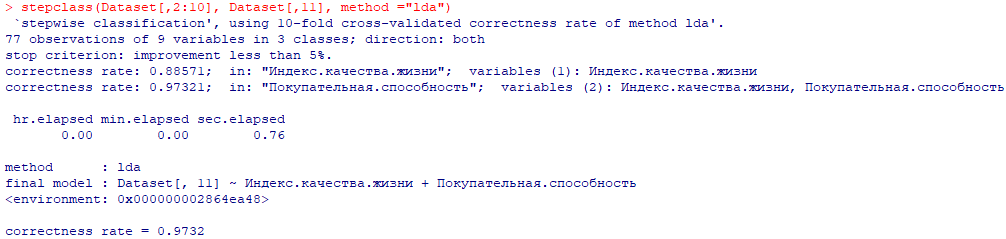


Рисунок 4 – Результат выбора переменных

**7. Построим дискриминантную модель с выбранными переменными, составим уравнение дискриминантной функции.**

dataset.lda <- lda (Dataset.train[, 2:3], Dataset.train[,11])

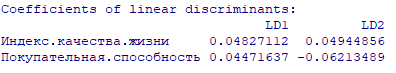


Рисунок 5 – Коэффициенты линейных дискриминантов

Исходя из результатов получаем две дискриминантные функции:

1)

2)

**8. Вывести показатели оценки качества построенной модели: матрица неточностей, ошибку распознавания, расстояние Махалонобиса.**

dataset.ldap <- predict(dataset.lda, Dataset.unknow [, 2:3])$class

table (dataset.ldap, Dataset.unknow[,11])

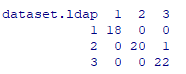


Рисунок 6 – Матрица неточностей

Err\_S <- mean (Dataset.unknow[,11] != dataset.ldap)



Рисунок 7 – Ошибка распознавания

mahDist <- dist(dataset.lda$means %\*% dataset.lda$scaling)

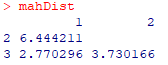


Рисунок 8 – Расстояние Махалонобиса

ВЫВОД

В ходе выполнения лабораторной работы была создана тренировочная выборка, выполнен дискриминантный анализ и проведена классификация, а для ее проверки была построена матрица неточностей. По полученной матрице видно, что тренировочная выборка привела к построению гипотезы, по которой 2 объекта первого класса и 1 объект из второго класса попали не в «свою» группу.

Во второй части работы была проведена шаговая процедура выбора переменных для построения дискриминантной модели с помощью функции stepclass() из пакета klaR. Построена дискриминантная модель с переменными “Индекс качества жизни” и “Покупательная способность”, составлены уравнения дискриминантных функций. По матрице неточностей видно, что выборка привела к построению гипотезы, по которой всего 1 объект третьего класса попал не в «свою» группу, а ошибка распознавания равна 0.01639.