Overview and Practical Application of Machine Learning Methods in Pricing

Lewis Maddock

22 November 2018

Willis Towers Watson

Contacts

Willis Towers Watson

Lewis Maddock
Director, Nordic P&C
Investment, Risk & Reinsurance

Lästmakargatan 22
Stockholm, 111 44
Sweden

+46 (0) 721 633 592
lewis.maddock@willistowerswatson

Confidentiality Statement

- This presentation has been provided for the sole and exclusive use of the members of the Swedish Actuarial Society.
- This document is for discussion purposes only with other members of the actuarial society and not to be relied upon for any other purpose.
- Distribution or disclosure of, or quotation from, or reference to this document to any other party, is prohibited without the prior written consent of Willis Towers Watson.
- For questions or further information about the content of this presentation please use contact details shown

Willis Towers Watson

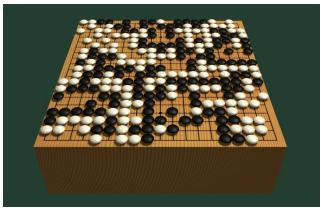
Introduction

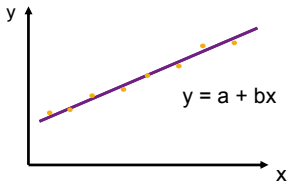
What I promised to do…

"We take a look at the common machine learning methods in use today and try to extract the value from the hype around this field"

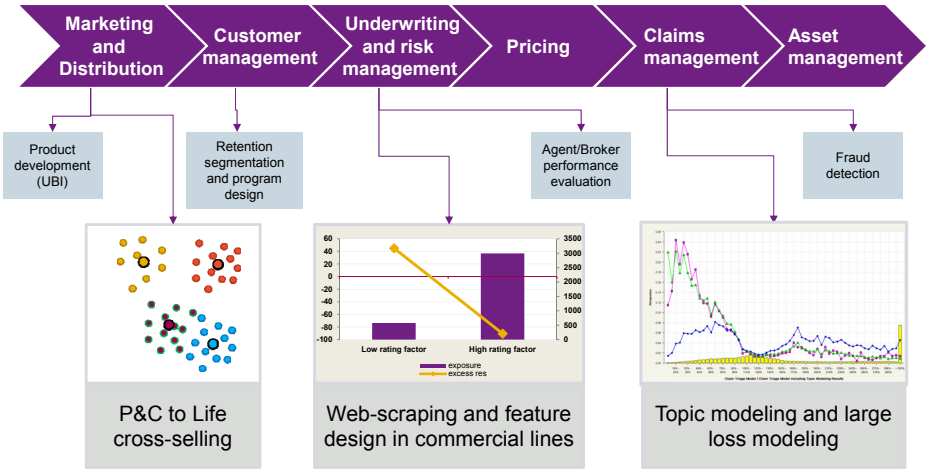
- What is needed to make them work?
- What are the strengths and weaknesses of these methods?
- Where does the role of the pricing actuary and the actuarial profession fit in?

Who's interested in what?



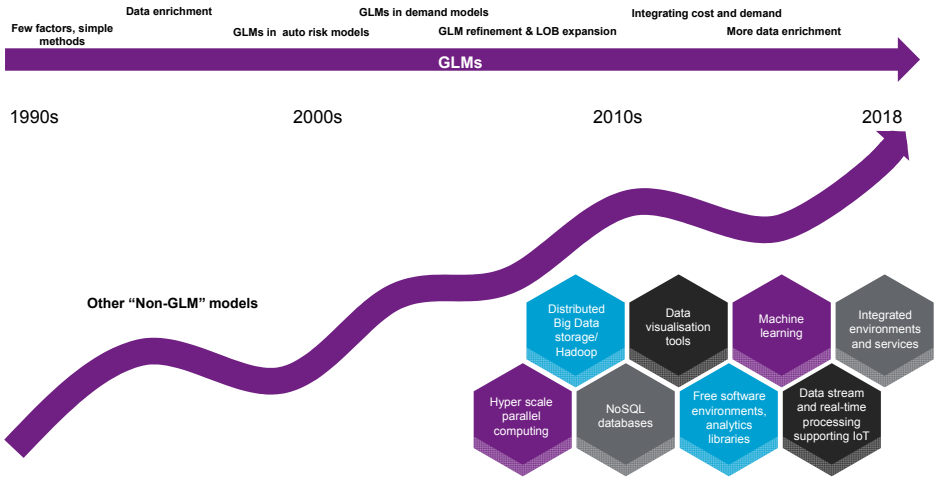


Applications of machine learning in the insurance sector



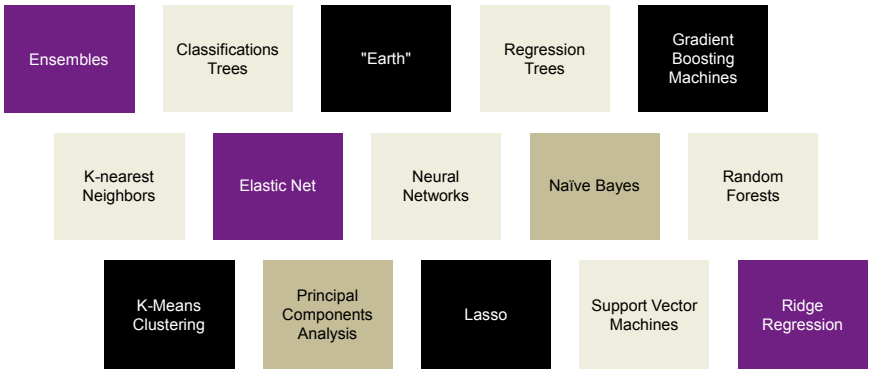
Willis Towers Watson

This is not new….



Willis Towers Watson

What are these machine learning methods?

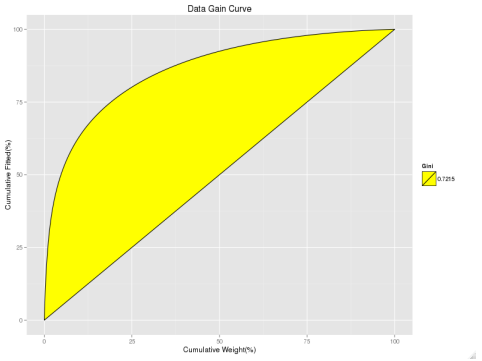


Willis Towers Watson

How do you know if a method works?

Willis Towers Watson

How do you measure value?



- 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)

Willis Towers Watson

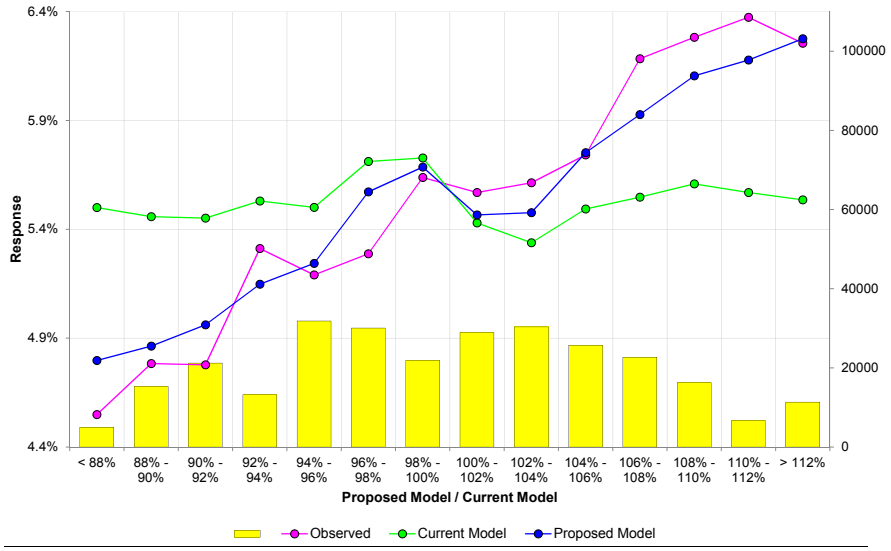
But…

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

- …and you get the same Gini coefficient!

Willis Towers Watson

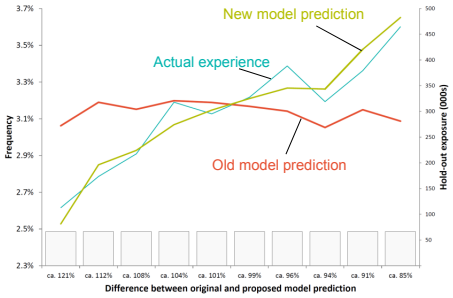
Double lift chart



Willis Towers Watson

Financial value estimate

- Errors in insurance pricing are not symmetrical
- Financial benefit can be estimated
- Consider ***actual experience*** in out of sample data for each percentile of old vs new model fitted values
- Estimate financial benefit that would have been attained
  - given an assumed elasticity
  - given business rules such as an assumed cap/floor approach



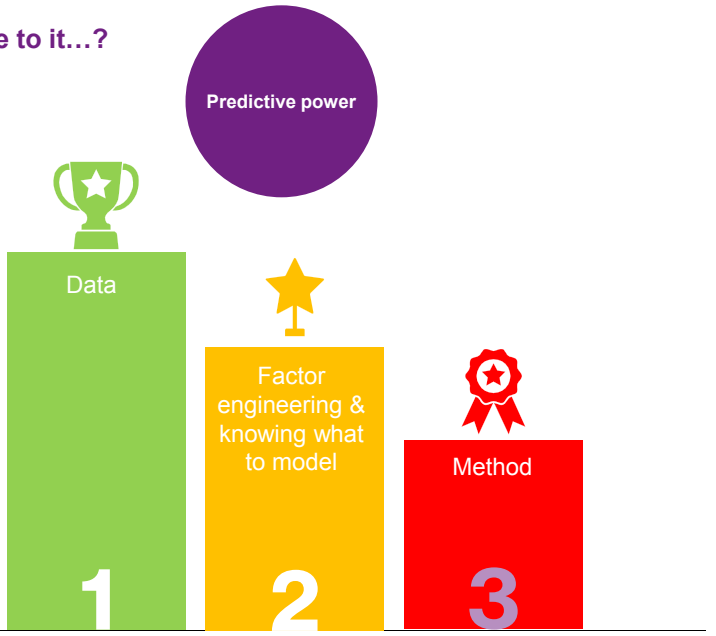
Simple formula

Old/New	New premium	Expected volume	Actual claims	Increased profit
121%	P_1	V_1	C_1	X_1
…	P_2	V_2	C_2	X_2
…	…	…	…	…
…	P_{99}	V_{99}	C_{99}	X_{99}
85%	P_{100}	V_{100}	C_{100}	X_{100}
			Value created	$ X

2.1% on loss ratio…

Willis Towers Watson

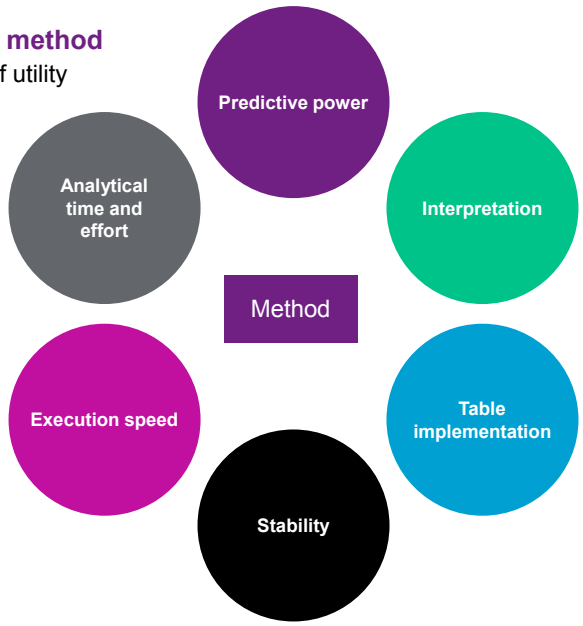
Is there more to it…?



