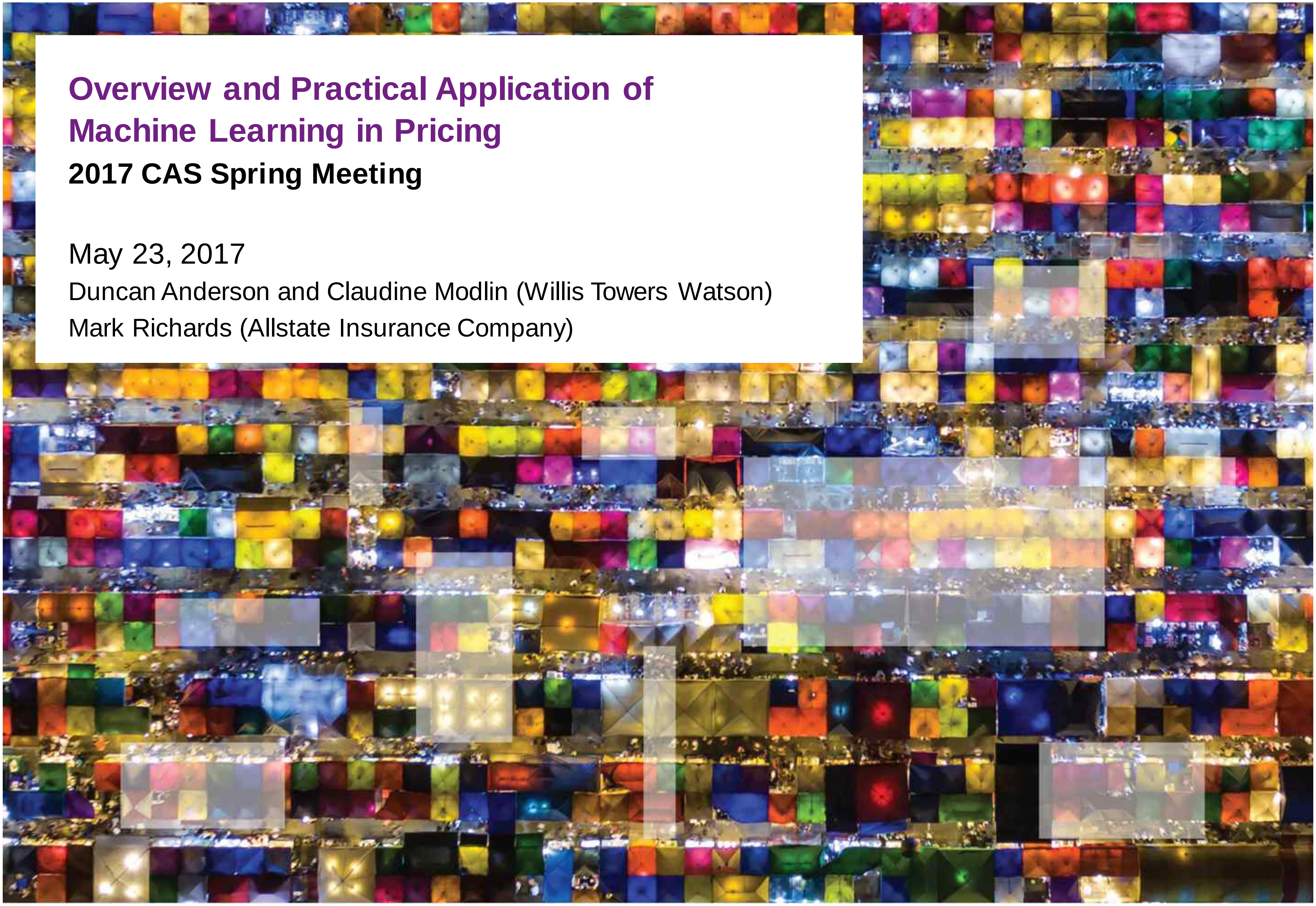
Overview and Practical Application of Machine Learning in Pricing

2017 CAS Spring Meeting

May 23, 2017

Duncan Anderson and Claudine Modlin (Willis Towers Watson)
Mark Richards (Allstate Insurance Company)



Agenda

Agenda
Context of machine learning in pricing
Penalized regression methods
Trees, random forests and GBMs
Conclusions
Q&A

Objective: to give you a working knowledge of some machine learning methods that may be used to improve GLM results and/or offer valuable insights in their own right in the field of P&C insurance pricing

Applications of machine learning in the insurance sector

Marketing and Distribution

Customer management

Underwriting and risk management

Pricing

Claims management

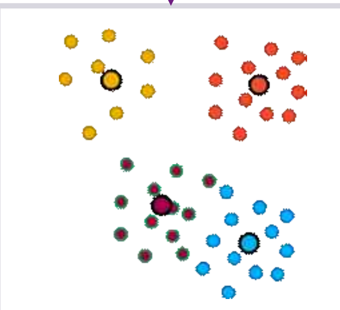
Asset management

Product development (UBI)

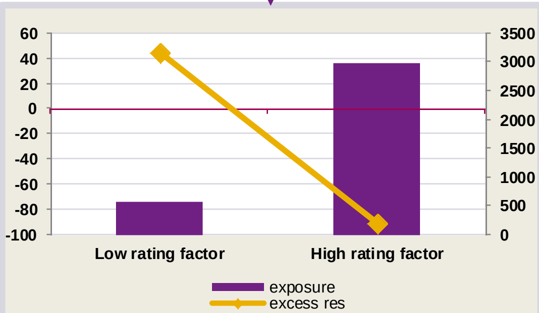
Retention segmentation and program design

Agent/Broker performance evaluation

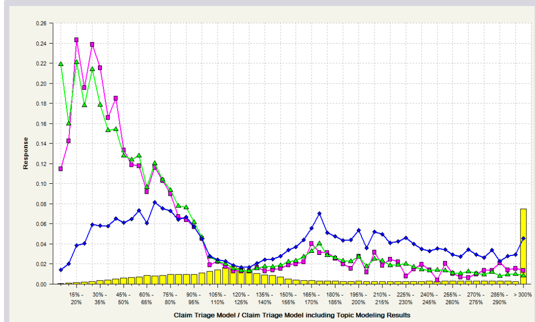
Fraud detection



P&C to Life cross-selling



Web-scraping and feature design in commercial lines

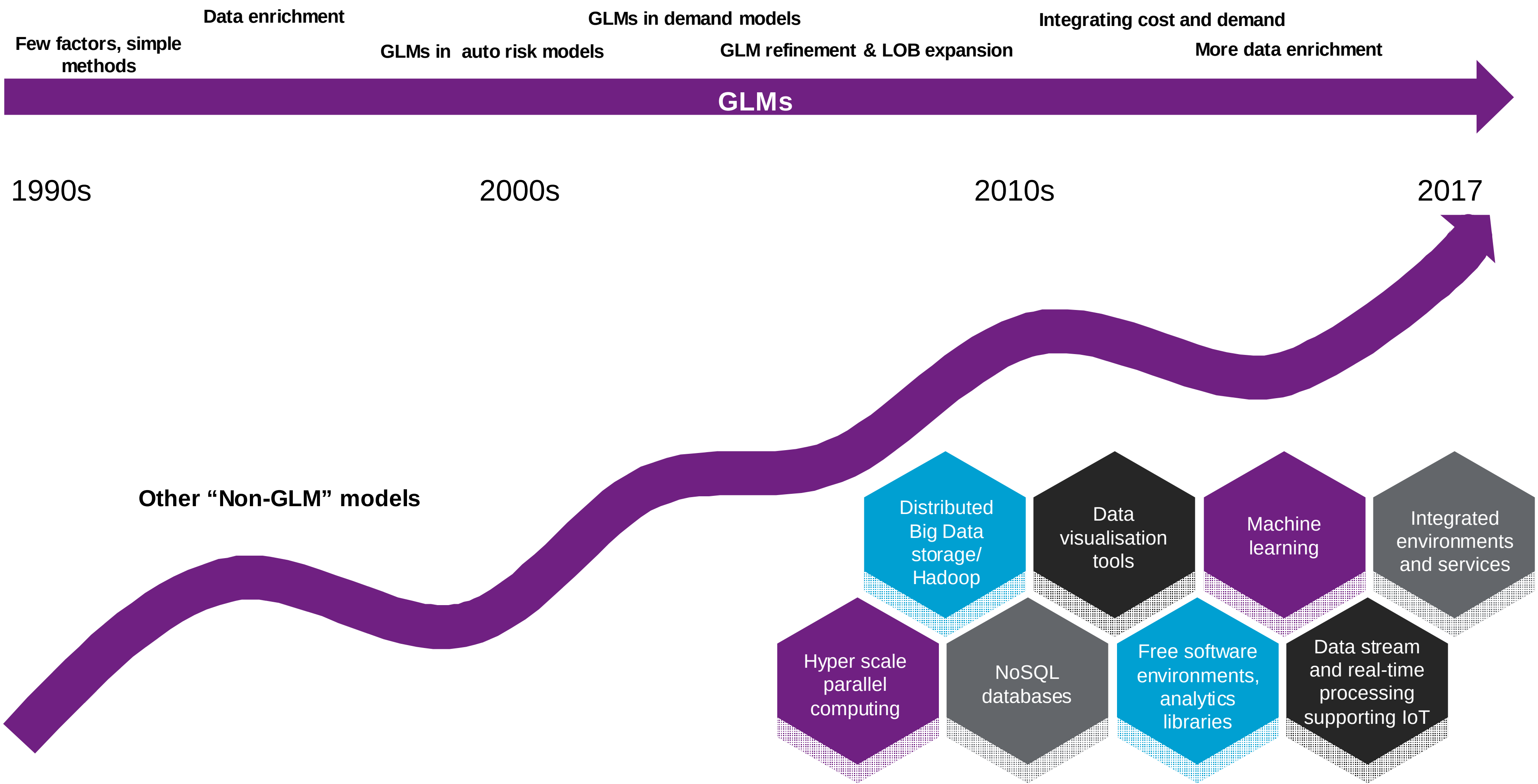


Topic modeling and large loss modeling

Focus of today's talk

Pricing

This is not new….



What are these methods?

Ensemble Methods

Classifications Trees

"Earth"

Regression Trees

Gradient Boosting Machines

K-nearest Neighbors

Elastic Net

Neural Networks

Naïve Bayes

Random Forests

K-Means Clustering

Principal Components Analysis

Lasso

Support Vector Machines

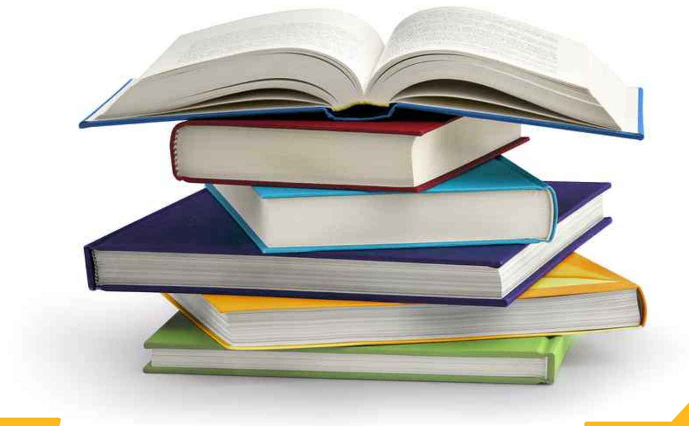
Ridge Regression

Does it work?

