A Framework for
Scorecard Modelling using 

Gero Szepannek

Statistics, Business Mathematics and Machine Learning

School of Business Studies

Short Bio

Gero Szepannek
Statistics, Business Mathematics
and Machine Learning
Dept. Economics
Stralsund University of Applied Sciences

1976 — 1997 — 2009 — 2016

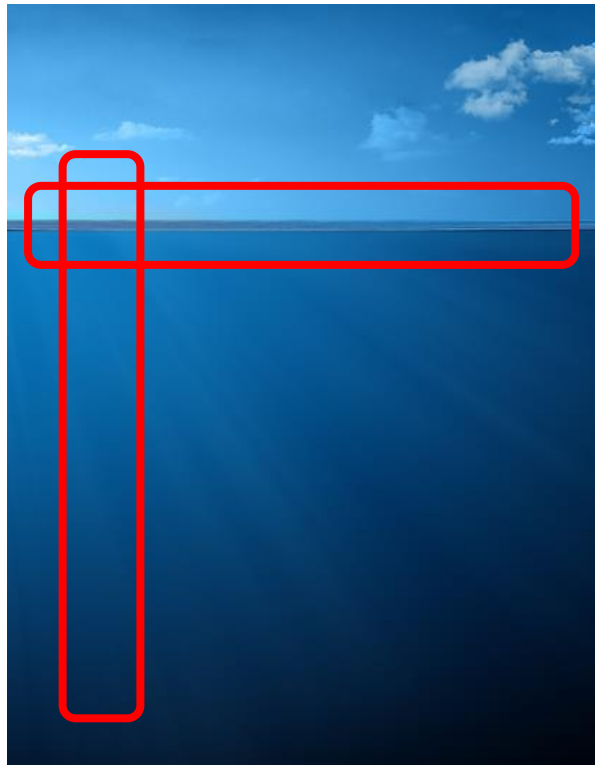
technische universität dortmund

Fraunhofer IDMT

Santander CONSUMER BANK

HOST Hochschule Stralsund

Two years ago… (Bischl, Szepannek and Kühn, 2016)



Typical **business applications** (cf. Szepannek et al., 2017)

- In depth understanding single tasks.
- High degree of business knowledge integration.
- Low degree of automatization.

Typical ML **benchmark studies** focus on

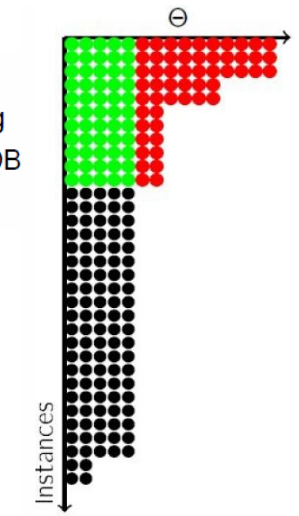
- …testing many algorithms
- …on many data sets
- High degree of automatization.
- Low degree of business understanding.

(…cf. e.g. Baesens et al., 2003, Lessmann et al., 2015, Brown and Mues, 2012, Szepannek et al., 2008, …)

- Class imbalance
- Systematic hyperparameter tuning
- Benchmark experiment on large DB
- Realistic evaluation

mlR

(Szepannek et al., 2010, Bischl et al., 2016)



Gero Szepannek
Stralsund University of
Applied Sciences

Idea of the Project

offers:

