ACTUARIAL DATA SCIENCE WORKING GROUP

Material for actuarial applications in industry

The Data Science Working Group has been again very active in 2021. We have significantly improved our available training material and the 2-day block course in October was the motivation to create and publish improved selflearning material in the area of machine learning for actuarial applications.

Selflearning material (tutorials)

By the time of writing, we have published ten tutorials introducing several methods, techniques and applications of machine learning for actuarial applications. The tutorials give a practical understanding of machine learning techniques exemplified on an actuarial application and a corresponding publicly available data set. The main benefits of the tutorials are the following:

- addressing an actuarial problem (e.g. pricing, mortality modeling),
- a short article providing the required mathematical theory,
- the example uses publicly available data,
- the code is available in form of a notebook (code with detailed comments).

An overview of the covered topics of the ten tutorials is shown below. All tutorials are selfcontained and hence you can start with the one you are most interested in.

How to start with them? That is really easy, just go to our website (see below), there you find the link to the article and the corresponding R or Python notebook. Then download the corresponding notebook and go through it on your computer:

- Use your browser to see the html-version, without executing the code by yourself.
- Use the Jupyter Notebook (www.jupyter.org) application to run the code by yourself.

An example of a notebook is shown on the next page.

A notebook contains live code, equations, visualizations and narrative text. And hence it is easy to follow and understand the steps and its rationales as well as the code. Lastly, download the notebook and apply it to your data!

1  Recap of GLMs, introduction to regression trees and neural networks

2  Best practice step-by-step guide to fit neural networks for actuarial applications

3  How to embed a GLM into a neural network and explore additional structure

4  Overview of boosting algorithms

5  Unsupervised learning: Clustering data into homogeneous groups

6  Recurrent neural networks (RNNs) are used on a mortality rate prediction problem

7  Overview of the three approaches to preprocess text data with NLP

8  Overview of tools for explaining and interpreting black box machine learning models like boosted trees or deep neural networks for P&C pricing

9  Convolution Neural Networks (CNN) to detect anomalies in mortality rates

10  LocalGLMnet: a deep learning architecture for actuaries, allowing variable selection and nice interpretations

In case that our learning material does not fit your needs, then we recommend to check the following free online learning material, as well targeted for actuaries:

- «Introduction to Machine Learning», Michael Mayer, Actuary SAA
- «Insurance Data Science: Use and Value of Unusual Data», University of Lausanne and Swiss Association of Actuaries
- «Insurance Analytics, A Primer», University of Lausanne and Swiss Association of Actuaries

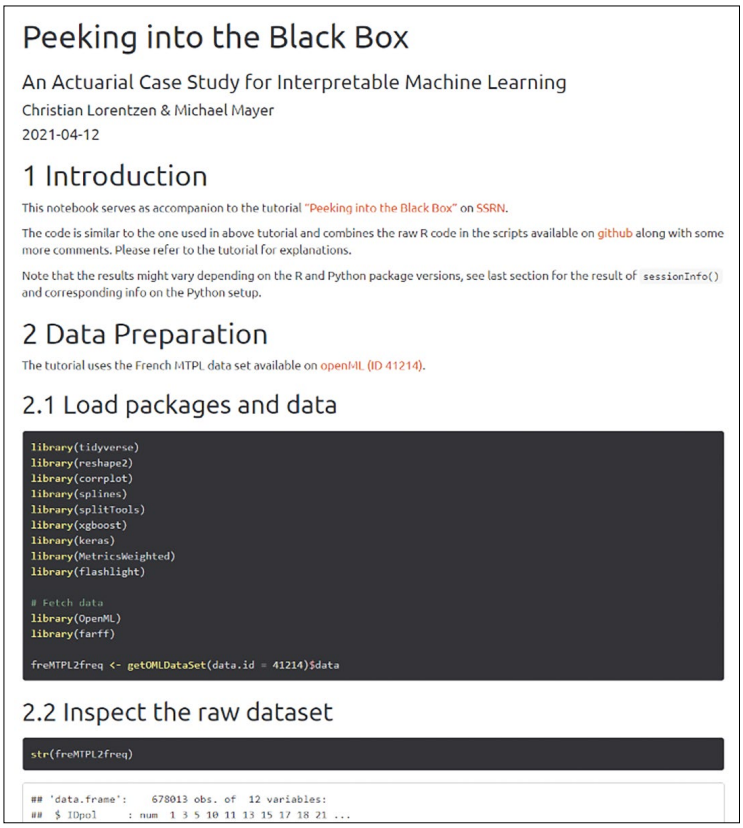
Actuarial Modeling vs. Standard Machine Learning

In industry, we have observed that there is sometimes confusion about the similarities and differences between actuarial modeling and standard machine learning approaches.

Due to these discussions, we have worked out an overview table as a foundation for discussions between actuaries, data scientists, quantitative specialists and the management. We are keen to hear your comments and improvements!

Visit our website www.actuarialdatascience.org and discover the tutorials and much more material about Actuarial Data Science.

Jürg Schelldorfer



Peeking into the Black Box

An Actuarial Case Study for Interpretable Machine Learning

Christian Lorentzen & Michael Mayer

2021-04-12

1 Introduction

This notebook serves as accompanion to the tutorial "Peeking into the Black Box" on SSRN.

The code is similar to the one used in above tutorial and combines the raw R code in the scripts available on github along with some more comments. Please refer to the tutorial for explanations.

Note that the results might vary depending on the R and Python package versions, see last section for the result of `sessionInfo()` and corresponding info on the Python setup.

2 Data Preparation

The tutorial uses the French MTPL data set available on openML (ID 41214).

2.1 Load packages and data

```
library(tidyverse)
library(reshape2)
library(corrplot)
library(splines)
library(splitTools)
library(xgboost)
library(keras)
library(MetricsWeighted)
library(flashlight)

# Fetch data
library(OpenML)
library(farff)

freMTPL2freq <- getOMLDataSet(data.id = 41214)$data
```

2.2 Inspect the raw dataset

```
str(freMTPL2freq)
```

```
## 'data.frame':    678013 obs. of  12 variables:
##  $ IDpol     : num  1 3 5 10 11 13 15 17 18 21 ...
```

A notebook to a tutorial looks as follows. Use the QR code for more information.

	Insurance Risk Modeling	Standard Machine Learning
Foundation	Distribution and uncertainty	Point estimate and algorithm
Mathematical foundation	Statistical model	Numerical optimization
Modeling target	Probabilistic forecast	Point forecast
Statistical distributions	Non-Gaussian (asymmetric, skewed)	Gaussian (symmetric)
Signal-to-noise (SNR) ratio	Small	High
Mathematical model selection «criteria»	• **Predictability** (insample) • **Stability and robustness** (long-term) • Smoothness • Parsimony • Interpretability / explainability • -	• **Predictability** (out-of-sample) • **Stability and robustness** (short-term) • - • Anti-parsimony • Black-box • Computability
Non-mathematical model selection «criteria»	• Causality / truth between predictors and predictant • Inclusion of expert knowledge • Human adjustability of models	• Correlation, train/test paradigm • - • -
Non-technical considerations	• Regulatory framework • Political and social aspects	• Ethics and fairnesss • Accountability and transparency
Professional associations	• Professional standards	• Ethical codes of conduct