AIC

Gini

RMSE

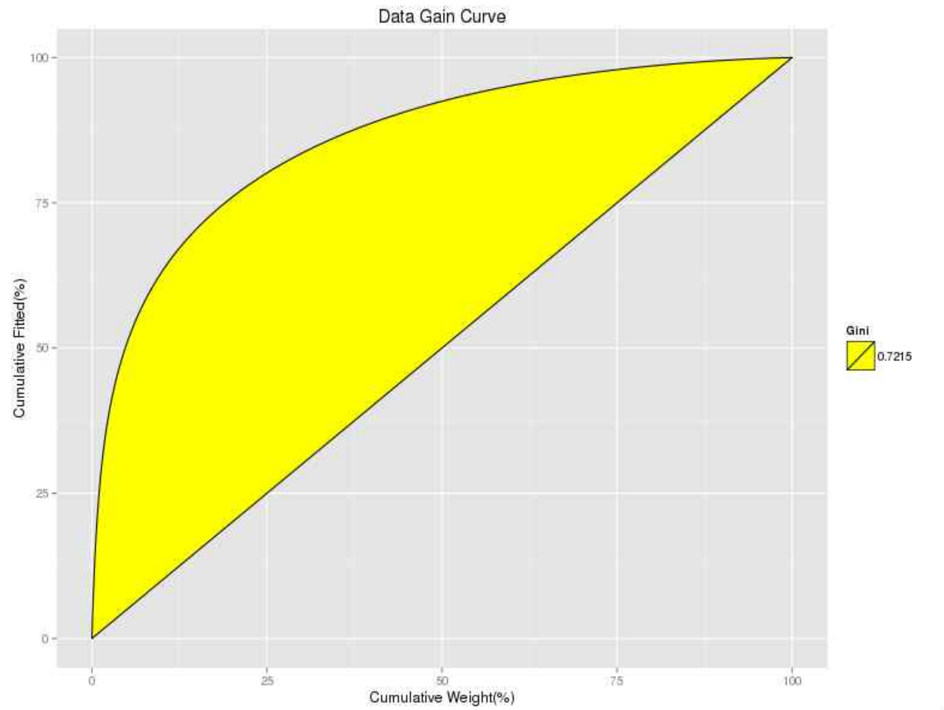


MAE

Log
loss

How do you measure value?

Gini



- Rank hold out observations by their **fitted values** (high to low)
- **Plot cumulative response** by cumulative exposure
- A **better model** will explain a **higher proportion of the response** with a **lower proportion of exposure**
- …and will give a **higher Gini coefficient** (yellow area)

Example results

Model	Gini
GLM	0.327

Example results

Model	Gini
GLM	0.327
New Model	0.330

Example results

Model	Gini	Gini improvement
GLM	0.327	0.0%
New Model	0.330	1.0%

Example results

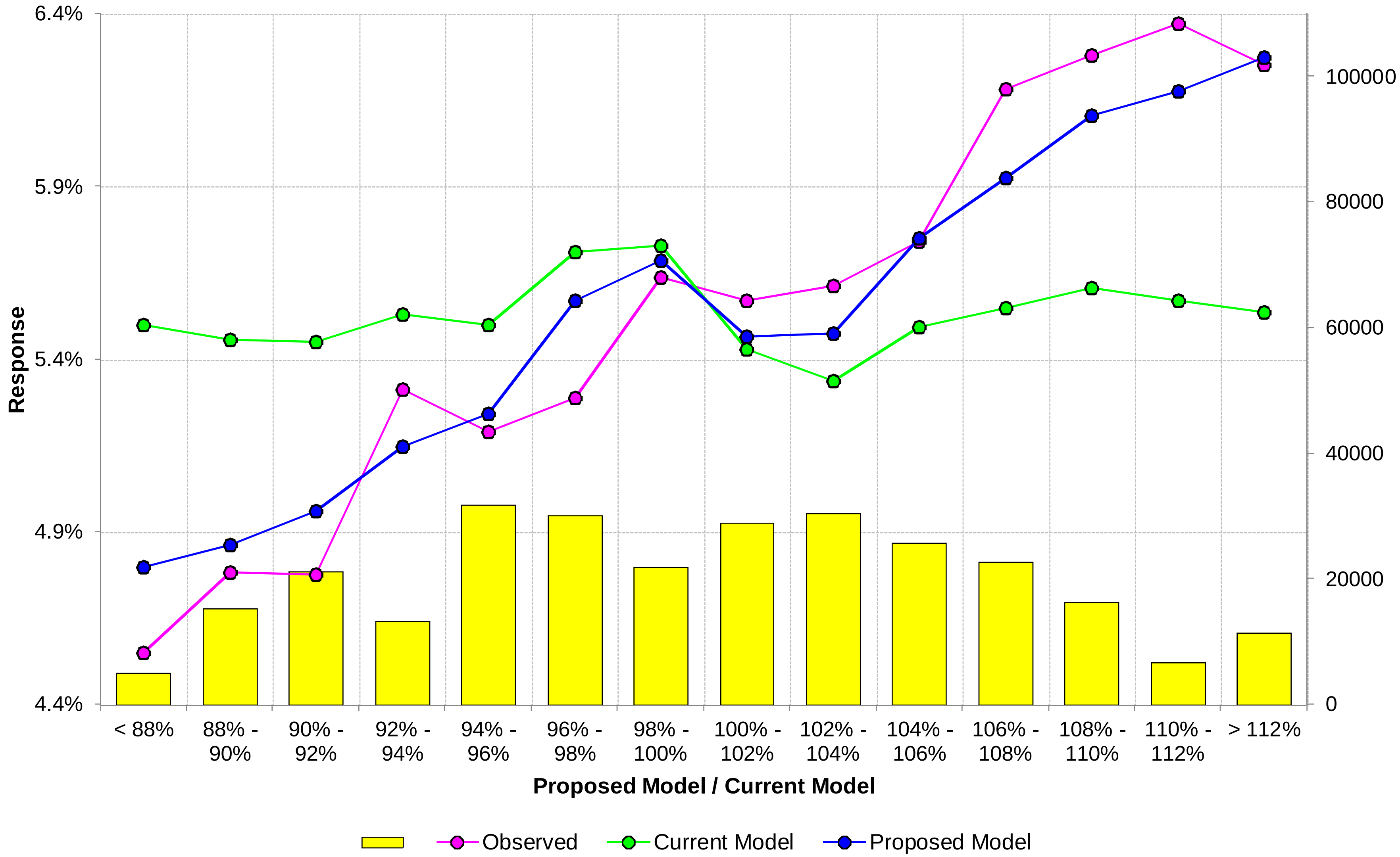
Model	Gini	Gini improvement	Gini rank
GLM (main factor removed)	0.318	-2.6%	4
GLM (minor factor removed)	0.322	-1.3%	3
GLM	0.327	0.0%	2
New Model	0.330	1.0%	1

But…

- Think of a model…
- Multiply it by 123
- Square it
- Add 74½ billion

- …and you get the same Gini coefficient!

Double lift chart



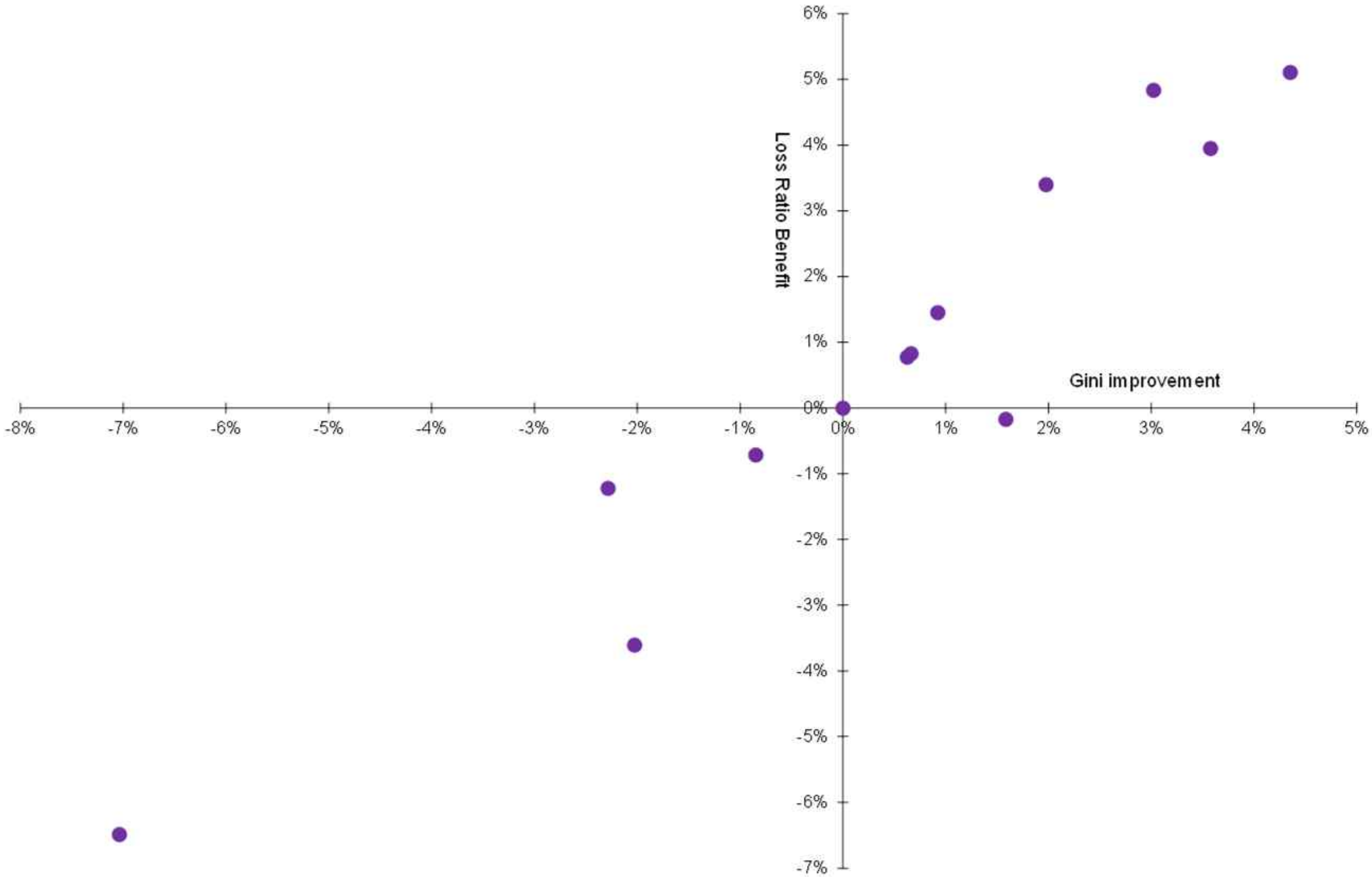
Financial value estimate

- Errors in insurance pricing are not symmetrical
- Financial benefit can be estimated

Example results

Results redacted

Financial value vs Gini



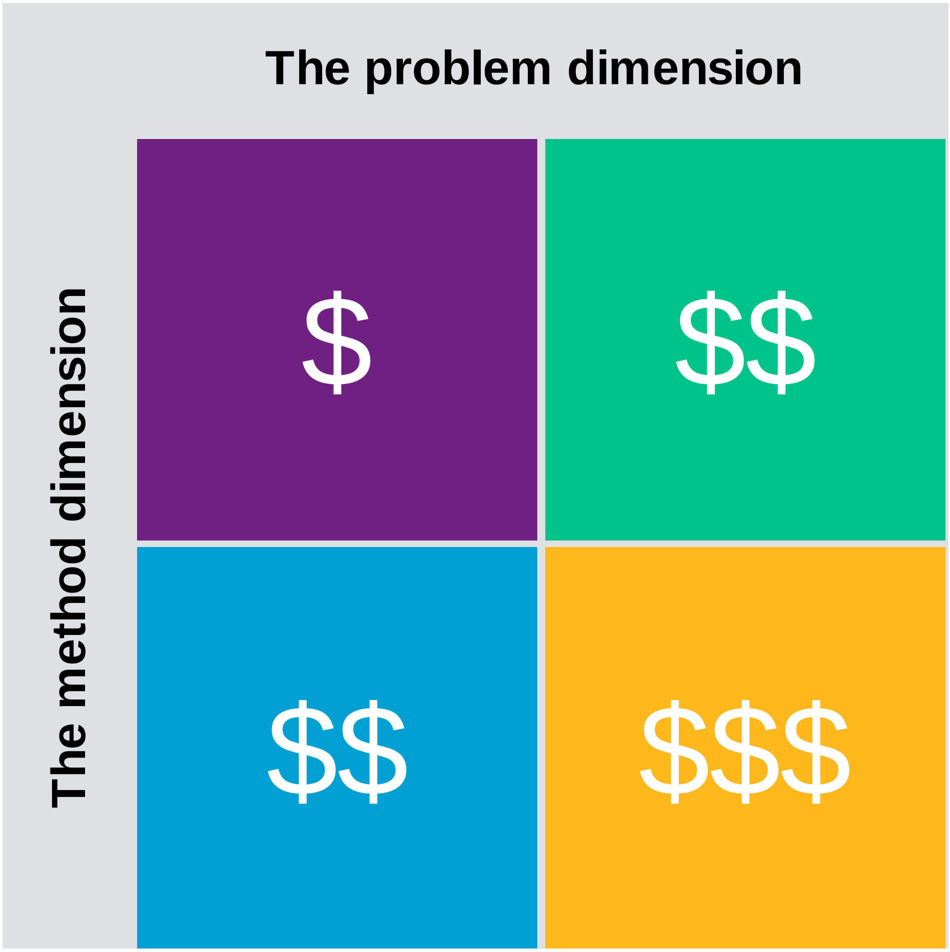
Is it really all about the method?