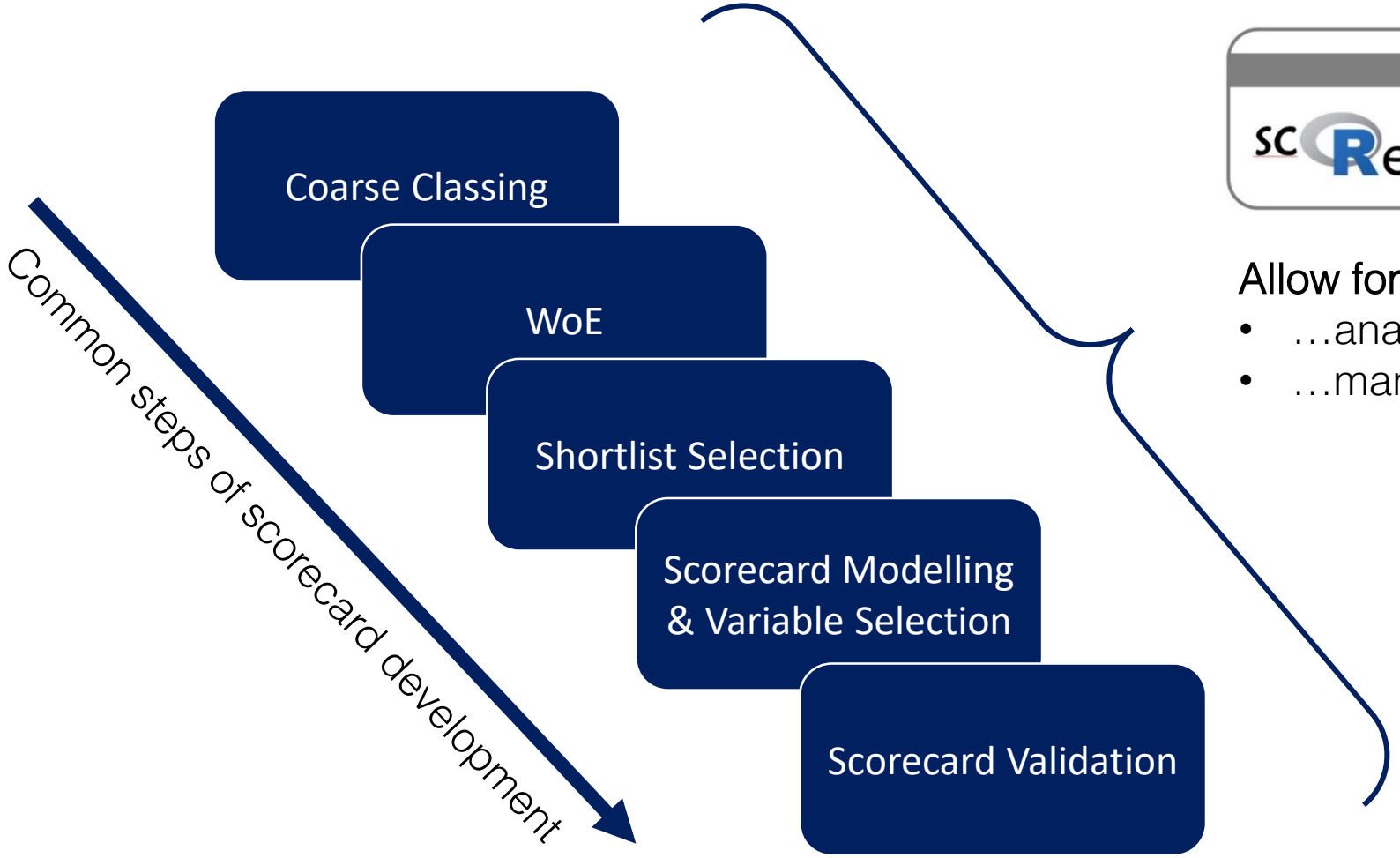
…free access to cutting edge methodology from statistics / ML community

Business Requirement

- Integration of expert business knowledge in modelling process
- Proper comparison with existing standards

Gero Szepannek
Stralsund University of Applied Sciences

The scoRe Package: Basic Structure



Allow for:

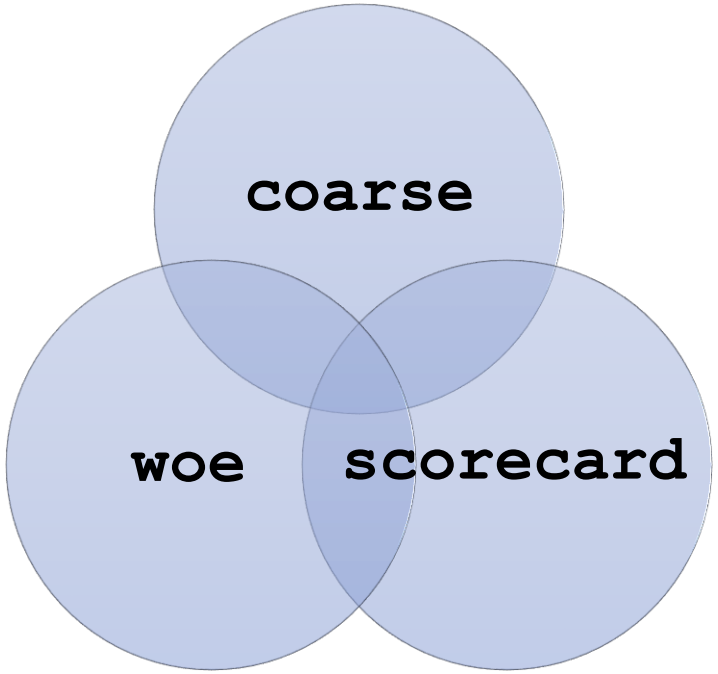
- …analysis & visualization
- …manual interaction

Gero Szepannek
Stralsund University of
Applied Sciences

Heart of the Package: 3 Classes

Common misunderstanding:

Preprocessing builds 1st part of the model!



Requirement of proper definition of objects and corresponding `predict()` methods!

…not given by existing implementations.

…several summary and plot methods to allow for an expert analysis & knowledge integration.

