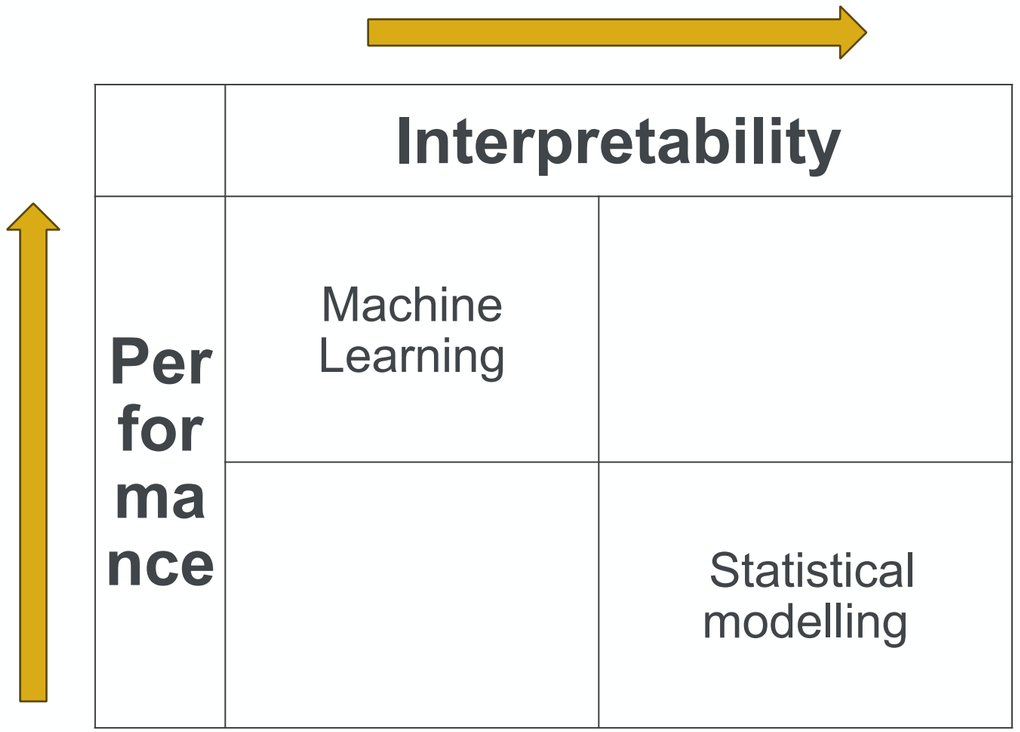
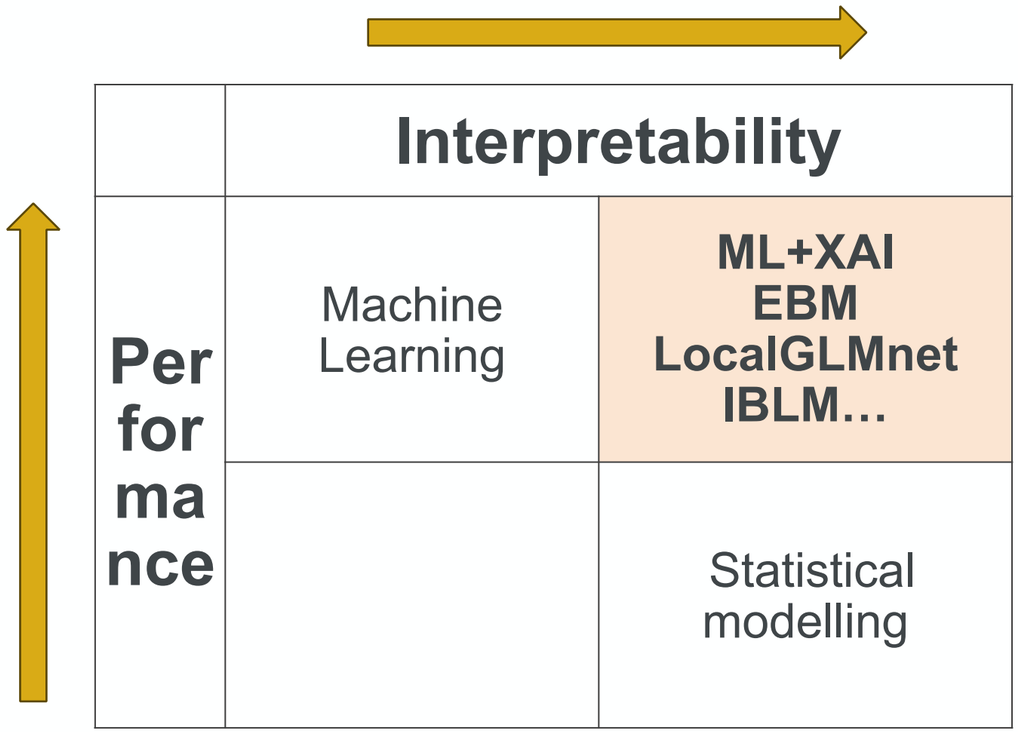
Glassbox Models:
Closing the Gap Between Transparency and Performance