The problem dimension

The method dimension

Original problem, solved with original methods	Newly defined problem, solved with original methods
Original problem, solved with new techniques	Newly defined problem, solved with new techniques

Where is the value?



The problem dimension

The method dimension	The problem dimension	
	Original problem, solved with original methods	Newly defined problem, solved with original methods
	Original problem, solved with new techniques	Newly defined problem, solved with new techniques

What are these methods?

Ensemble Methods

Classifications Trees

"Earth"

Regression Trees

Gradient Boosting Machines

K-nearest Neighbors

Elastic Net

Neural Networks

Naïve Bayes

Random Forests

K-Means Clustering

Principal Components Analysis

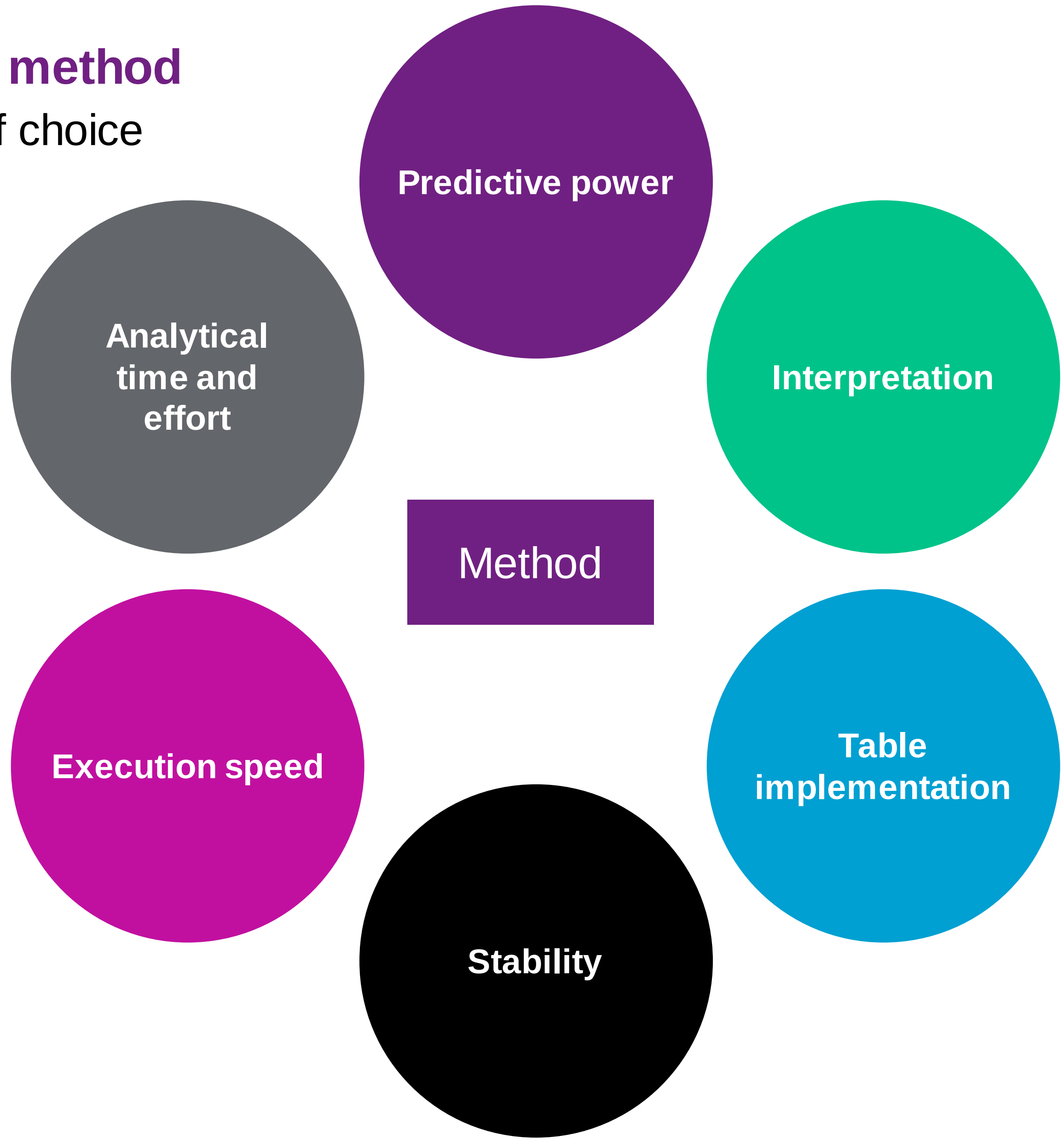
Lasso

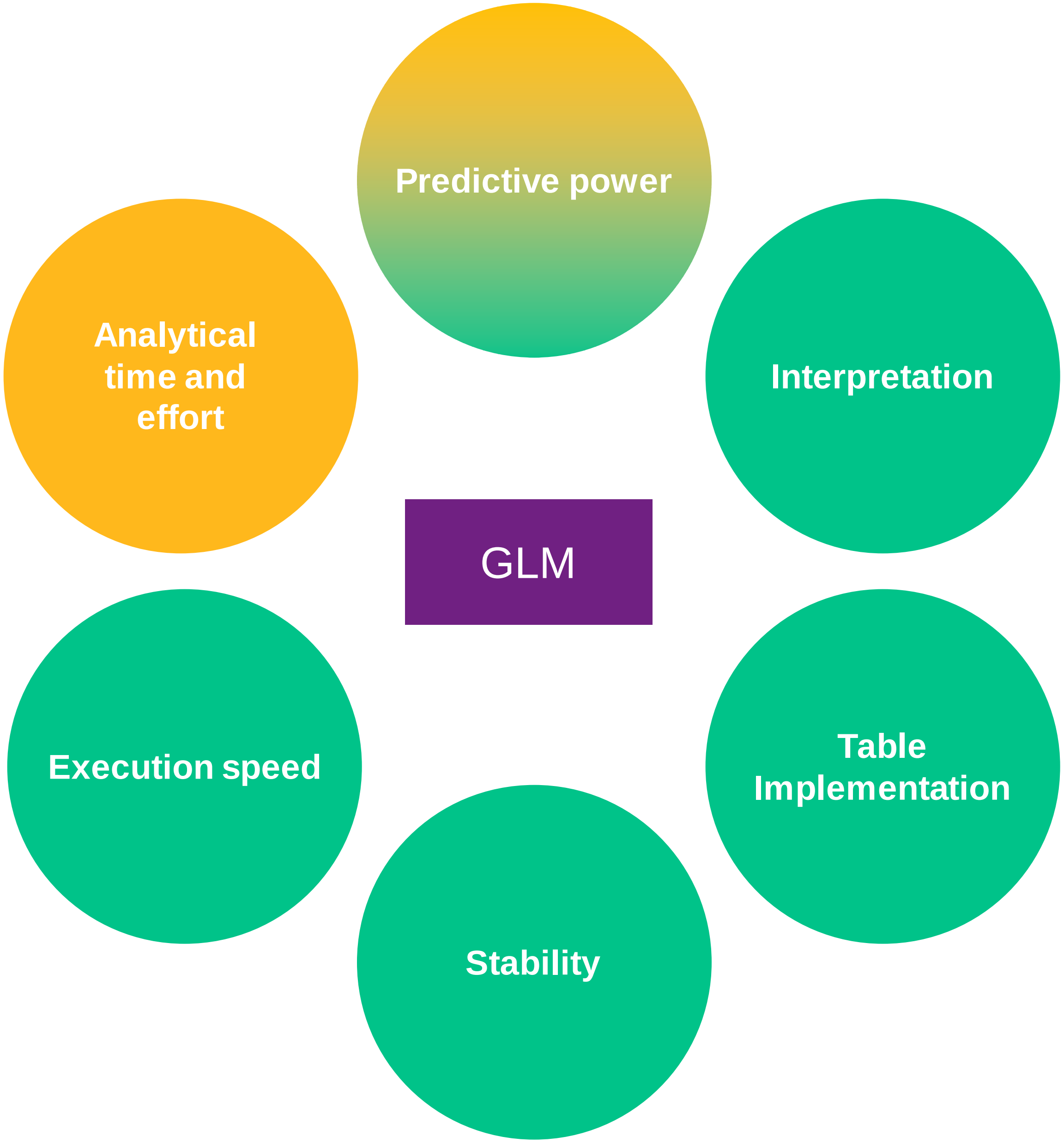
Support Vector Machines

Ridge Regression

Choosing a method

Dimensions of choice




Predictive power
Analytical time and effort
Interpretation
GLM
Execution speed
Table Implementation
Stability

What are these methods?

Ensemble Methods	Classifications Trees	"Earth"	Regression Trees	Gradient Boosting Machines
K-nearest Neighbors	Elastic Net	Neural Networks	Naïve Bayes	Random Forests
K-Means Clustering	Principal Components Analysis	Lasso	Support Vector Machines	Ridge Regression

Penalized Regression

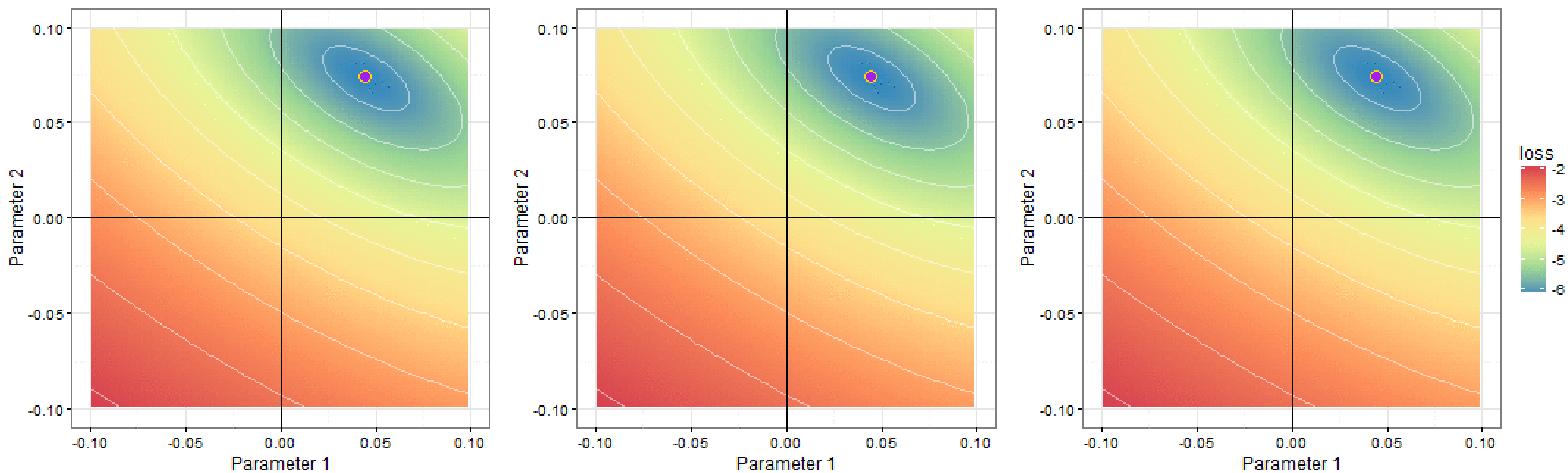
GLM | Lasso | Ridge

f(x) = $g^{-1}(\mathbf{X}.\underline{\beta})$ where $\underline{\beta}$ estimated by minimizing

$$L(\beta|X,y) + \underbrace{\boxed{\lambda_1 \sum_i |\beta_i|}_{\text{Lasso}} + \boxed{\lambda_2 \sum_i \beta_i^2}_{\text{Ridge}}}_{\text{Elastic Net}}$$

