Insurance analytics

Tree-based machine learning methods - putting it all together

Katrien Antonio
LRisk - KU Leuven and ASE - University of Amsterdam

May 26, 2019

Acknowledgement

- Some of the figures in this presentation are taken from *An Introduction to Statistical Learning, with applications in R* (Springer, 2013) with permission from the authors: G. James, D. Witten, T. Hastie and R. Tibshirani.

- Some of the figures in this presentation are from *Boosting insights in insurance tariff plans with tree-based machine learning* (available on arxiv, April 2019), written by Roel Henckaerts, Marie-Pier Côté, Katrien Antonio and Roel Verbelen.

The R universe

What is out there?

Model	Poisson	Gamma
Generalized linear model	✓ `stats`	✓ `stats`
Generalized additive model	✓ `mgcv`	✓ `mgcv`
Regression tree	✓ `rpart`	✗ -
Random forest	✗ -	✗ -
Gradient boosting machine	✓ `gbm`	✓ harrysouthworth/`gbm`

The R universe

Filling the gaps

Model	Poisson	Gamma
Generalized linear model	✓ `stats`	✓ `stats`
Generalized additive model	✓ `mgcv`	✓ `mgcv`
Regression tree	✓ `rpart`	✓ `rpart*`
Random forest	✓ `rpart*`	✓ `rpart*`
Gradient boosting machine	✓ `gbm`	✓ harrysouthworth/`gbm`

Interpretation tools

- Classical statistical methods are highly interpretable:
  - coefficients in a GLM
  - splines in a GAM.
- Not the case for machine learning methods:
  - \+ regression trees
  - − bagged trees/random forests
  - − boosted trees
- There is a need for interpretation tools!

Interpretation tools

- Variable importance plots to select relevant variables.
- Partial dependence plots to interpret the effect of a variable on the outcome.
- Individual conditional expectation plots to detect interaction effects.

Interpretation tools

Variable importance

- Measure the importance of feature x_ℓ in a tree t:

$$\mathcal{I}_\ell(t) \;\; = \;\; \sum_{j=1}^{J-1} \mathbb{I}\{v(j) = \ell\}(\Delta\mathcal{L})_j,$$

where

- sum is over $J-1$ internal nodes
- only consider nodes where splitting variable is x_ℓ
- $(\Delta\mathcal{L})_j$ is difference in loss before and after the split.

Interpretation tools

Variable importance

- Important variables appear often and high in the tree t, and the $\mathcal{I}_\ell(t)$ grows largest for these.
- When using an ensemble of trees:

$$\mathcal{I}_\ell = \frac{1}{T}\sum_{t=1}^{T}\mathcal{I}_\ell(t),$$

with the sum over all trees in the RF or GBM.

Interpretation tools

Partial dependence plots

- Partial dependence plots (PDPs) show the marginal effect of a variable on the predictions obtained from a model.

- Evaluate the prediction function in specific values of x_ℓ, while averaging over a range of values of the other variables $\boldsymbol{x}^\star$:

$$\bar{f}_\ell(x_\ell) = \frac{1}{n}\sum_{i=1}^{n} f_{\text{model}}(x_\ell, \boldsymbol{x}_i^\star).$$

- Interaction effects between x_ℓ and other variables in $\boldsymbol{x}^\star$ can distort the effect.

Interpretation tools

Individual conditional expectation plots

- Individual conditional expectations (ICEs) show the effect of a variable on the predictions, but on an individual level:

$$\tilde{f}_{\ell,i}(x_\ell) = f_{\text{model}}(x_\ell, \boldsymbol{x}_i^\star).$$

- Effect for each observation i, allows to detect interaction effects when some observations show different behavior compared to others.

