Institute
and Faculty
of Actuaries

Machine Learning and Fairness in Commercial Insurance

Paul Bassan and Oliver Laslett

Cytora

26 October 2018

Introductions

Paul Bassan, Lead Actuary

- London Market Pricing Actuary
- Research Scientist: applied machine learning
- PhD in Physics

Oliver Laslett, Senior Data Scientist

- Insurance Data Scientist
- PhD in Physics

Agenda

1. Machine learning primer
2. Explainable machine learning
3. Machine learning and fairness
4. Strategies for fair modelling
5. Key takeaways

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Cytora

Machine learning primer

17 July 2018

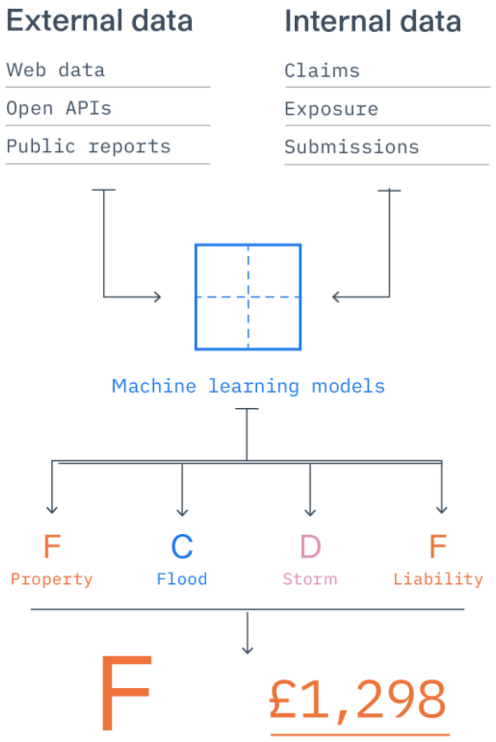
What is machine learning?

Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" (e.g. progressively improve performance on a specific task) with data, without being explicitly programmed.

Topical use cases:

- Driverless cars
- Forecasting in equity prices in financial markets

How can ML be applied to actuarial pricing?



- ML models can be leveraged to predict expected frequency and severity for a risk through the ingestion of a multitude of data sources covering historical losses and risk features
- Robustness of ML models enable relaxation of model assumptions
- Risk premiums are directly derived from an ML model's loss cost prediction

Institute and Faculty of Actuaries

Current pricing vs machine learning

- Rating tables are created typically from risk modelling using generalised linear models (GLMs)
- Data requirements:
  - Policy list with meta information (inception dates etc.)
  - Exposure and covers (e.g. property with excesses & sum insured values)
  - Claims list with meta information
- Rating factors are extracted and dependant factors are encoded using interaction terms
  - Difficult to find and encode every combination of rating factors that have conditional relationship

Institute and Faculty of Actuaries

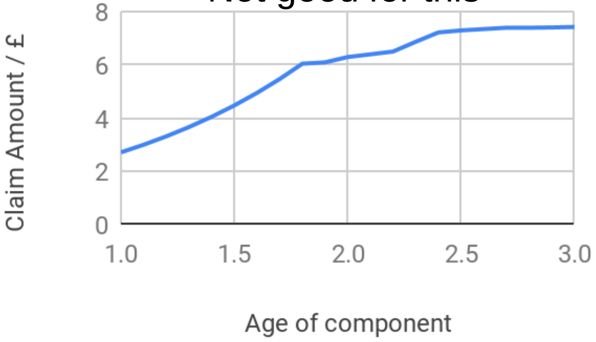
Limitations of linear modelling

$$y = m\mathbf{X} + c$$

Linear refers to the co-efficient

Good for this

Not good for this



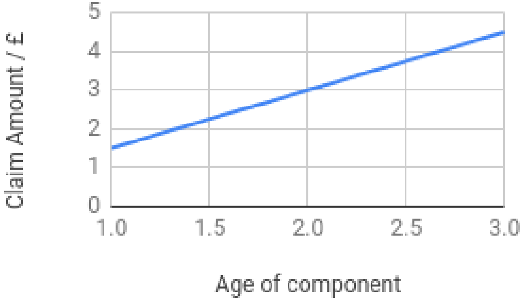
Institute and Faculty of Actuaries

Limitations of generalised linear modelling

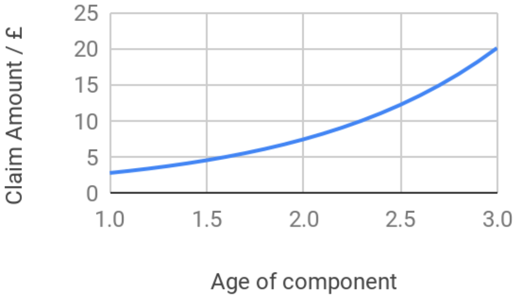
$E[Y] = g^{-1}(\eta)$ $\qquad$ $\eta = \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \ldots.$