Karol Gawlowski, William Flaherty

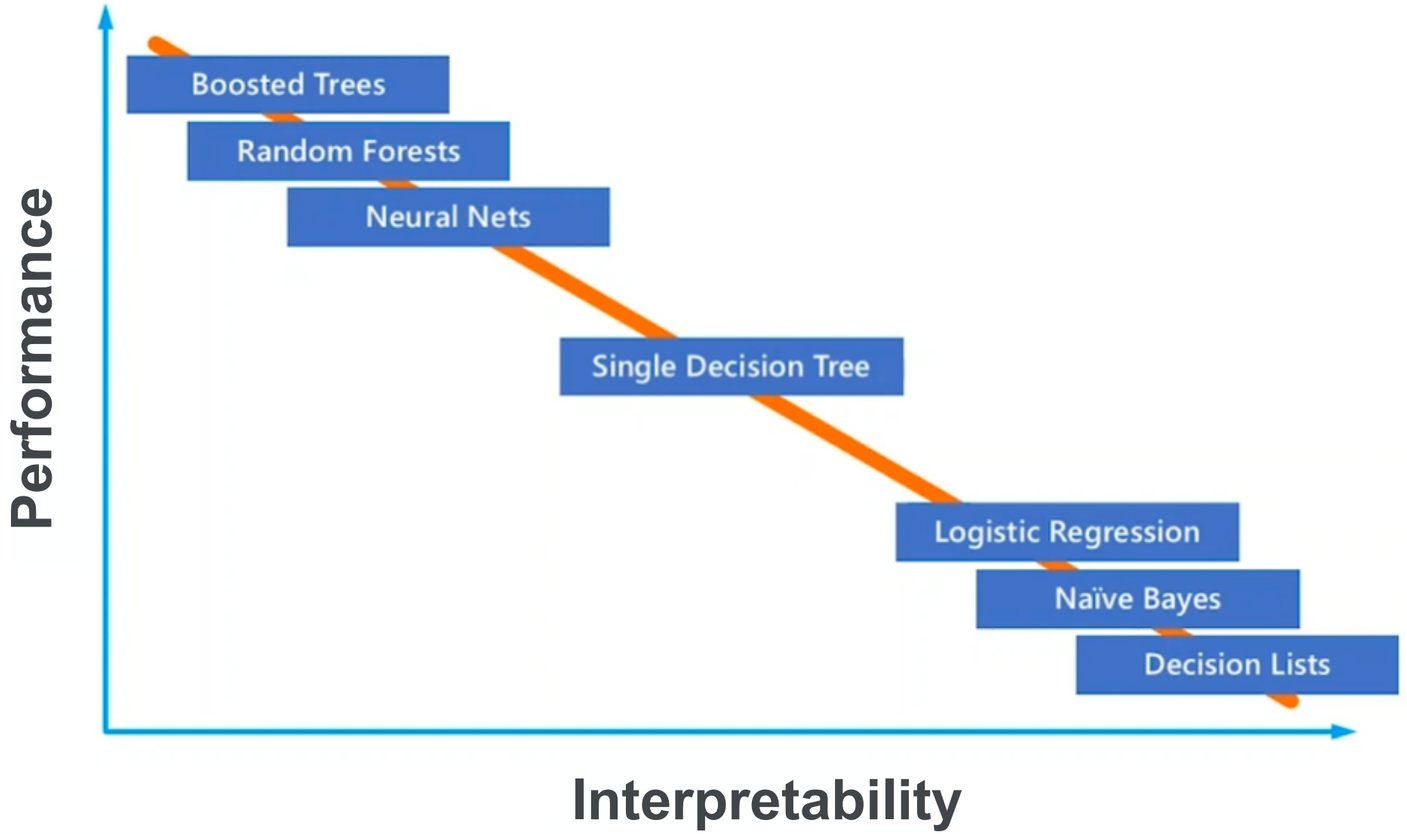
The tradeoff in predictive modelling



Performance vs Interpretability



Performance vs Interpretability



Source: Rich Caurana, Microsoft Research,

Model Performance Benchmark

CV	GLM	GLM*	GAM	LocalGLMnet	EBM	IBLM	XGB
CV_1	3.80%	8.10%	4.10%	9.90%	11.00%	12.10%	12.00%
CV_2	3.20%	7.00%	3.80%	9.60%	10.10%	11.40%	11.70%
CV_3	4.00%	8.10%	4.20%	10.60%	11.40%	12.70%	13.00%
CV_4	3.60%	7.90%	4.00%	10.70%	11.10%	12.40%	12.50%
CV_5	3.40%	7.60%	3.90%	9.40%	9.90%	11.30%	11.60%
Overall	**3.6%**	**7.8%**	**4.0%**	**10.0%**	**10.7%**	**12.0%**	**12.2%**

$$Poisson\ Deviance\text{:}\ D^2 = 1 - \frac{D_m}{D_0}$$

Model Performance Benchmark

CV	GLM	GLM*	GAM	LocalGLMnet	EBM	IBLM	XGB
CV_1	3.80%	8.10%	4.10%	9.90%	11.00%	12.10%	12.00%
CV_2	3.20%	7.00%	3.80%	9.60%	10.10%	11.40%	11.70%
CV_3	4.00%	8.10%	4.20%	10.60%	11.40%	12.70%	13.00%
CV_4	3.60%	7.90%	4.00%	10.70%	11.10%	12.40%	12.50%
CV_5	3.40%	7.60%	3.90%	9.40%	9.90%	11.30%	11.60%
Overall	**3.6%**	**7.8%**	**4.0%**	**10.0%**	**10.7%**	**12.0%**	**12.2%**

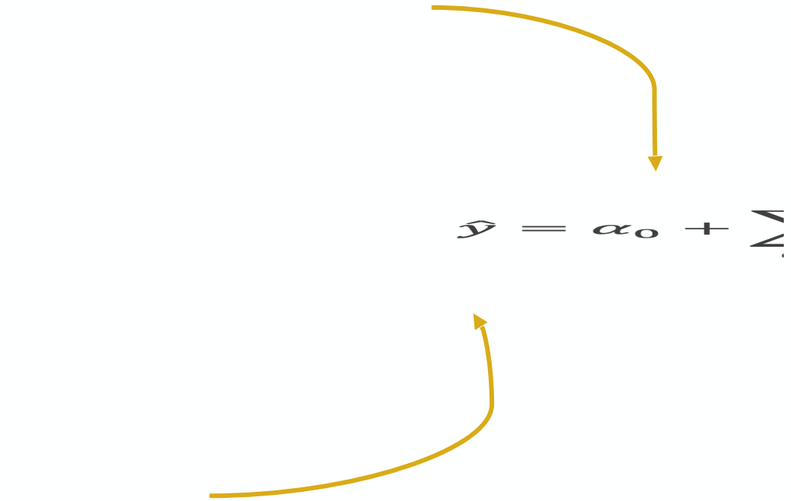
$$Poisson\ Deviance\text{:}\ D^2 = 1 - \frac{D_m}{D_0}$$

