

Оценка оптимального объёма выборки для задач классификации*

Мурлатов А. С., Катруца А. М.¹, Стрижов В. В.² stanislav.murlatov@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 [1].

Определяется понятие достаточности объёма выборки [2]. Показывается, какими методами возможно выбрать оптимальное количество объектов, обеспечивающее необходимую точность классификации объектов. В работе рассматривается, применение каких критериев выявляет наилучшее качество классификации [3]. Приводится теоретическое и практическое обоснование предложенных критериев.

Ключевые слова: определение оптимального объёма выборки.

Литература

[1] Motrenko Anastasiya, Strijov Vadim V., Weber Gerhard-Wilhelm. Sample size determination for logistic regression // J. Computational Applied Mathematics, 2014. Vol. 255. P. 743–752.

[2] Tseng Sheng-Tsaing, Tang Jen, Lin Chien-Hua. Sample size determination for achieving stability of double multivariate exponentially weighted moving average controller // Technometrics, 2007. Vol. 49. No. 4. P. 409–419.

[3] Martini Daniele De. Evaluating the risk in sample size determination // Communications in Statistics - Simulation and Computation, 2008. Vol. 37. No. 9. P. 1776–1784.

*Работа выполнена при финансовой поддержке РФФИ, проект No00-00-00000.
Научный руководитель: Стрижов В. В. Консультант: Катруца А. М.

Машинное обучение и анализ данных, 2018. Том 00, No 00.