Elastic Net

Ridge $\sum_i \beta_i^2$ | **Elastic Net** | **Lasso** $\sum_i |\beta_i|$



Heavily penalize large parameters, but does not reduce parameters to zero

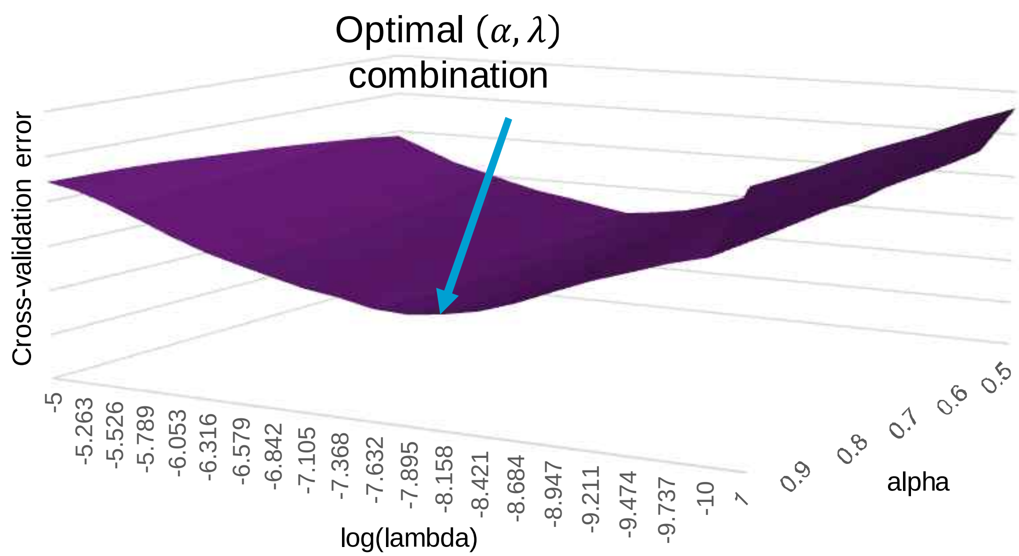
Mix of the two

Penalty reduces insignificant parameter values to zero - useful for variable selection

Penalized Regression

Parameter selection

- Minimize: $L(\beta|X,y) + \lambda_1 \sum_i |\beta_i| + \lambda_2 \sum_i \beta_i^2$
- Penalty parameters can be re-written: $\lambda_1 = \lambda\alpha,\quad \lambda_2 = \lambda\left(\frac{1-\alpha}{2}\right)$
- α controls the mixture between Lasso ($\alpha = 1$) and Ridge ($\alpha = 0$)
- λ controls the overall size of the penalty
- λ, α selected using cross-validation
- Factors automatically selected from initial set!



Penalized Regression

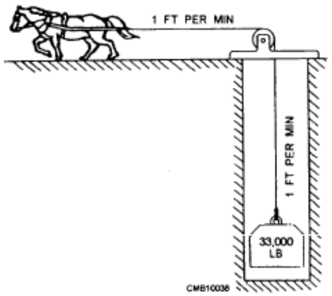
Case study – vehicle classification

Physical facticity

E.g., height, length, weight

Mechanical nature

E.g., engine size, fuel type



Qualitative descriptors

E.g., body type, model range

Performance

E.g., maximum speed, torque, BHP

Penalized Regression

Case study – vehicle classification

Results redacted

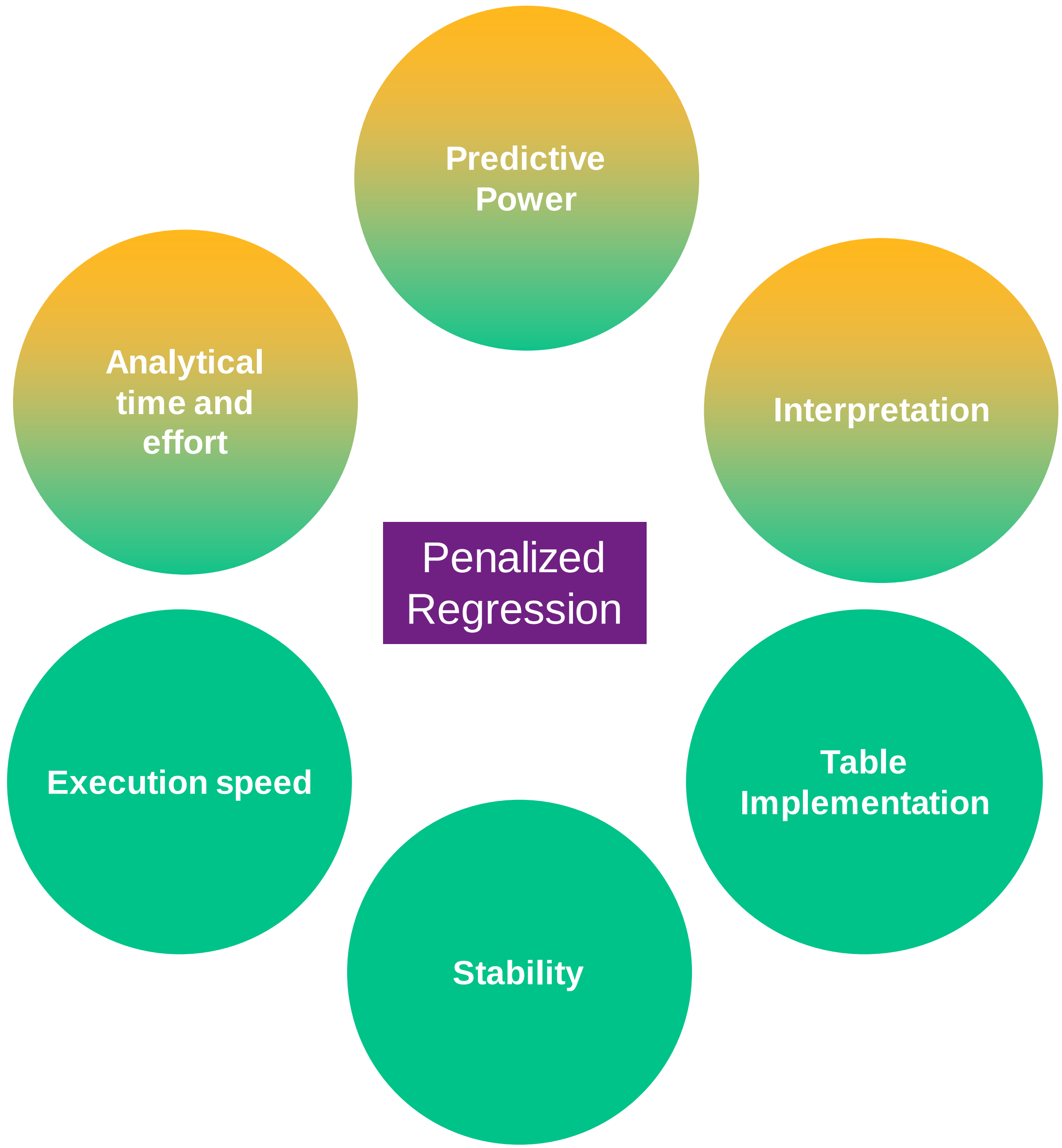
Deploying Penalized Regression

Same as GLMs!

Age	Multiplier
<20	2.12
20-25	1.74
25-30	1.09
30-39	1.00
40-49	0.95
50+	0.06

Vehicle Group	Multiplier
1	0.83
2	0.91
3	0.96
4	1.00
5	1.05
6	1.17
7	1.25
8	1.42
9	1.89

Gender	Multiplier
Male	1.00
Female	0.97


Predictive Power
Analytical time and effort
Interpretation
Penalized Regression
Execution speed
Table Implementation
Stability

What are these methods?

Ensemble Methods

Classifications Trees

"Earth"

Regression Trees

Gradient Boosting Machines

K-nearest Neighbors

Elastic Net

Neural Networks

Naïve Bayes

Random Forests

K-Means Clustering

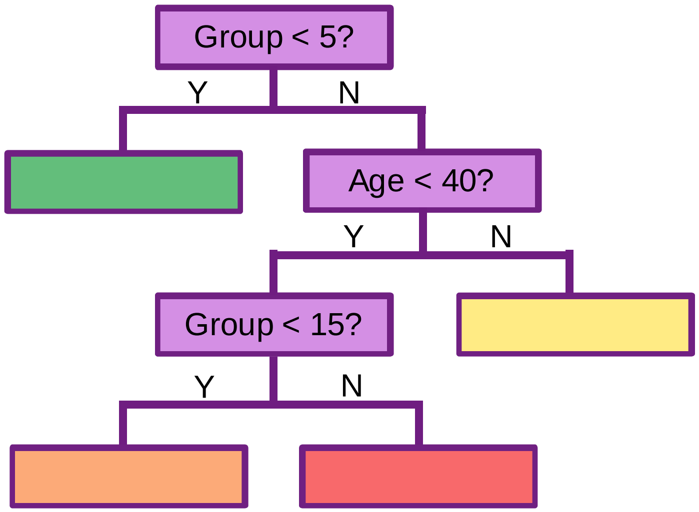
Principal Components Analysis

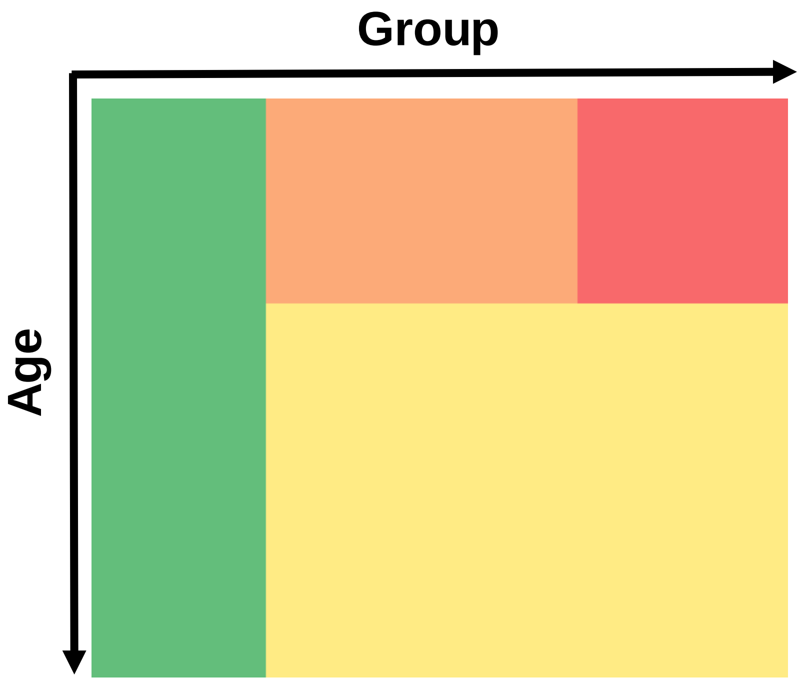
Lasso

Support Vector Machines

Ridge Regression

Decision Trees

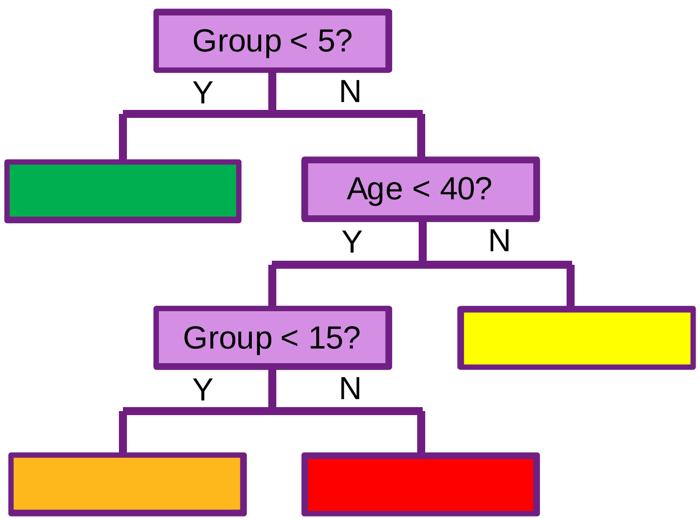




Random Forests

A tree

$$f_i(x)$$



A random forest

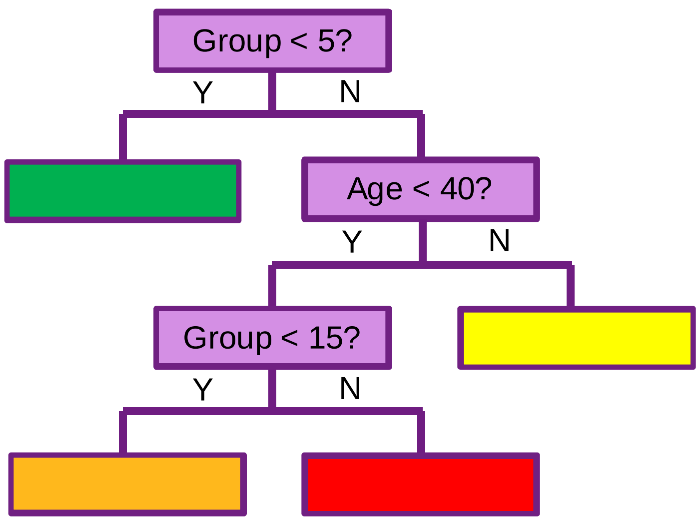
$$f(x) = \frac{1}{N}\sum_{n=1}^{N} f_n(x)$$

$$\frac{1}{N}$$

Gradient Boosting Machine or “GBM”

