

Оценка оптимального объёма выборки для задач классификации*

Харатьян А. С., Катруца А. М.¹, Стрижов В. В.²

haratyan.as@phystech.edu; aleksandr.katrutsa@phystech.edu;

strijov@phystech.edu

¹Московский физико-технический институт, Москва, Россия; ²Вычислительный центр им. А. А. Дородницына ФИЦ ИУ РАН, Москва, Россия

В статье рассматривается задача выбора оптимального числа объектов выборки для их классификации. Исследуется использование порождающих и разделяющих вероятностных моделей бинарной классификации. Обсуждается проблема медицинской диагностики пациентов. Определяется понятие достаточности объёма выборки. Показывается, какими методами возможно выбрать оптимальное количество объектов, обеспечивающее необходимую точность классификации объектов. В работе рассматривается, применение каких критериев выявляет наилучшее качество классификации. Приводится теоретическое и практическое обоснование предложенных критериев.

Ключевые слова: *определение оптимального объёма выборки.*

*Работа выполнена при финансовой поддержке РФФИ, проект №00-00-00000. Научный руководитель: Стрижов В. В. Консультант: Катруца А. М.