Gero Szepannek
Stralsund University of
Applied Sciences

Coarse Classing

- Initial automatic coarse classing as a starting point
  - …by either optimizing IVs
  - …or χ^2 significance tests on differences in BR.
- Output object of class coarse stored.
- …ready for subsequent analysis and expert changements.

coarse {scoRe} R Documentation

Automatic coarse classing

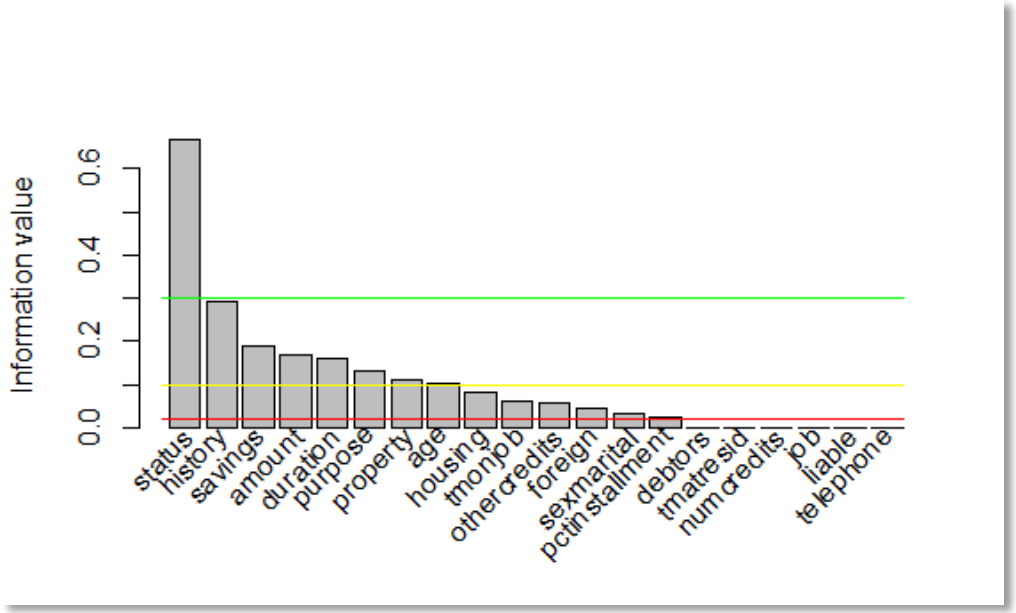
Description

Automatic coarse classing.

Usage

```
coarse(x, ...)

## S3 method for class 'matrix'
coarse(x, y, weights = NULL, pos = "1",
  nom.meth = "chisq", num.meth = "IV", vars = NULL, minsize = 0.05,
  siglev = 0.05, minimp = 0.1, miniv = 0.005, adj = 0.1,
  other = "average")
```



Gero Szepannek
Stralsund University of
Applied Sciences

Analyzing the Binning

Summary of the results:

```
####################################
summary of classing for each variable:
####################################

 status IV: 0.666

          tot good bad  %pop  %def
[-Inf,0)  274  139 135 0.274 0.493
[0,200)   269  164 105 0.269 0.390
[200,Inf)  63   49  14 0.063 0.222
none      394  348  46 0.394 0.117
```

Gero Szepannek
Stralsund University of
Applied Sciences

Analyzing the Binning

Summary of the results:

```
####################################
Summary of classing for each variable:
####################################

 status IV: 0.666

          tot good bad  %pop  %def
[-Inf,0)  274  139 135 0.274 0.493
[0,200)   269  164 105 0.269 0.390
[200,Inf)  63   49  14 0.063 0.222
none      394  348  46 0.394 0.117
```