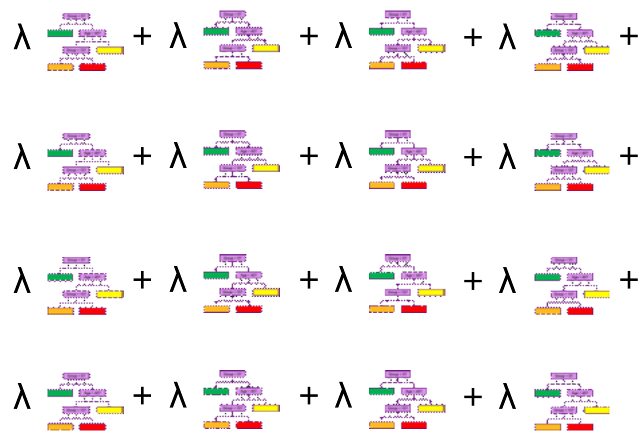
A tree

$$f_i(x)$$



A GBM

$$f(x) = \lambda \sum_{n=1}^{N} f_n(x)$$



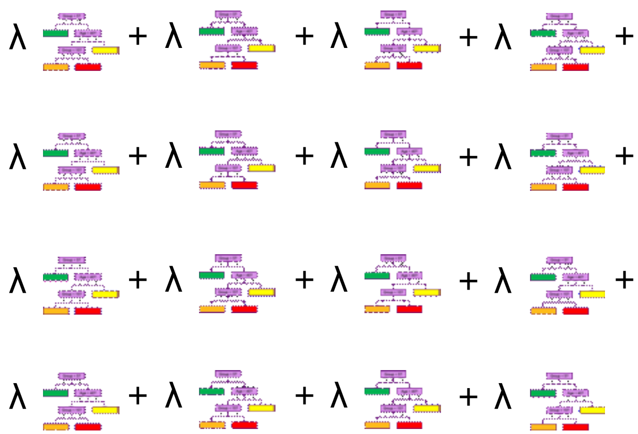
Gradient Boosted Machine or "GBM"

Four main assumptions

- λ **Learning rate / "shrinkage"**
  - Amount by which the old model predictions are varied for the next model iteration
  - New model =
    Old + (Prediction x Learning rate)
- **Interaction depth**
  - Number of splits allowed on each tree
    (or the number of terminal nodes – 1)
- **N Number of trees** (iterations) allowed
- **Bag fraction**
  - Trees are fitted to a subset of the data (the bag fraction) on a randomized basis
  - Additional noise-reduction can be achieved by using a random subset of the available factors at each iteration

A GBM

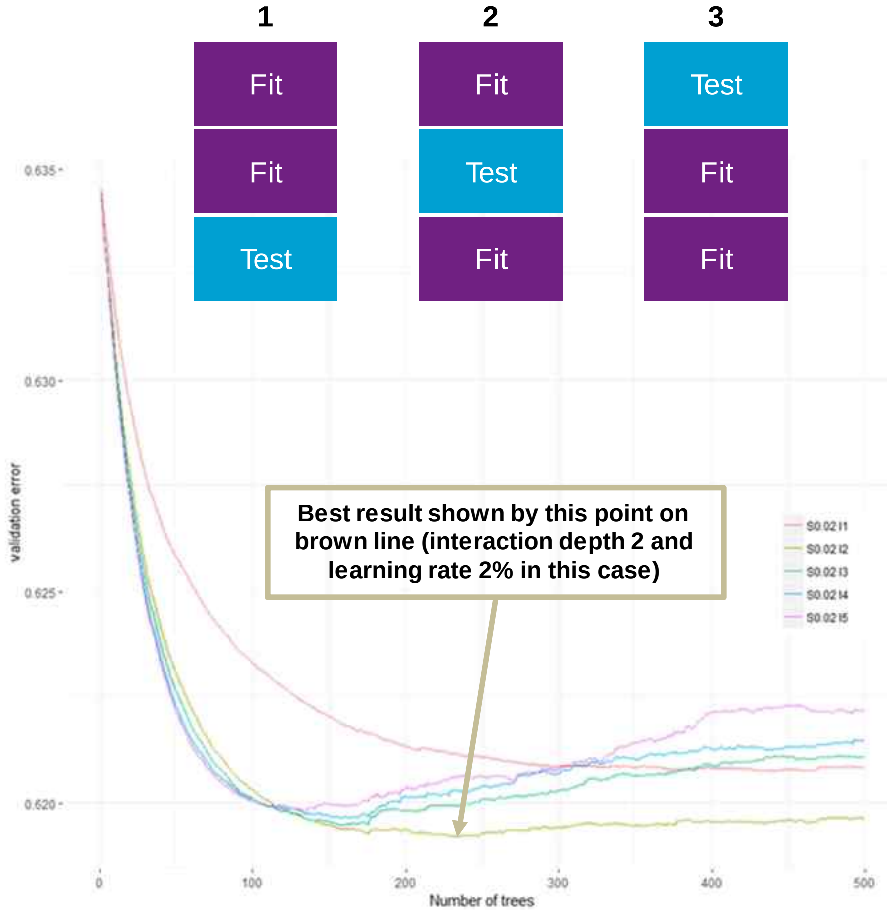
$$f(x) = \lambda \sum_{n=1}^{N} f_n(x)$$



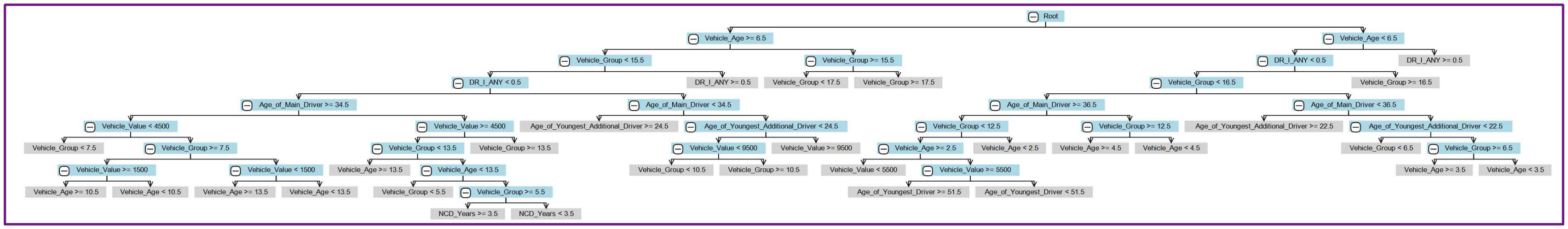
Gradient Boosted Machine or “GBM”

Four main assumptions

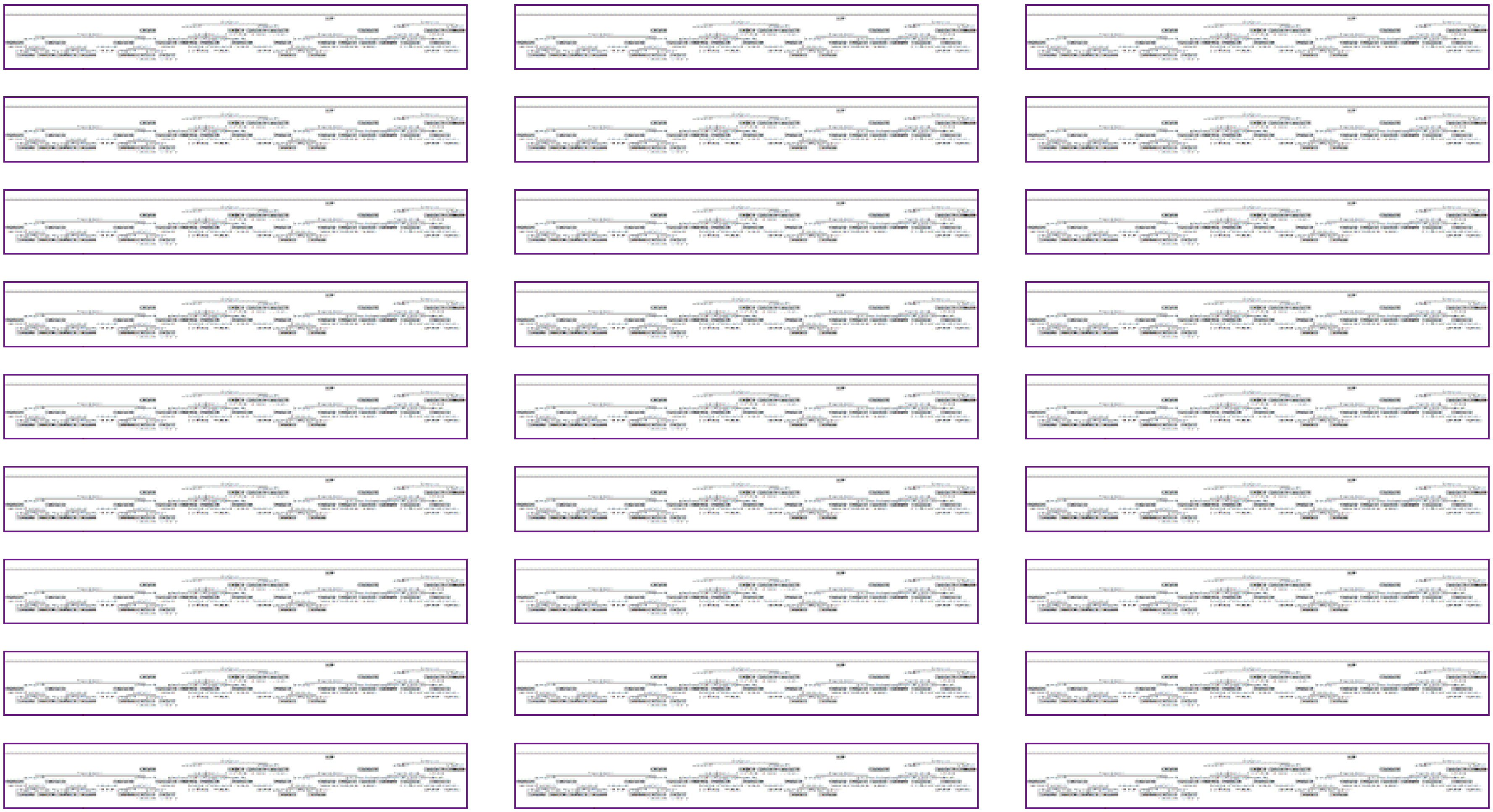
- λ **Learning rate / “shrinkage”**
  - Amount by which the old model predictions are varied for the next model iteration
  - New model =
    Old + (Prediction x Learning rate)
- **Interaction depth**
  - Number of splits allowed on each tree
    (or the number of terminal nodes – 1)
- **N Number of trees** (iterations) allowed
- **Bag fraction**
  - Trees are fitted to a subset of the data (the bag fraction) on a randomized basis
  - Additional noise-reduction can be achieved by using a random subset of the available factors at each iteration