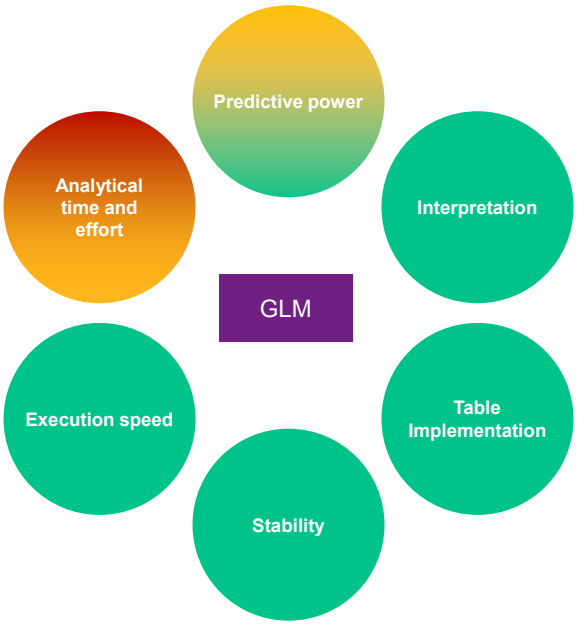
Willis Towers Watson

Choosing a method

Dimensions of utility

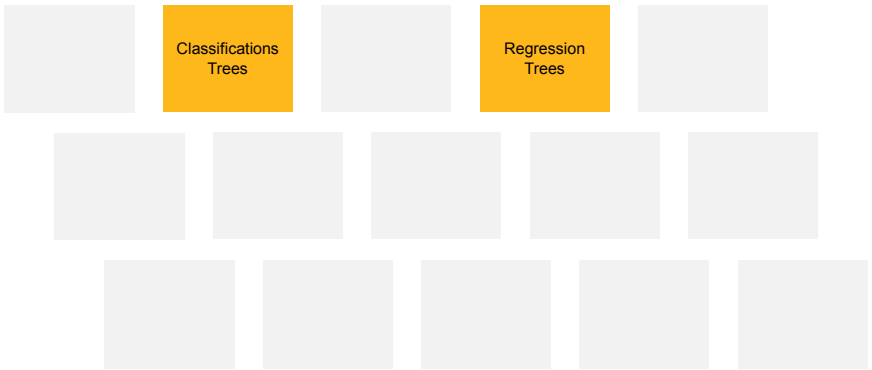


Willis Towers Watson



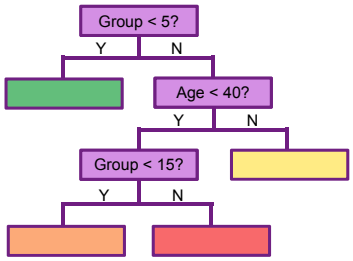
Willis Towers Watson

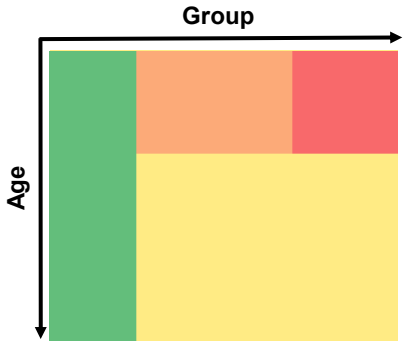
Focus on Trees



Willis Towers Watson

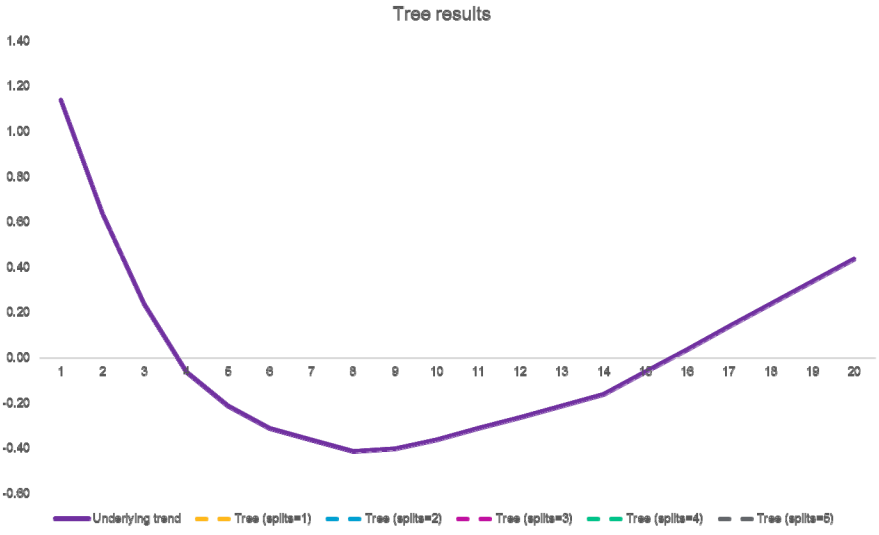
Decision Trees





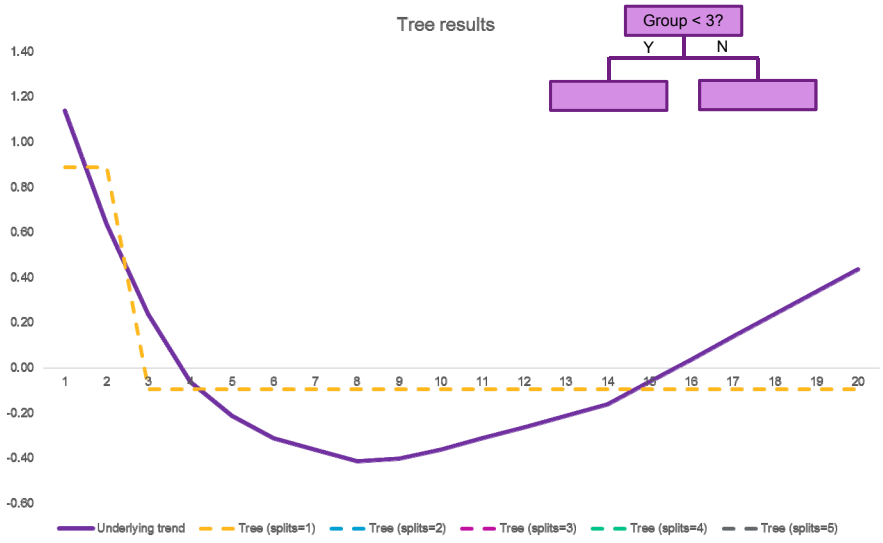
Willis Towers Watson

A simple Tree example



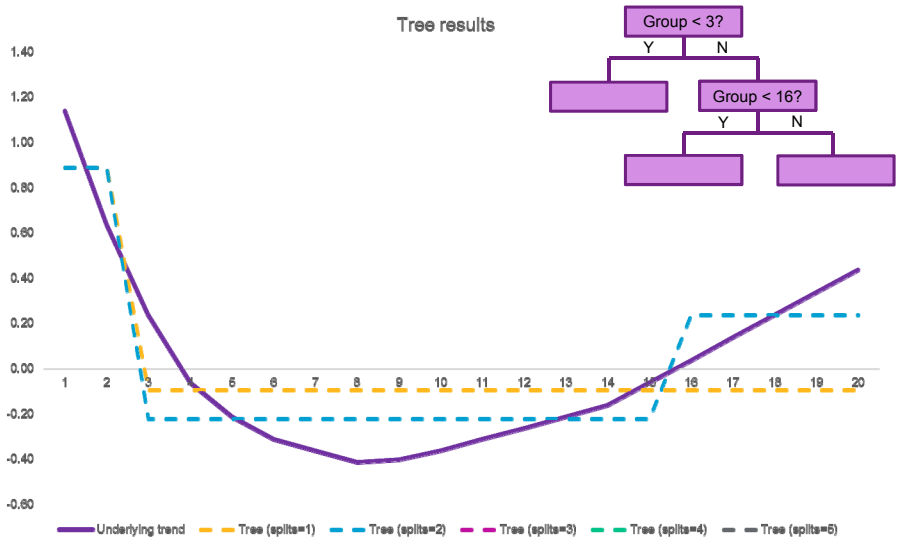
Willis Towers Watson

A simple Tree example



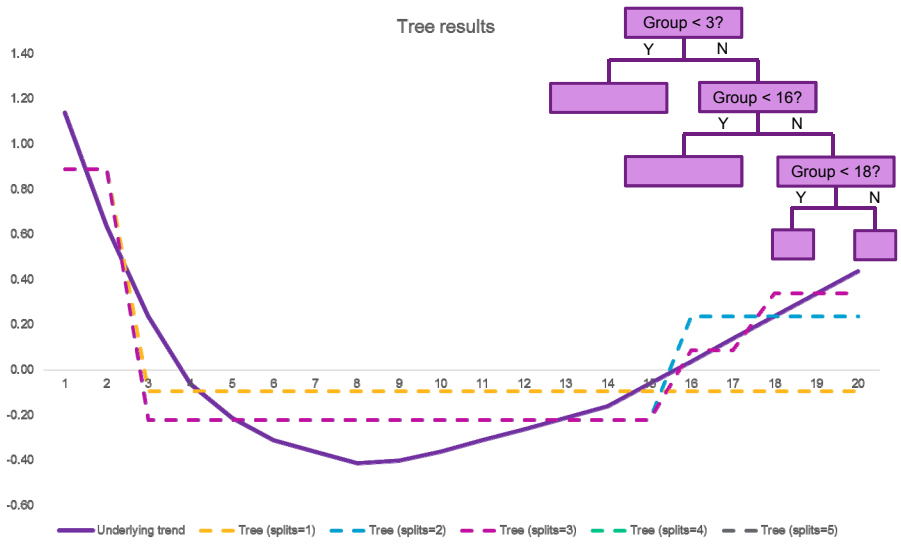
Willis Towers Watson

A simple Tree example



Willis Towers Watson

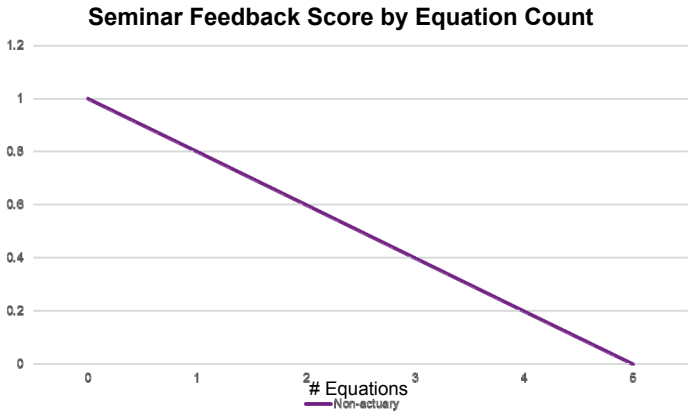
A simple Tree example



Willis Towers Watson

Trees are greedy

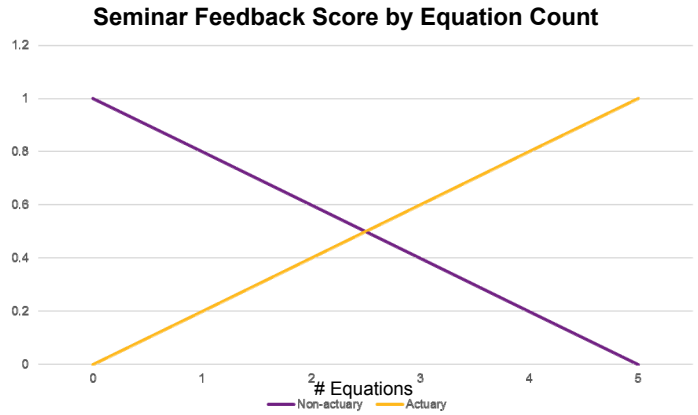
Let's build a simple model:



Willis Towers Watson

Trees are greedy

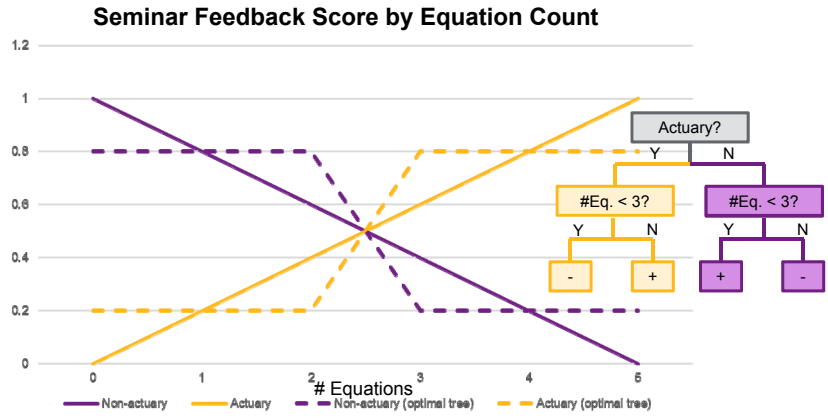
Let's build a simple model:



Willis Towers Watson

Trees are greedy

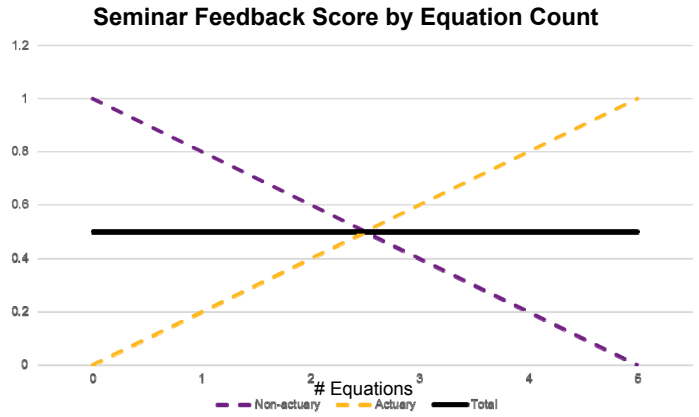
Let's build a simple model:



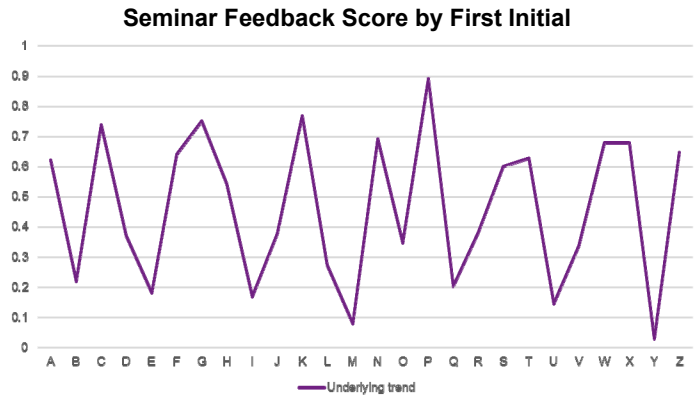
Willis Towers Watson

Trees are greedy

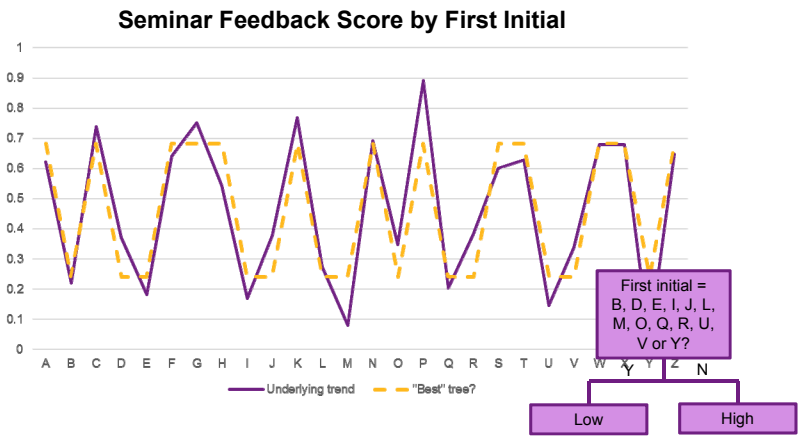
Let's build a simple model:



