

Секция «Математика и механика»

Порождающий и разделяющий подходы к решению задачи классификации
малых выборок

Мотренко Анастасия Петровна

Студент

*Московский физико-технический институт, управления и прикладной математики,
Долгопрудный, Россия*

E-mail: pastt.petrovna@gmail.com

В задачах классификации требуется, по набору известных реализаций случайных величин, определить метку класса зависящей от них случайной величины. Для этого строится параметрическая модель, параметры которой оцениваются по обучающей выборке прецедентов [1, 2]. В случае, когда процедура оценки параметров включает восстановление условной плотности переменной класса при заданном значении независимой переменной, модель называется разделяющей. При порождающем подходе вначале оценивается функция совместного распределения зависимых и независимых переменных. Затем, с помощью формулы Байеса, выводится условное распределение зависимой переменной. Хотя разделяющий подход асимптотически доставляет большую точность прогнозирования [3], порождающий подход быстрее выходит на асимптоту [4]. То есть при малом объеме выборки точность прогнозирования порождающих алгоритмов может превысить точность разделяющих.

В качестве примера рассмотрим задачу прогнозирования инфаркта. Сложность решаемой задачи состоит в малом объеме выборки, приводящем к переобучению модели. Задача решается с помощью разделяющих и порождающих алгоритмов классификации, а также с использованием комбинированного подхода. Комбинированный подход заключается в использовании выпуклой линейной комбинации функций условного и совместного правдоподобия как функционала качества. Решаются задачи оценки параметров условного и совместного распределений, а также коэффициентов линейной комбинации.

Литература

1. McCullagh P., Nelder J. A. Generalized Linear Models. 2nd edition. Chapman and Hall, London. 1989.
2. Hosmer D., Lemeshow S. Applied Logistic Regression, 2nd Edition, John Wiley and Sons, 2000.
3. Efron, B. The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis. // Journal of the American Statistical Association. 1975. V. 70, No. 352, P. 892-898.
4. Ng, A. Y. and Jordan, M. I. On Discriminative vs. Generative Classifiers: A comparison of logistic regression and Naive Bayes. // Advances in Neural Information Processing Systems. 2002. V. 14, P. 841-848.