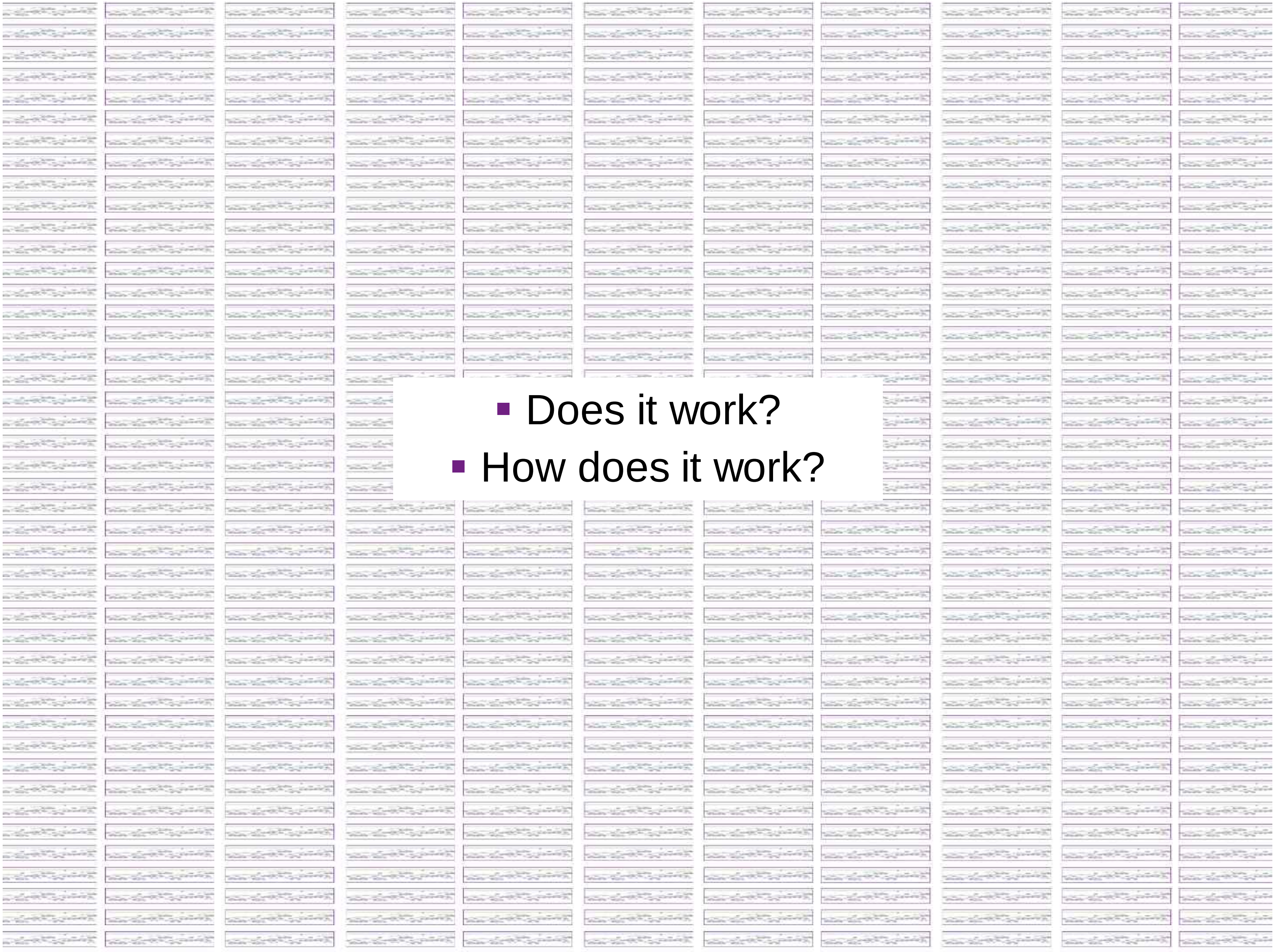


What does a GBM look like?



What does a GBM look like?





- Does it work?
- How does it work?

GBM – value add

Results redacted

What are these methods?

Ensemble Methods

Classifications Trees

"Earth"

Regression Trees

Gradient Boosting Machines

K-nearest Neighbors

Elastic Net

Neural Networks

Naïve Bayes

Random Forests

K-Means Clustering

Principal Components Analysis

Lasso

Support Vector Machines

Ridge Regression

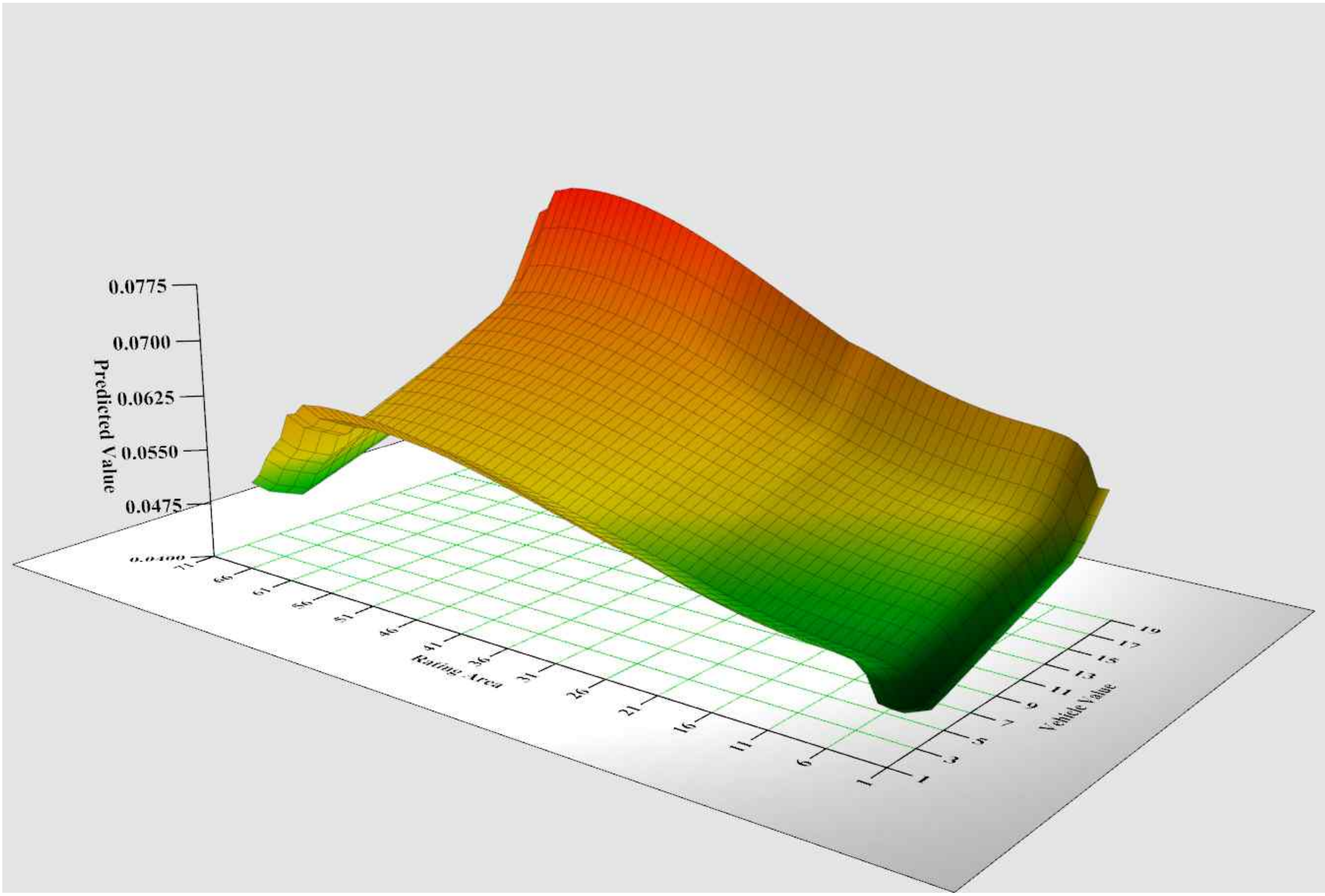
Adding an ensemble

Results redacted

What about a modelled down GBM?

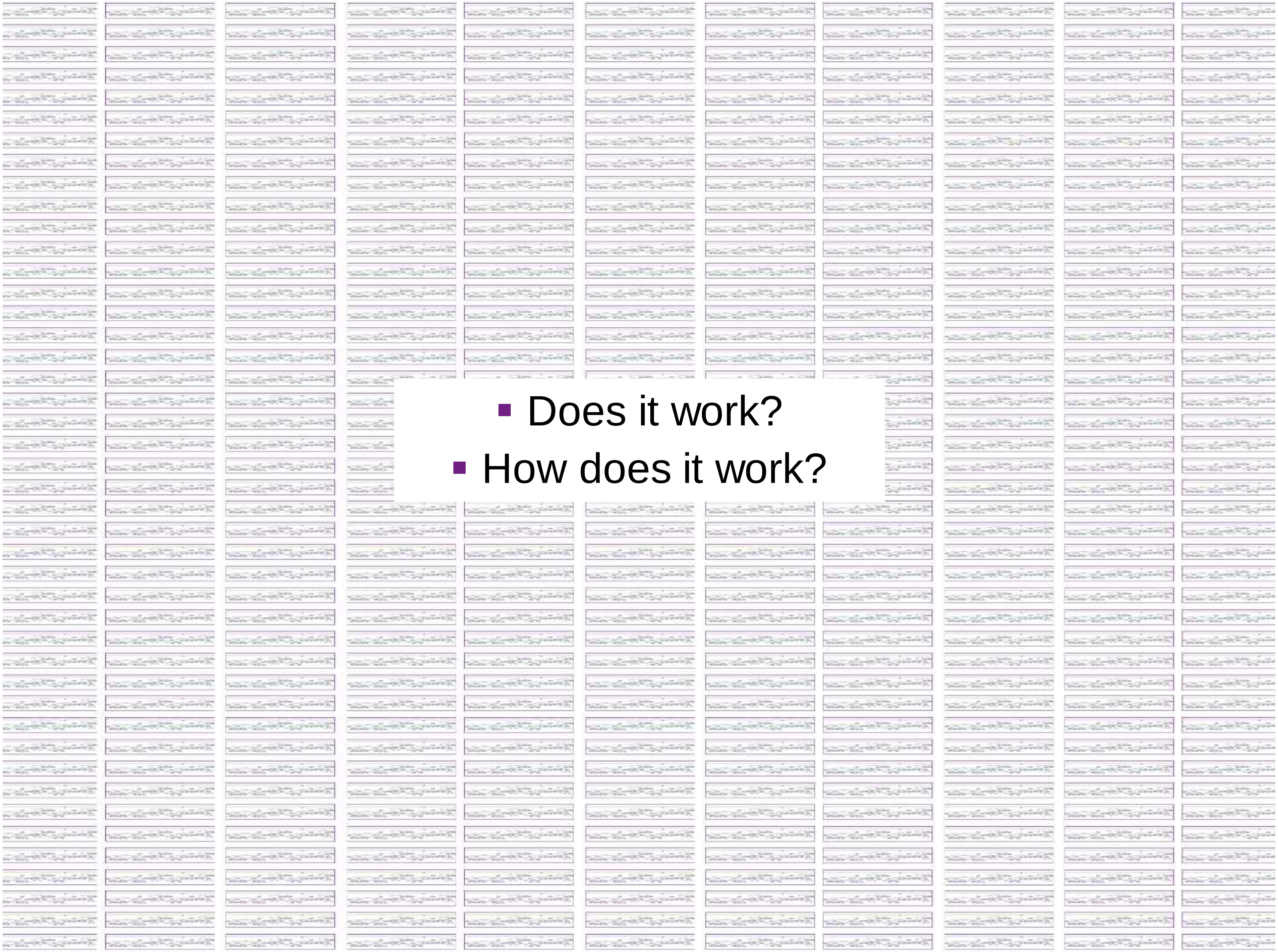
Results redacted

What about an automated GLM?



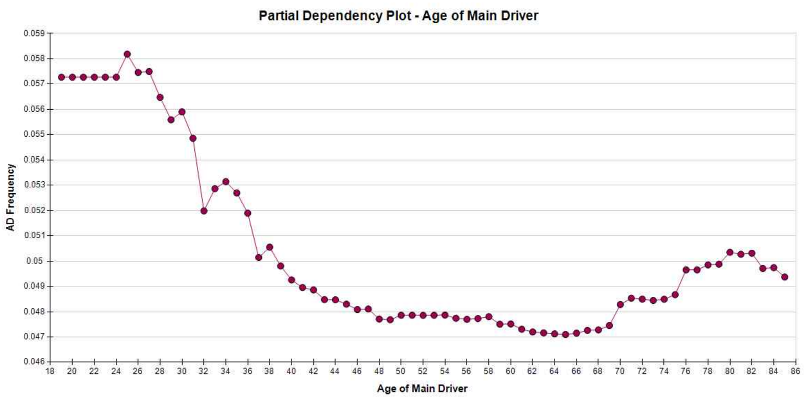
What about an automated GLM?