Trees are bad at categorical variables

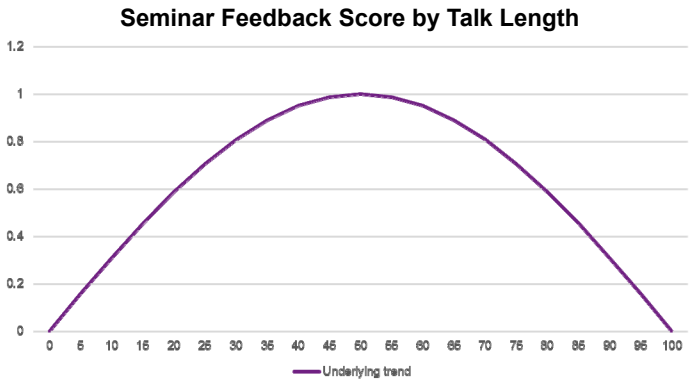


Trees are bad at categorical variables



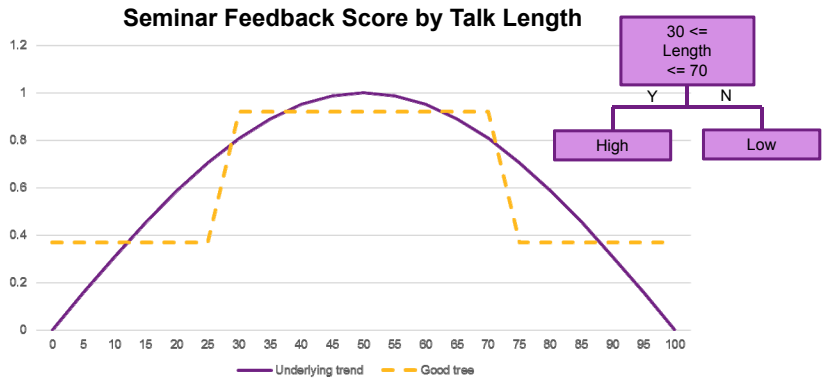
Willis Towers Watson

Trees are bad at turning points



Willis Towers Watson

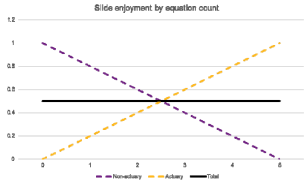
Trees are bad at turning points



Shortcomings of using trees

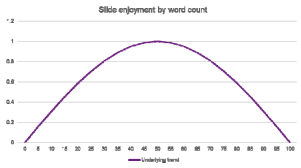
Summary

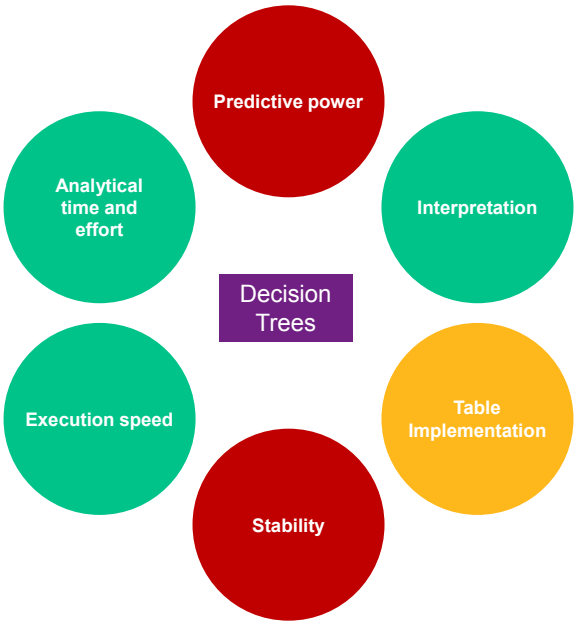
They may miss interactions…



… they may struggles with categorical variables….

…and they can be bad at turning points





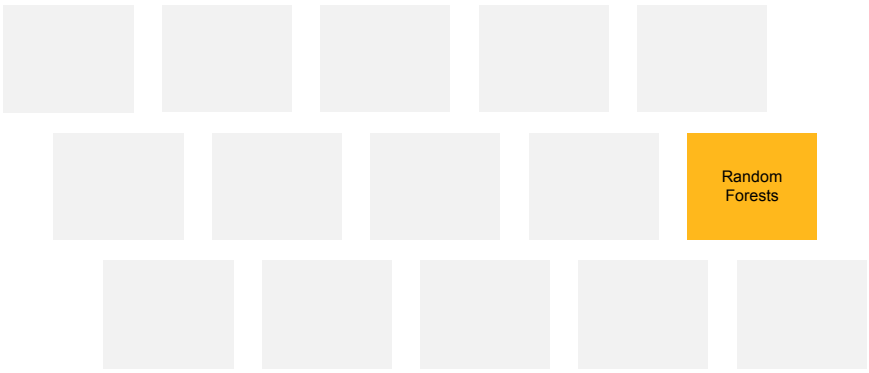
Willis Towers Watson

Some machine learning methods

Ensembles
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

Willis Towers Watson

Focus on Random Forests



Willis Towers Watson

Random Forests

Tree 1: Prediction 1 = Signal 1 + Noise 1
Tree 2: Prediction 2 = Signal 2 + Noise 2
Tree 3: Prediction 3 = Signal 3 + Noise 3
…
Tree 1000: Prediction 1000 = Signal 1000 + Noise 1000

Random Forest:

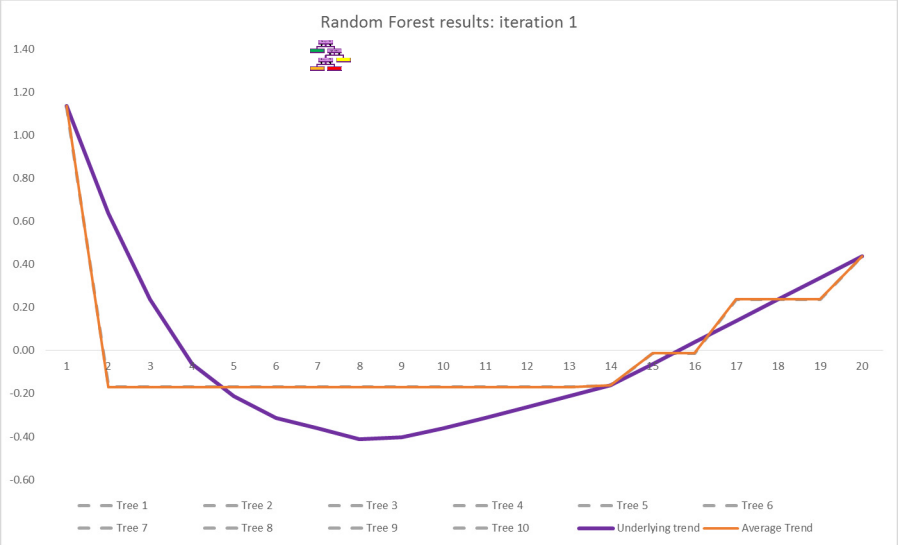
Prediction = AVERAGE(Tree Predictions)

= AVERAGE(Tree Signal) + AVERAGE(Tree Noise)

- Average Noise → 0 if the trees are independent
- Independence of trees achieved by fitting each tree to:
  - Random subset of data (bootstrap sample)
  - Random subset of factors
- Average Signal → Underlying trend, provided trees are complex enough to represent it
- This is bagging (**b**ootstrap **ag**gregation) – fit **lots** of independent models and take an average

