Institute
and Faculty
of Actuaries

Machine Learning Modelling on Triangles

April Lu, John McCarthy

15 November 2021

Agenda

This presentation builds on previous work presented at the 2021 IFoA Spring Conference* and is aimed at those relatively new to machine learning

- Reminder of machine learning framework for modelling triangle data
- Data
- Results
- Diagnostic charts
- Next steps
- Q&A

*https://institute-and-faculty-of-actuaries.github.io/mlr-blog/post/f-mlr3example/

Machine Learning in Reserving Working Party

- Who are we?
  - International group of actuaries, data scientists and academics from diverse backgrounds, chaired by Sarah MacDonnell

- What are our aims?
  - Learn how machine learning (ML) can be used in non-life reserving
  - Carry out research on the use of ML in reserving

- Our workstreams
  - Foundations
  - Literature Review
  - Survey
  - Data
  - Research

Find us at https://institute-and-faculty-of-actuaries.github.io/mlr-blog/

Institute and Faculty of Actuaries

Framework

Incremental loss triangle

Available data

To be predicted

Development period

Accident period	1	2	3	4	5
1	47,401	554,104	741,907	908,901	915,388
2	111,019	364,224	715,306	958,406	
3	48,953	357,500	794,413		
4	57,679	418,971			
5	68,434				

Incremental loss data table

Training data

Test data

Accident period	Development period	Incremental loss
1	1	47,401
1	2	554,104
1	3	741,907
1	4	908,901
1	5	915,388
2	1	111,019
2	2	364,224
2	3	715,306
2	4	958,406
2	5	
3	1	48,953
3	2	357,500
3	3	794,413
3	4	
3	5	
4	1	57,679
4	2	418,971
4	3	
4	4	
4	5	
5	1	68,434
5	2	
5	3	
5	4	
5	5	

Framework

X = “Features” or “Predictors” or “Inputs” or “Independent variables”

$$Y \approx f(X)$$

Accident period	Development period	Incremental loss
1	1	47,401
1	2	554,104
1	3	741,907
1	4	908,901
1	5	915,388
2	1	111,019
2	2	364,224
2	3	715,306
2	4	958,406
2	5	
3	1	48,953

$$\left(Y - f(X)\right)^2$$

Y = “Target” or “Output” or “Response” or “Dependent variable”

