

Использование модели CART в задаче двухфакторной биометрической аутентификации

Цель работы

Исследовать модель системы двухфакторной бинарной классификации «Свой-Чужой» с использованием аппарата деревьев регрессии и классификации (CART).

Постановка задачи и сведения из теории

Постановка задачи.

1. В соответствии с номером варианта (см. лаб. работу 2) сгенерировать искусственные выборки объектов двух классов («Свой», «Чужой») с коррелированными параметрами на основе моделирования соответствующих двумерных эмпирических функций распределения вероятностей.

2. Провести классификацию методом CART. Глубина дерева решений – 2 (рис. ##.1). Пороги классификатора (рис. ##.2) рассчитать по одному из критериев информативности в соответствии с номером варианта (см. табл. ##.1).

Критерий Джини (Gini impurity):

$$E = -(p_1 \log p_1 + p_2 \log p_2).$$

Критерий энтропии (Entropy impurity):

$$G = 1 - p_1^2 - p_2^2.$$

Эмпирическая вероятность p_i — относительное количество элементов в узле, помеченном как класс $i \in \{1, 2\}$ в соответствии с критерием голосования (большинства).

3. Построить результирующие матрицы ошибок. Сравнить результаты с многоклассовой классификацией по правилам слияния AND/OR с порогами на уровне равных ошибок (EER).

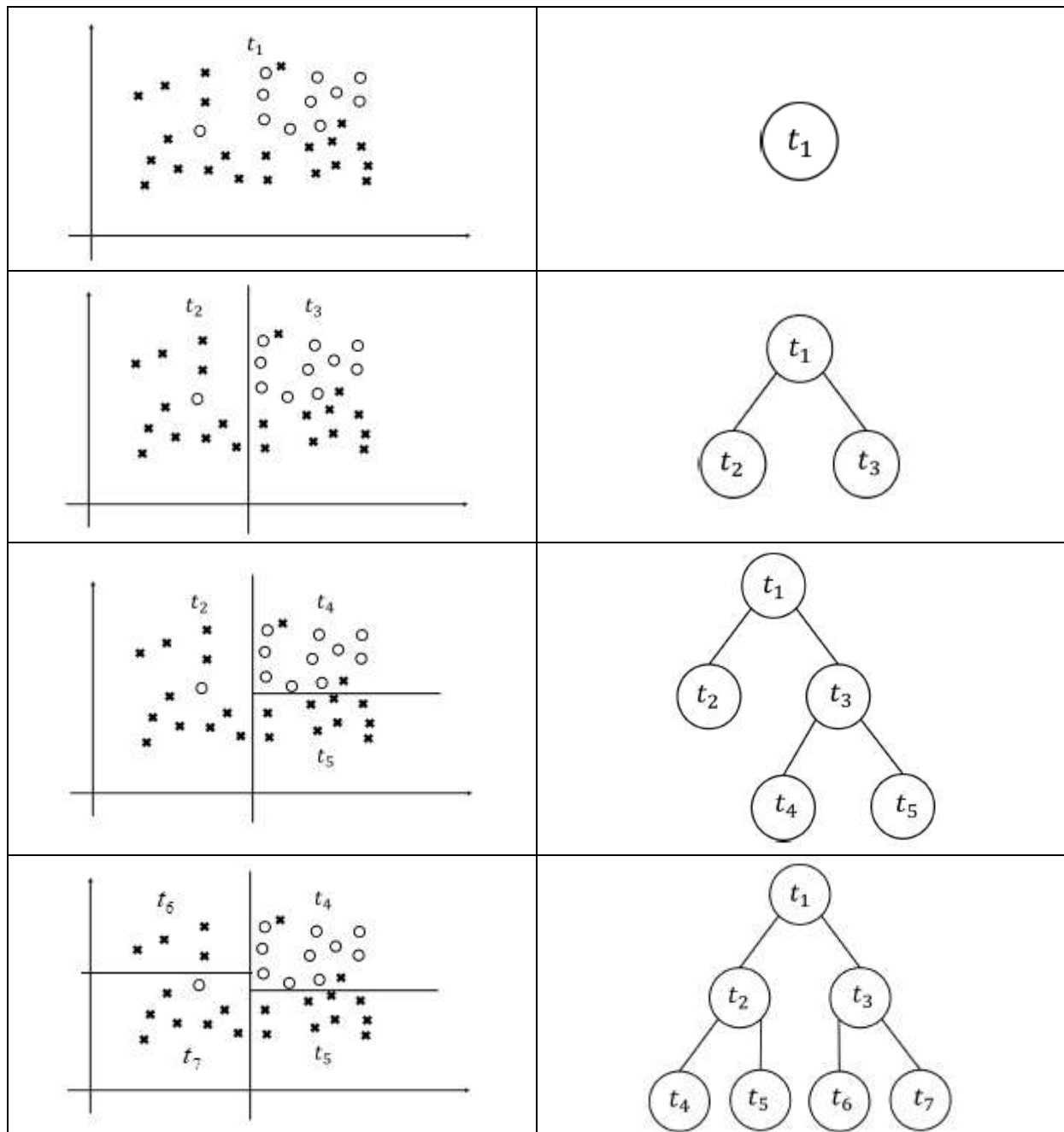


Рис. ##.1. Общая схема классификации «Свой», «Чужой» для дерева CART глубины 2

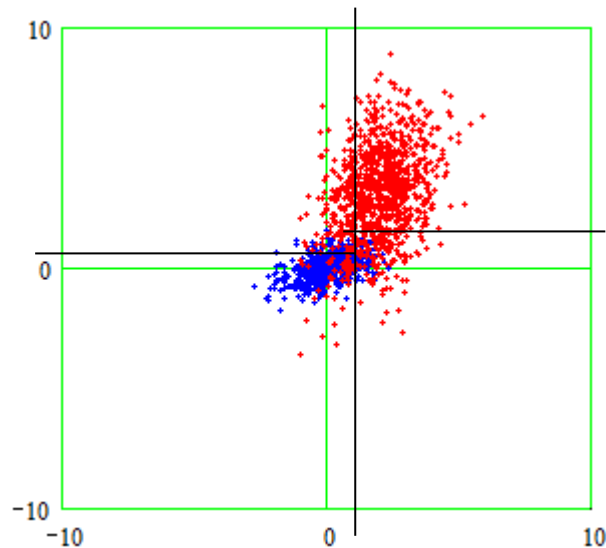


Рис. ##.2. Распределения классов «Свой», «Чужой» и построенные пороги для дерева CART глубины 2

Варианты работы

Таблица ##.1. Варианты критерия информативности

№ вар.	Критерий
$\text{mod}(n,2)=1$	Энтропия
$\text{mod}(n,2)=0$	Джини

Требования к отчету

Отчет должен содержать: титульный лист; цель работы; постановку задачи; законы распределений классов; распределения образов классов и пороги классификации, приведенные на одном графике; матрицы ошибок.