

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 данной статье рассматривается задача подбора оптимального числа объектов выборки для их классификации. Проблема находит применение в ряде научных областей. Например, в медицине постановка диагноза болезни пациентам основывается на результатах анализов биомедицинских данных, получение которых либо несёт много расходов, либо крайне затруднительно. В таком случае необходимо подобрать оптимальное количество образцов, достаточного для обеспечения необходимой точности постановки диагноза. При увеличении этого числа процент правильной классификации объектов меняется незначительно, а при уменьшении заметно сокращается. В качестве другого примера можно привести число подопытных животных, необходимых для проверки научных гипотез. В этом случае также требуется минимизировать количество образцов данных для достижения достаточной точности. В работе рассмотрены различные критерии и показывается, применение каких из них выявляет наилучшее качество классификации.

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.