Features

X

Y	1.30	3.00	8.30	6.30	11.00
Angle_h	45	90	255	195	330
Angle_m	180	0	180	180	0

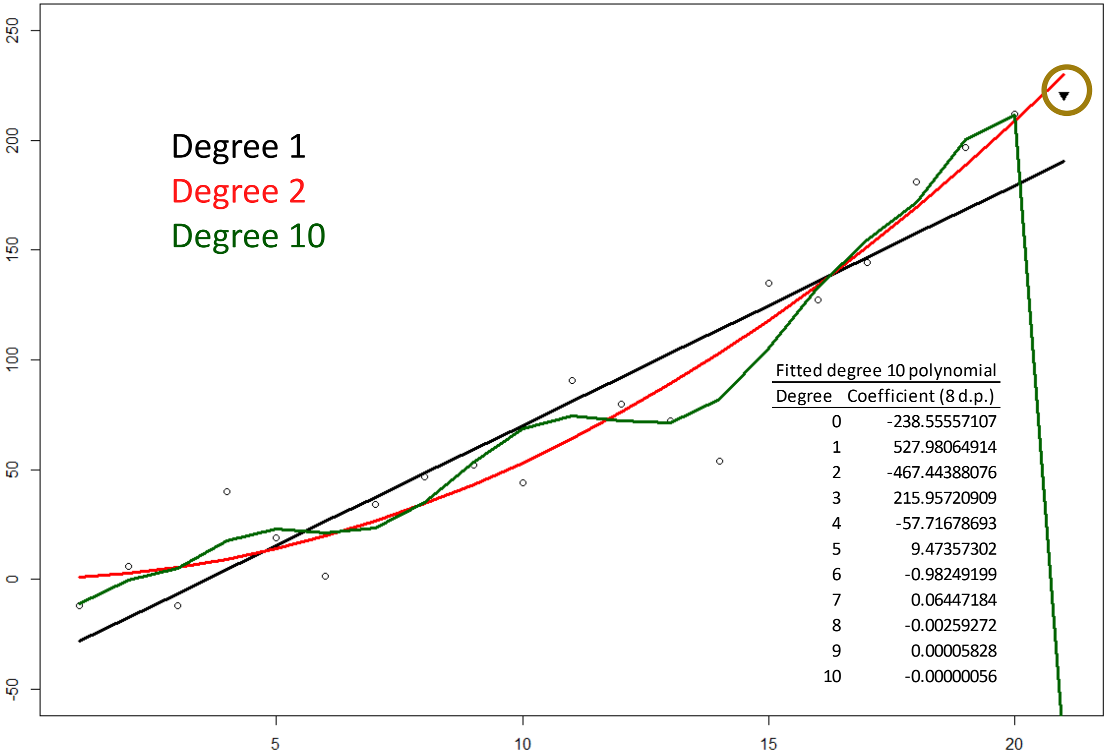
E.g. calendar period

Institute and Faculty of Actuaries

*Based on an example from the book “Deep Learning with R” by Francois Chollet with J.J. Allaire

Hyperparameters and tuning

Polynomials of degree 1, 2 and 10 fitted to 20 x-y pairs of a quadratic (plus noise) and used to predict value at x = 21



Fitted degree 10 polynomial

Degree	Coefficient (8 d.p.)
0	-238.55557107
1	527.98064914
2	-467.44388076
3	215.95720909
4	-57.71678693
5	9.47357302
6	-0.98249199
7	0.06447184
8	-0.00259272
9	0.00005828
10	-0.00000056

- Example – quadratic plus random noise
- Fit a polynomial using first 20 points (training data)
- Predict the value at x = 21 (test data)
- Degree of polynomial is a hyperparameter

Cross validation

Acc	Dev	Incremental loss	**Cross validation fold**
1	1	47,401	**2**
1	2	554,104	**2**
1	3	741,907	**1**
1	4	908,901	**3**
1	5	915,388	**2**
2	1	111,019	**1**
2	2	364,224	**1**
2	3	715,306	**3**
2	4	958,406	**3**
2	5		**N/A**
3	1	48,953	**3**
3	2	357,500	**3**
3	3	794,413	**2**
3	4		**N/A**
3	5		**N/A**
4	1	57,679	**1**
4	2	418,971	**2**
4	3		**N/A**
4	4		**N/A**
4	5		**N/A**
5	1	68,434	**1**
5	2		**N/A**
5	3		**N/A**
5	4		**N/A**
5	5		**N/A**

- Withhold some training data from fitting process
- Use this data to estimate performance out-of-sample for candidate hyperparameter
- Random 3-fold cross validation: fit model to data using folds 1 and 2 and predict target in fold 3. Compare prediction to (known) actual in fold 3 to estimate out-of-sample performance.

Institute and Faculty of Actuaries

LASSO

Select λ → Fit model (fits a β_i for each feature x_i)

$$e^{\beta_0+\beta_1 x_1+\cdots\beta_p x_p}$$

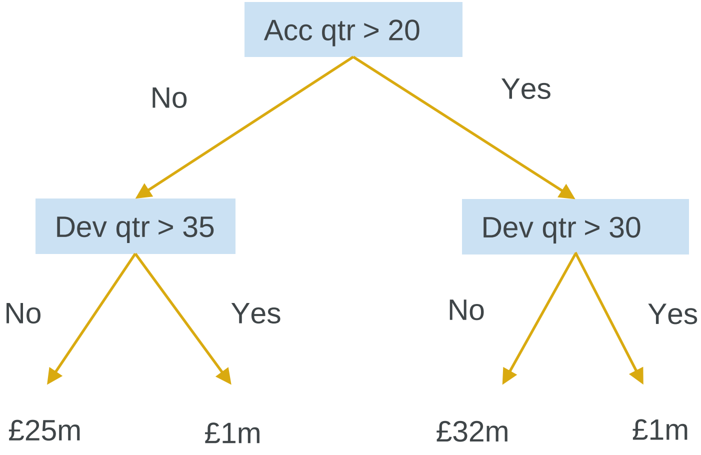
Minimise the expression below:

$$-\sum_{m=1}^{n} l(y_m;\hat{\beta}) + \lambda \sum_{r=1}^{p} |\hat{\beta}_r|$$

Hyperparameter

Institute and Faculty of Actuaries

XGBoost



- Individual decision tree model typically performs poorly
- XGBoost outputs a collection of decision trees – combined prediction much better
- Several hyperparameters control how the collection of decision trees is constructed – number of trees to use, rate of adjustment from one tree to the next, tree depth and many more
- Outstanding track record in data science prediction competitions
- Not easy to grasp the details behind fitting procedure