XGB + SHAP

SHAP

Additive Explanations

Baseline Prediction



Model output

SHAP

Additive Explanations

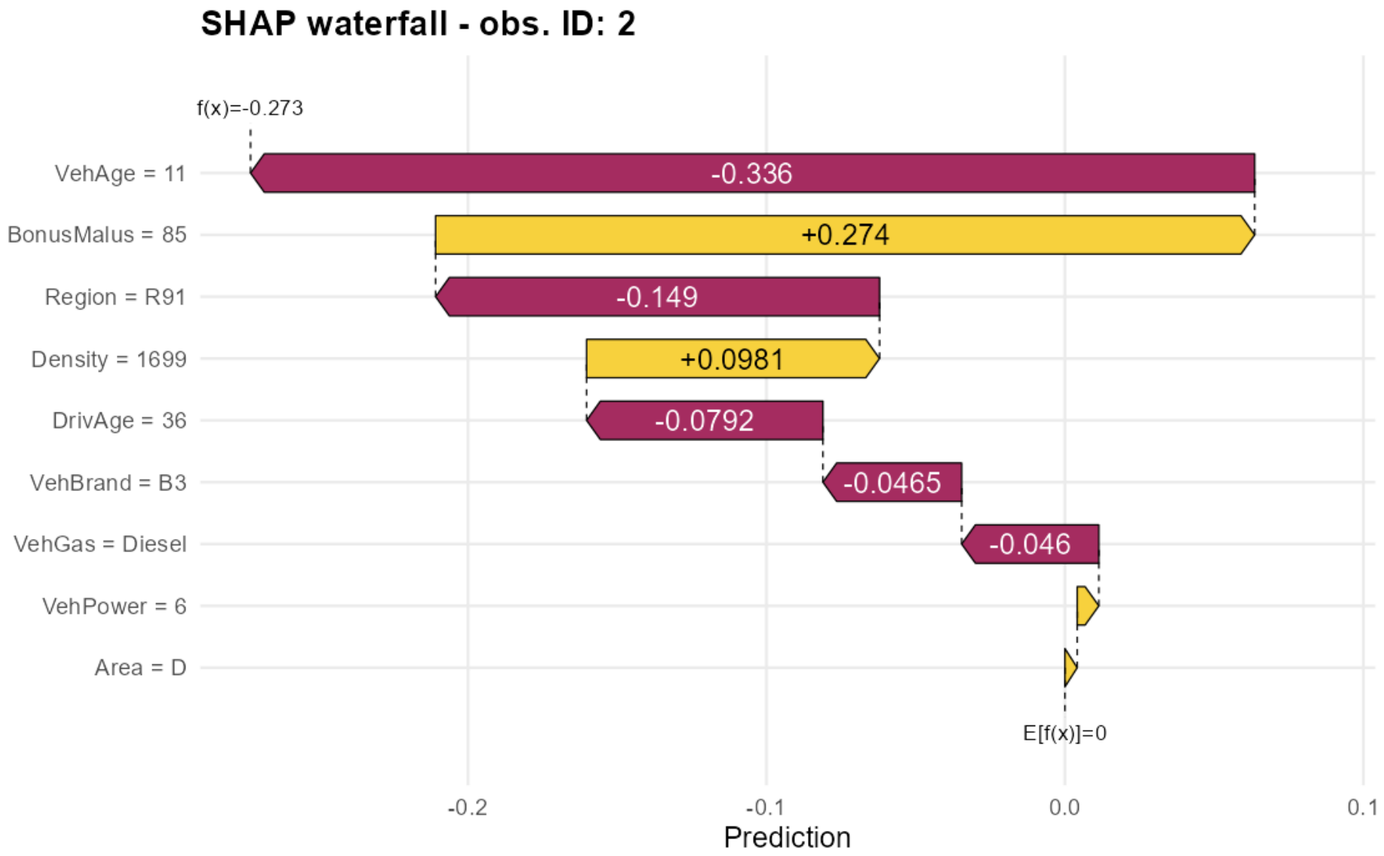
Baseline Prediction

$$\hat{y} = \alpha_0 + \sum_d \alpha_i$$

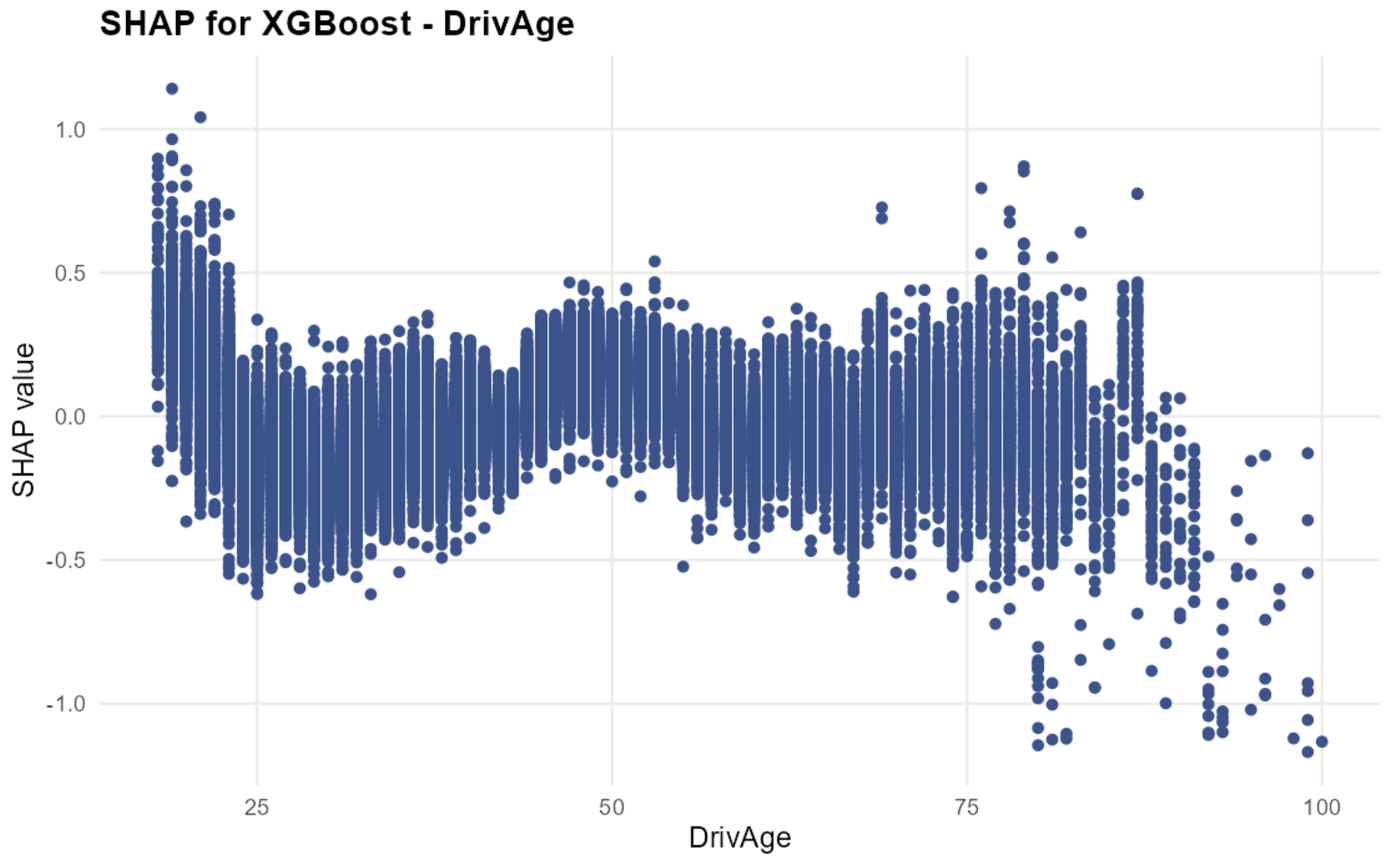
Model output

Contribution of i-th predictor