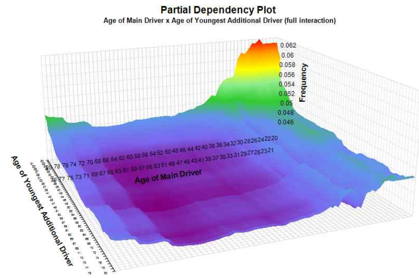
Results redacted

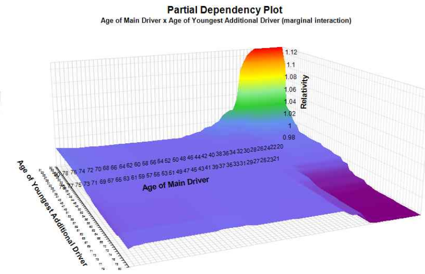


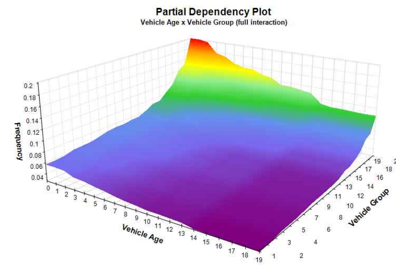
- Does it work?
- How does it work?

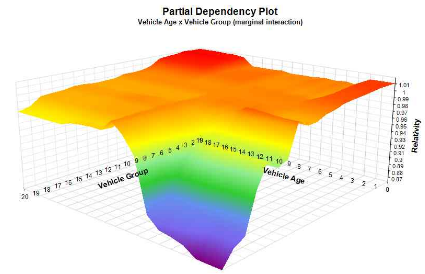
Partial dependency plots etc

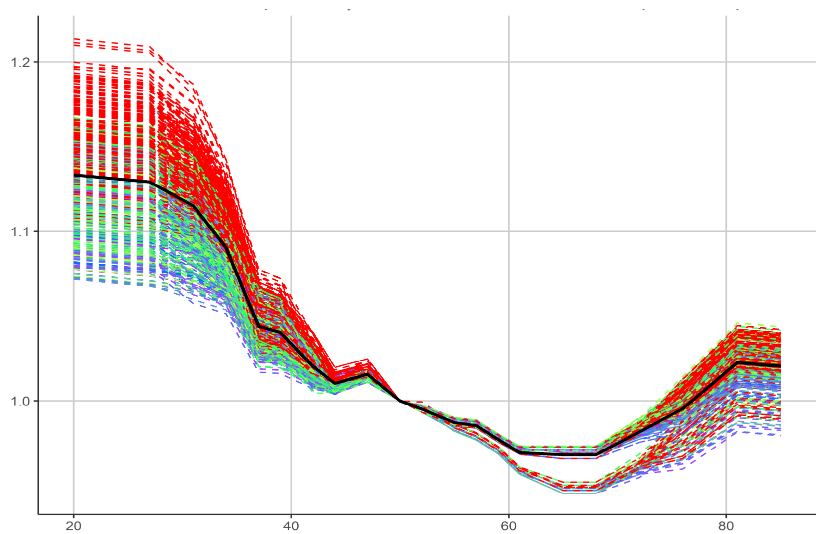


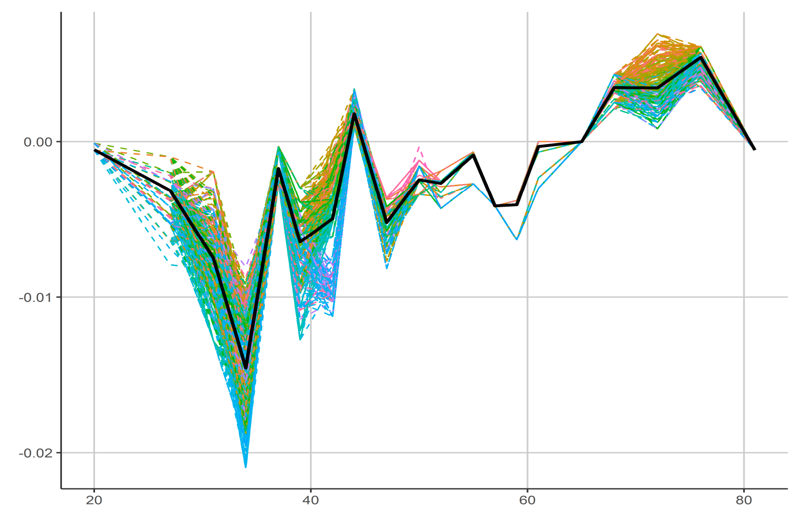












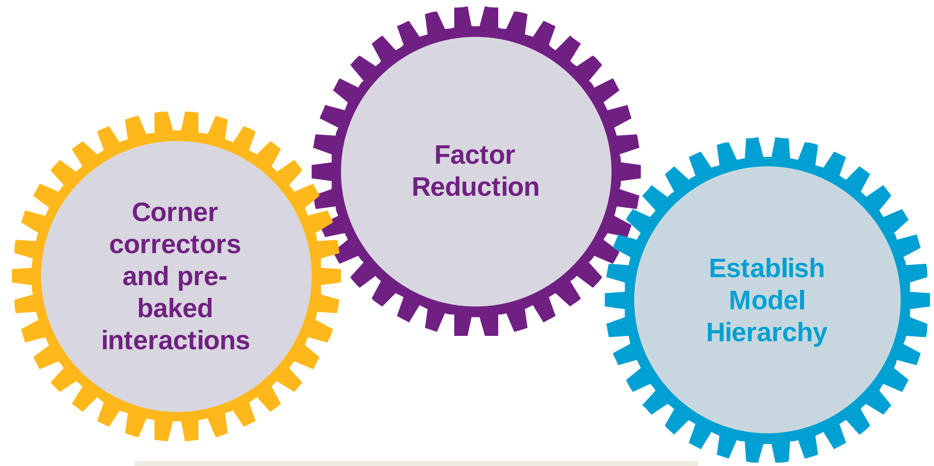
Deploying GBMs

Model down into multiplicative tables via GLMs

	Age	Exposure	Burning Cost
1	<=20	1,720	179
2	21-30	34,893	122
3	31-50	118,182	102
4	51+	127,054	70
5	Age Total	281,849	91

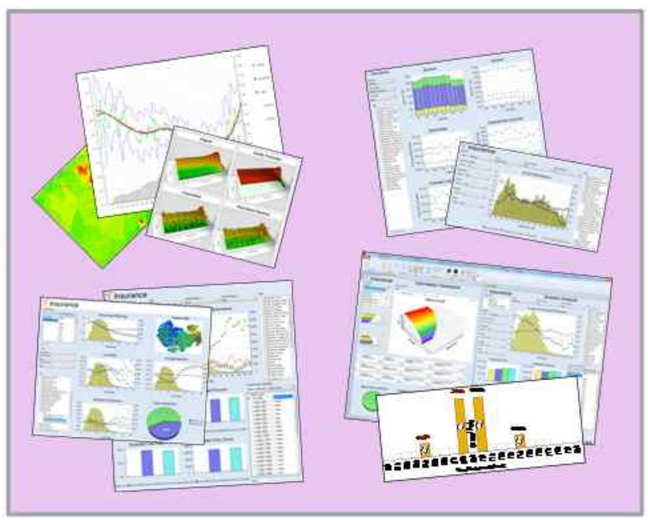
	Vehicle Group	Exposure	Burning Cost
1	1-10	164,107	77
2	11-14	84,859	101
3	15-18	28,952	116
4	19-20	3,931	272
5	VG Total	281,849	91

	Gender	Exposure	Burning Cost
1	Male	197,339	92
2	Female	84,510	87
3	Gender Total	281,849	91