Data

The SynthETIC R package* implements a simulation machine for claims data using the methodology described by Avanzi et al, 2020.

Four interesting environments are already in the public domain**

We simulated twenty triangles for each environment

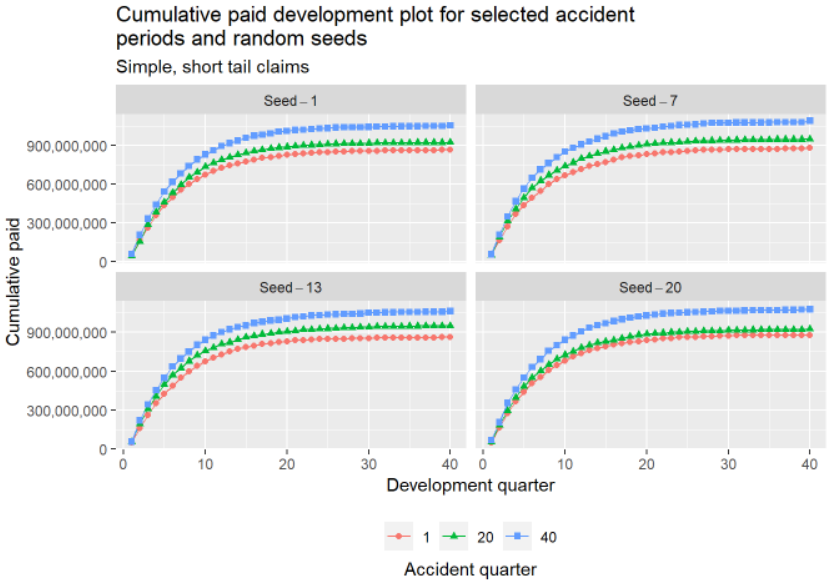
Institute and Faculty of Actuaries

*https://institute-and-faculty-of-actuaries.github.io/mlr-blog/post/synthetic/ for background

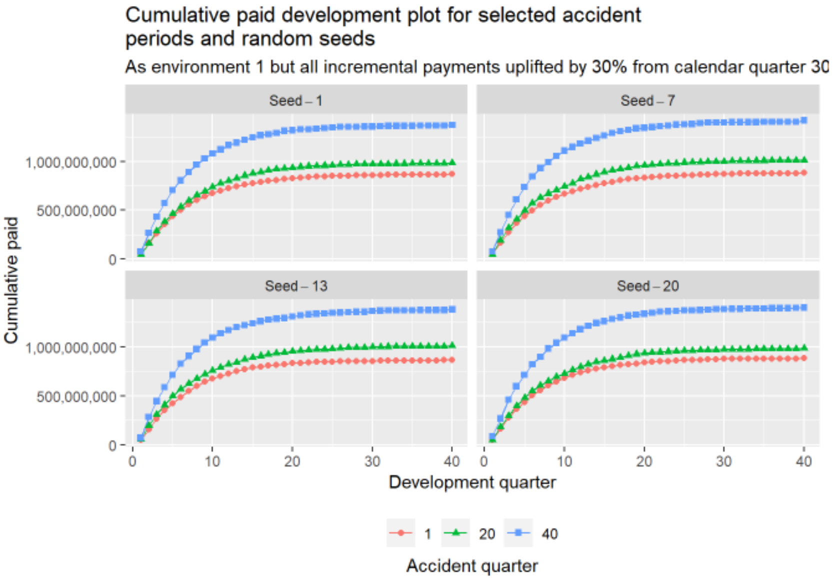
**https://github.com/agi-lab/reserving-MDN-ResMDN