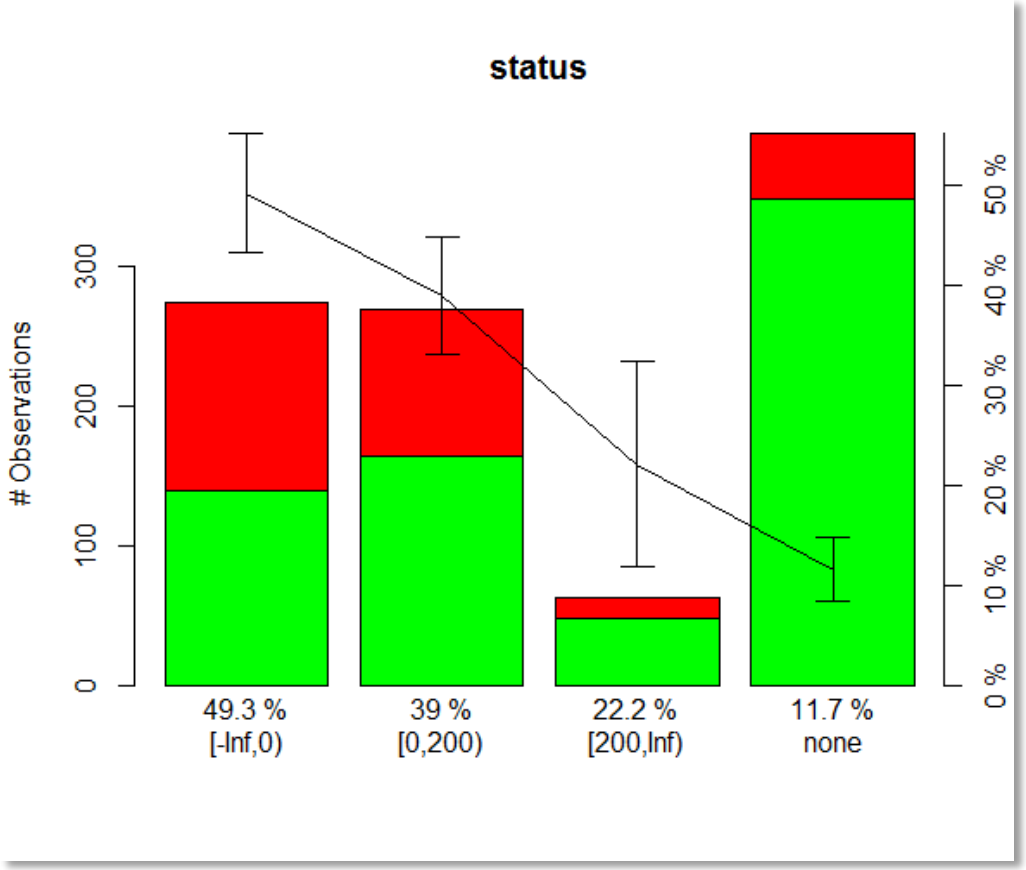
Further insights:

```
> res$tmonjob$trace
                lev1       lev2    pvalue merged       br1       br2 size1 size2
1            [0, 1) unemployed 0.7298022   TRUE 0.4069767 0.3709677   172    62
2 [0, 1),unemployed     [1, 4) 0.0310736  FALSE 0.3974359 0.3067847   234   339
3            [1, 4)   [7, Inf) 0.1787006   TRUE 0.3067847 0.2529644   339   253
4   [1, 4),[7, Inf)     [4, 7) 0.1441674   TRUE 0.2837838 0.2241379   592   174
>
> res$tmonjob$desc
                      tot good bad  %pop  %def
[0, 1),unemployed     234  141  93 0.234 0.397
[1, 4),[7, Inf),[4, 7) 766  559 207 0.766 0.270
```

Example of levels with implicit order. Only manual check can indentify inconsistencies here…

Gero Szepannek
Stralsund University of
Applied Sciences

Analyzing and Modifying Binning



Several utilities to manipulate an existing coarse classing model:

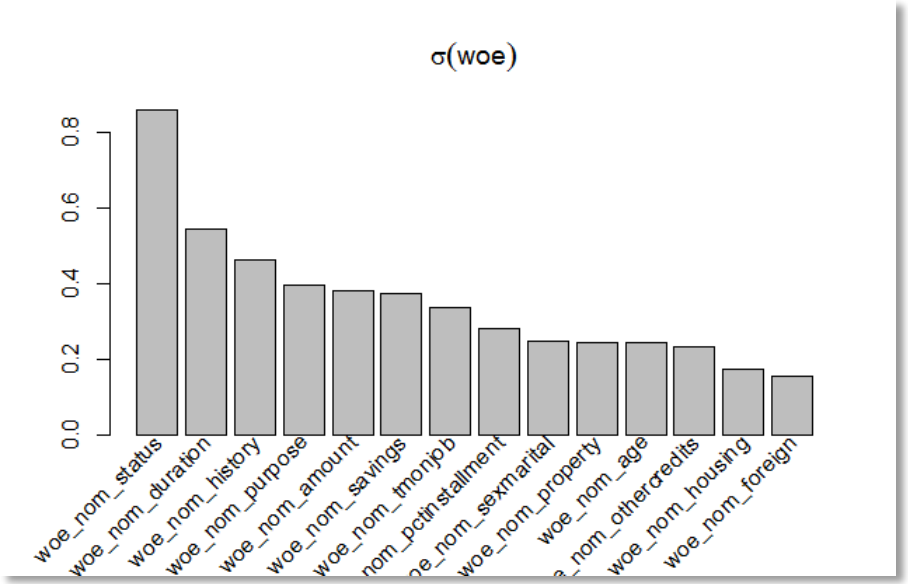
- merge
- split
- rename
- autoimpute and
- restore (!).

