

On the application of machine learning methods for motor vehicle insurance tariffs

Andreas Christmann

University of Dortmund, Department of Statistics, D-44221 Dortmund, Germany

Abstract

The goals of the talk are twofold: we describe common features in data sets from motor vehicle insurance companies and we investigate a general strategy which exploits the knowledge of such features. The results of the strategy are a basis to develop insurance tariffs. The strategy is applied to a data set from 15 motor vehicle insurance companies. We use a nonparametric approach based on a combination of kernel logistic regression and ε -support vector regression. Both methods belong to the class of statistical machine learning methods based on convex risk minimization. Some recent results of robustness properties of such methods are also given.

Key words: Classification; Data Mining; Insurance tariffs; Kernel logistic regression; Machine learning; Regression; Robustness; Simplicity; Support Vector Machine; Support Vector Regression.

References:

- Christmann, A. (2004). *On a strategy to develop robust and simple tariffs from motor vehicle insurance data*. University of Dortmund. SFB-475. Preprint.
- Christmann, A., Rousseeuw, P.J. (2001). Measuring overlap in logistic regression. *Computational Statistics and Data Analysis*, **37**, 65-75.
- Christmann, A., Steinwart, I. (2003). *On robust properties of convex risk minimization methods for pattern recognition*. University of Dortmund, SFB-475, Technical Report 15/2003.
- Rousseeuw, P.J., Christmann, A. (2003). Robustness against separation and outliers in logistic regression. *Computational Statistics & Data Analysis*, **43**, 315-332.