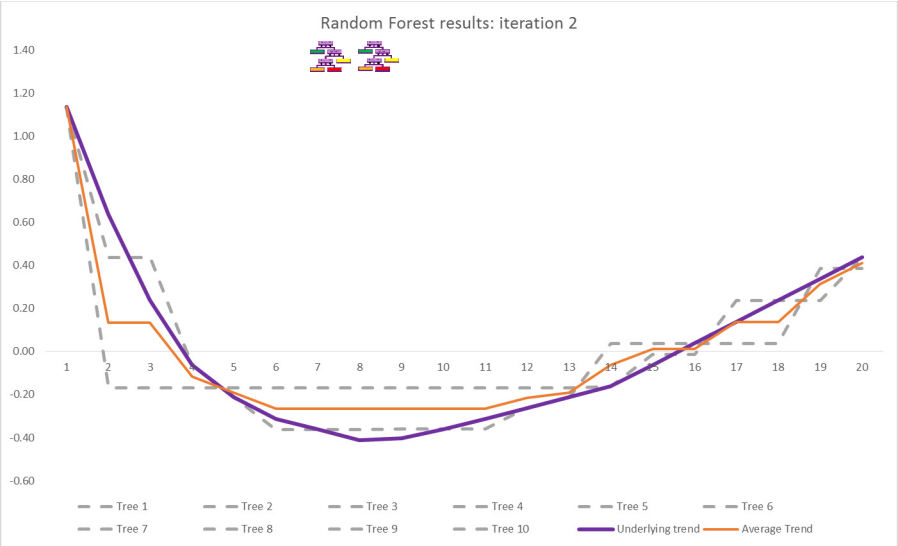
Willis Towers Watson

A simple Random Forest example



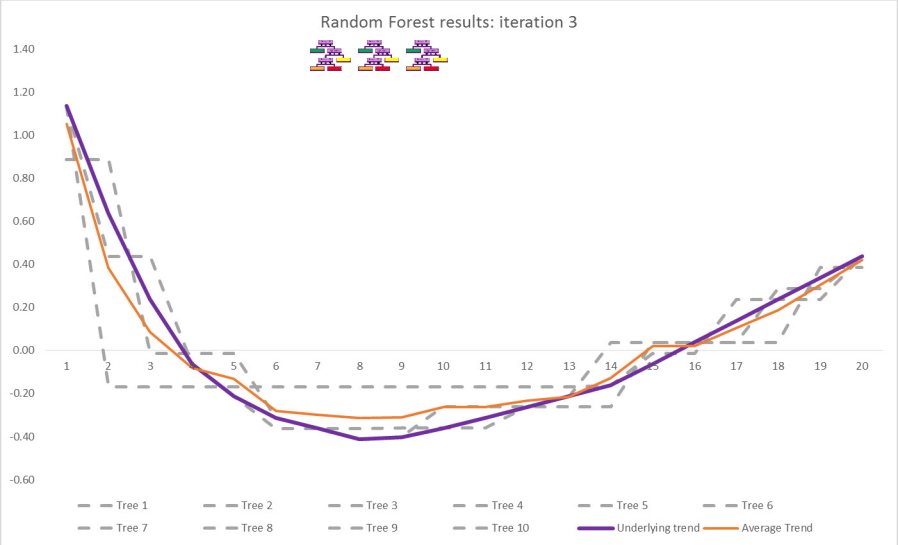
Willis Towers Watson

A simple Random Forest example



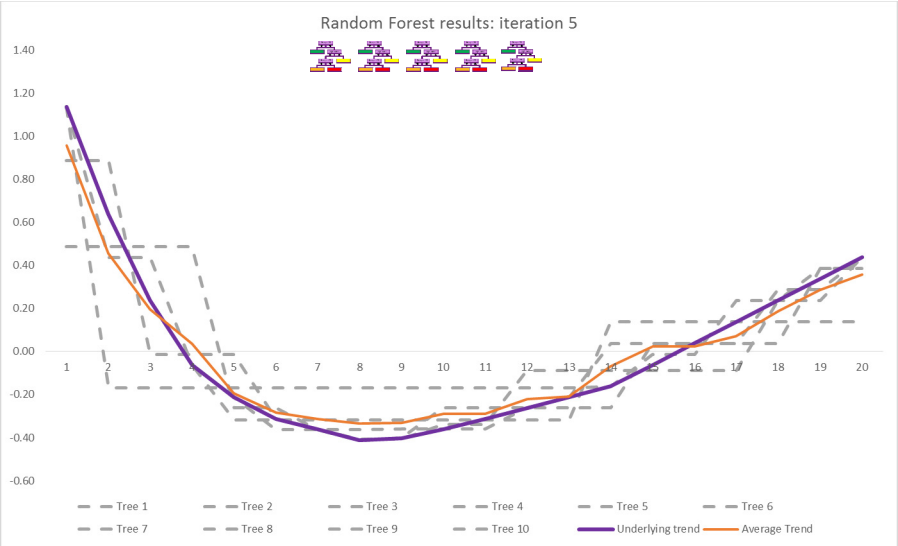
Willis Towers Watson

A simple Random Forest example



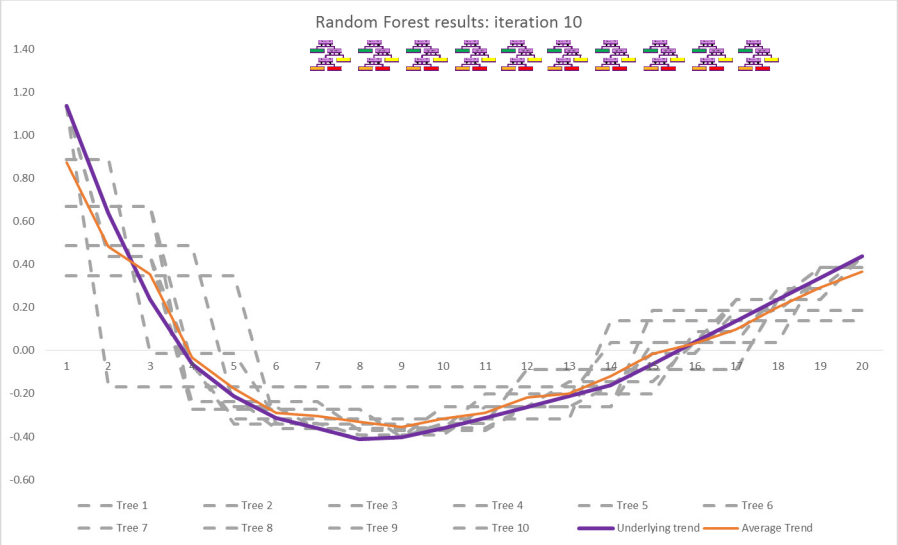
Willis Towers Watson

A simple Random Forest example

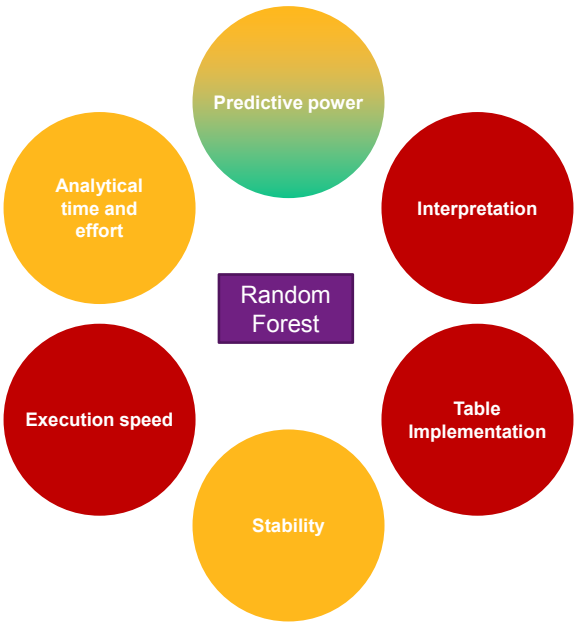


Willis Towers Watson

A simple Random Forest example

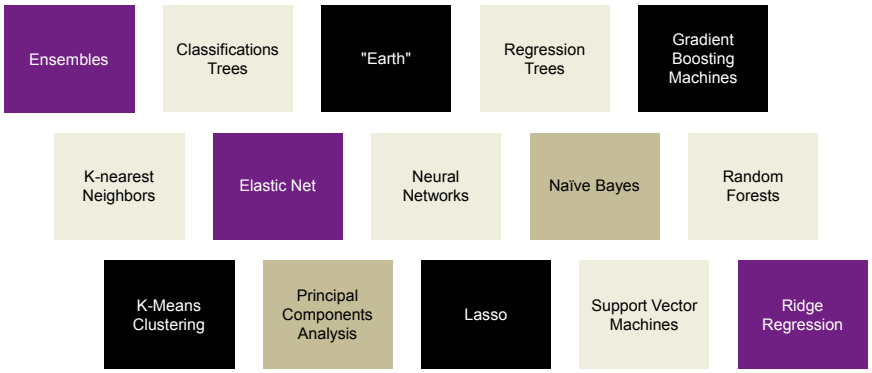


Willis Towers Watson



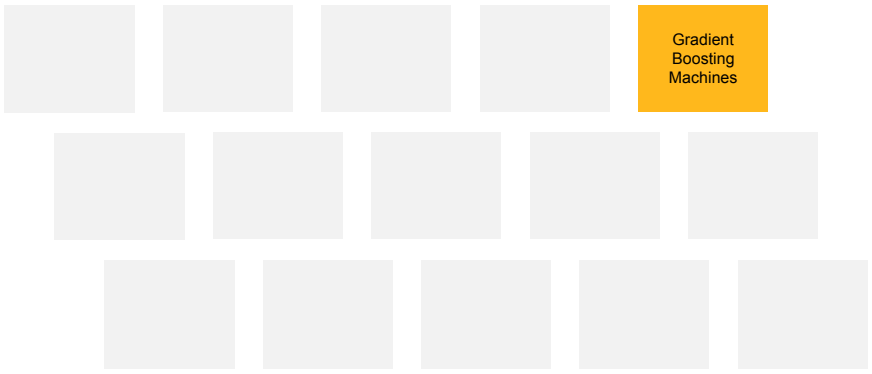
Willis Towers Watson

Some machine learning methods



Willis Towers Watson

Focus on Gradient Boosting Machines

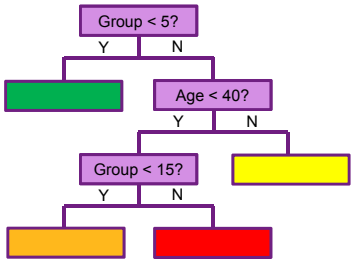


Willis Towers Watson

Gradient Boosting Machine or "GBM"