Linear refers to the coefficients

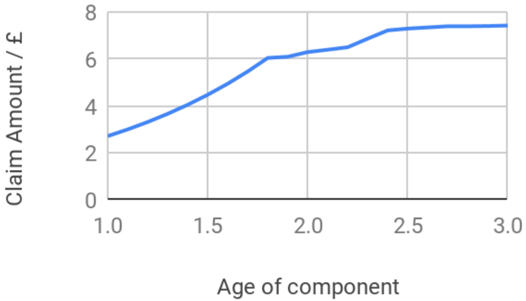
Good for this



Good for this

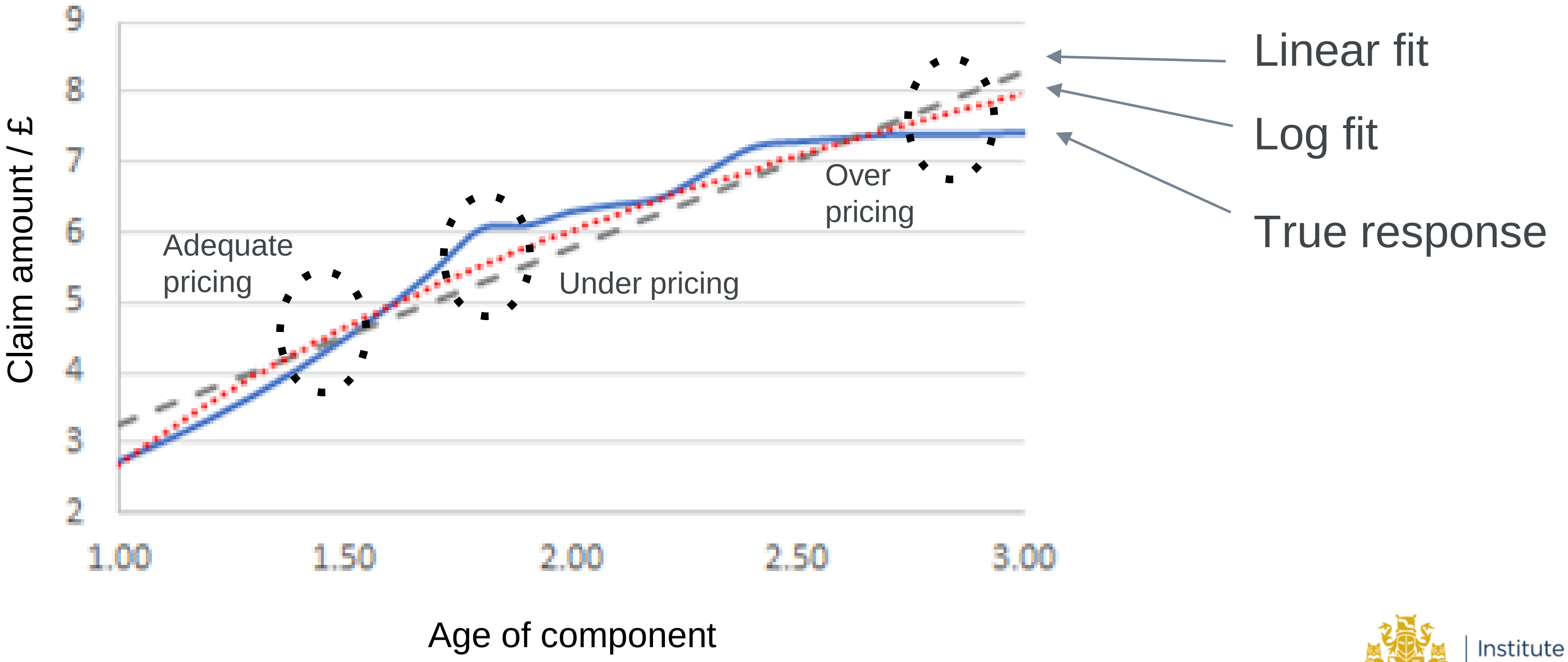


Not good for this



Institute and Faculty of Actuaries

GLM limitations and competitive pricing



Institute and Faculty of Actuaries

Complex relationships in data: implications for rating & pricing — interacting terms

Take 3 rating factors:

- Trade (categorical: Restaurant, Office, Shop etc.)
- Post code (categorical EC1, M20 etc.)
- Distance to fire station (continuous number: 0.2km, 3.14km etc.)

Independent rating factors

Trade: Restaurant	Trade: Office	Trade: Shop	Postcode: EC1	Postcode: M20	Distance to fire station
1	0	0	1	0	0.2
0	1	0	0	1	3.14

Institute and Faculty of Actuaries

Complex relationships in data: implications for rating & pricing (2)

y ~ Trade + Postcode + Distance + Trade*Postcode

Trade: Restaurant	Trade: Office	Trade: Shop	Postcode : EC1	Postcode : M20	Distance to fire station	Trade: Restaurant * Postcode EC1	Trade: Restaurant * Postcode M20	Trade: Office * Postcode EC1	etc.
1	0	0	1	0	0.2				
0	1	0	0	1	3.14				