Gero Szepannek
Stralsund University of
Applied Sciences

WoEs and Shortlist Selection

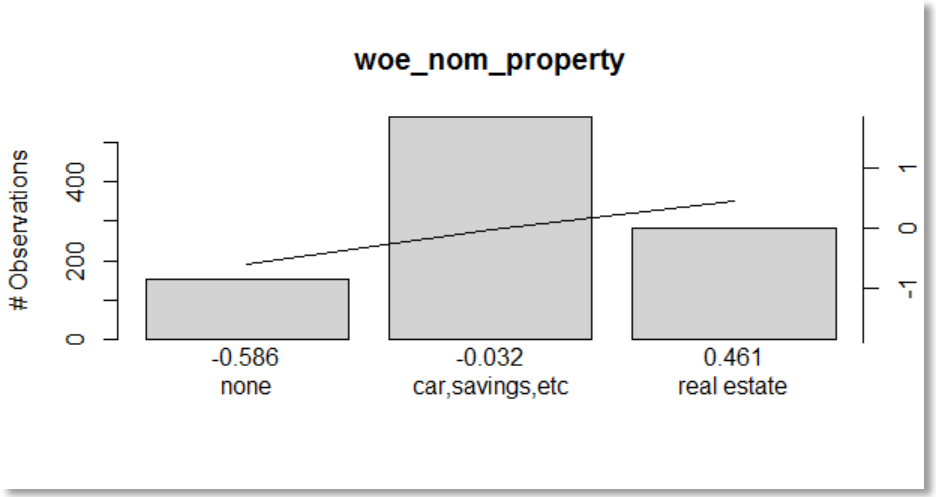
Weights of Evidence:

- It's common practise to use WoEs
  (cf. e.g. Siddiqi, 2006, Finlay 2012, Thomas et al., 2002)
- …stand alone R implementation in Package `klaR` (Roever et al., 2014)
- Implementation similar to coarse.



Short List:

- Preliminary variable exclusions,e.g. by
  - IV
  - PSI
  - Correlations (according to Breiter et al., 2009, Kuhn, 2016)



Gero Szepannek
Stralsund University of
Applied Sciences

Scorecard Objects

- (Still) gold standard: Logistic Regression (→ `glm()` in R, cf. e.g. Sharma, 2008)
- `Scorecard`: conveniance wrapper around glm object.
- Corresponding `predict()` and `plot()` methods.

⇨ all advantages of glm, including variable selection (…e.g. BIC).

- Further: use of continuous characteristics.
- allows for expert analysis (!)
  (several checks implemented)

level	points	pcttot	pctdef	avpd	checks
Total population	NA	1.00	0.3	0.3	OK
(Intercept)	412	NA	NA	NA	OK
woe_nom_status					
none	26	0.394	0.12	0.118	OK
[200,Inf)	9	0.063	0.22	0.209	OK
[-Inf,200)	-14	0.543	0.44	0.443	OK
woe_nom_duration					
(-Inf,33.3]	3	0.83	0.26	0.263	OK
(33.3,43.9]	-9	0.1	0.42	0.409	OK
(43.9, Inf]	-19	0.07	0.57	0.58	OK
woe_nom_history					
critical running	14	0.293	0.17	0.176	OK
positive running,past delays	-2	0.618	0.32	0.314	OK
rest	-24	0.089	0.6	0.608	OK

Gero Szepannek
Stralsund University of
Applied Sciences

Marginal Information

- Alternative VS based on marginal information values (Scallan, 2011)
- Choice between variable candidates → **expert knowledge integration**.
- Tracing the selection process: identifies relevant candidates
- …allows for an easy and intuitive step by step model updating.

```
> autoglm$trace
```

	step 1	step 2	step 3	step 4	step 5	step 6	step 7	step 8
woe_nom_status	0.639	0.000	0.000	0.001	0.000	0.000	0.001	0.001
woe_nom_history	0.292	0.212	0.000	0.000	0.000	0.001	0.001	0.002
woe_nom_amount	0.170	0.139	0.124	-0.001	-0.001	-0.001	0.000	0.001
woe_nom_savings	0.191	0.122	0.122	0.122	-0.001	-0.001	-0.001	-0.002
woe_nom_purpose	0.132	0.097	0.092	0.090	0.087	0.000	0.000	0.000
woe_nom_duration	0.160	0.132	0.114	0.059	0.056	0.061	0.000	0.000
woe_nom_property	0.113	0.096	0.084	0.056	0.056	0.052	0.044	0.000
woe_nom_age	0.062	0.038	0.032	0.035	0.032	0.036	0.036	0.039
woe_nom_foreign	0.044	0.045	0.041	0.038	0.038	0.040	0.039	0.036
woe_nom_sexmarital	0.031	0.023	0.020	0.028	0.028	0.029	0.032	0.034
woe_nom_pctinstallment	0.024	0.024	0.025	0.035	0.037	0.038	0.036	0.034
woe_nom_tmonjob	0.063	0.038	0.034	0.036	0.032	0.030	0.033	0.031
woe_nom_othercredits	0.058	0.049	0.030	0.029	0.030	0.029	0.030	0.027
woe_nom_housing	0.083	0.059	0.049	0.042	0.042	0.041	0.039	0.026