SHAP

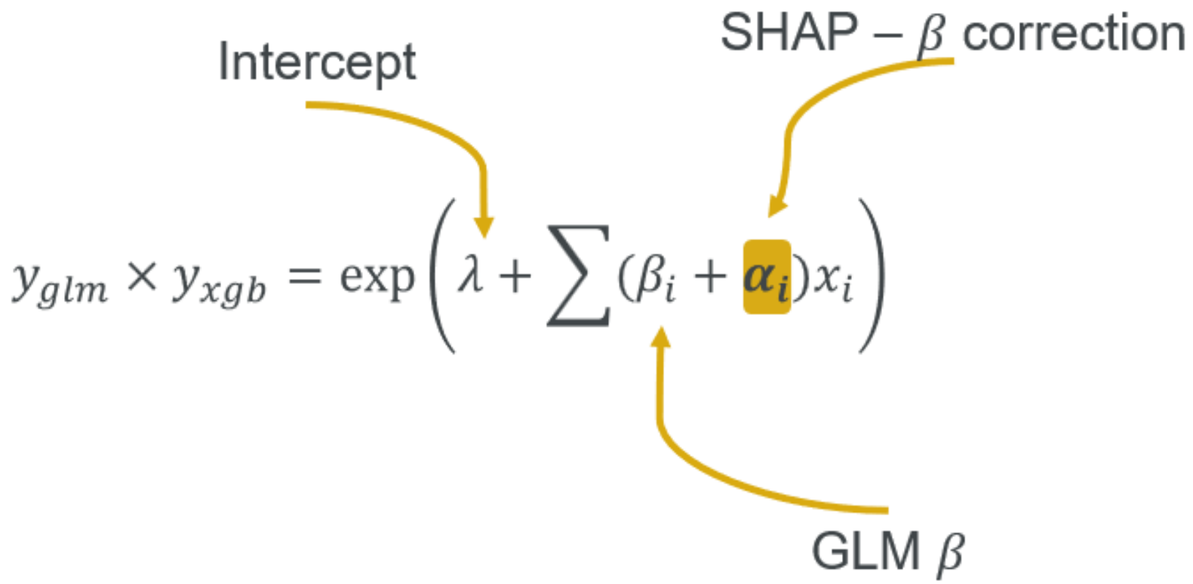

SHAP waterfall - obs. ID: 2
f(x)=-0.273
VehAge = 11
BonusMalus = 85
Region = R91
Density = 1699
DrivAge = 36
VehBrand = B3
VehGas = Diesel
VehPower = 6
Area = D
-0.336
+0.274
-0.149
+0.0981
-0.0792
-0.0465
-0.046
E[f(x)]=0
-0.2
-0.1
0.0
0.1
Prediction