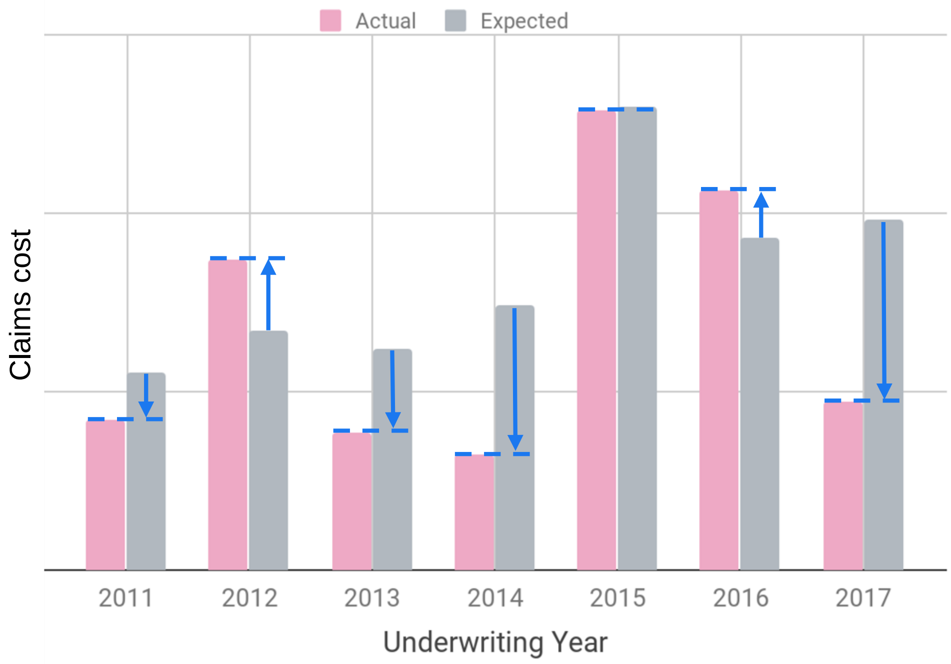
Institute and Faculty of Actuaries

Complex relationships in data: implications for rating & pricing (3)

- What if there are hundreds of rating factors, which are categorical and continuous in nature?
- What if the interactions are not easily identifiable?
- What if 5 rating factors, which have low predictable power when used independently, are actually very strong predictors when used in a joint (interacting) way?
- How feasible is it to assess every combination of rating factor interaction terms?
- Interactions are easier to deal with using machine learning methods.

Institute and Faculty of Actuaries

Complex relationships in data: implications for rating & pricing (4)

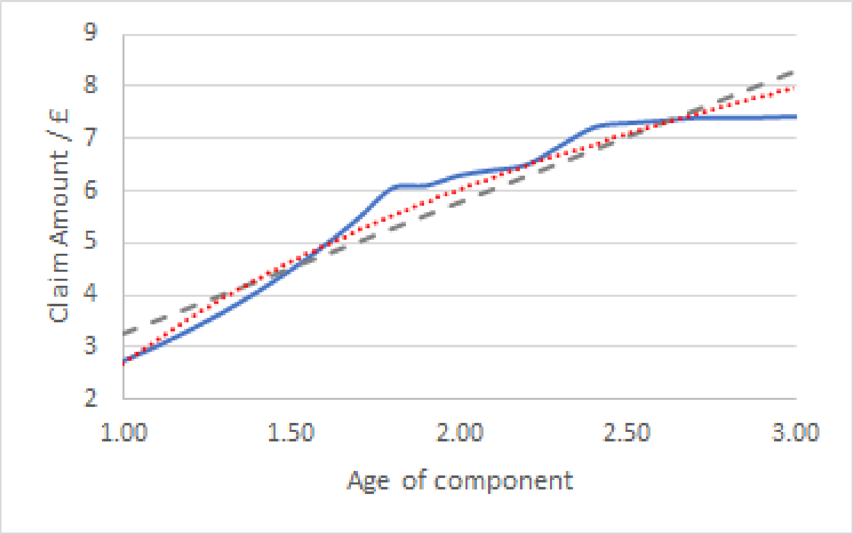


Non-linear machine learning techniques narrow the gap between actual vs expected

Institute and Faculty of Actuaries

Complex relationships in data: implications for rating & pricing (5)

- Machine learning makes it easier to deal with complex data with underlying structure
- No knowledge of the underlying structure is needed
- No interaction terms need to be "built".



Institute and Faculty of Actuaries

Machine learning primer: key takeaway

Machine learning is the next generation of regression methods, which can be applied to pricing.

- Is machine learning model behaviour reliable?
- Is machine learning a black box?
- Is it possible to audit a machine learning model?
- Is machine learning fair?