Data

Environment 1



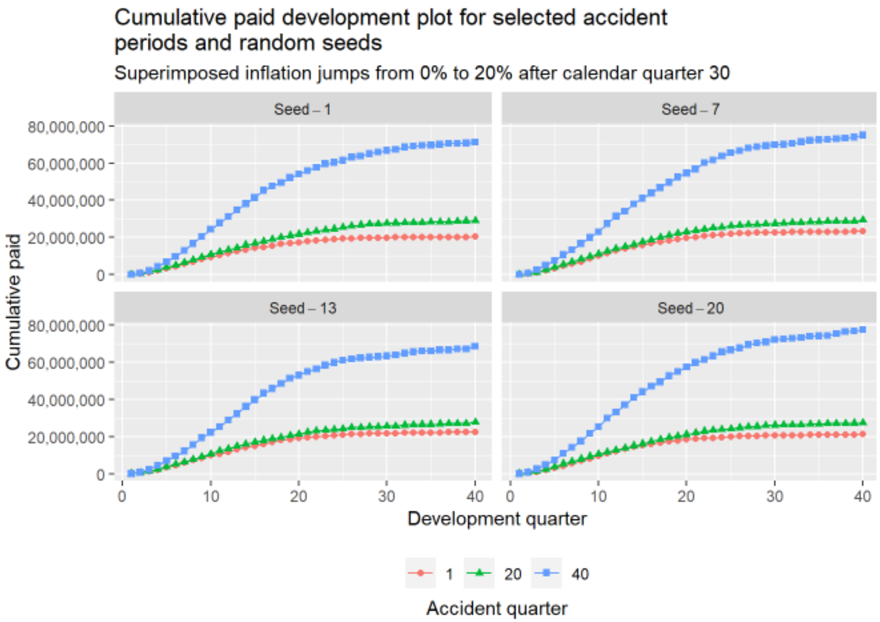
Environment 2



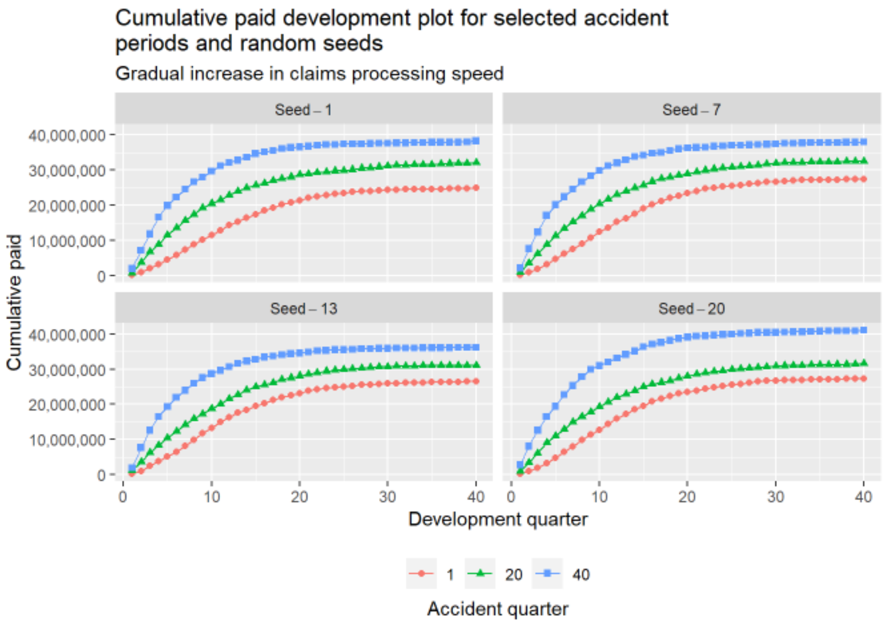
Institute and Faculty of Actuaries

Data

Environment 3



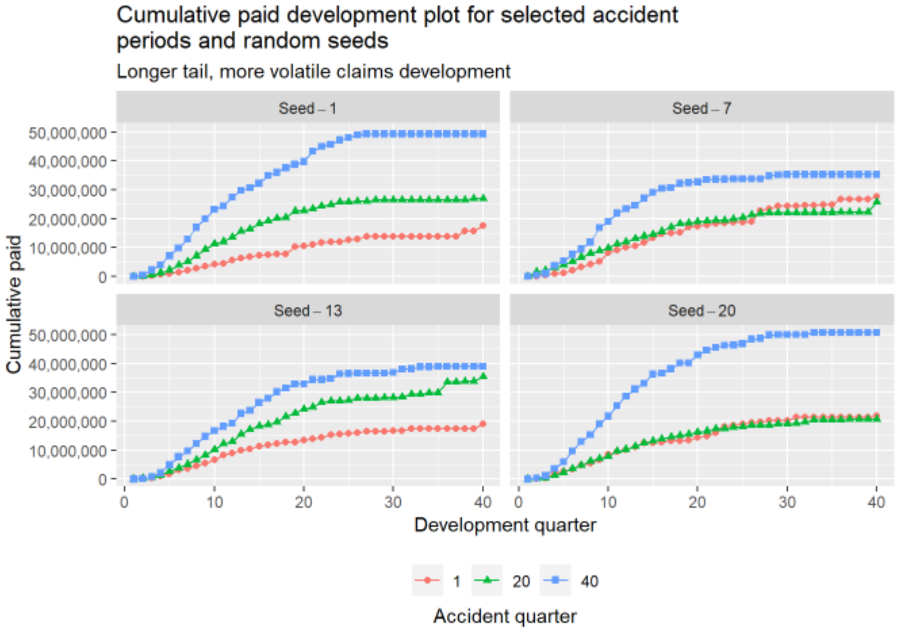
Environment 4



Institute and Faculty of Actuaries

Data

Environment 5



Summary of modelling approach

- 20 simulations of 40 x 40 triangle of accident x development quarter.
- Training data is calendar quarter <= 40, test data is calendar quarter>40
- Chain ladder (volume all), LASSO and XGBoost fit using accident and development quarter factors as features ("_Basic" models)
- 5-fold random cross validation
- LASSO lambda tuned per blog post* and XGBoost n_rounds tuned
- Additional features engineered based on LASSO blog post* to capture interactions and calendar/accident/development period trends. LASSO and XGBoost fitted to this data ("_Extra" models)

*https://institute-and-faculty-of-actuaries.github.io/mlr-blog/post/f-lasso/

Caveat

- The examples here are intended to be instructional rather than conclusive
- We make no claims about the superiority/inferiority of any individual machine learning method for reserving in general.
- Real world data will introduce more problems
- Better performance in our examples may be possible with more time to tune the hyperparameters/different cross validation approach/different loss function
- Full code will be released on the blog site (soon)

Institute and Faculty of Actuaries

Results

Average reserve error [(predicted future paid / actual future paid) – 1] across all 20 random seeds