SHAP


SHAP for XGBoost - DrivAge
SHAP value
1.0
0.5
0.0
-0.5
-1.0
25
50
75
100
DrivAge

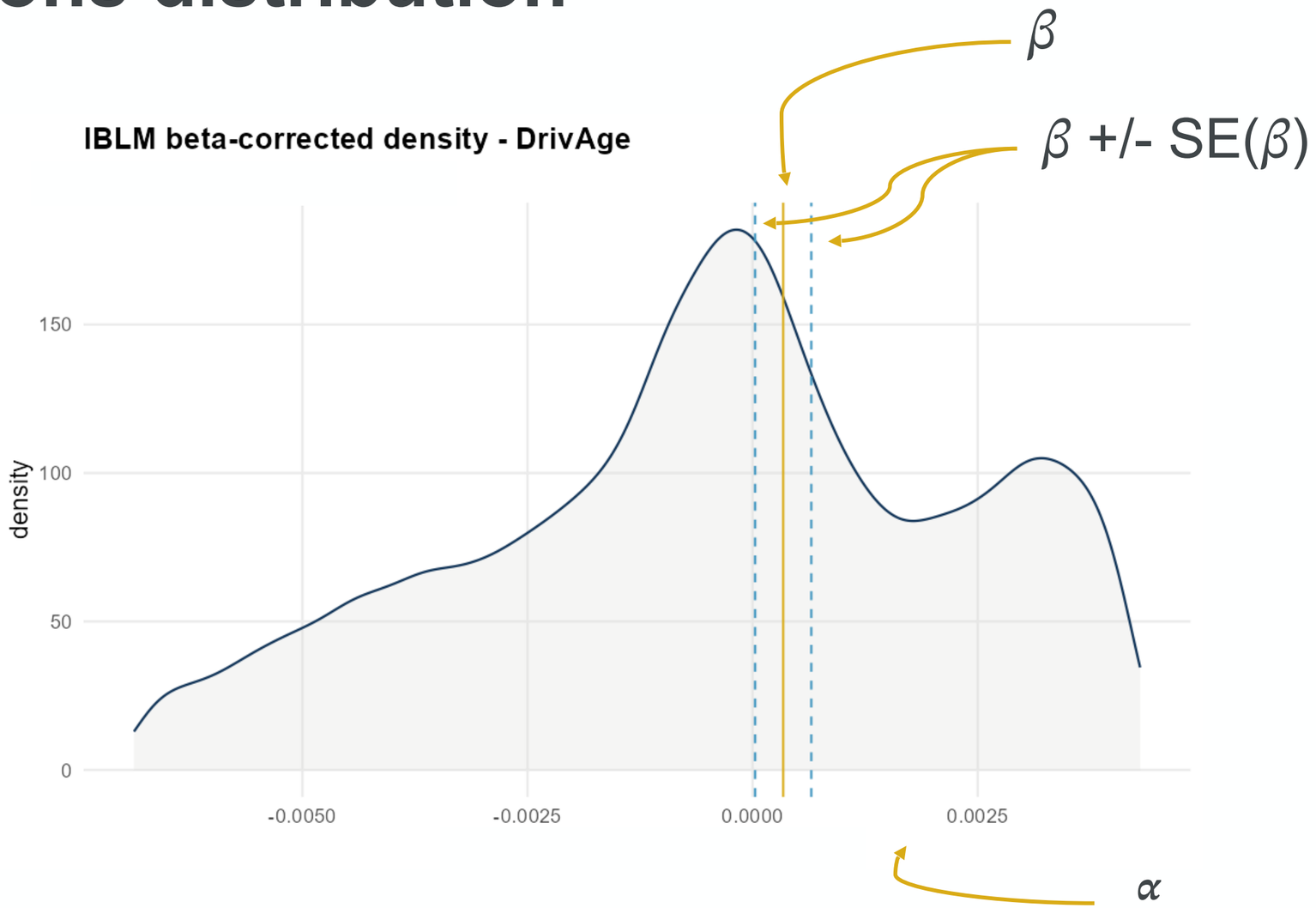
Interpretable Boosted Linear Model (IBLM)