Use insights to guide GLM

Analytical environ-ment

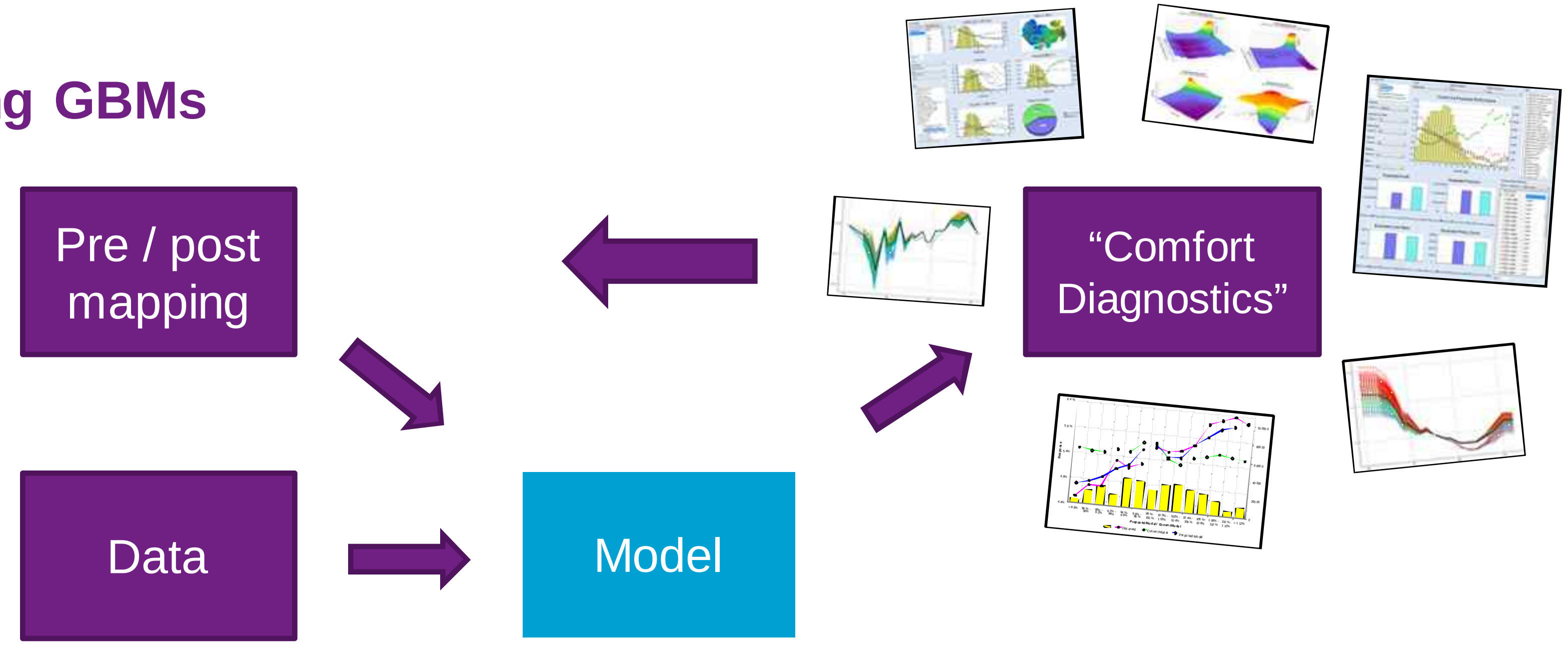


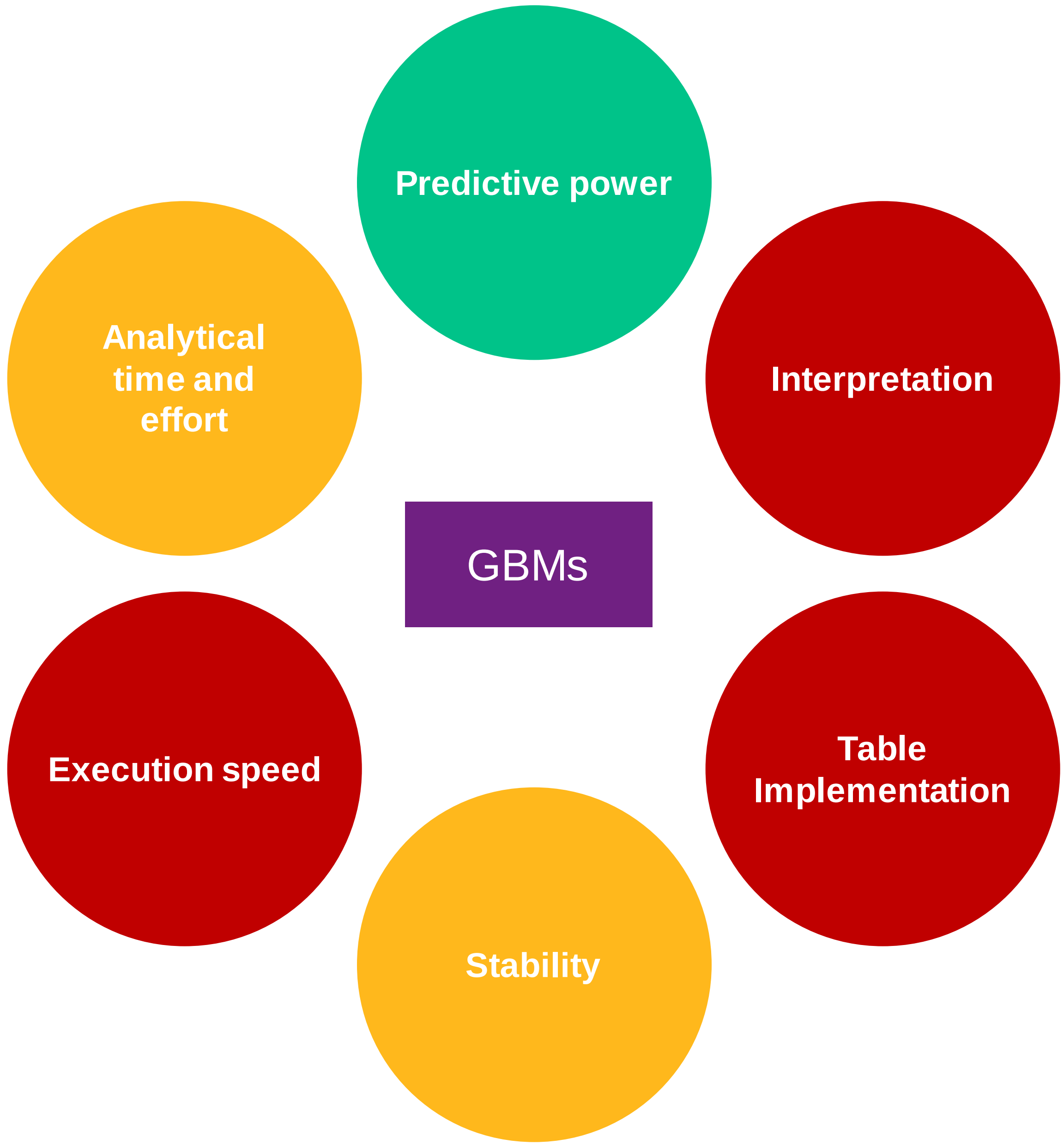
Next gen rating engine

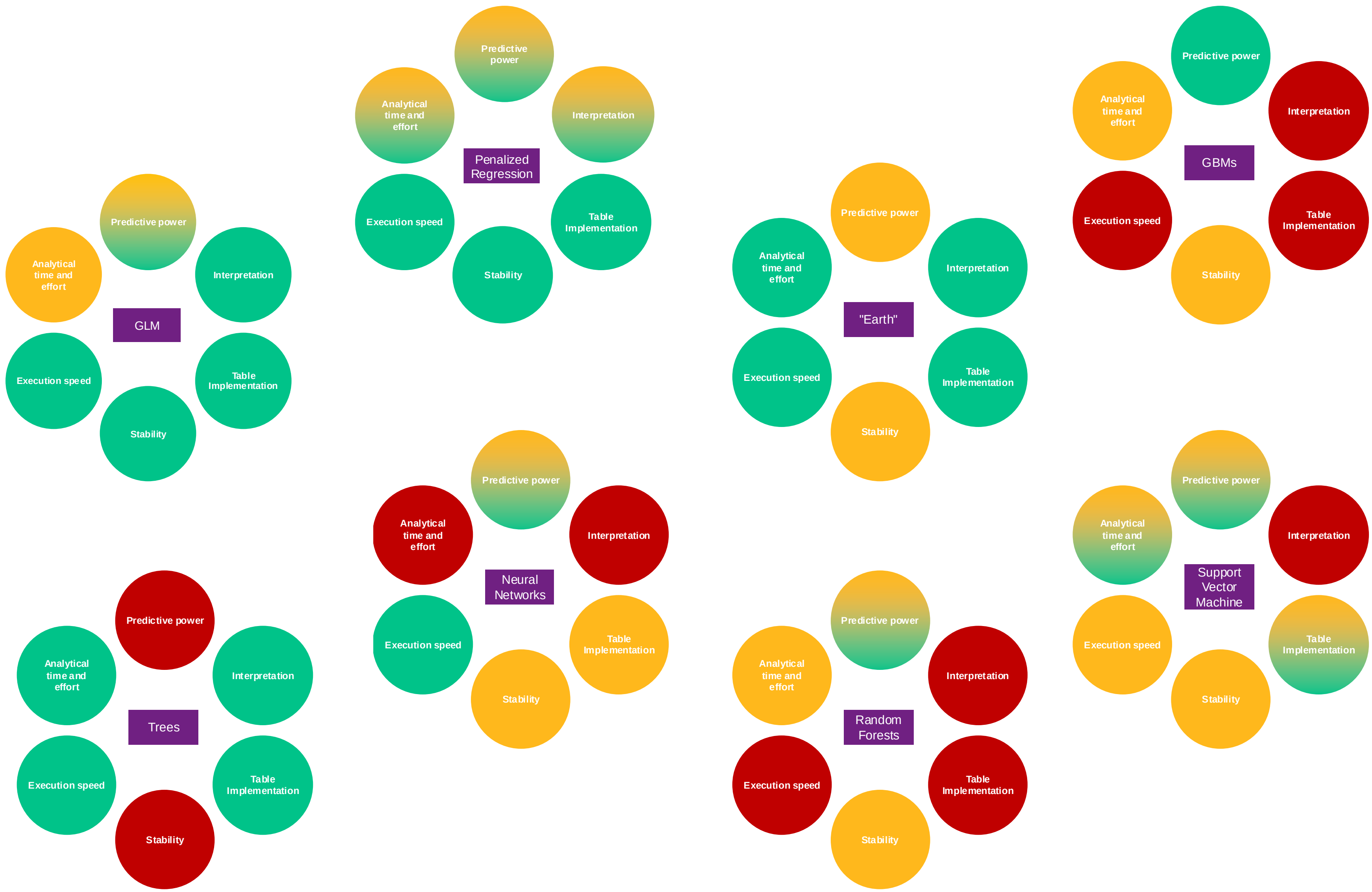
Main Policy Admin System

Deploy directly

Deploying GBMs



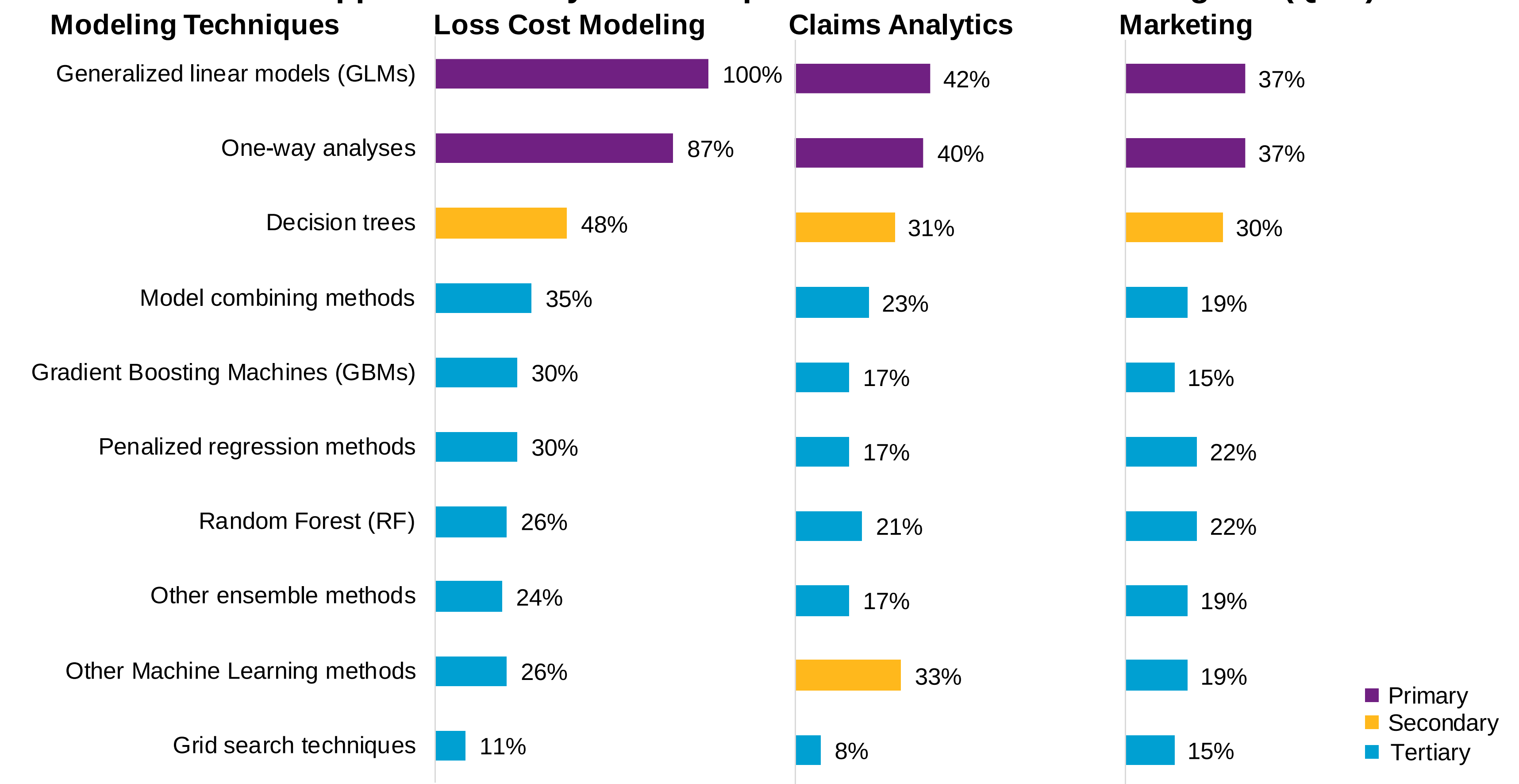

Predictive power
Analytical time and effort
Interpretation
GBMs
Execution speed
Table Implementation
Stability


Penalized Regression
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
GBMs
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
GLM
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
"Earth"
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
Neural Networks
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
Support Vector Machine
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
Trees
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability
Random Forests
Predictive power
Analytical time and effort
Interpretation
Execution speed
Table Implementation
Stability

How is the North American market doing with machine learning?

Methods used

For which business applications do you use or plan to use these methodologies? (Q.13)

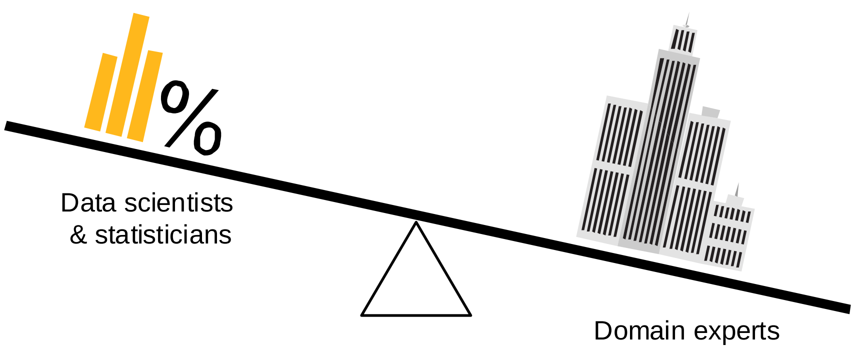


Base: U.S. respondents who use or plan to use the methodology for the application specified (Loss Cost Modeling n = 46, Claims Analytics n = 48, Marketing n = 27).

Machine learning in personal lines pricing: evolution or revolution?

Conclusions

- It's not all about methods
  - Domain expertise in formulating the problem can be more important
  - New (wider) data generally adds more lift than new methods
- There are practical ways to assess model improvement as well as statistical
- Predictions:
  - Many methods can augment the traditional GLM modeling process – in particular with growing datasets
  - Others (e.g., GMBs) can improve prediction in own right but these require different approaches in interpretation and technology/deployment
  - Methods support predictions in new areas
    - Very wide datasets
    - Market rate analysis
    - Cross-selling and other customer behaviors
    - UW, claims, etc



- But it's not all about predictive power
  - Fast investigation of new or "messy" data
  - Quick assessment of emerging experience
  - Operational efficiency
- Industry (and Profession) has work to do in developing machine learning skills integrated with domain expertise

Contacts

Claudine Modlin
Director, Risk Consulting & Software
Willis Towers Watson

Claudine.Modlin@willistowerswatson.com +1 805 499 2164

Duncan Anderson
Managing Director, Risk Consulting & Software
Willis Towers Watson

Duncan.Anderson@willistowerswatson.com +44 1737 274 080

Mark Richards
Director, Allstate Personal Lines
D^3: Data, Discovery & Decision Science

Mark.Richards@allstate.com +1 630 415 9547