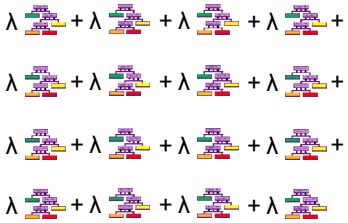
A tree

$f_i(x)$



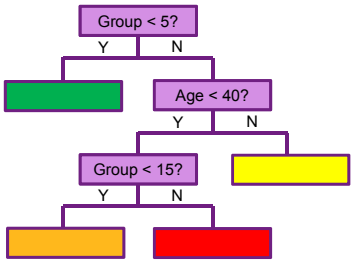
A GBM

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

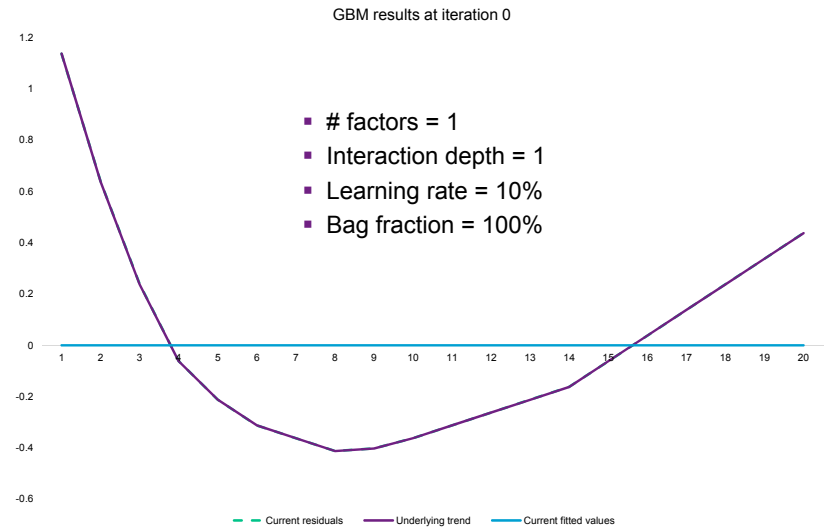


Four main assumptions

- 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

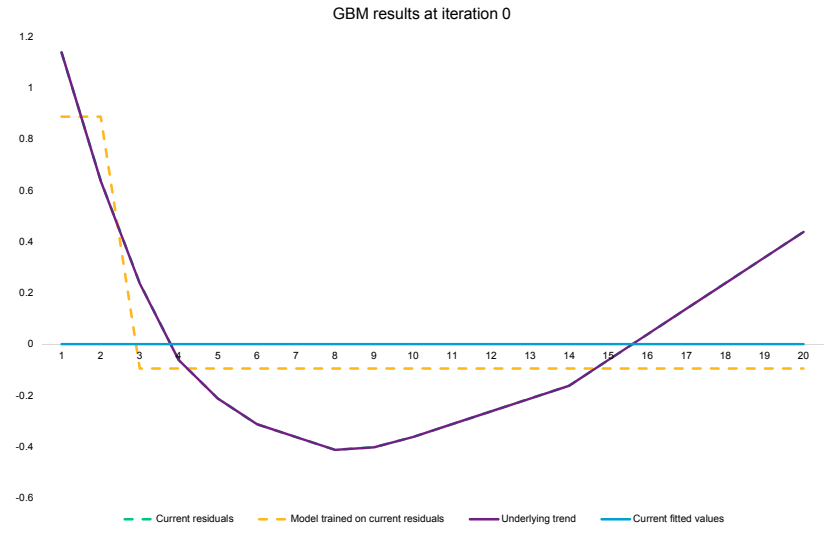


A simple GBM example



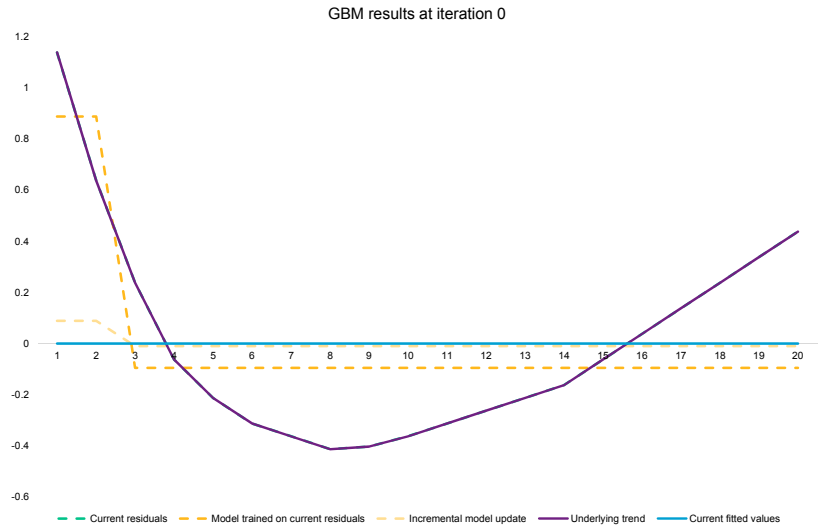
Willis Towers Watson

A simple GBM example



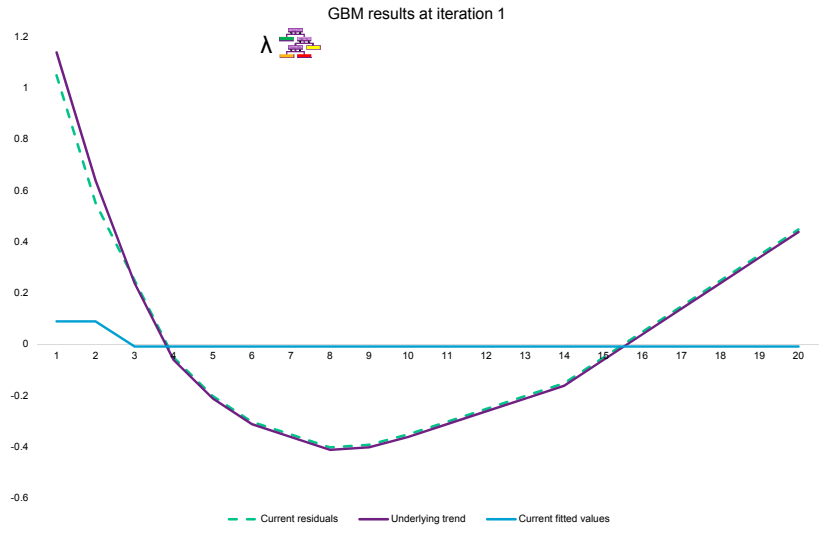
Willis Towers Watson

A simple GBM example



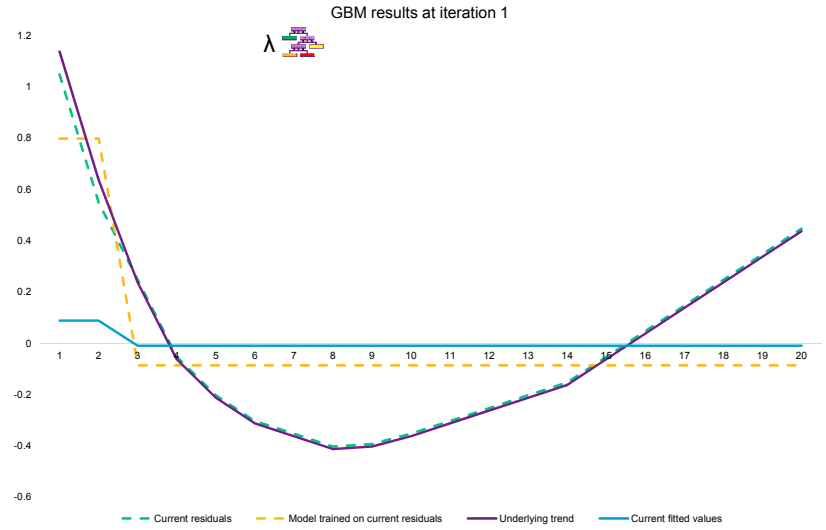
Willis Towers Watson

A simple GBM example



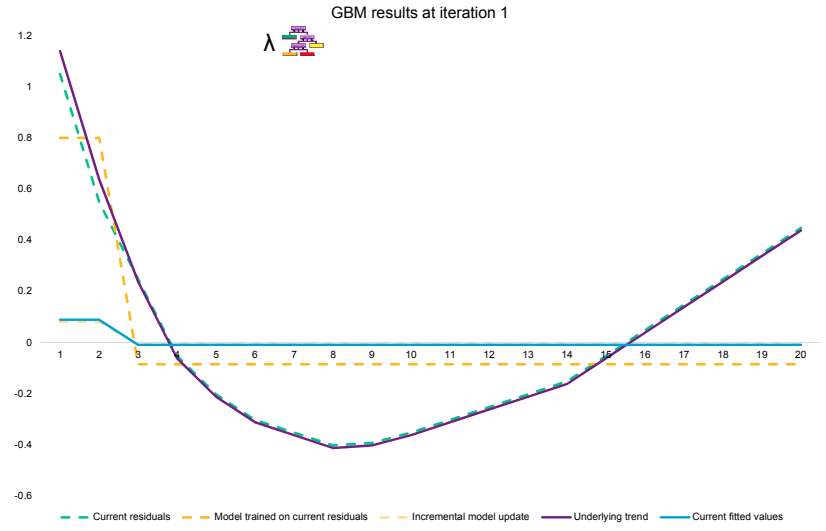
Willis Towers Watson

A simple GBM example



Willis Towers Watson

A simple GBM example

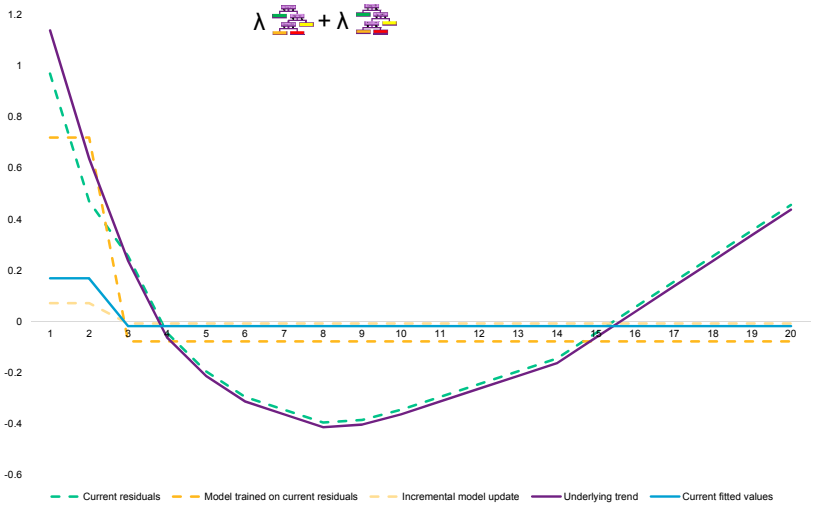


Willis Towers Watson

A simple GBM example

GBM results at iteration 2

λ + λ

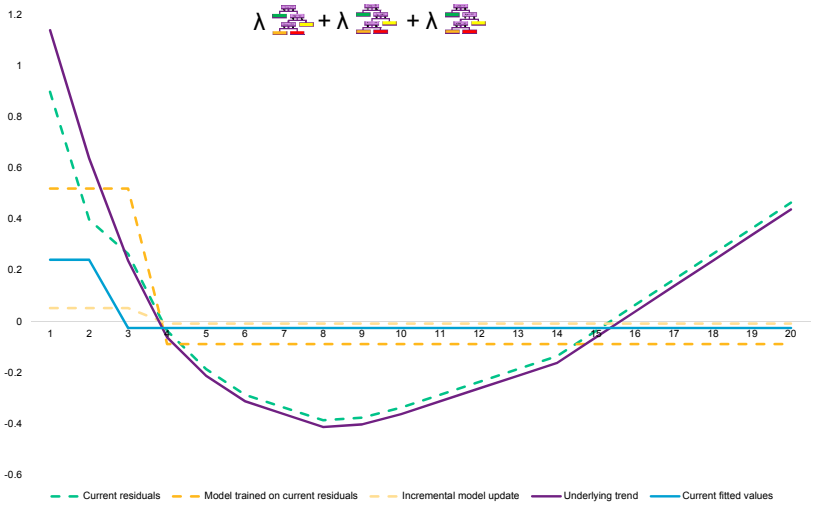


Willis Towers Watson

A simple GBM example

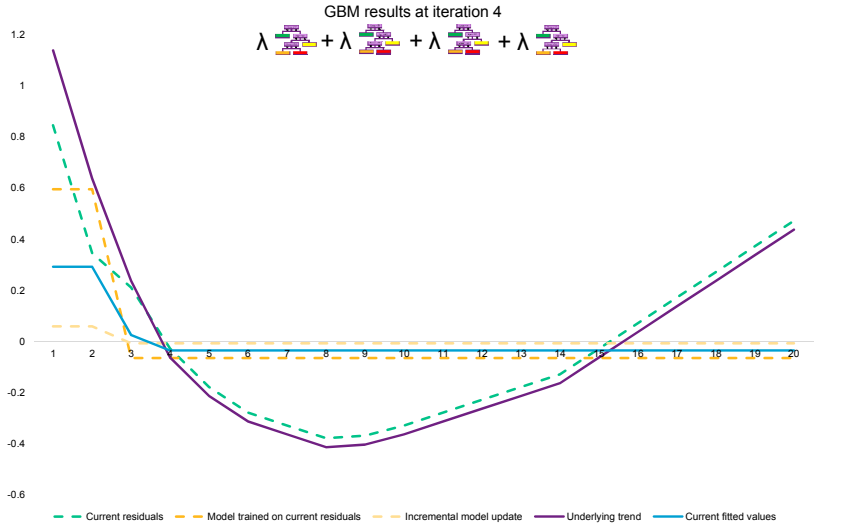
GBM results at iteration 3

λ + λ + λ



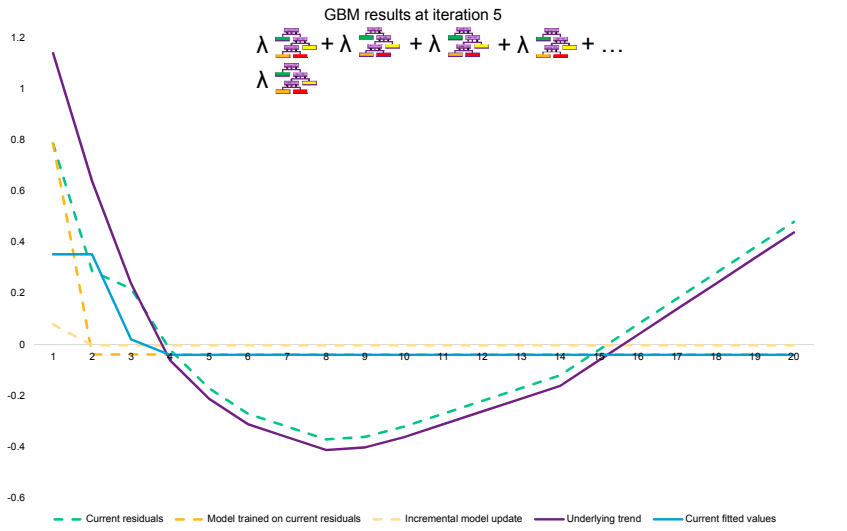
Willis Towers Watson

A simple GBM example



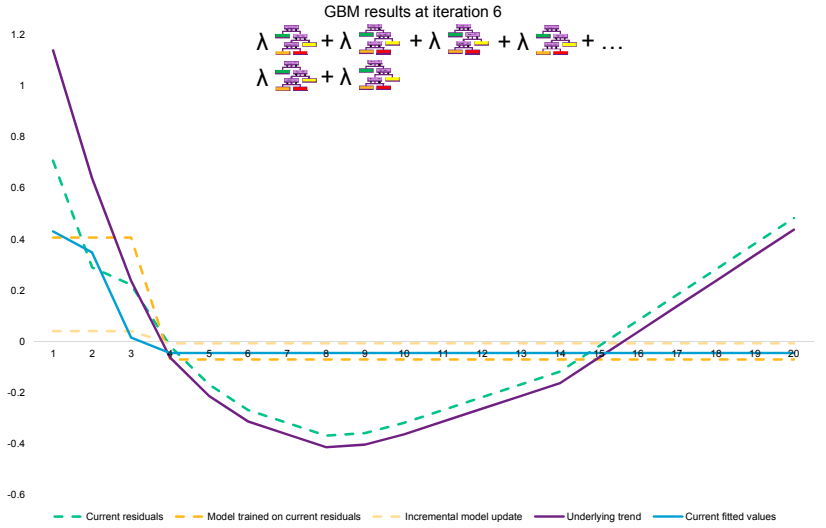
Willis Towers Watson

A simple GBM example



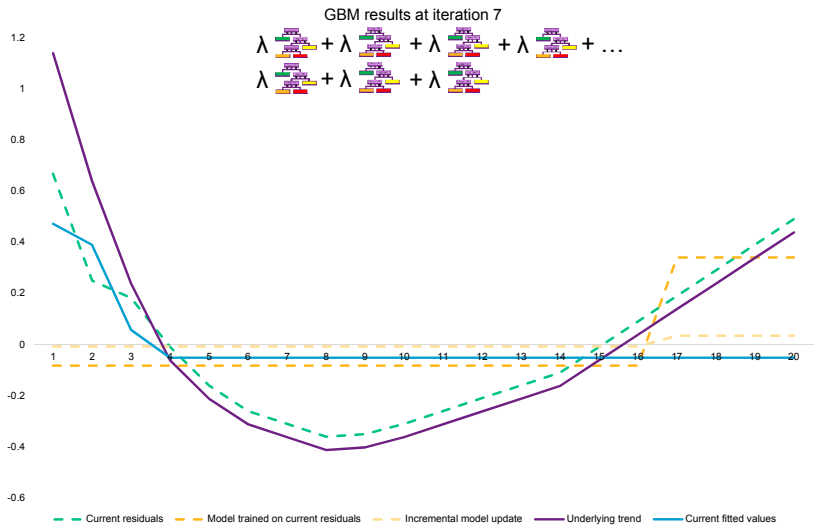
Willis Towers Watson

A simple GBM example



