МОНОГРАФИИ НГТУ

Т. А. ГУЛЬТЯЕВА, А. А. ПОПОВ,
А. С. САУТИН

МЕТОДЫ СТАТИСТИЧЕСКОГО ОБУЧЕНИЯ В ЗАДАЧАХ РЕГРЕССИИ И КЛАССИФИКАЦИИ

Т. А. ГУЛЬТЯЕВА, А. А. ПОПОВ,
А. С. САУТИН

МЕТОДЫ СТАТИСТИЧЕСКОГО ОБУЧЕНИЯ В ЗАДАЧАХ РЕГРЕССИИ И КЛАССИФИКАЦИИ

НОВОСИБИРСК
2016

УДК 519.217+519.233.5
Г 944

Рецензенты:
д-р техн. наук, доцент *В.Б. Бериков*
д-р техн. наук, профессор *Б.Ю. Лемешко*

Гультяева Т.А.

Г 944 Методы статистического обучения в задачах регрессии и классификации: монография / Т.А. Гультяева, А.А. Попов, А.С. Саутин. – Новосибирск : Изд-во НГТУ, 2016. – 323 с. (серия «Монографии НГТУ»).

ISBN 978-5-7782-2817-7

В монографии рассматриваются вопросы статистического обучения в задачах построения регрессии по методу опорных векторов и в задачах классификации с использованием скрытых марковских моделей (СММ). Для решения задачи устойчивого оценивания модели регрессии по методу опорных векторов (SVM) в условиях зашумленных данных с помехой, имеющей распределение с большим эксцессом или имеющей характер асимметричного засорения, предложено использовать адаптивные и асимметричные функции потерь. Приводятся формулировки двойственных задач квадратичного программирования для этих случаев. Описывается метод квантильной регрессии на основе SVM для произвольной функции потерь. На его основе рассмотрен метод построения доверительных интервалов для отклика, а также непараметрический метод оценки неизвестной дисперсии ошибок наблюдений. Для построения компактной модели регрессии в условиях работы с выборками большого объема предлагаются алгоритмы построения разреженных решений в SVM. Показывается их эффективность в сравнении с классическим методом построения разреженных решений на основе функции нечувствительности Вапника. Описывается модификация SVM, позволяющая строить разреженные решения в условиях гетероскедастичности ошибок наблюдений. Приводятся результаты экспериментальных исследований по построению регрессионных моделей с использованием SVM при мультиколлинеарности данных, автокорреляции и гетероскедастичности ошибок наблюдений.

Приводятся результаты исследования подхода к решению задачи классификации наблюдаемых последовательностей, представленных скрытыми марковскими моделями, с использованием инициированных этими моделями признаков. С использованием метода статистического моделирования рассматривается поведение нескольких классификаторов, когда наблюдаемые последовательности подвергались искажению действием на них различных помех. Также проанализированы случаи, когда нарушены одни из априорных представлений либо о наблюдаемых последовательностях, либо о структуре скрытых марковских моделей, описывающих эти последовательности.

Книга будет полезна научным сотрудникам и специалистам, сталкивающимся в своей деятельности с необходимостью решения задач построения зависимостей и классификации последовательностей, а также студентам и аспирантам.

УДК 519.217+519.233.5

ISBN 978-5-7782-2817-7

T. A. GULTYAEVA, A. A. POPOV,
A. S. SAUTIN

METHODS OF STATISTICAL LEARNING IN SOLVING REGRESSION AND CLASSIFICATION PROBLEMS

NOVOSIBIRSK
2016

УДК 519.217+519.233.5
Г 944

Reviewers:
Prof *V.B. Berikov*, D. Sc. (Eng.),
Prof *B.Yu. Lemeshko*, D. Sc. (Eng.)

Gultyaeva T.A.

Г 944 Methods of statistical learning in solving regression and classification problems: Monograph / T.A. Gultyaeva, A.A. Popov, A.S. Sautin. – Novosibirsk: NSTU Publisher, 2016. – 323 pp. ("NSTU Monographs" series).

ISBN 978-5-7782-2817-7

The book deals with statistical learning in the construction of Models of regression and support vector classification problems with the use of Hidden Markov Models (HMM). To solve the problem of estimating the regression model by using the sustainable Support Vector Machines (SVM) technology under noisy data with an interference having a distribution with large kurtosis or having an asymmetric clogging character, the use of adaptive and asymmetric loss functions is proposed. The formulations of the dual quadratic programming problems for these cases are given. The quantile regression method based on SVM for arbitrary loss functions is described. Based on it the method of constructing confidence intervals for the response as well as a non-parametric method for estimating the unknown variance of observation errors are considered. To construct a compact regression model to deal with a large volume of samples the algorithms for constructing sparse solutions in SVM are proposed. Their effectiveness in comparison with the classical method of constructing sparse solutions based on the function of the Vapnik insensitivity is shown. The modification of SVM which gives possibilits to build sparse solutions under heteroscedasticity of observation errors described. The results of experimental studies on the construction of regression models using SVM with data multicollinearity, autocorrelation and heteroscedasticity of observation errors are given.

The results of studying the approach to solving the problem of classification of the observed sequences represented by HMM using features initiated by these models are presented. Using the method of statistical modeling the behavior of classifiers is considered when observed sequences are subject to distortion caused by various disturbances. Cases are also analyzed when some of the a priori representations of either any observed sequence or the structure of the HMM describing these sequences are violated.

The book will be of interest to researchers and specialists who have to solve problems of constructing dependences and sequence classifications as well as to undergraduate, graduate and postgraduate students.

УДК 519.217+519.233.5

ISBN 978-5-7782-2817-7

ОГЛАВЛЕНИЕ

CONTENTS