Institute and Faculty of Actuaries

Cytora

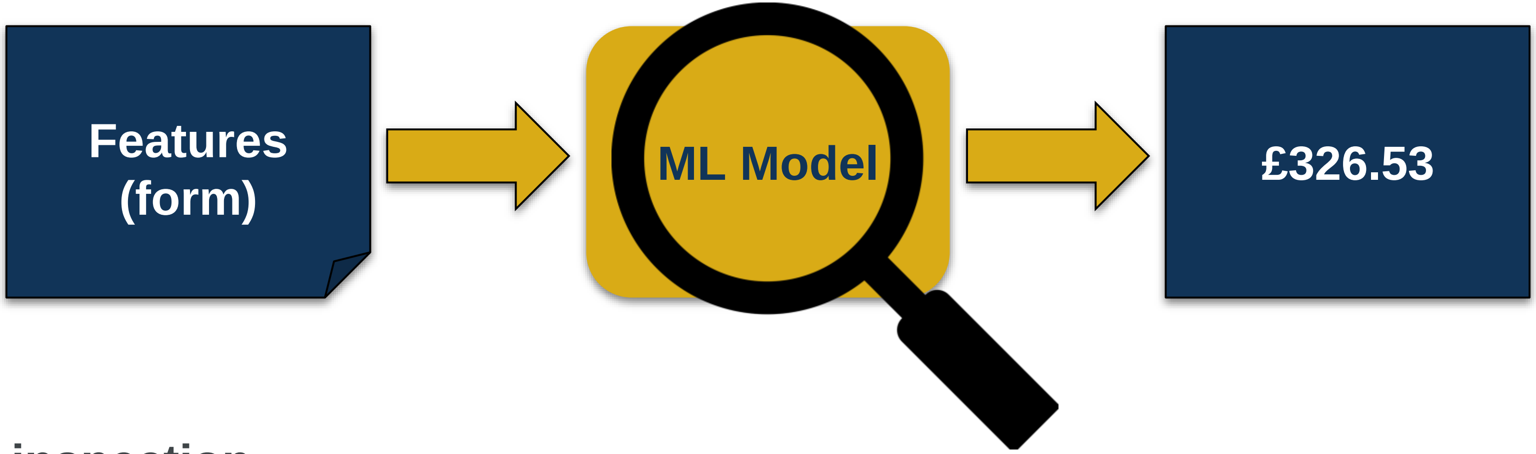
Explainable machine learning

Risk models must be explainable

How/why did we form a decision from the data?

Easy for rating tables

Risk models must be explainable

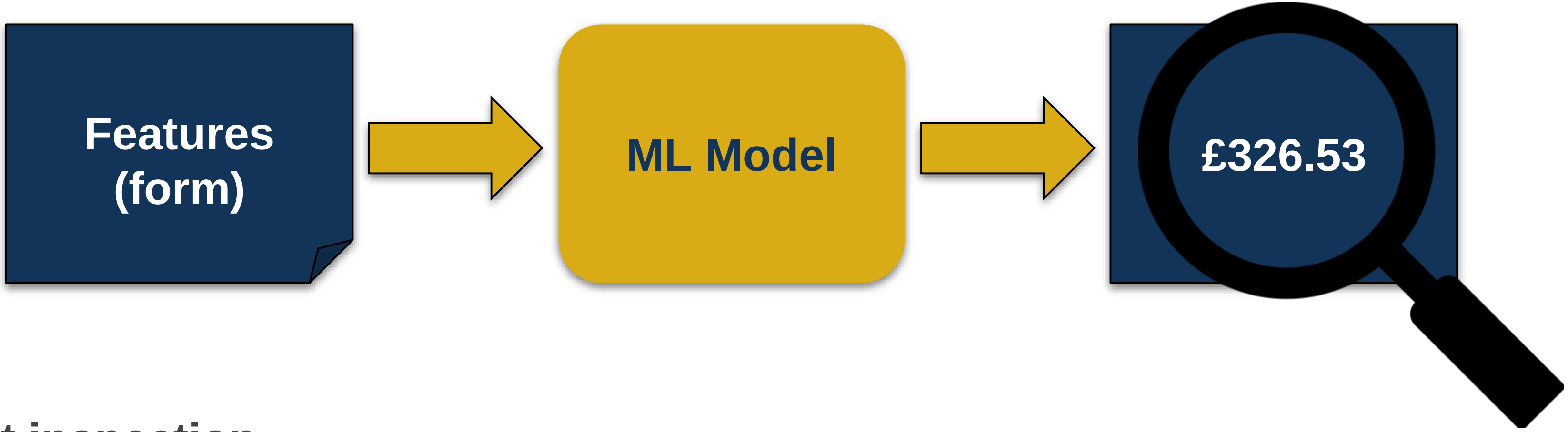


Model inspection

- How is it structured? Coded? Defined?
- What are the key features? Trends? Inputs? Outputs?

Institute and Faculty of Actuaries

Risk models must be explainable



Output inspection

- How did we arrive at this specific output given the inputs?
- How much did each feature contribute?

Institute and Faculty of Actuaries

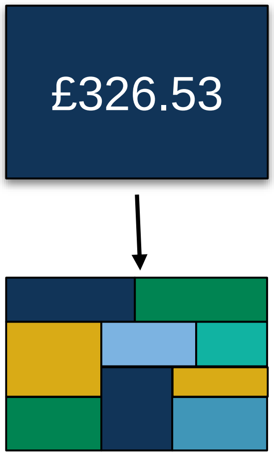
Linear models have excellent explainability

- The effect of every variable is quantified
- Confidence bounds can be estimated
- Predicted value can be decomposed into feature contributions
- One-way and two-way analyses are easily visualised
- Understand coefficient and its effect