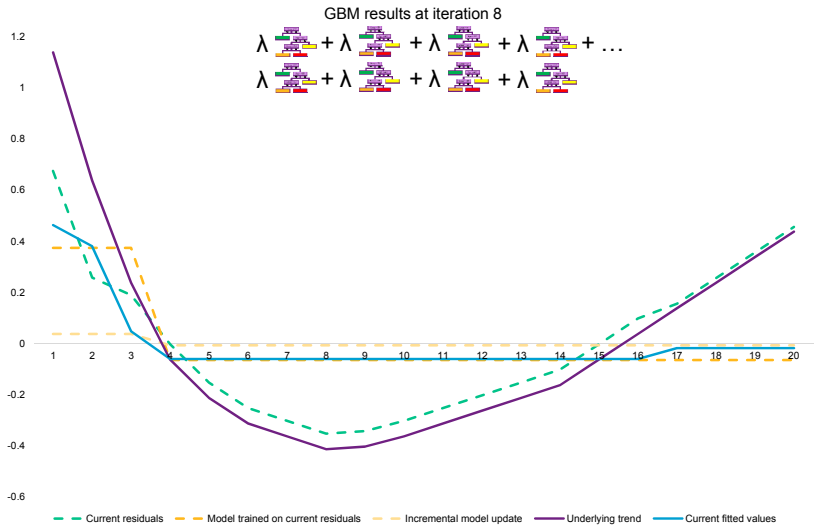
Willis Towers Watson

A simple GBM example



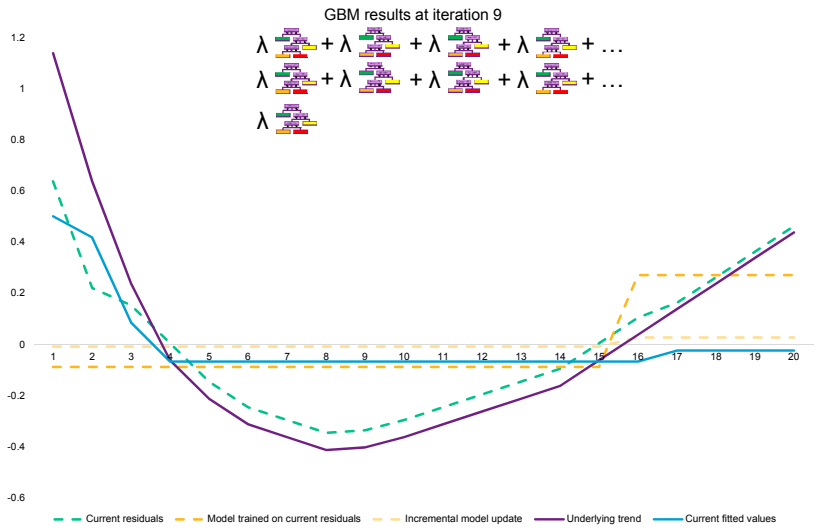
Willis Towers Watson

A simple GBM example



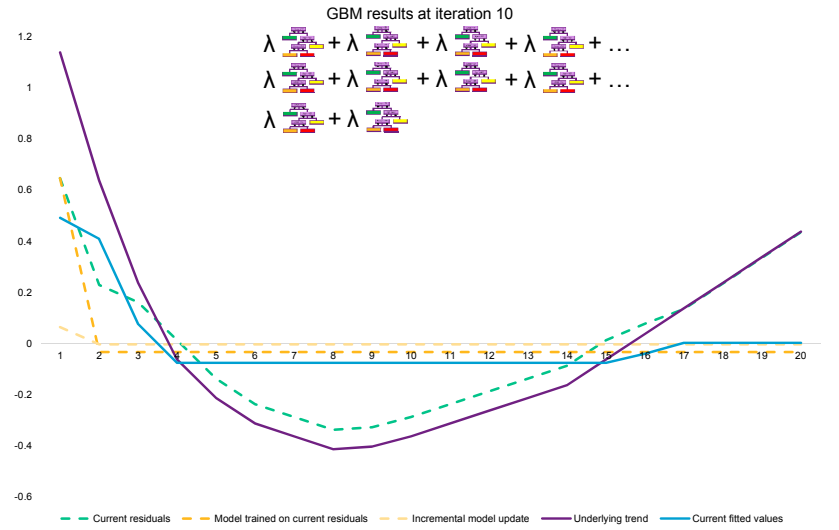
Willis Towers Watson

A simple GBM example



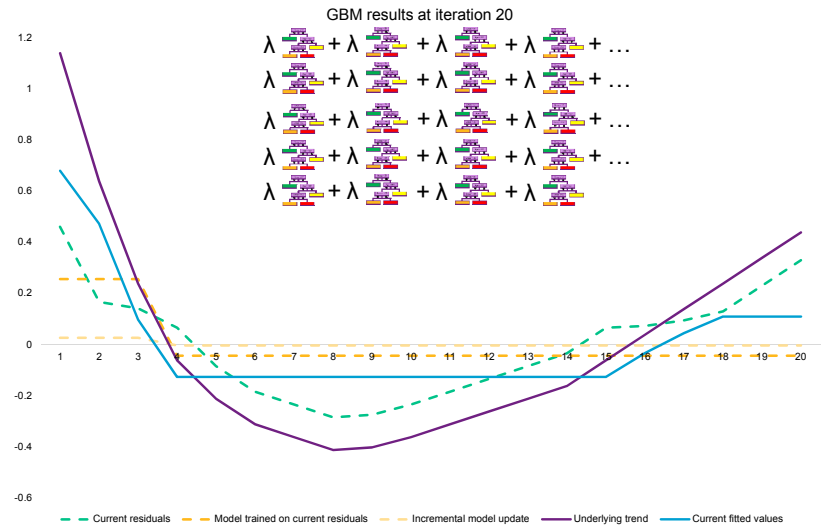
Willis Towers Watson

A simple GBM example



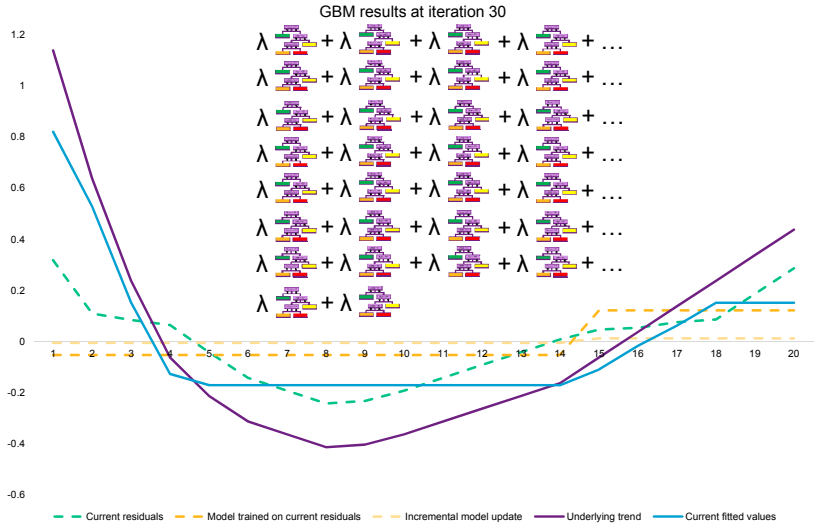
Willis Towers Watson

A simple GBM example



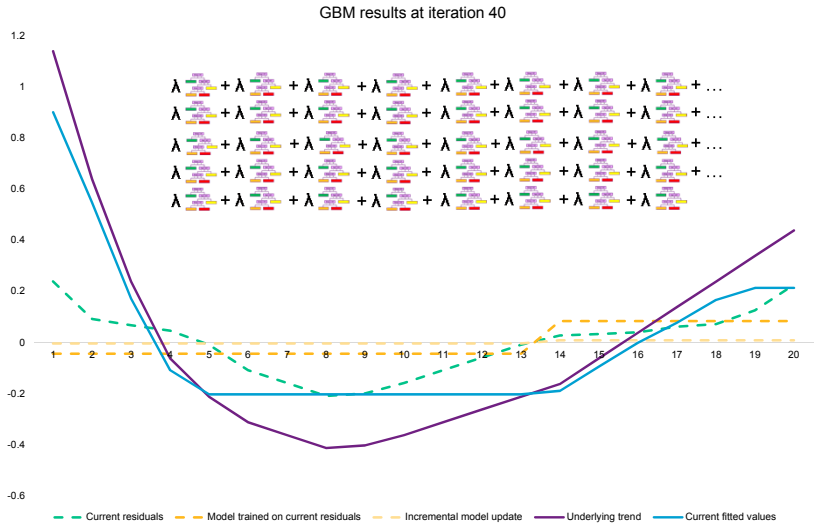
Willis Towers Watson

A simple GBM example



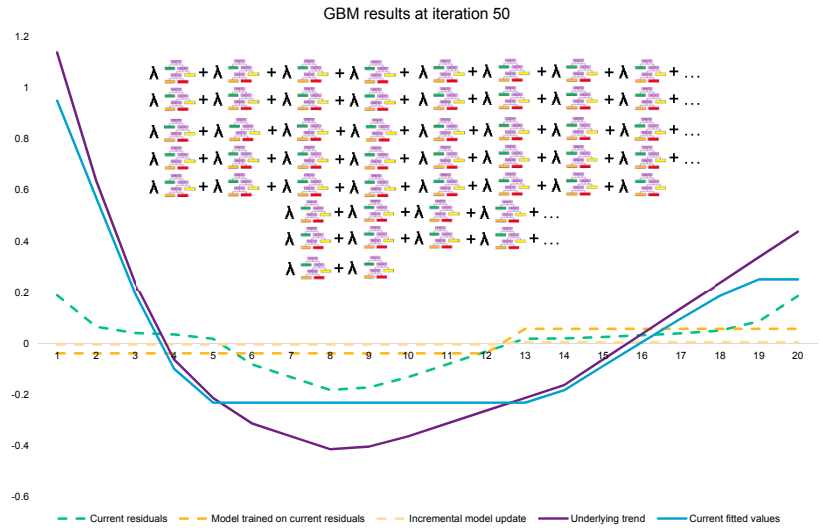
Willis Towers Watson

A simple GBM example



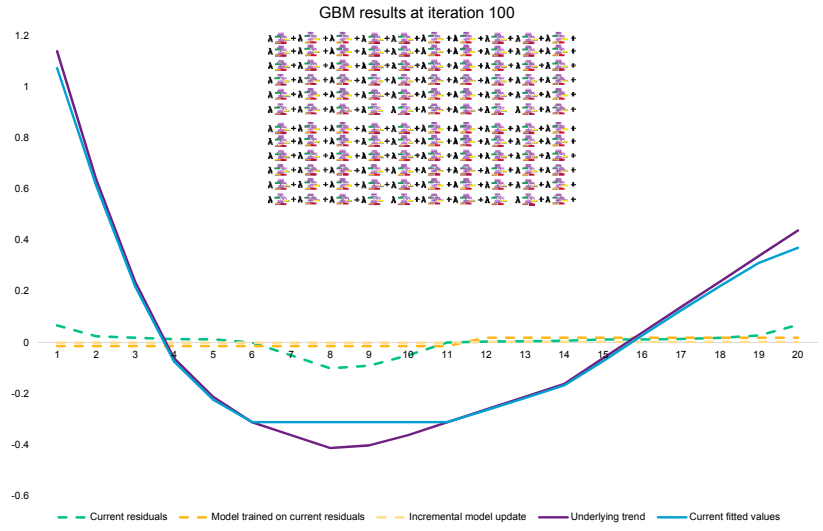
Willis Towers Watson

A simple GBM example



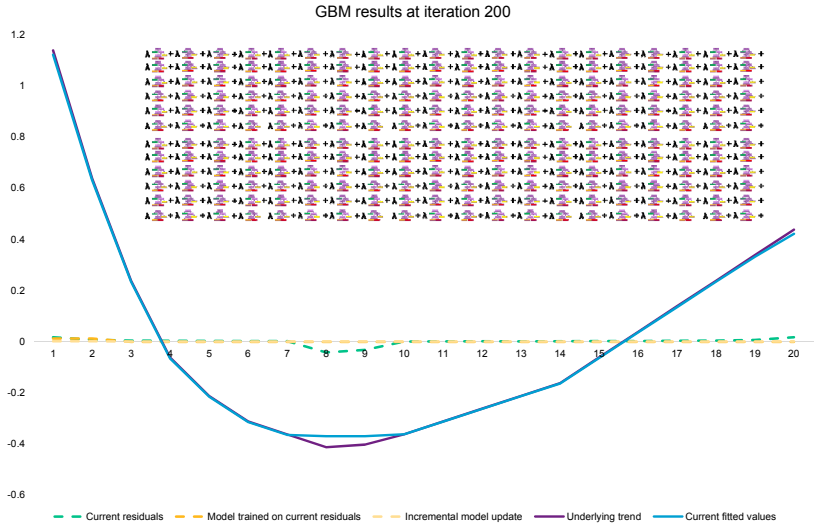
Willis Towers Watson

A simple GBM example



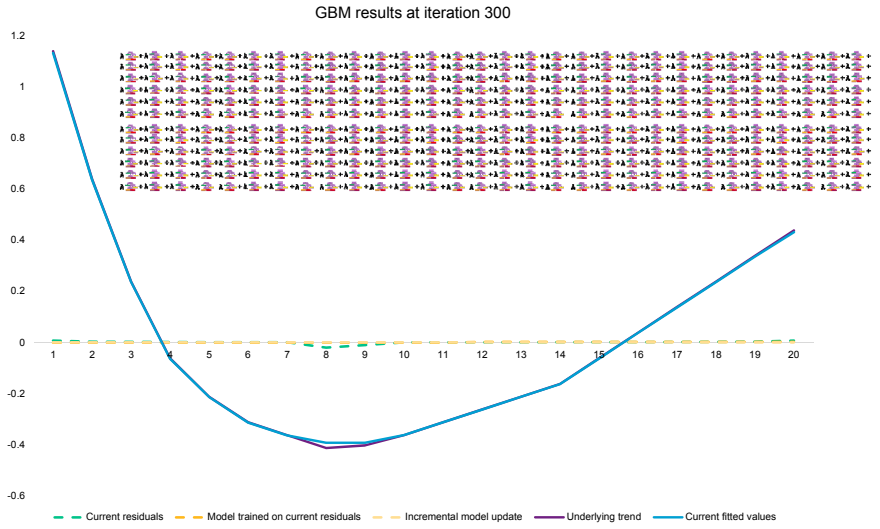
Willis Towers Watson

A simple GBM example



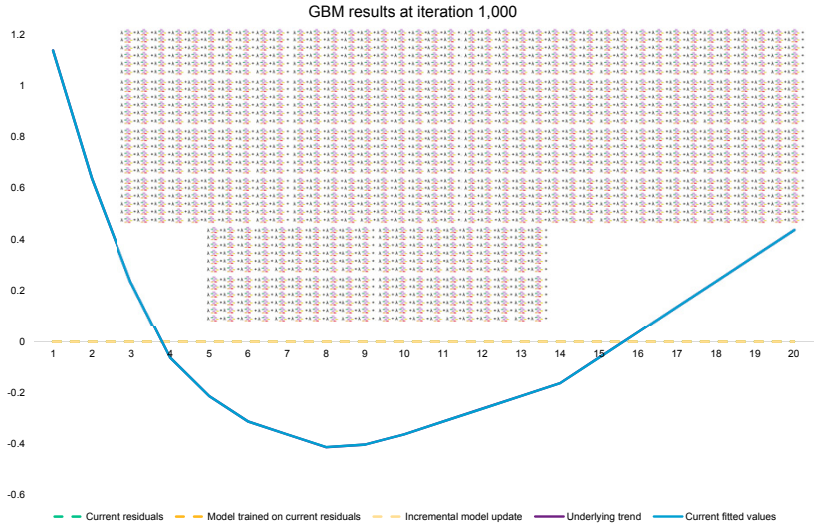
Willis Towers Watson

A simple GBM example



Willis Towers Watson

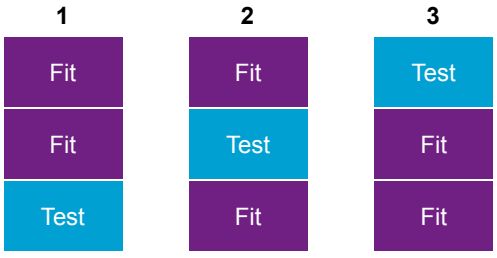
A simple GBM example



Willis Towers Watson

Calibrating the assumptions

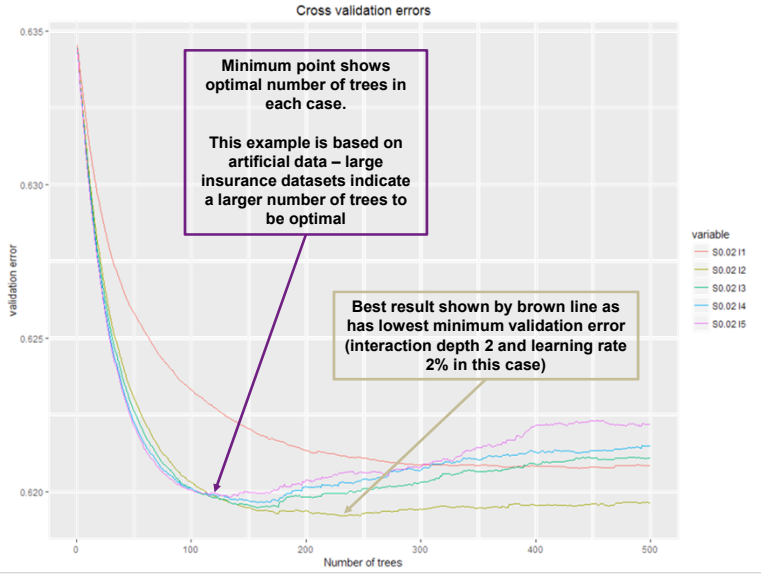
- *n*-fold cross validation used to develop the interaction depth and learning rate assumptions
  - Eg for 3-fold validation, split into 3, fit on purple, test on blue parts, take average