Environment	Description	Chain ladder	LASSO_Basic	LASSO_Extra	XGBoost_Basic	XGBoost_Extra
1	Simple, short tail	1%	13%	0%	2%	-3%
2	30% uplift to incremental paid from cal qtr 30 onwards	9%	21%	1%	6%	0%
3	Superimposed inflation jumps to 20% after cal qtr 30	-33%	-39%	-3%	-54%	-25%
4	Gradual increase in claims processing speed	95%	111%	2%	65%	9%
5	Longer tail, more volatile claims development	53%	3%	23%	-21%	-25%

Results

- Shiny app walkthrough

Conclusion

- In simulated data, ML methods were able to reproduce CL results on simple development data and pick up on calendar / accident period trends that cause CL problems
- R Shiny is a useful environment for analysing diagnostic plots that support results interpretation
- Lots more work to do!

Further work on triangles

- Rolling origin cross validation
- Loss function – claims development result
- Real-world data
- Further model interpretation and diagnostics

Questions

Comments

The views expressed in this [publication/presentation] are those of invited contributors and not necessarily those of the IFoA. The IFoA do not endorse any of the views stated, nor any claims or representations made in this [publication/presentation] and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this [publication/presentation].

The information and expressions of opinion contained in this publication are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this [publication/presentation] be reproduced without the written permission of the IFoA [*or authors, in the case of non-IFoA research*].

Institute and Faculty of Actuaries