

1. Исследование метрик и характеристик систем биометрической аутентификации на основе принципов ROC-анализа

Цель работы

Исследовать принципы тестирования систем бинарной классификации «Свой-Чужой» по уровню ошибок I и II рода с использованием технологии ROC-анализа (англ. Receiver Operating Characteristic).

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 эмпирических функций распределения вероятностей одним из численных методов.

Например, если класс имеет функцию плотности вероятности, соответствующую нормальному гауссовому закону:

$$N(\mu_i, \sigma_i^2) : \text{pdf}_i(x) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}, \quad i = 1, 2, \quad (1.1)$$

моделирование можно осуществить с помощью метода Бокса–Мюллера (Box–Muller).

Необходимо для обоих классов выбрать различные значения математических ожиданий и дисперсий (рекомендуется дисперсию выбрать так, чтобы класс «Свой» имел более высокую локализацию, либо был менее размытым, чем класс «Чужой»). Математические ожидания классов необходимо выбрать так, чтобы осуществить моделирование для различных случаев, включая крайние случаи наложения классов (математические ожидания совпадают) и ситуацию разнесенных классов, когда функции плотности вероятности не пересекаются (либо пересекаются по уровню порядка 0.01);

Построить гистограммы распределений на одном графике.

2. Для различных равноотстоящих значений порогов, начиная от крайнего нижнего значения до крайнего верхнего, рассчитать компоненты матриц ошибок (Confusion Matrix): TP, TN, FP, FN, а также чувствительность ($Sp = TPR$) и специфичность ($Sp = 1 - FPR$).

3. По полученным значениям TPR и FPR построить ROC-кривую и рассчитать площадь под ней (Area Under Curve - AUC).

Пример

На рис. 1.1 представлены гистограммы 2-х классов с гауссовыми функциями плотности распределения $N_1(0,4)$, $N_2(8,9)$, а также огибающие, построенные по формуле (1.1). Количество экземпляров каждого класса: 500 и 1000 соответственно.

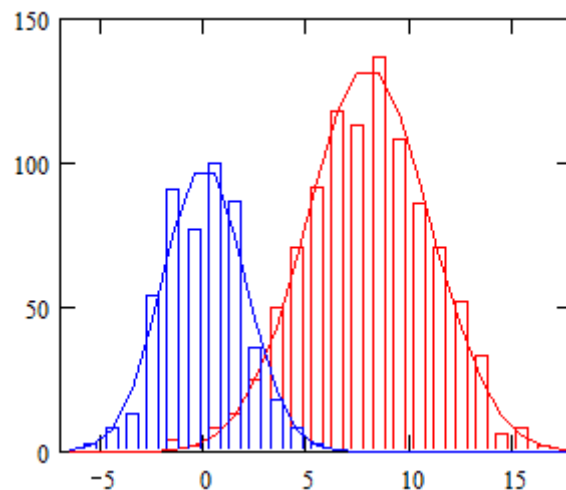


Рис. 1.1. Гистограммы классов «Свой» - $N_1(0,4)$, «Чужой» - $N_2(8,9)$

На рис. 2 показана ROC-кривая, построенная для 26 равноотстоящих значений порога, взятых от -7 до 18.

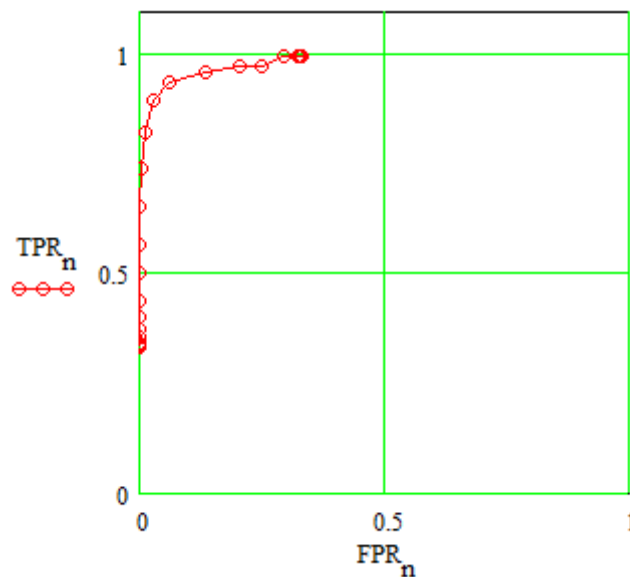


Рис. 1.2. ROC-кривая для порогового разделения классов на рис. 1.1

Варианты работы

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

	Треугольное $\max(0, 1 - x)$	Гауссово $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$	Рэлея $\frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$	Ирвина-Холла ($n = 3$)	Гамма $\Gamma(n, 1)$ ($n > 1$)
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; гистограммы распределения образов обоих классов, приведенные на одном графике; матрицы ошибок, полученные для различных пороговых величин; таблицу с вычисленными значениями показателей TPR, FPR; построенные ROC-кривые с рассчитанными значениями AUC.