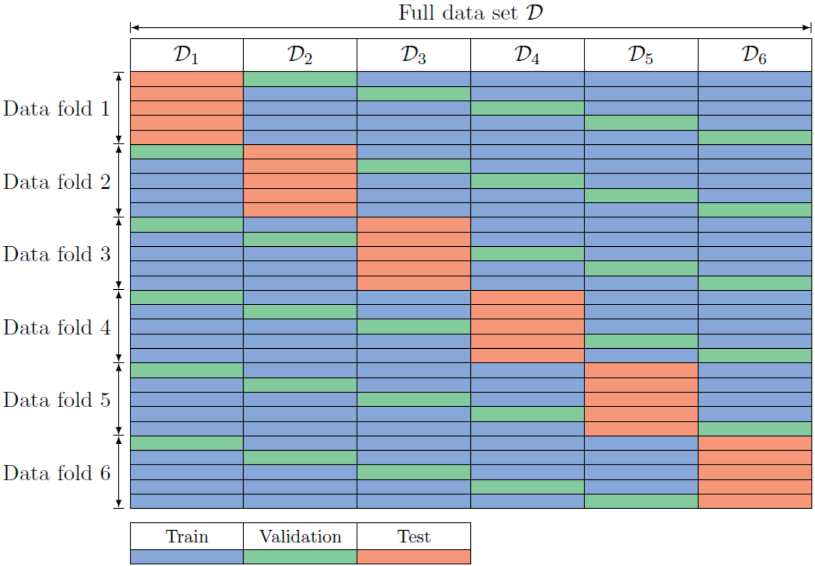
- ICEs also picture the uncertainty of the effect of variable x_ℓ on the prediction outcome.

More interpretation tools

- More on interpretable machine learning:
  - `https://christophm.github.io/interpretable-ml-book/`
  - *Boosting insights in insurance tariff plans with tree-based machine learning* (available on arxiv).

