

## **2. Исследование модели системы биометрической аутентификации с использованием методов кластерного анализа**

### **Цель работы**

Исследовать модель системы бинарной классификации «Свой-Чужой» с использованием технологий статистического моделирования и кластерного анализа данных.

### **Постановка задачи и сведения из теории**

#### Постановка задачи.

1. В соответствии с номером варианта, с помощью генератора псевдослучайных чисел необходимо сгенерировать искусственные выборки объектов двух классов («Свой», «Чужой») на основе моделирования соответствующих *двумерных* эмпирических функций распределения вероятностей.

Например, если класс имеет двумерную функцию плотности вероятности, соответствующую нормальному гауссовому закону, в случае *некоррелированных* параметров его двумерная функция плотности вероятности имеет вид:

$$N(\mu_x, \mu_y; \sigma_x^2, \sigma_y^2) : \text{pdf}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left(\frac{(x-\mu_x)^2}{2\sigma_x^2} + \frac{(y-\mu_y)^2}{2\sigma_y^2}\right)}. \quad (2.1)$$

Построить сгенерированные точки на одном графике.

2. Рассчитать центроиды обоих классов.

3. Сгенерировать новые тестовые выборки обоих классов аналогично п. 1.

4. Для каждого образа тестовой выборки оценить его принадлежность одному из классов методами:

- ближайшего соседа (нечетные варианты);
- k-средних (четные варианты).

5. Провести аналогичные пп. 1-4 расчеты для случая *коррелированных параметров*. Для этого необходимо осуществить поворот каждой из главных осей распределения на произвольный угол.

6. Рассчитать эмпирические коэффициенты корреляции каждого класса, построить корреляционную матрицу.

7. Вычислить расстояние Махаланобиса до центра каждого класса для произвольного образца, сравнить его с евклидовым расстоянием.

8a. Выбрав угол наклона дискриминантной прямой, как у биссектрисы между главными полуосями функций распределения двух классов, смоделировать дискретные значения порогов принятия решения между двумя классами аналогично одномерному случаю. Для каждого значения порога рассчитать ошибки 1 и 2 рода классификации объектов тестовой выборки, построить ROC-кривую.

8b. Для каждого значения угла  $\varphi_j = \frac{\pi}{M} j$ ,  $j = 0, 1, \dots, M - 1$ , построить семейство  $N$  эквидистантных параллельных дискриминантных прямых, расположенных между центрами обоих классов. Для каждого семейства рассчитать ошибки 1 и 2 рода классификации объектов тестовой выборки, построить ROC-кривую.

### Пример

На рис. 2.1 представлены гистограммы 2-х классов с гауссовыми функциями плотности распределения  $N_1(0, 0; 1, 0.25)$ ,  $N_2(2, 3; 4, 1)$ . Количество экземпляров каждого класса: 500 и 1000 соответственно.

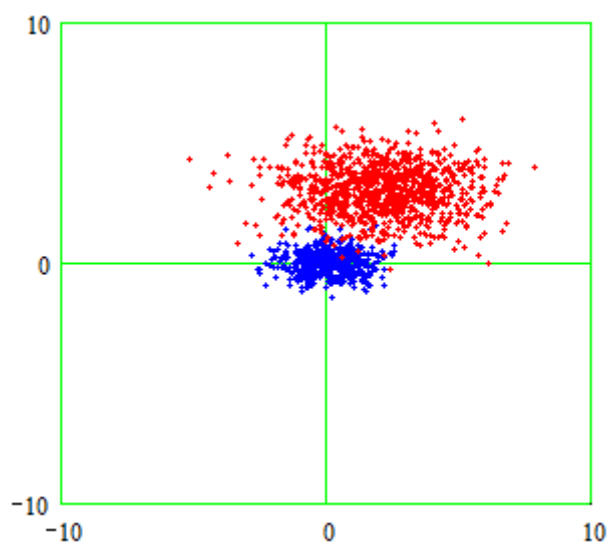


Рис. 2.1. Гистограммы классов «Свой» -  $N_1(0, 0; 1, 0.25)$ , «Чужой» -  $N_2(2, 3; 4, 1)$  с некоррелированными параметрами

На рис. 2.2 приведены гистограммы тех же классов, с коррелированными параметрами (углы ориентаций главных осей эллипса рассеивания:  $22.3^\circ$ ,  $76.4^\circ$  соответственно).

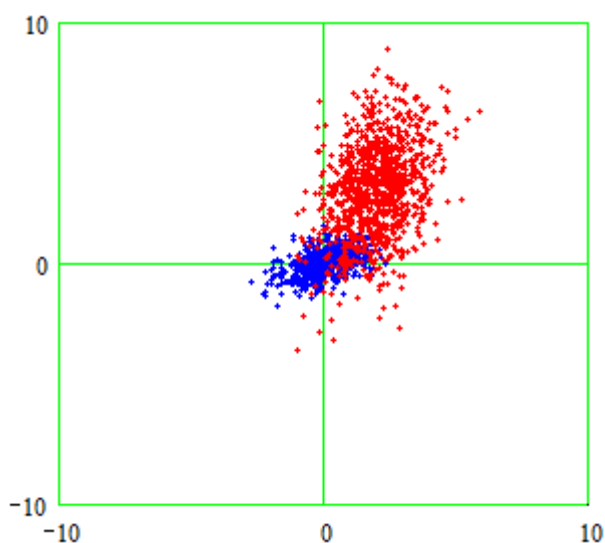


Рис. 2.1. Гистограммы классов «Свой», «Чужой» с коррелированными параметрами

## Варианты работы

Таблица 1. Варианты функций распределений образов классов «Свой», «Чужой»

	Треугольное	Гауссово	Пуассоново	Ирвина-Холла	Гамма
Треугольное	1	2	3	4	5
Гауссово	6	7	8	9	10
Пуассоново	11	12	13	14	15
Ирвина-Холла	16	17	18	19	20
Гамма	21	22	23	24	25

## Требования к отчету

Отчет должен содержать: титульный лист; цель работы; постановку задачи; выбранные законы распределений классов; описание использованных алгоритмов моделирования дискретных законов соответствующих распределений; распределения образов обоих классов, приведенные на одном графике для случаев коррелированных и некоррелированных параметров; коэффициенты дискриминантных прямых и соответствующие матрицы ошибок; ROC-кривые.