Institute and Faculty of Actuaries

Additive feature explanations

Decompose the predicted output into an individual contribution for each feature in the model



$$y = f(x) = \phi_0 + \sum_{i=0}^{N} \phi_i x_i$$

e.g. of the £345 risk premium, £59 is attributed to flood risk

Institute and Faculty of Actuaries

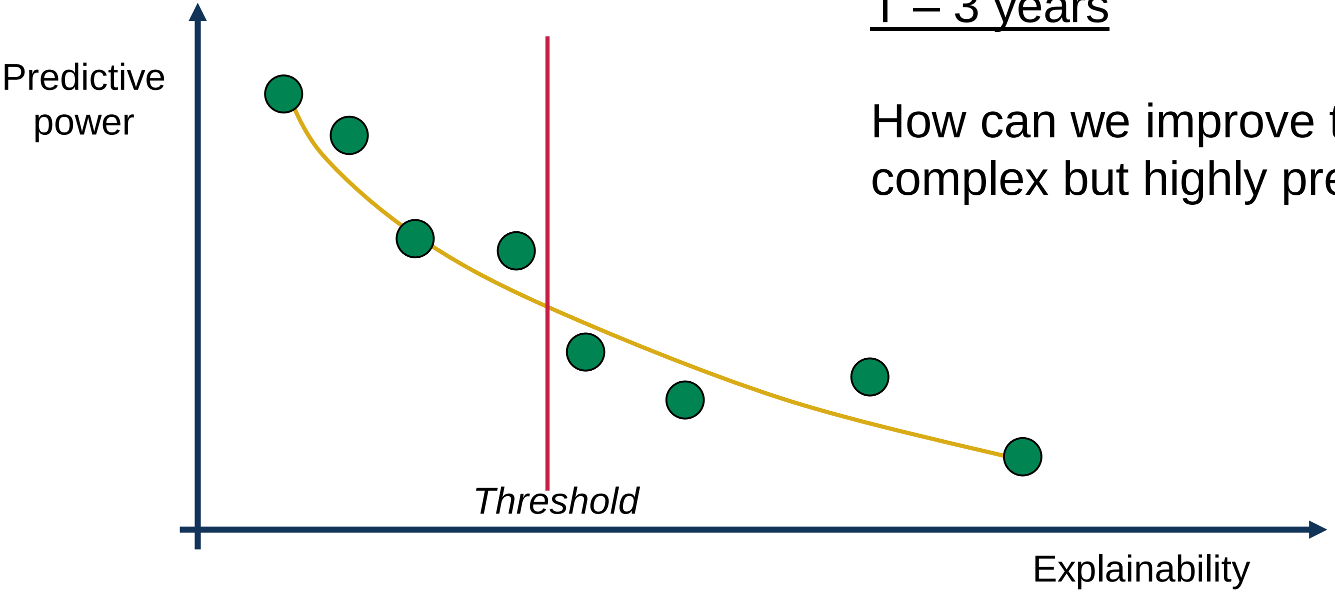
Explainability of common models

Model name	Model explanations	Prediction explanations
GLM	Yes	Yes
GAM	Yes	Yes
Graphical models	Yes	Approximate*
Random forests & GBMs	No	Approximate*
Deep learning	No	Approximate*

* Additive feature contributions can be approximated using linearisation techniques

Institute and Faculty of Actuaries

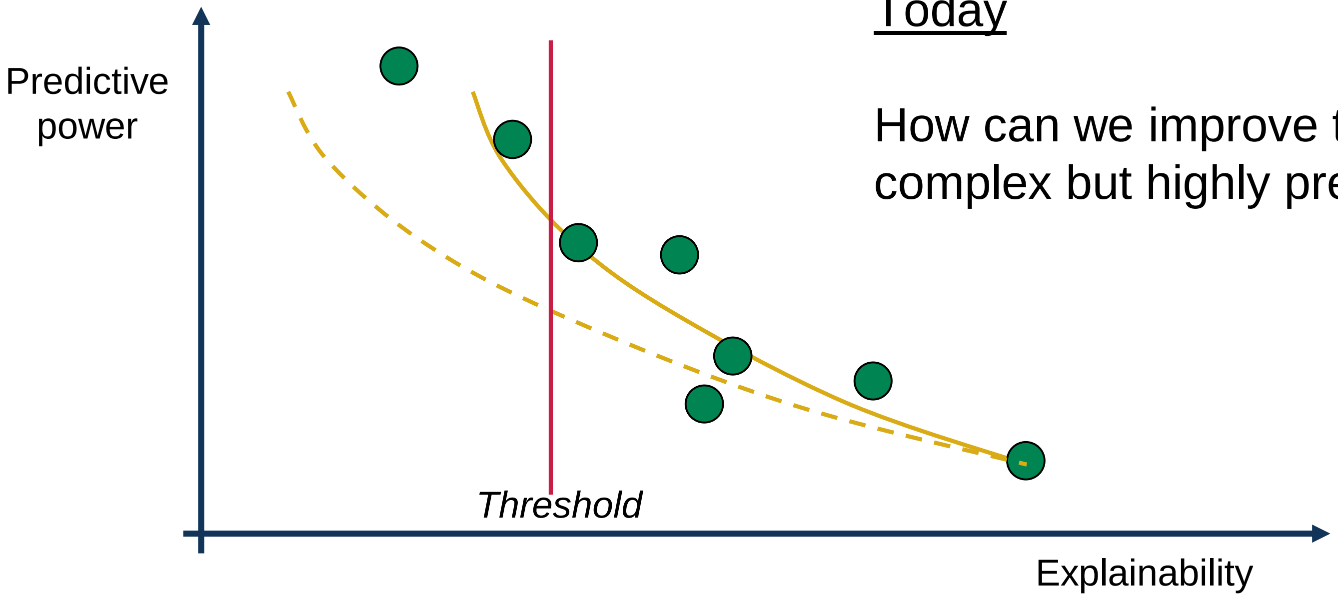
The Explainability Frontier



<u>T – 3 years</u>

How can we improve the explainability of complex but highly predictive ML algos?

