Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет Информационных технологий и управления

Кафедра Интеллектуальных информационных технологий

Индивидуальная практическая работа №2

По дисциплине статические основы индукционного вывода

на тему:

ДИСКРИМИНАНТНЫЙ АНАЛИЗ С ГОТОВЫМИ КЛАСТЕРАМИ В MS EXCEL

Выполнил Мулярчик Д.С.

Проверил Ефремов А.А

Минск 2023

Задача

1. Подобрать в открытых источниках data set, состоящий из не менее чем 4 переменных. Количество наблюдений – не менее 100. Наблюдения должны быть предварительно распределены по кластерам. Например, если речь кластеризации моделей ноутбуков, то варианты кластеров: «Для геймеров», «Для офисных работников», «Для путешествий», «Для пенсионеров» и т.п.
2. Разделить исходную выборку на 2 части: обучающую и тестовую.
3. Выполнить дискриминантный наблюдений из тестовой выборки, пользуясь методическими указаниями из примера ниже.
4. В отчёте представить: постановку задачи с описанием переменных (А), фрагмент таблицы с исходными данными (Б), расстояния до кластеров (В), оценку точности анализа через расчёт процента ошибок (Г).

Оценка типа звезды на основе дискриминативного анализа

В качестве dataseta выбран https://www.kaggle.com/datasets/deepu1109/star-dataset

1. Тип звезды
2. Температура - x1
3. Яркость - x2
4. Радиус - x3
5. Абсолютная величина - x4

Показатели потенциальной звезды x1 = 6133.3, x2= 0.057916, x3=0.170004, x4 = 13.8.

Требуется:

1. построить множественную дискриминантную модель и с ее помощью отнести потенциальный цветок к одному из трёх классов.

2. Построить регрессионную дискриминантную модель, найти граничное значение и отнести цветок к одному из классов.

Temperatu	Luminosity	Radius(R/F	Absolute r	Star type
3068	0.0024	0.17	16	1
3042	0.0005	0.1542	16	1
2600	0.0003	0.102	18	1
2800	0.0002	0.16	16	1
3600	0.0029	0.51	10	2
3129	0.0122	0.3761	11	2
3134	0.0004	0.196	13	2
25000	0.056	0.0084	10	3
7740	0.00049	0.01234	14	3
7220	0.00017	0.011	14	3

Пункт 1

Рассчитаем центроиды для каждого типа звезды:

1. x1 = 2877.5, x2 = 0.00085, x3= 0.13155, x4= 16.5

   o1 = 0.36935, o2 = 0.15384, o3 = 0.63517, o4 = 0.14724

2. x1 = 3287.667, x2 = 0.000516, x3= 0.3607, x4= 11.33

o1 = 0.44274, o2 = 0.54163, o3 = 0.73925, o4 = 0.27495

3. x1 = 13320, x2 = 0.0188, x3= 0.01058, x4= 12.66

o1 = 0.92745, o2 = 0.27495, o3 = 0.57632, o4 = 0.19472

Для оценки коэффициентов проверки работоспособности модели проведём используем её для тестовых данных и сравним результаты с их классами, для этого посчитаем расстояние этого цветка до каждого класса по следующей формуле:

$$D(x, \overline{x^s}) = \sqrt{(\frac{x_1 - \overline{x_1^s}}{\sigma_1^s})^2 + (\frac{x_2 - \overline{x_2^s}}{\sigma_2^s})^2 + (\frac{x_3 - \overline{x_3^s}}{\sigma_3^s})^2 + (\frac{x_4 - \overline{x_4^s}}{\sigma_4^s})^2}$$

D1 = 88.14%

D2 = 64.27%

D3 = 77.44%

Потенциальная звезда относят к группе 2 так как значение D2 < D3 < D1.