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
  - Product development (UBI)
- Customer management
  - Retention segmentation and program design
- Underwriting and risk management
  - Agent/Broker performance evaluation
- Pricing
- Claims management
  - Fraud detection
- Asset management
  - P&C to Life cross-selling
  - Web-scraping and feature design in commercial lines
  - Topic modeling and large loss modeling

25 April 2019
This is not new….

- Hyper scale parallel computing
- Distributed Big Data storage/ Hadoop
- Data visualisation tools
- Free software environments, analytics libraries
- Machine learning
- Data stream and real-time processing supporting IoT
- Integrated environments and services
- NoSQL databases
- Few factors, simple methods
- GLMs in auto risk models
- GLM refinement & LOB expansion
- More data enrichment
- GLMs
- 1990s
- 2000s
- 2010s
- 2019
- Other “Non-GLM” models
- Other “Non-GLM” models
What are these 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
Choosing a method
Dimensions of choice

- Analytical time and effort
- Predictive power
- Interpretation
- Execution speed
- Table implementation
- Stability
Focus on Trees

- Ensembles
  - Classifications Trees
  - "Earth"
  - Regression Trees
  - Boosting

- K-nearest Neighbours
- Elastic Net
- Neural Networks
- Naïve Bayes
- Random Forests

- K-Means Clustering
- Principal Components Analysis
- Lasso
- Support Vector Machines
- Ridge Regression

25 April 2019
Decision Trees

Group < 5?
  Y
  N
  Age < 40?
    Y
    N
  Group < 15?
    Y
    N
Decision Trees

- Group < 5?
  - Y
  - N
    - Age < 40?
      - Y
      - N
- Group < 15?
  - Y
  - N
A simple Tree example
A simple Tree example

Tree results
A simple Tree example

Tree results

Group < 3?
Y N

Group < 16?
Y N
A simple Tree example

Tree results

Group < 3?
Y  N

Group < 16?
Y  N

Group < 19?
Y  N

25 April 2019
Shortcomings of using trees

They may miss interactions…

… they may struggle with categorical variables….

…and they can be bad at turning points
Analytical time and effort

Predictive power

Interpretation

Decision Trees

Execution speed

Stability

Table Implementation

25 April 2019
Focus on Random Forests

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

25 April 2019
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 $\rightarrow 0$ 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 $\rightarrow$ Underlying trend, provided trees are complex enough to represent it
- This is bagging (bootstrap aggregation) – fit lots of independent models and take an average

25 April 2019
A simple Random Forest example
A simple Random Forest example
A simple Random Forest example
A simple Random Forest example

Random Forest results: Iteration 5

Tree 1  Tree 2  Tree 3  Tree 4  Tree 5  Tree 6  Tree 7  Tree 8  Tree 9  Tree 10  Underlying trend  Average Trend
A simple Random Forest example
Random Forest

- Predictive power
- Interpretation
- Table Implementation
- Execution speed
- Stability
- Analytical time and effort
Focus on Gradient Boosting Machines
Gradient Boosted Machine or “GBM”

A tree

\[ f_i(x) \]

A GBM

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

25 April 2019
Four main assumptions

- **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)

- **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

GBM results at iteration 0

- # factors = 1
- Interaction depth = 1
- Learning rate = 10%
- Bag fraction = 100%

[Graph showing GBM results at iteration 0 with current residuals, underlying trend, and current fitted values.]
A simple GBM example

GBM results at iteration 0

Current residuals
Model trained on current residuals
Underlying trend
Current fitted values

25 April 2019
A simple GBM example

GBM results at iteration 0

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

25 April 2019
A simple GBM example

GBM results at iteration 1

Current residuals Underlying trend Current fitted values
A simple GBM example

GBM results at iteration 1

- Current residuals
- Model trained on current residuals
- Underlying trend
- Current fitted values

25 April 2019
A simple GBM example

GBM results at iteration 1

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

λ

25 April 2019
A simple GBM example

GBM results at iteration 2

-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values
A simple GBM example

GBM results at iteration 3

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

λ
λ
λ

25 April 2019
A simple GBM example

GBM results at iteration 4

λ + λ + λ + λ

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

25 April 2019
A simple GBM example

GBM results at iteration 5

Current residuals Model trained on current residuals Incremental model update Underlying trend Current fitted values

\[ \lambda + \lambda + \lambda + \lambda + \ldots \]

25 April 2019
A simple GBM example

GBM results at iteration 6

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

\[ \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \ldots \]

\[ \lambda_1 + \lambda_2 \]

25 April 2019
A simple GBM example

GBM results at iteration 7

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

\[ \lambda + \lambda + \lambda + \lambda + \ldots \]

\[ \lambda + \lambda + \lambda + \ldots \]

25 April 2019
A simple GBM example

GBM results at iteration 8

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

λ + λ + λ + λ + λ + ...