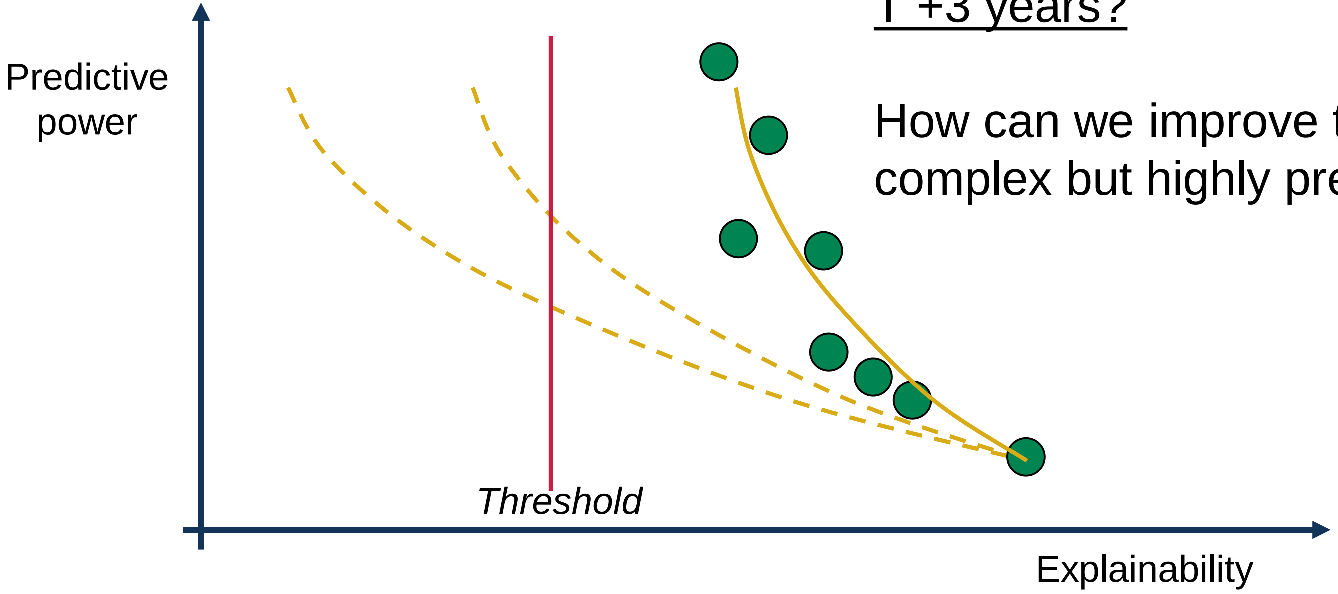
The Explainability Frontier



Today

How can we improve the explainability of complex but highly predictive ML algos?

The Explainability Frontier



<u>T +3 years?</u>

How can we improve the explainability of complex but highly predictive ML algos?

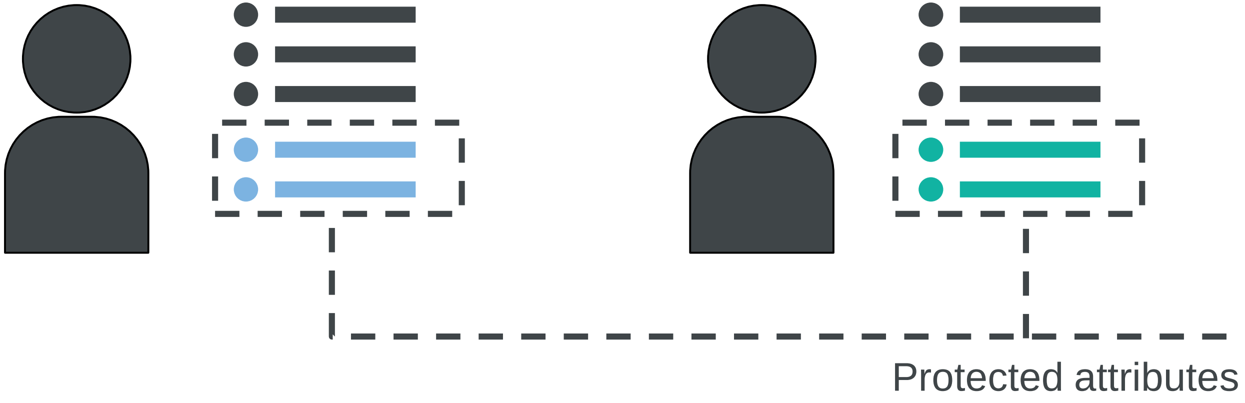
Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Cytora