Gero Szepannek
Stralsund University of
Applied Sciences

Marginal Information

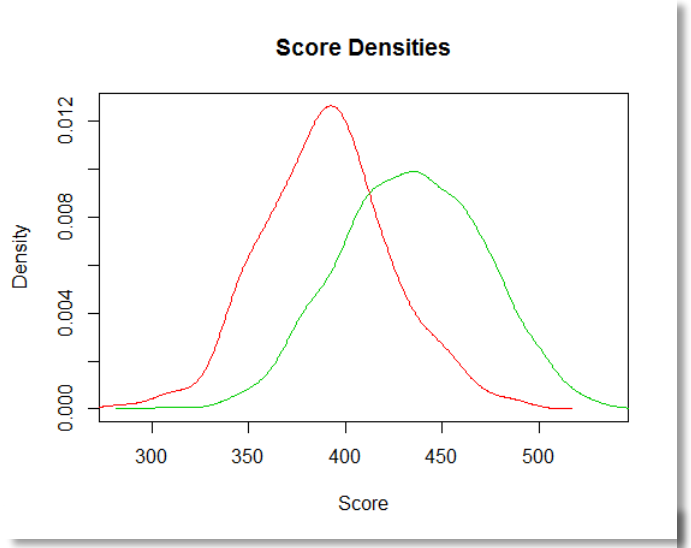
- Alternative VS based on marginal information values (Scallan, 2011)
- Choice between variable candidates → **expert knowledge integration**.
- Tracing the selection process: identifies relevant candidates
- …allows for an easy and intuitive step by step model updating.

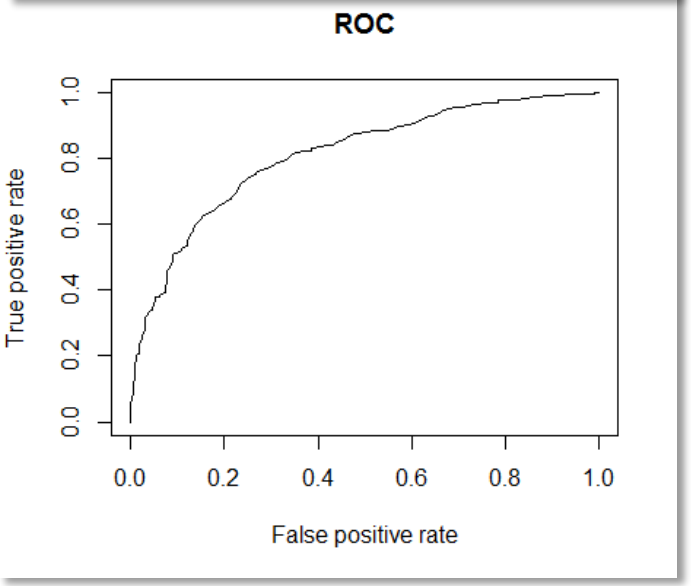
```
> autoglm$trace
                       step 1 step 2 step 3 step 4 step 5 step 6 step 7 step 8
woe_nom_status          0.639  0.000  0.000  0.001  0.000  0.000  0.001  0.001
woe_nom_history         0.292  0.212  0.000  0.000  0.000  0.001  0.001  0.002
woe_nom_amount          0.170  0.139  0.124 -0.001 -0.001 -0.001  0.000  0.001
woe_nom_savings         0.191  0.122  0.122  0.122 -0.001 -0.001 -0.001 -0.002
woe_nom_purpose         0.132  0.097  0.092  0.090  0.087  0.000  0.000  0.000
woe_nom_duration        0.160  0.132  0.114  0.059  0.056  0.061  0.000  0.000
woe_nom_property        0.113  0.096  0.084  0.056  0.056  0.052  0.044  0.000
woe_nom_age             0.062  0.038  0.032  0.035  0.032  0.036  0.036  0.039
woe_nom_foreign         0.044  0.045  0.041  0.038  0.038  0.040  0.039  0.036
woe_nom_sexmarital      0.031  0.023  0.020  0.028  0.028  0.029  0.032  0.034
woe_nom_pctinstallment  0.024  0.024  0.025  0.035  0.037  0.038  0.036  0.034
woe_nom_tmonjob         0.063  0.038  0.034  0.036  0.032  0.030  0.033  0.031
woe_nom_othercredits    0.058  0.049  0.030  0.029  0.030  0.029  0.030  0.027
woe_nom_housing         0.083  0.059  0.049  0.042  0.042  0.041  0.039  0.026
```

Gero Szepannek
Stralsund University of
Applied Sciences

List of Implemented Validation Measures

- Currently 13 different performance measures (not the meaningless standard ones like error rate, precision, …) implemented for both rank ordering and calibration, e.g.:
  - Gini / AUC
  - CLs for the Gini (Henking et al., 2006)
  - KS
  - H-Measure (Hand, 2009)
  - partial AUC (Robin et al., 2011)
  - EMP (Verbraken et al., 2014)
  - Hosmer-Lemeshow test
  - Spiegelhalter test
  - Brier score
  - …
- …reflects observations weights.
- …easily extendable by other R packages.
- Additional wrappers for comparing several measures and models…