λ + λ + λ + λ + λ

25 April 2019
A simple GBM example

GBM results at iteration 9

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

\[ \lambda \]

\[ + \lambda \]

\[ + \lambda \]

\[ + \lambda \]

\[ + \ldots \]

\[ \lambda \]

25 April 2019
A simple GBM example

GBM results at iteration 10

Current residuals Model trained on current residuals Incremental model update Underlying trend Current fitted values

\[ \lambda + \lambda + \lambda + \lambda + \ldots \]

\[ \lambda + \lambda + \lambda + \lambda + \ldots \]

\[ \lambda + \lambda \]

25 April 2019
A simple GBM example

GBM results at iteration 20

-0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1 1.2

λ +λ +λ +λ +λ +…

λ +λ +λ +λ +λ +…

λ +λ +λ +λ +λ +…

λ +λ +λ +λ +λ +…

λ +λ +λ +λ +λ +…

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

25 April 2019
A simple GBM example

GBM results at iteration 30

- Current residuals
- Model trained on current residuals
- Incremental model update
- Underlying trend
- Current fitted values

25 April 2019
A simple GBM example

GBM results at iteration 40

- Current residuals
- Model trained on current residuals
- Incremental model update
- Underlying trend
- Current fitted values

25 April 2019
A simple GBM example

GBM results at iteration 50

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

25 April 2019
A simple GBM example

GBM results at iteration 100

- Current residuals
- Model trained on current residuals
- Incremental model update
- Underlying trend
- Current fitted values

25 April 2019
A simple GBM example

GBM results at iteration 200

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

25 April 2019
A simple GBM example

GBM results at iteration 300

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

25 April 2019
A simple GBM example

GBM results at iteration 1,000

Current residuals
Model trained on current residuals
Incremental model update
Underlying trend
Current fitted values

25 April 2019
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 gold, test on blue parts, take average

- Resulting plots can be used to determine the optimal assumption choice
  - Including how many trees to run
Example 5-fold cross validation

Minimum point shows optimal number of trees in each case.
This example is based on artificial data – large insurance datasets indicate a larger number of trees to be optimal.

Best result shown by brown line as has lowest minimum validation error (interaction depth 2 and learning rate 2% in this case)
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?
The relative influence of a factor can be measured as the total reduction in error attributable to splits by that factor, across all trees in the GBM.
Partial dependency plots

Example

Use the model to make a prediction for observation 1 (Factor = 10).
Partial dependency plots

Example

Vary the value of Factor only for observation 1 and make a range of alternative predictions.

This gives the Individual Conditional Expectation of observation 1 across Factor.
Partial dependency plots

Example

Repeat for all observations.
Partial dependency plots

Example

Repeat for all observations.
Partial dependency plots

Example

Repeat for all observations.
Partial dependency plots

Example

The full picture of the variation in predictions for all observations is the Individual Conditional Expectation (or ICE) plot.

Repeat for all observations.
Partial dependency plots

Example

Take the average prediction for each level of Factor.

The average variation across the factor gives the Partial Dependency Plot.
Partial dependency plots

Example

Rebasing all lines to pass through a single point gives a sense of the interactions present in the model.

This is a Centered PDP/ICE plot (c-PDP/c-ICE)
Partial dependency plots

Example

Colouring the c-ICE plots by each observation’s value of a secondary factor can help locate the interaction.
Partial dependency plots etc
Partial dependency plots

- Advantages
  - Qualitative description of properties of relationships
  - Most revealing of additive and multiplicative relationships

- Disadvantages
  - “GLM view of a non-GLM thing”
  - Interaction effects outside of the chosen subset may be obfuscated
  - eg if X1X2 is important and X2 is averaged out in the partial dependence plot, X1 may show as being heterogeneous, thus obfuscating the complexity of the modelled relationships
Model build process

Data → Model → "Comfort Diagnostics" → Deploy directly

Pre / post mapping
Deploying GBMs

Model down into multiplicative tables via GLMs

### Age Exposure Burning Cost
1 <=20 1,720 179 1 -10 164,107 77
2 21-30 34,893 122 2 11-14 84,859 101
3 31-50 118,182 102 3 15-18 28,952 116
4 51+ 127,054 70 4 19-20 3,931 272
5 Age Total 281,849 91

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

Analytical environment

---

New Business Price £
Current Rate
-2.0% +2.0%
-5.0% +5.0%
10%
40%
40%
10%

Next gen rating engine

Deploy directly

Main Policy Admin System

25 April 2019
GBMs

- Predictive power
- Interpretation
- Implementation in modern rating engines
- Table Implementation
- Stability
- Execution speed
- Analytical time and effort

25 April 2019
A summary...

25 April 2019
Machine Learning in Pricing

Conclusions

- There are many forms of ML models
- New data and feature/response engineering generally add more value than new methods BUT we need to continuously explore which methods work on which problems
- Traditional measures of prediction value may not reflect applications in insurance
- And it’s not all about predictive power anyway – other criteria are important

- GBMs can provide predictive lift benefits by capturing higher order effects … BUT
  - Can you cope with not seeing the model and instead use broad diagnostics
  - Effort is required to expose/understand higher order effects in an expeditious manner
  - How will business leaders and regulators respond to this method?
  - Do you have the software and hardware to fit to large dataset
  - Do you have a rating engine that can implement a GBM
Practical applications of tree based methods in pricing

Corner correctors and pre-baked interactions

Factor Reduction

Establish Model Hierarchy
Thank you