Machine learning and fairness

Fairness means treating individuals from different groups equally



Institute and Faculty of Actuaries

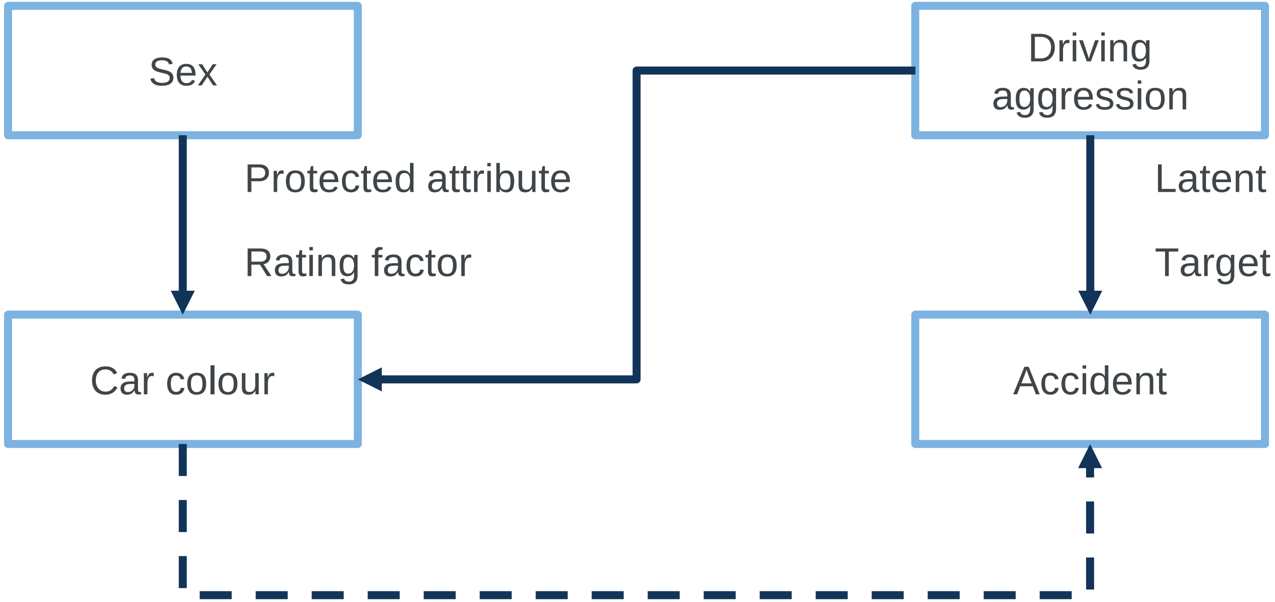
Fairness through unawareness is <u>not sufficient</u> to guarantee equal treatment for individuals in protected groups

Careful consideration is necessary when designing decision systems

Data	**Modelling**
- Inherent data biases	- Quantify feature contributions
- Reasoned vetting of variables	- Tune for fairness
- True measures of latent risk	- Bias in, bias out
- Measure the protected attribute	

Institute and Faculty of Actuaries

Protected attributes encoded in “harmless” rating factors



Kusner, Loftus, Russell, Silva (2018). Counterfactual Fairness.

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Cytora

Strategies for fair modelling

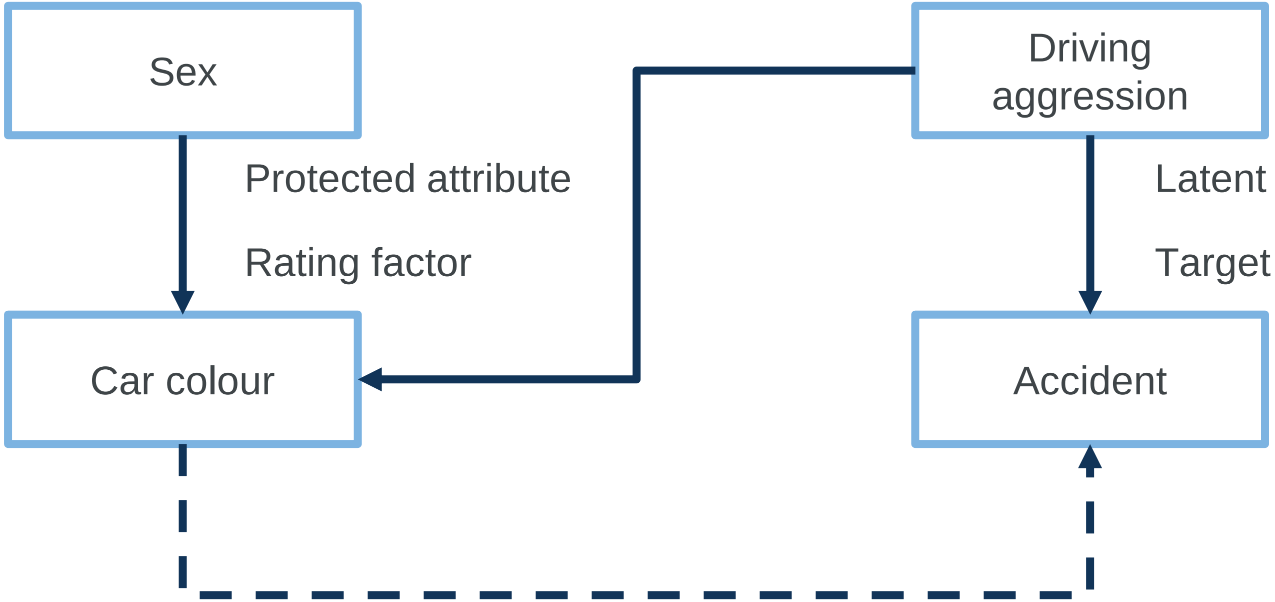
Three strategies for fairness

Fairness is achievable in machine learning models, but we need to be **active** about seeking it out:

1. Observe relevant rating factors
2. Adjust premiums to optimise metrics of fairness
3. Design and train algorithms with fairness baked in

Institute and Faculty of Actuaries

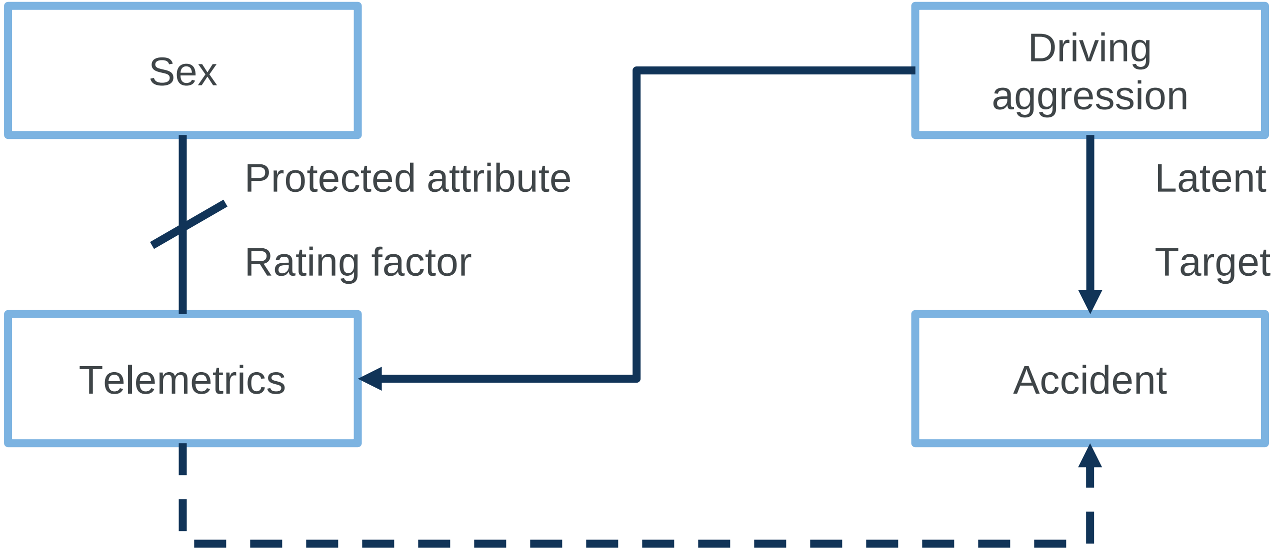
Protected attributes encoded in “harmless” rating factors



Kusner, Loftus, Russell, Silva (2018). Counterfactual Fairness.

Institute and Faculty of Actuaries

1. Observe relevant rating factors



2. Adjust premiums to optimise metrics of fairness

- Profit
- False positive rate, equal opportunity
- False negative rate
- Equalised odds
- Equality of opportunity
- Calibration
- Demographic parity

Predicted

Actual	No loss	Loss
No loss	TN	FP
Loss	FN	TP

How does the confusion matrix compare between groups?

Berk, Heidari, Jabbari, Kearns, Roth (2017). Fairness in Criminal Justice Risk Assessments: The State of the Art.

Kleinberg, Mullainathan, Raghavan (2016). Inherent Trade-Offs in the Fair Determination of Risk Scores.

Institute and Faculty of Actuaries

3. Design and train algorithms with fairness baked-in

- Structural models
  - Kilbertes *et al.* (2018) "Avoiding discrimination through causal reasoning"
  - Kusner, Loftus, Russell, Silva (2018) "Counterfactual fairness"
- Penalised / constrained loss functions
  - Zafar *et al.* (2017) "Fairness beyond disparate treatment and disparate impact"
  - Zhao *et al.* (2017) "Men also like shopping: reducing gender bias amplification using corpus-level constraints"
- Model inspection
  - Tan, Caruana, Hooker, Lou (2018) "Detecting bias in black-box models using transparent model distillation"

Institute and Faculty of Actuaries

Key takeaways

Summary

1. Machine learning uses statistical techniques to give computer systems the ability to "learn" with data, without being explicitly programmed
2. Machine learning models can predict expected frequency and severity for a risk using a multitude of data sources covering historical losses and risk features
3. Machine learning makes it easier to deal with complex data with underlying structure, which can help rating and pricing
4. Machine learning is not a black box
5. Machine learning can be used as a tool to improve fairness (e.g. telemetrics, causal models, …)

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