IBLM – ensemlbing GLM with XGB



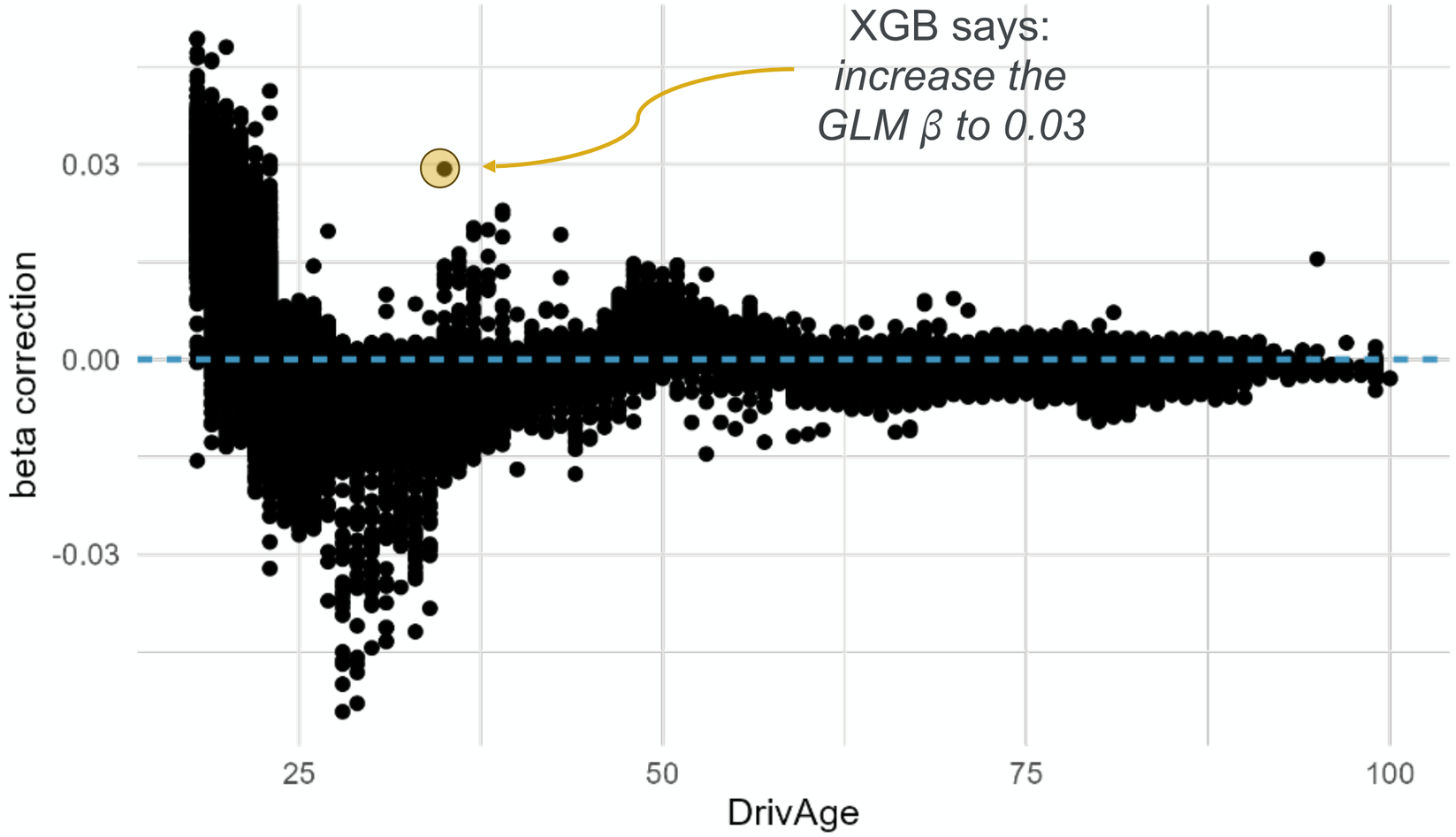
β corrections distribution

Driver Age



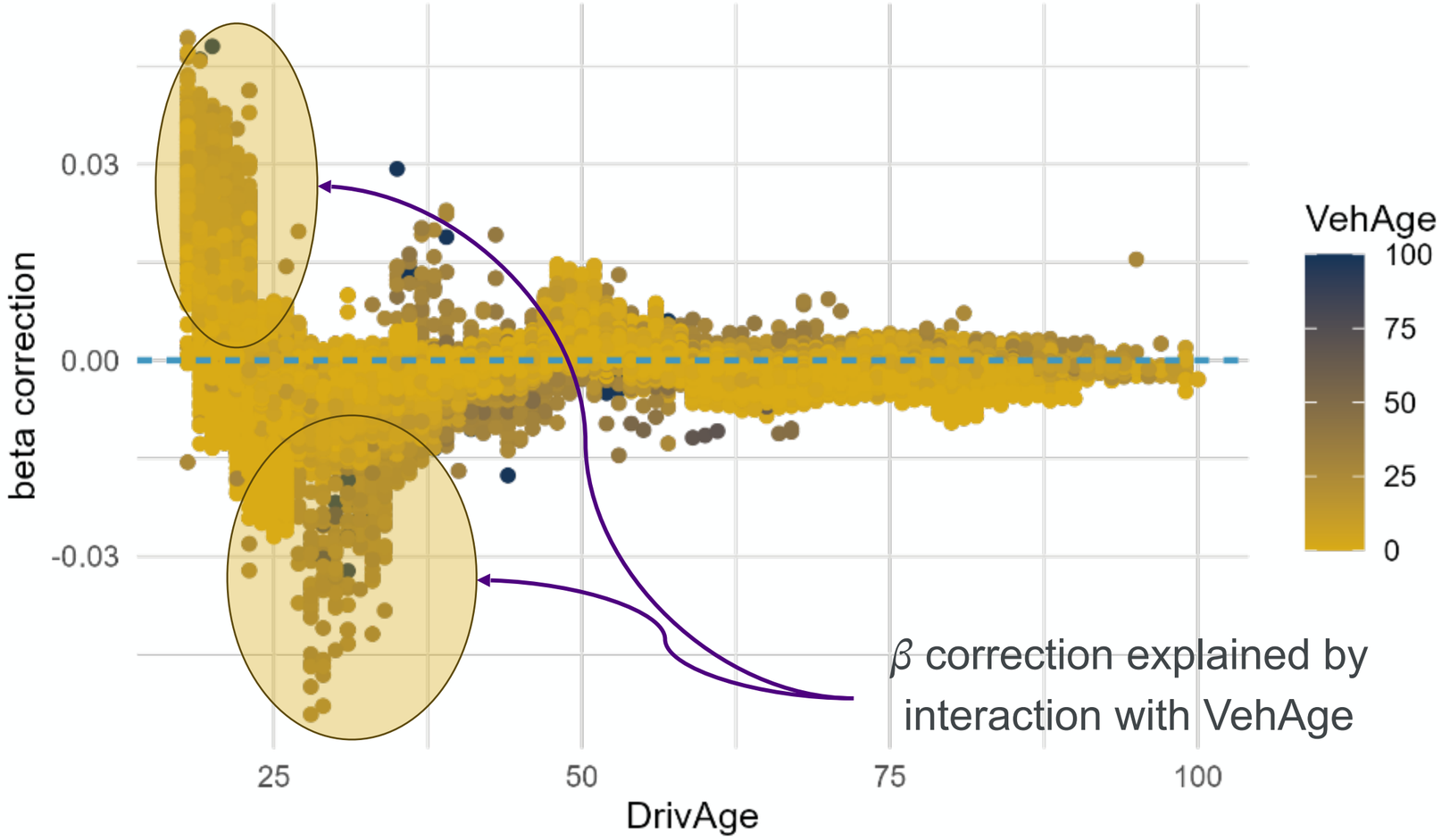
β corrections scatter

Driver Age



β corrections scatter

Driver Age



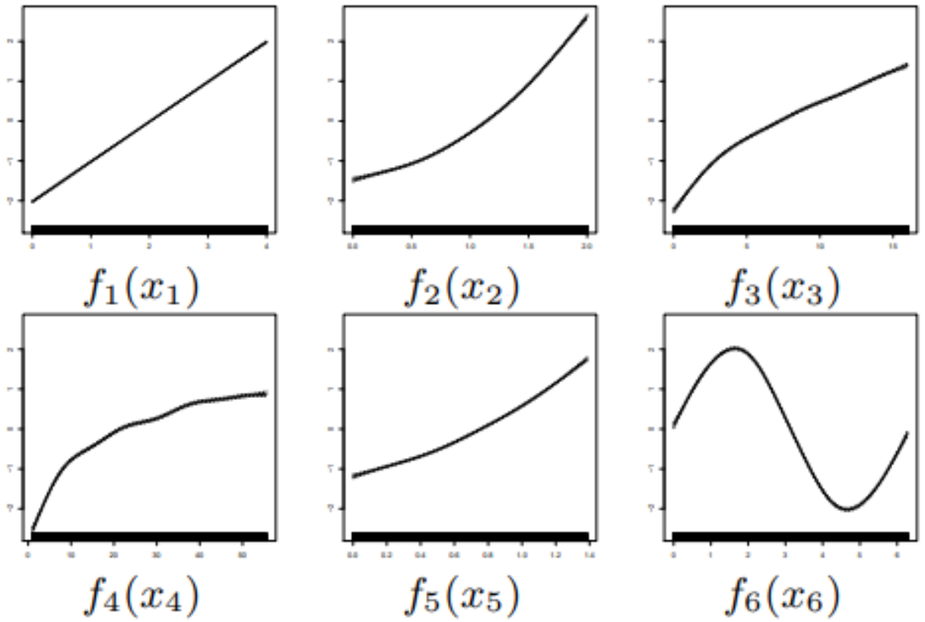
Explainable Boosting Machine (EBM)

Explainable Boosting Machine

$$g(\mathbb{E}\left[Y_i|\mathbf{X}_i = \mathbf{x}_i\right]) = \beta_0 + \sum_{j=1}^{p} f_j(x_{i,j}) + \sum_{j_1=1}^{p} \sum_{j_2 \neq j_1}^{p} f_{(j_1,j_2)}(x_{i,j_1}, x_{i,j_2})$$