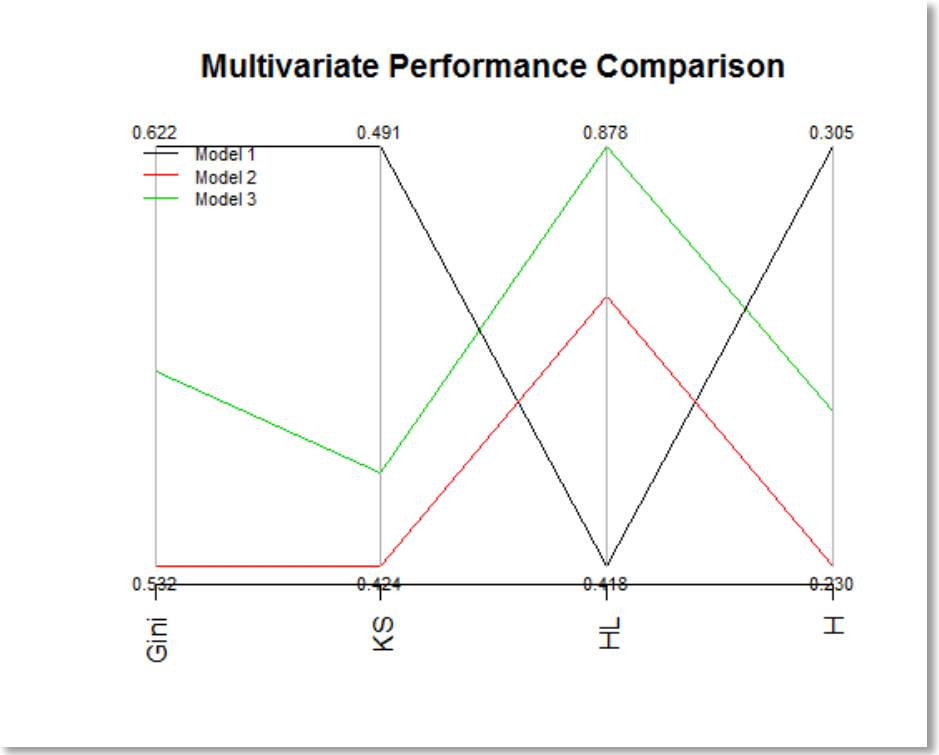


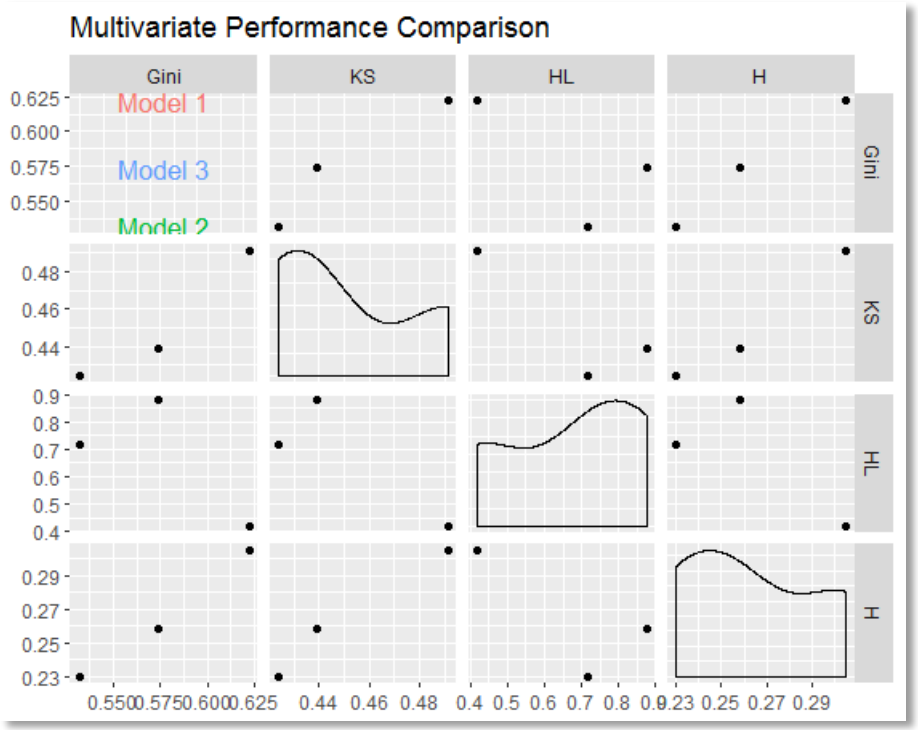


Gero Szepannek
Stralsund University of
Applied Sciences

Model Selection

…based on Multivariate Performance Analysis





Gero Szepannek
Stralsund University of
Applied Sciences

Lessons Learned from Existing Shortcomings…

- Proper `predict()` functions.
- Restore option for manual changes of coarse classing.
- Flexibility w.r.t using continuous characteristics in the model.
- Interface to arbitrary classification methods (from other packages), score scalings and validation measures
- Treatment of numerical attributes
- Reflection of weights (typical after RI) (AUC, KS, WoE)
- Properly documented
- …

Gero Szepannek
Stralsund University of
Applied Sciences

…finally some References

Bache, K. and Lichmann, M (2013): UCI Machine Learning Repository, http://archive.ics.uci.edu/ml, University of California, Irvine, School of Information and Computer Sciences

Baesens, B. van Gestel, T., Viane, S., Stepanova, M. Suykens, J. and Vanthienen, J. (2003): Benchmarking State-of-the-art Classification Algorithms for Credit Scoring, JORS 54(6), 627-635.