- Resulting plots can be used to determine the optimal assumption choice
  - Including how many trees to run

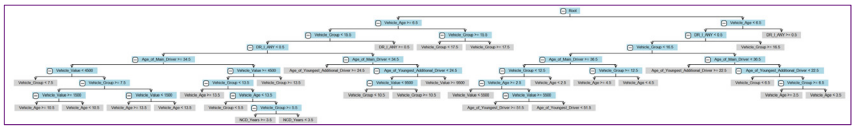
Willis Towers Watson

Example 5-fold cross validation

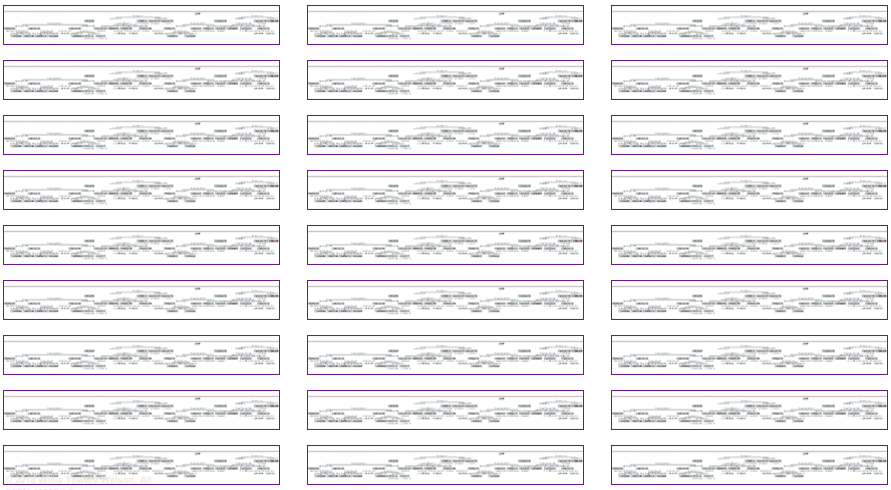


Willis Towers Watson

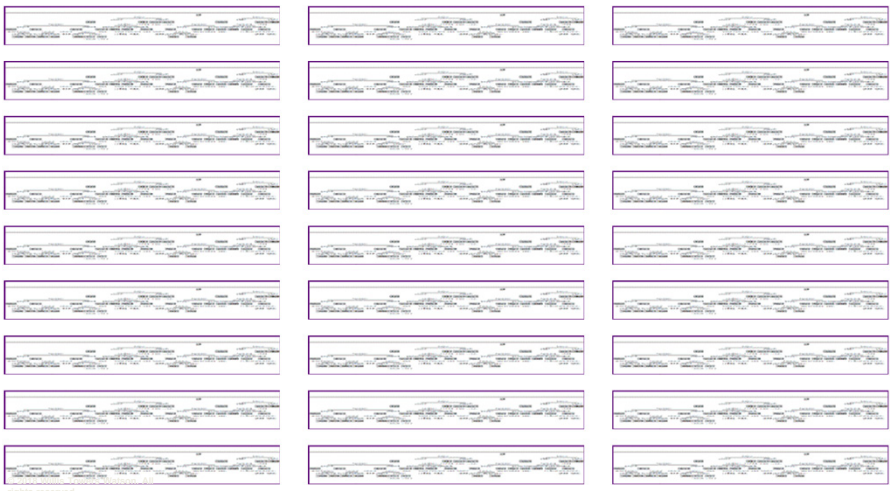
What does a GBM look like?

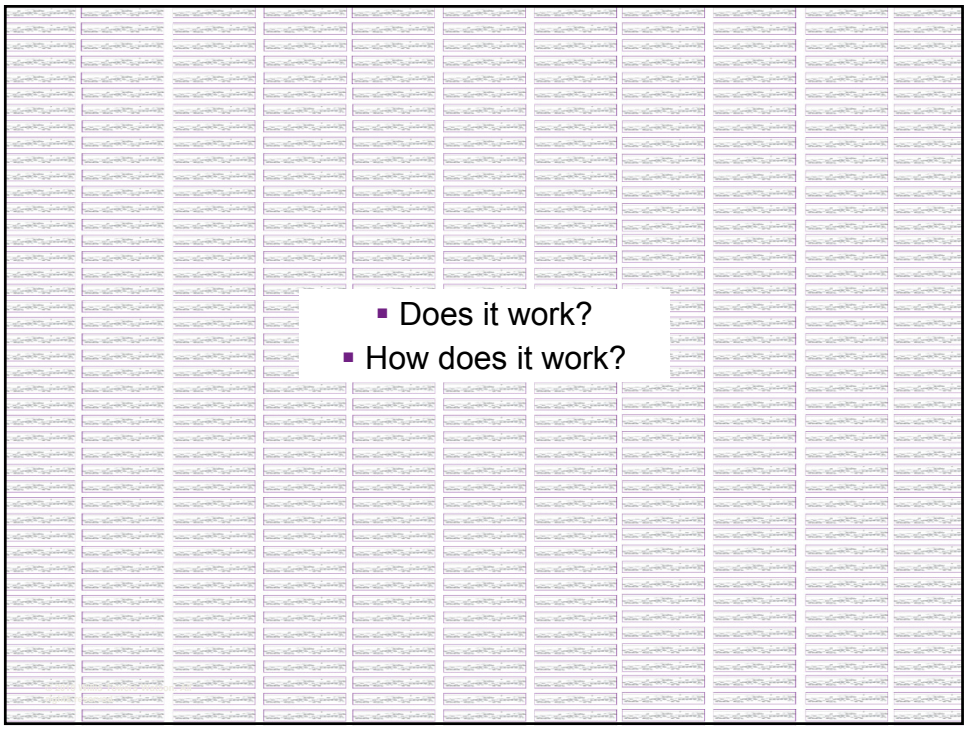


What does a GBM look like?



What does a GBM look like?




Does it work?
How does it work?

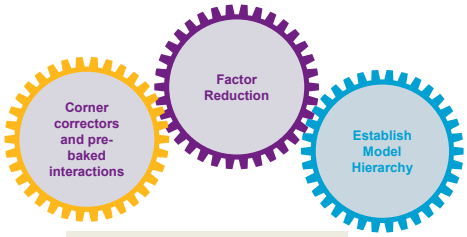
Deploying GBMs

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

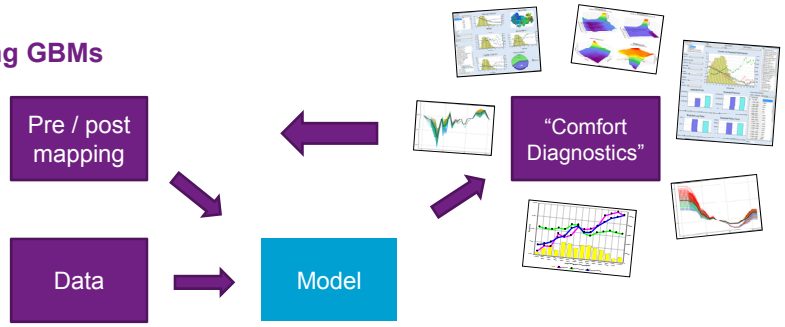


Use insights to guide GLM

Deploy directly

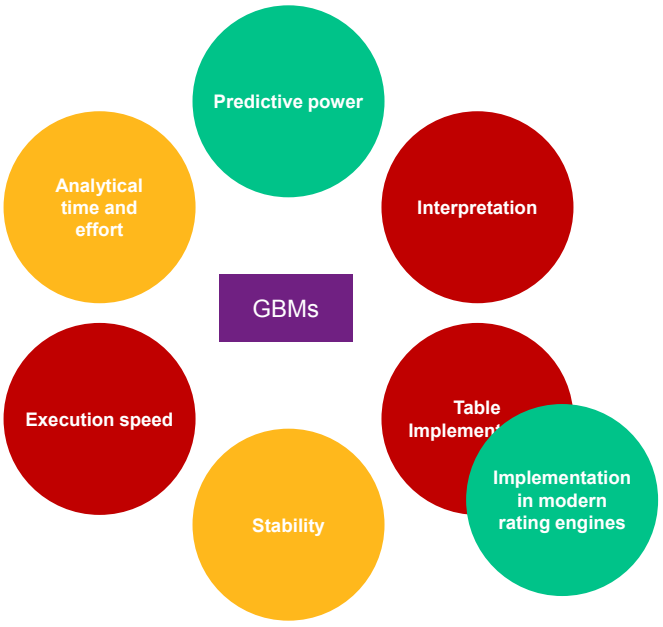
Willis Towers Watson

Deploying GBMs



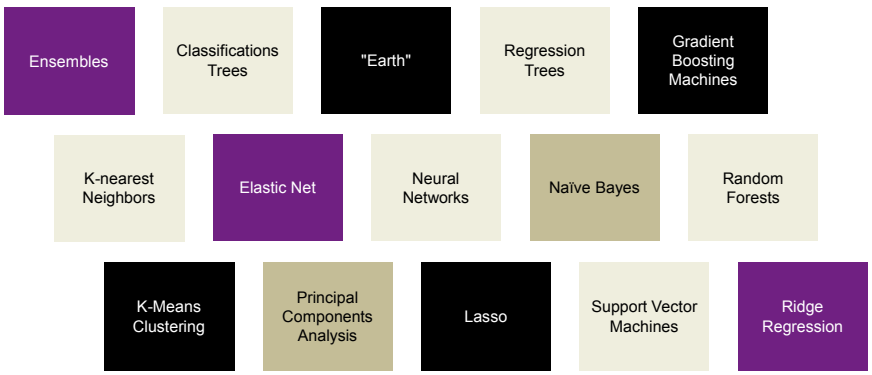
Deploy directly

Willis Towers Watson



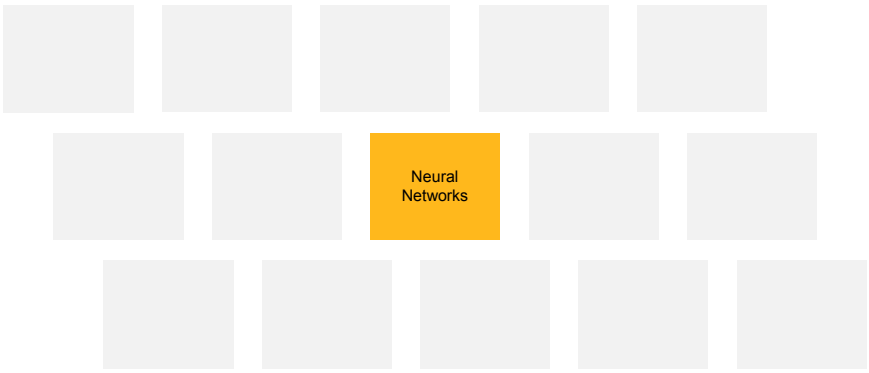
Willis Towers Watson

Some machine learning methods



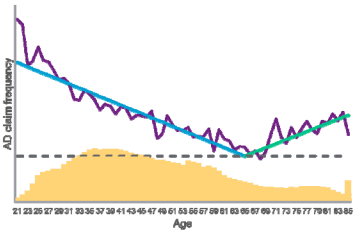
Willis Towers Watson

Focus on Neural Networks



Willis Towers Watson

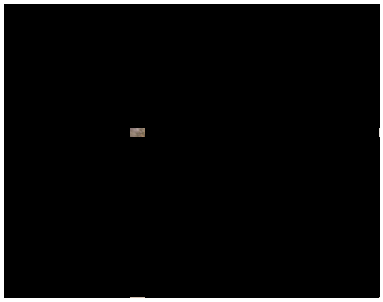
Where is the value?



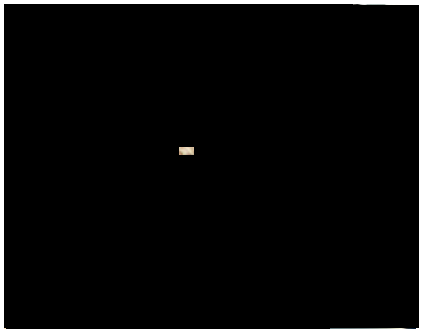
Which policyholder is more likely to make a claim?

Willis Towers Watson

Where is the value?