Explainable Boosting Machine

GAMs: smooth shape functions

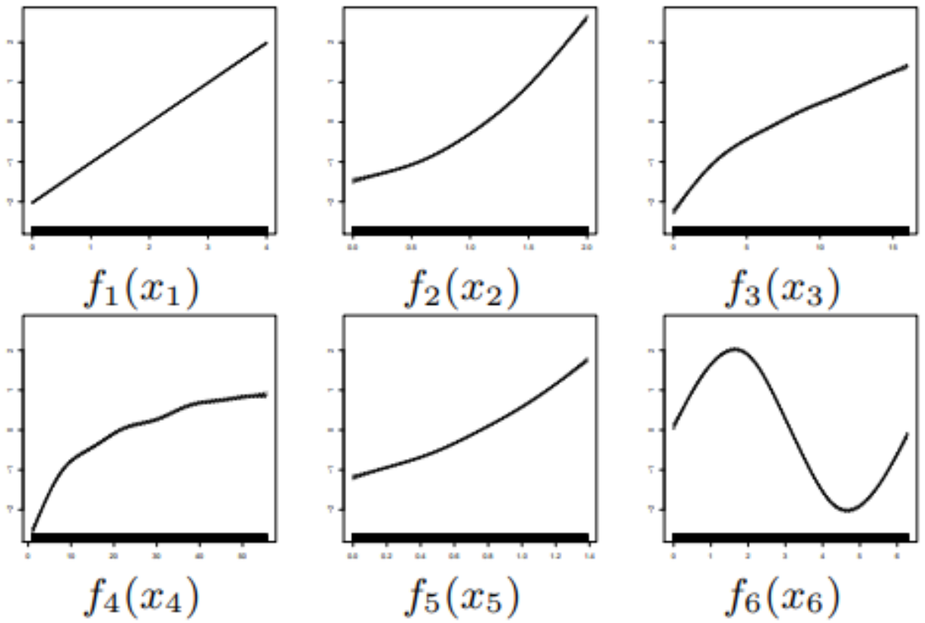


Sources: Lin Y. et al.: *Intelligible Models for Classification and Regression*

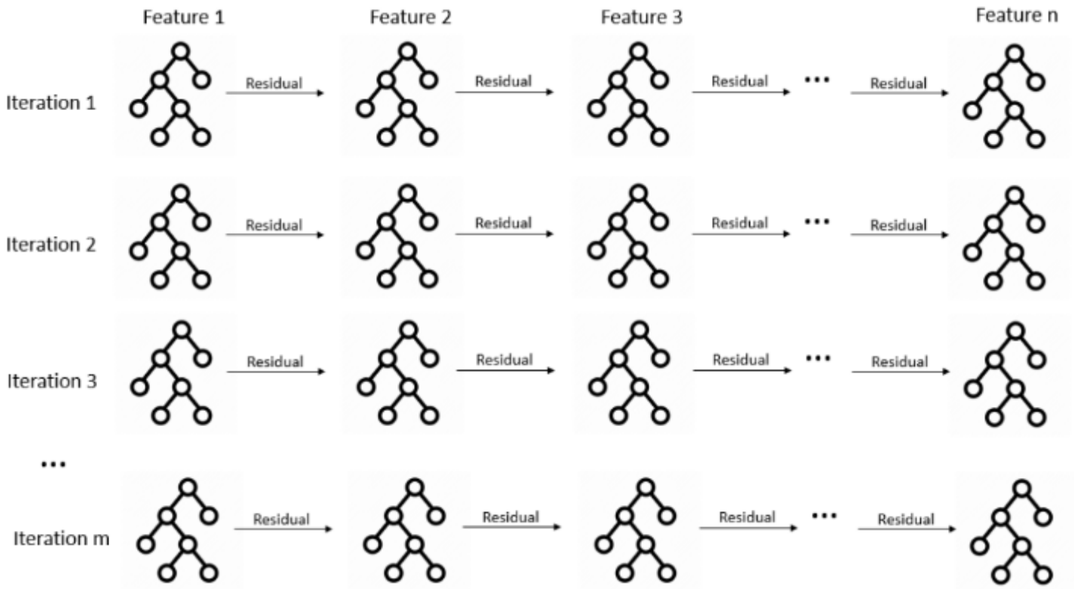
InterpretML: Another Way to Explain Your Model, Towards Data Science

Explainable Boosting Machine

GAMs: smooth shape functions



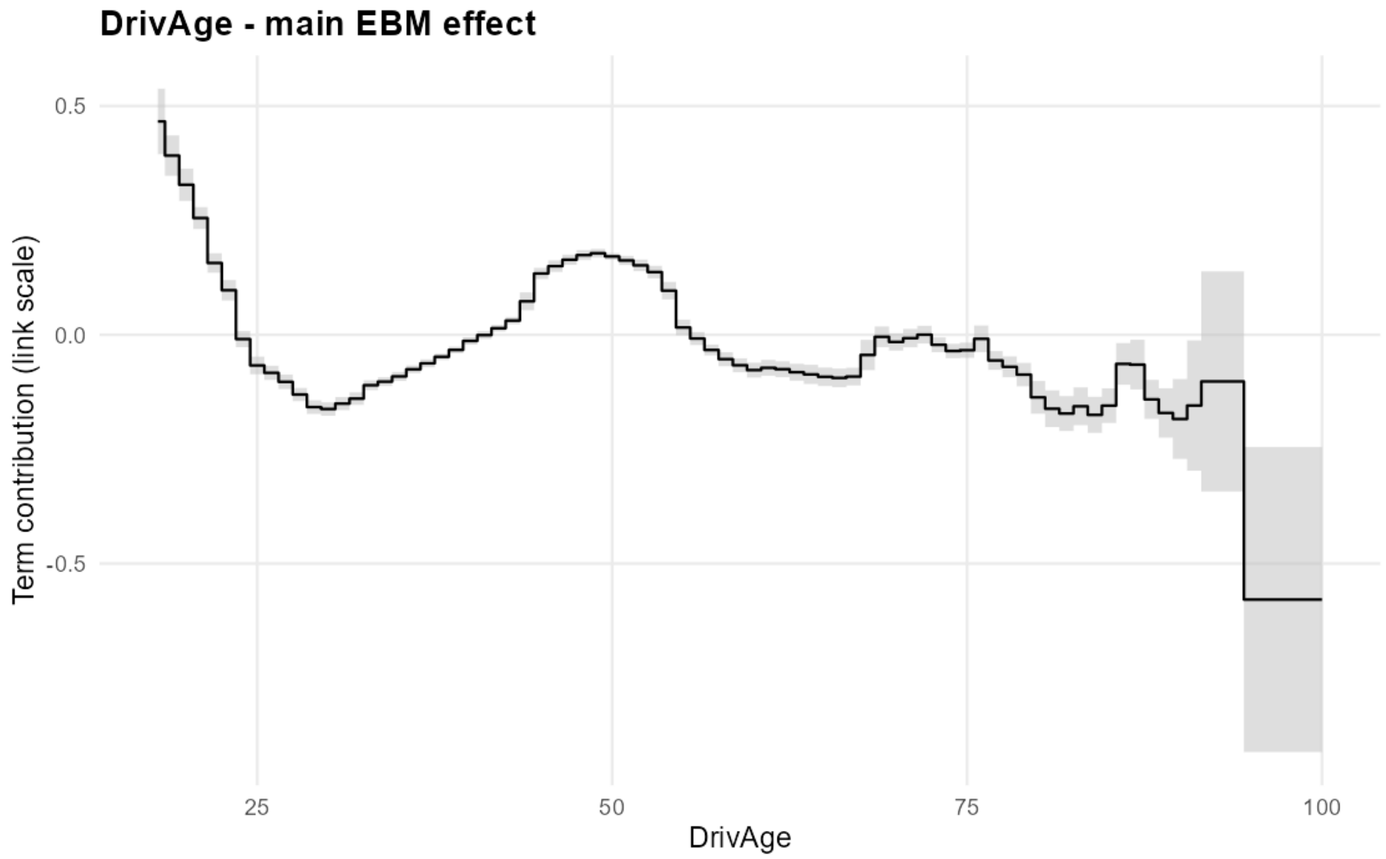
EBM: Tree-based shape functions



Sources: Lin Y. et al.: *Intelligible Models for Classification and Regression*

InterpretML: Another Way to Explain Your Model, Towards Data Science

Explainable Boosting Machine



LocalGLMnet

LocalGLMnet

Key Features:

- GLM structure - retains the linear predictor structure of a GLM, preserving interpretability by design
- Feature dependent coefficients - regression parameters are functions of the input, not fixed constants, allowing for interaction effects

The LocalGLMnet is defined by the function

$$x \mapsto \beta_0 + \langle \boldsymbol{\beta}(\boldsymbol{x}),\ \boldsymbol{x} \rangle$$

Where β are the attention weights of the model.

Institute
and Faculty
of Actuaries

Thank you