Bischl, B., Kühn, T. and Szepannek, G. (2016): On Class Imbalance Correction for Classification Algorithms in Credit Scoring, In: Lübbecke, M., Koster, A., Letmathe, P., Madlenerm R., Peis, B. und Walther, G. (Eds): Operations Research Proceedings 2014, 37-43, Springer.

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., Jones, Z. (2016): mlr: Machine Learning in R, JMLR 17(170):1–5.

Breiter, D., Wornowitzki, M., Schaltenbrand, S., Priefer, A., Bischl, B. and Szepannek, G. (2009): Data-Mining-Cup 2009 – Vorhersage von Buchabverkäufen. Research Paper 01/2009, Faculty of Statistics, Dortmund University of Technology.

Finlay, S. (2012): Credit Scoring, Response Modelling and Insurance rating. Palgarve MacMillan.

Hand, D. (2009): Measuring classifier performance: a coherent alternative to the area under the ROC curve. Machine Learning, 77,103–123.

Hastie, T., Tibshirani, R. and Friedman, J. (2001): The Elements of Statistical Learning. Springer.

Henking, A., Blum, C. and Fahrmeir, L. (2006): Kreditrisikomessung – Statistische Grundlagen, Methoden und Modellierung, Springer, Berlin.

Hoffmann, H. (1994): German Credit Data Set (Statlog), http://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29

Kuhn, M. (2016): caret: Classification and Regression Training, R package version 6.0-71, https://CRAN.R-project.org/package=caret.

Lessmann, S., Baesens, B., Seov, H. and Thomas, L. (2015): Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research, EJOR 247(1), 124-136.

Brown, I. and Mues, C. (2012): An Experimental Comparison of Classification Algorithms for Imbalanced Credit Scoring Data Sets. Expert Systems with Applications 39(3), 3446-3453.

Robin, X., Turck, N., Hainard, A., Tiberti, A., Lisacek, F., Sanchez, J. and Müller, M. (2011): pROC: an open-source package for R and S+ to analyze and compare ROC curves". BMC Bioinformatics, 12, p. 77.

Roever, C., Raabe, N., Luebke, K., Ligges, U., Szepannek, G. and Zentgraf, M. (2014): klaR: Classification and visualization, R package version 0.6-12, https://CRAN.R-project.org/package=klaR.

Scallan, G. (2011): Class(ic) Scorecards – Selecting Attributes in Logistic Regression. Talk @ CSCC 2011, http://www.business-school.ed.ac.uk/crc/conferences/conference-archive?a=46012.

Sharma, D. (2009): Guide to Credit Scoring in R. https://cran.r-project.org/doc/contrib/Sharma-CreditScoring.pdf

Siddiqi, N. (2006): Credit Risk Scorecards. Wiley.

Szepannek, G. (2011): Vortransformation in der Kreditantrags-Scoremodellierung, Data Mining Anwendertag, Heidelberg, 2011, http://www.sas.com/reg/offer/de/datamining_2011?page=download.

Szepannek, G. (2017): On the Relevance of Modern Machine Learning Algorithms for Credit Scoring Applications, WIAS report 29, 88-96.

Szepannek, G., Gruhne, M., Bischl, B., Krey, S., Harczos, T. Klefenz, F. and Weihs, C. (2010): Perceptually Based Phoneme Recognition in Popular Music, in Locareck-Junge, H. and Weihs, C. (eds.): Classification as a Tool for Research, Springer, 751-758.

Szepannek, G., Schiffner, J., Wilson, J. and Weihs, C. (2008): Local Modelling in Classification, in Perner, P. (ed.): Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects, Springer LNAI 5077, 154-163.

Thomas, L., Edelman, D. and Crook, J. (2002): Credit Scoring and its Applications. SIAM.

Verbraken, T., Bravo, C., Weber, R. and Baesens, B. (2014): Development and application of consumer credit scoring models using profit-based classification measures, EJOR 238(2), 505-513.

Gero Szepannek
Stralsund University of
Applied Sciences

Thank you!

Interested in the project? Please contact:

Gero Szepannek
Statistics, Business Mathematics and Machine Learning
School of Business Studies
Stralsund University of Applied Sciences
gero.szepannek@hochschule-stralsund.de

HOST
Hochschule Stralsund
University of
Applied Sciences

Gero Szepannek
Stralsund University of
Applied Sciences