A case study in the application of machine learning methods in non-life insurance

Philip Harmuth

2018-02-22

Contents

1 Motivation

2 Abstract

Xie (2015)

3 Software, languages and libraries

3.1 SAS - Statistical Analysis Software

3.2 Python

3.2.1 Libraries

3.3 R

3.3.1 Packages

4 Models

4.1 General machine learning heuristics

4.1.1 Bias-variance tradeoff

4.1.2 Hyper parameter tuning

4.1.3 Regularization and "the bet on sparsity"

4.1.4 Model validation

4.1.4.1 Cross-validation

4.1.4.2 Train / dev / test

4.1.4.3 Bayesian Approach

4.1.4.4 Bootstrapping

4.2 Neural networks

4.2.1 Vanilla-NN (NN basics)

4.2.1.1 Activation functions

4.2.1.2 Gradient descent for neural networks

4.2.2 Regularization

4.2.2.1 Norm regularization

4.2.2.2 Drop-out

4.2.2.3 Early stopping

4.3 Regularized general linear models

4.3.1 LASSO regression

4.3.2 Ridge regression

4.3.3 Elastic net regression

4.4 Boosting

4.4.1 Regularization

4.5 Meta ensemble methods

4.5.1 Stacking

5 Data

Description of data and insurance data in general

5.1 Exploratory data analysis

6 Predictive analysis

6.1 Performance measures

6.2 Neural network

7 Conclusions

8 Bibliography

9 Appendices

References

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.