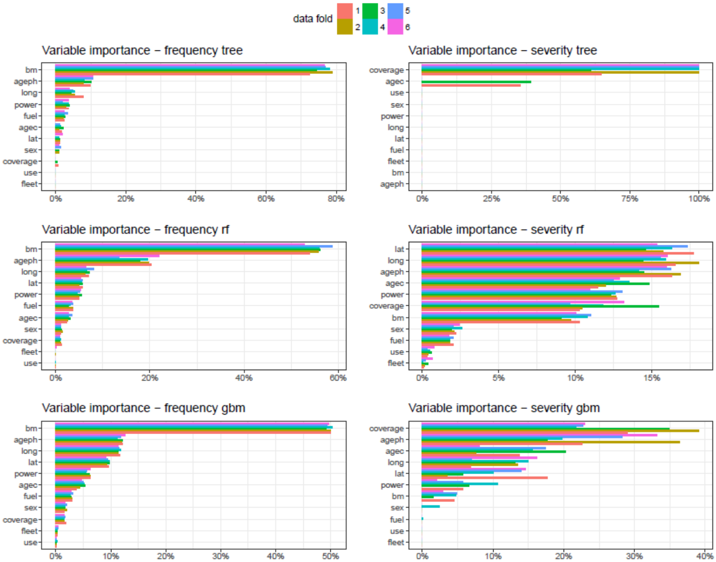
Comparison of pricing models

Set-up



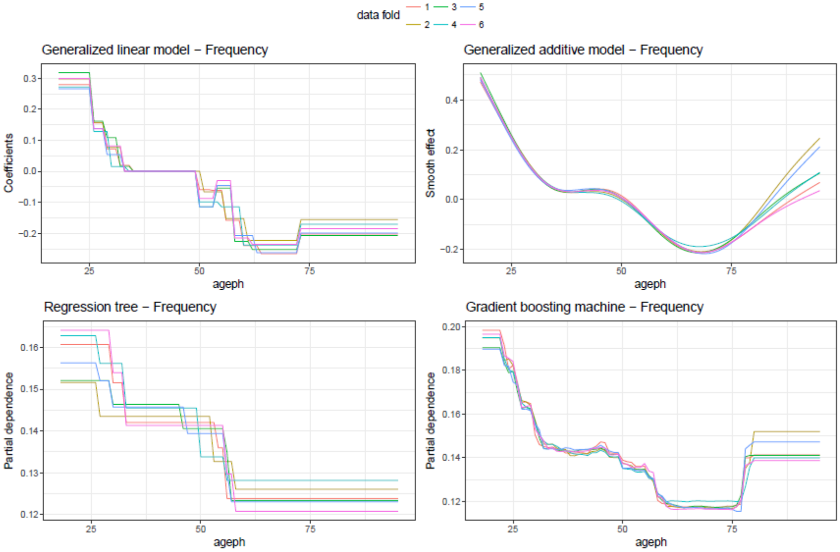
Comparison of pricing models

Variable importance plots



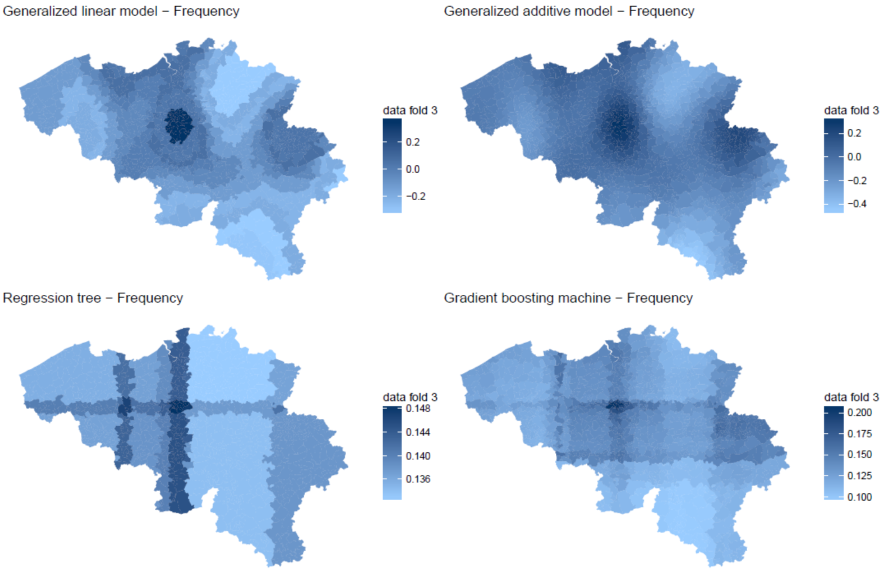
Comparison of pricing models

PDPs



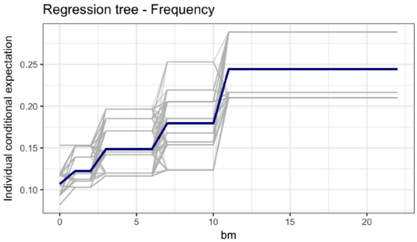
Comparison of pricing models

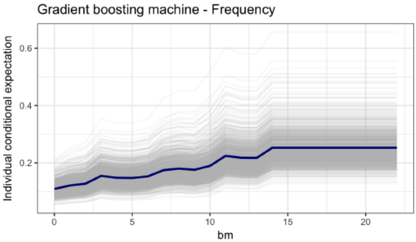
PDPs



Comparison of pricing models

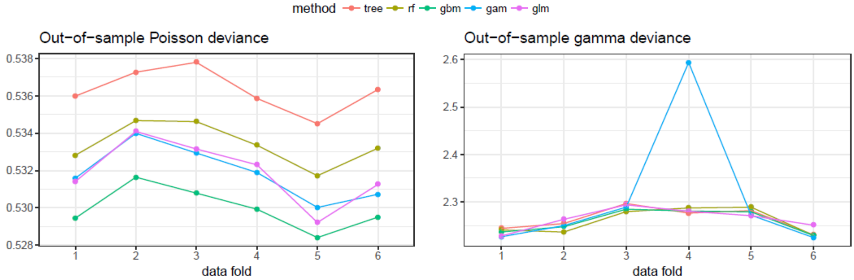
ICEs





Comparison of pricing models

Out–of–sample



More comparison tools in our paper.