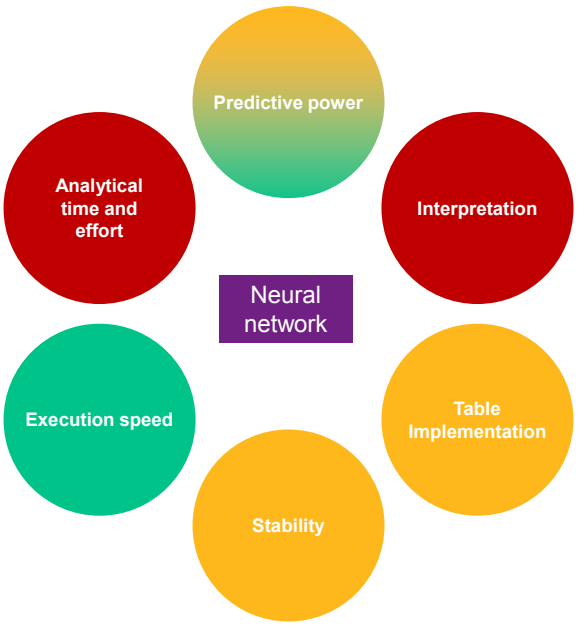
Which picture is more likely to be of a cat?



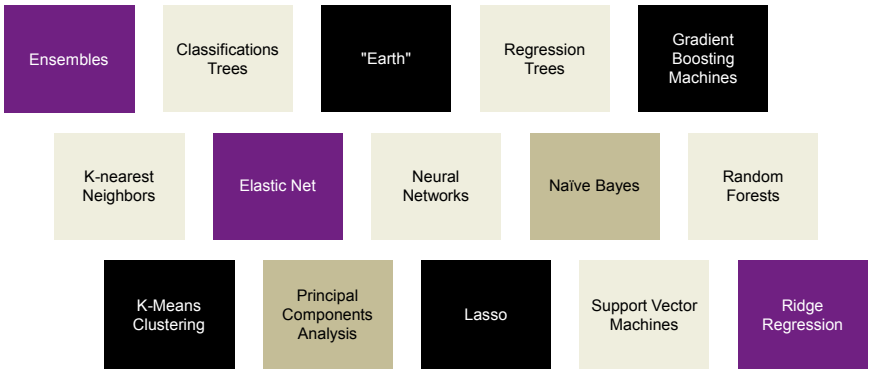
Where is the value?

Which picture is more likely to be of a cat?

Neural networks

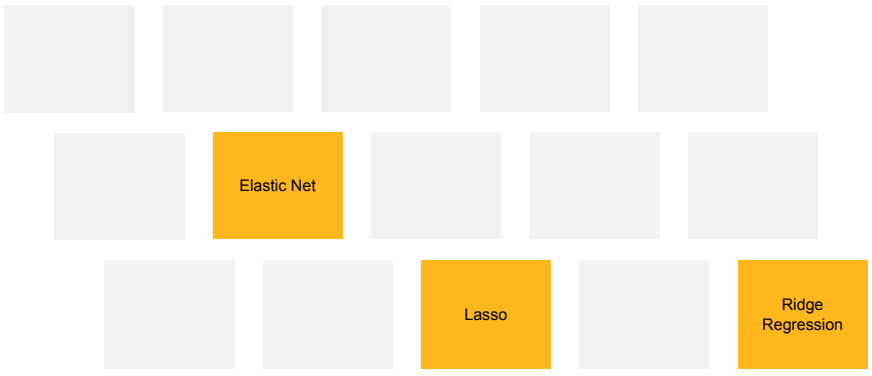


Some machine learning methods



Willis Towers Watson

Focus on Penalized Regression

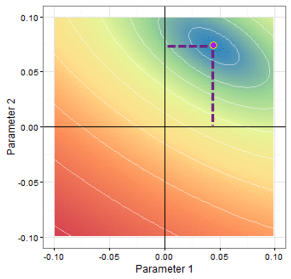


Willis Towers Watson

Penalized Regression

GLM

f(x) = g^{-1}(**X**.β) where β estimated by minimizing $L(\beta|X,y)$



Willis Towers Watson

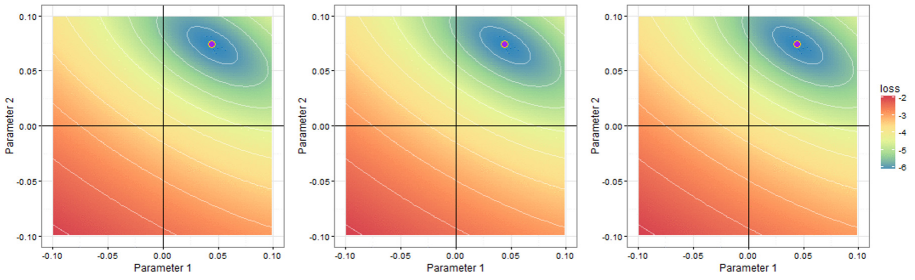
Penalized Regression

GLM **Lasso** **Ridge**

f(x) = g^{-1}(**X**.β) where β estimated by minimizing $L(\beta|X,y) + \lambda_1 \sum_i |\beta_i| + \lambda_2 \sum_i \beta_i^2$

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

Willis Towers Watson

Penalized Regression

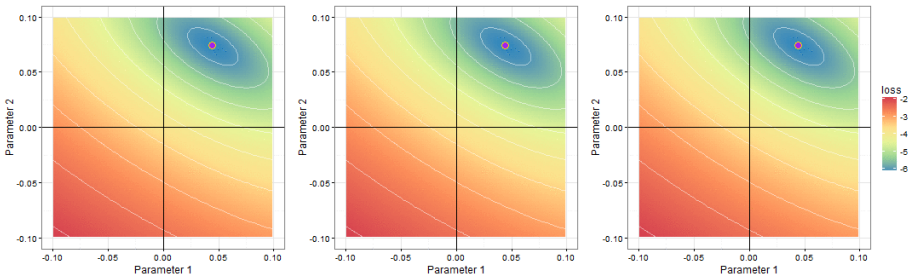
f(x) = g^{-1}(**X**.β) where β estimated by minimizing

GLM Lasso Ridge

$$L(\beta|X,y) + \lambda_1 \sum_i |\beta_i| + \lambda_2 \sum_i \beta_i^2$$

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

Willis Towers Watson

Deploying Penalized Regression

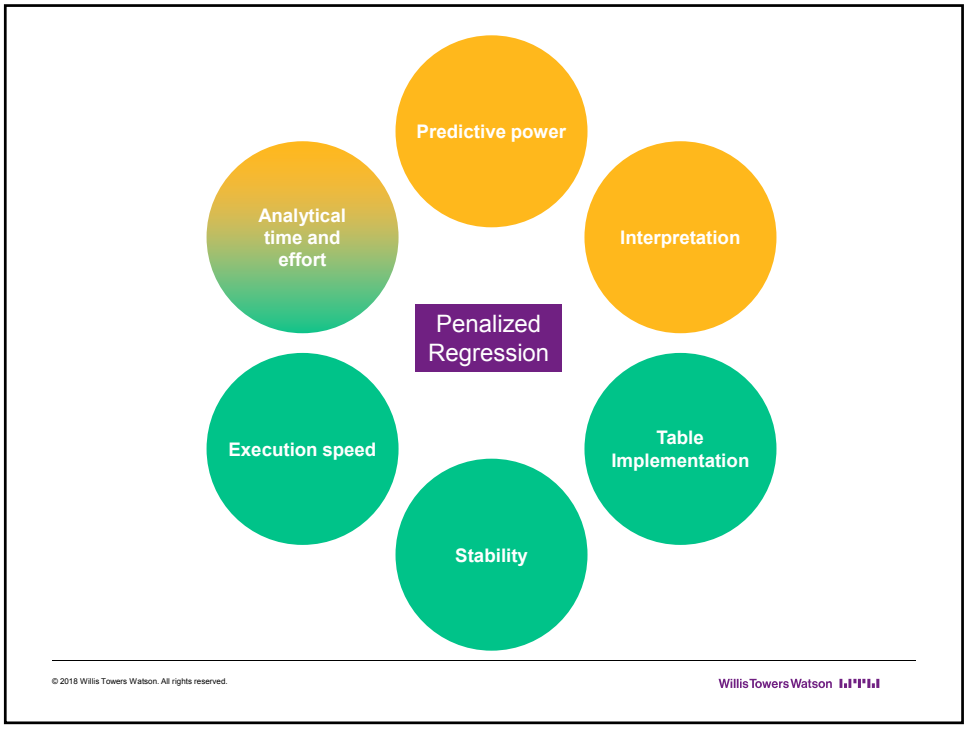
Same as GLMs!

	Age	Exposure	Loss 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

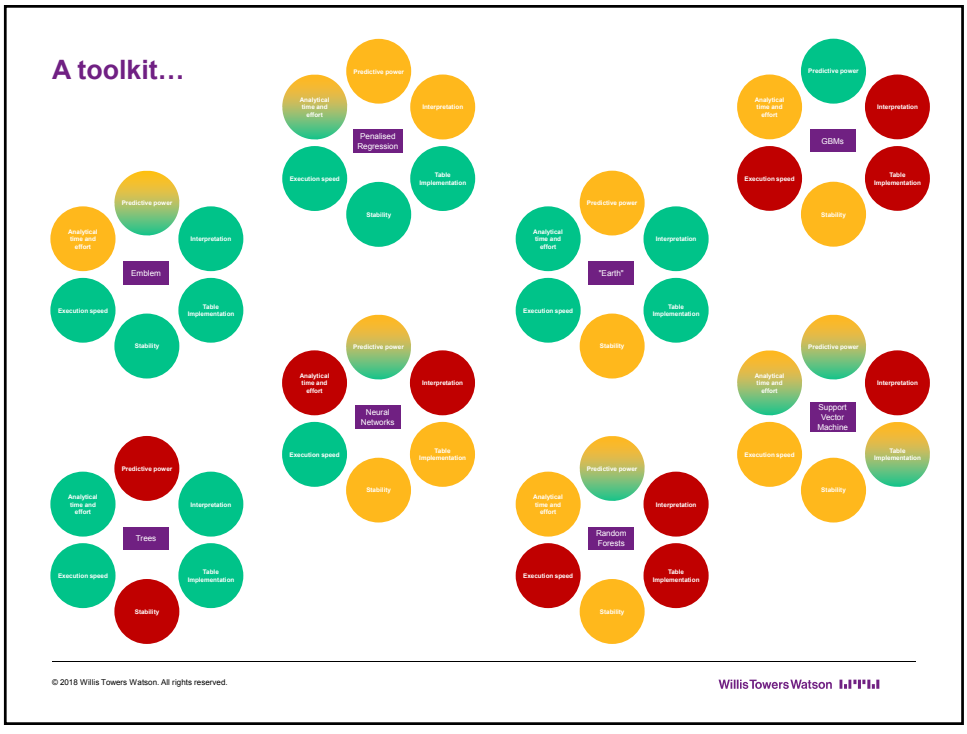
	Vehicle Group	Exposure	Loss Cost
1	1-10	164,107	77
2	11-14	84,859	101
3	15-18	28,952	116
4	19-20	3,931	272
5	VG Total	281,849	91

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

Willis Towers Watson

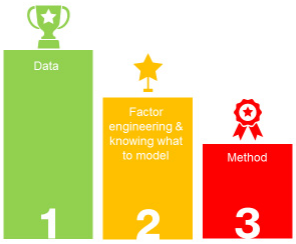


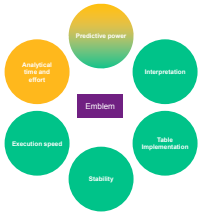
Willis Towers Watson



Willis Towers Watson

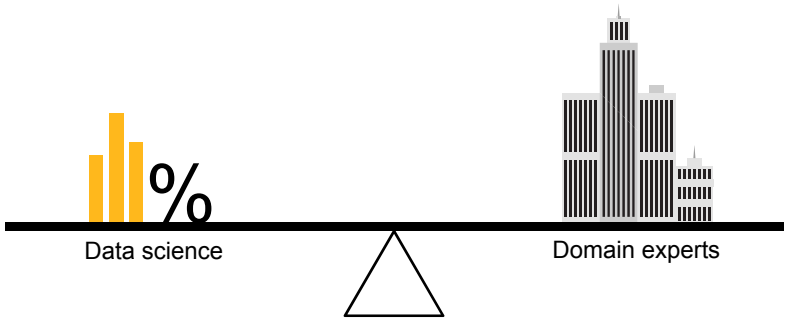
Conclusions





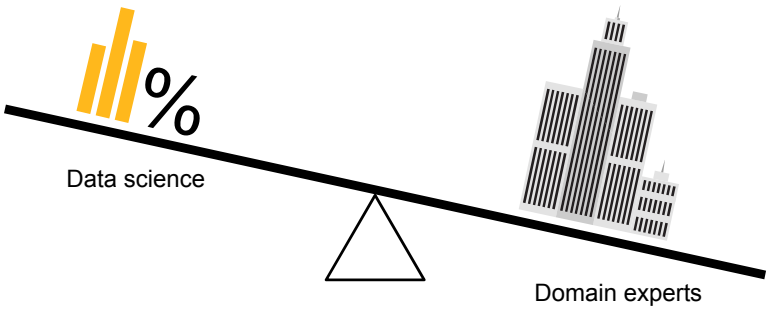
Willis Towers Watson

What do you use where?



Willis Towers Watson

It's domain expertise that helps decide



Willis Towers Watson

Issues for the Profession(s)

Role of the actuary

- Domain expertise matters (at least currently)
- Easier for an actuary to pick up machine learning than for a data scientist to understand insurance?
- Siloed teams don't work
- Familiarity and the right vernacular can help
- Scope of involvement?
  Pricing ✓ Reserving ✓ Claims analytics ✓
  Customer management ? Marketing ???

Training

- CAS, SOA ahead? (eg CSPA)
- IFoA playing catch up
- Is it too late?

Regulatory issues

- TAS: Judgement - what judgement?
- GDPR
- UK Government Select Committee (Science and Technology)

Willis Towers Watson

Questions and Discussion

Willis Towers Watson