AI in Actuarial Science

Ronald Richman

4 April 2019

Hosted by ACTUARIAL SOCIETY OF SOUTH AFRICA

SECTION COLLOQUIUM 2019

Disclaimer

This presentation is given solely in my personal capacity and the views expressed in these slides do not necessarily represent those of the AIG Group and its subsidiaries, nor the professional organizations to which I belong (the IFoA, ASSA or IRMSA).

Any software or code referred to in the presentation is provided as is for demonstration purposes only, without any implied warranty, and is licensed under the MIT License which can be viewed on the associated GitHub repository.

Agenda

- *Introduction*
- Machine Learning
- Deep Learning
- Applications in Actuarial Science
- Discussion and Conclusion

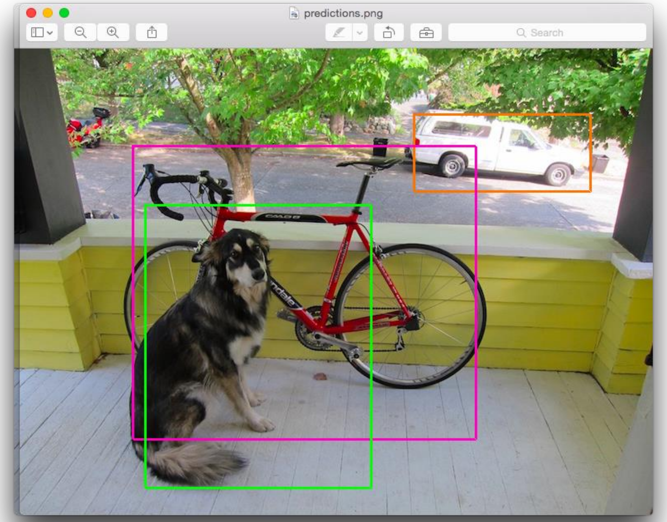
Introduction

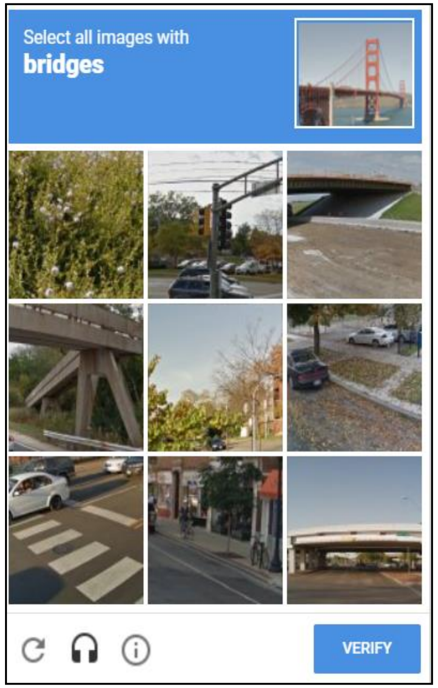
- This talk is about 3 things:
  - Provide context to understand deep learning
  - Discuss applications of deep learning in actuarial science
  - Provide code to experiment (see last slide)
- Inspiration of paper and talk:

"The future of insurance will be staffed by bots rather than brokers and AI in favor of actuaries"- Daniel Schreiber, CEO, Lemonade Inc.

Deep Learning in the Wild

- We all use Deep Learning today:
  - Google/Apple/Facebook/Instagram /Pinterest…
  - … and might use it more in the medium term (self-driving cars/medical applications)
- We help to train DL - Recaptcha
- DL is good enough to trick us
- But, are actuaries benefiting from Deep Learning?





Man from www.thispersondoesnotexist.com/

YOLO from https://github.com/pjreddie/darknet/wiki/YOLO:-Real-Time-Object-Detection

Practical Successes of Deep Learning

- Computer vision starting with AlexNet architecture of Krizhevsky, Sutskever and Hinton (2012)
- Speech recognition (Hannun, Case, Casper et al. 2014).
- Natural language processing, e.g. Google's neural translation machine (Wu, Schuster, Chen et al. 2016)
- Winning method in 2018 M4 time series forecasting competition (Makridakis, Spiliotis and Assimakopoulos 2018a).
- Analysis of GPS data (Brébisson, Simon, Auvolat et al. 2015)
- Analysis of tabular data (Guo and Berkhahn 2016) (plus other Kaggle competitions)

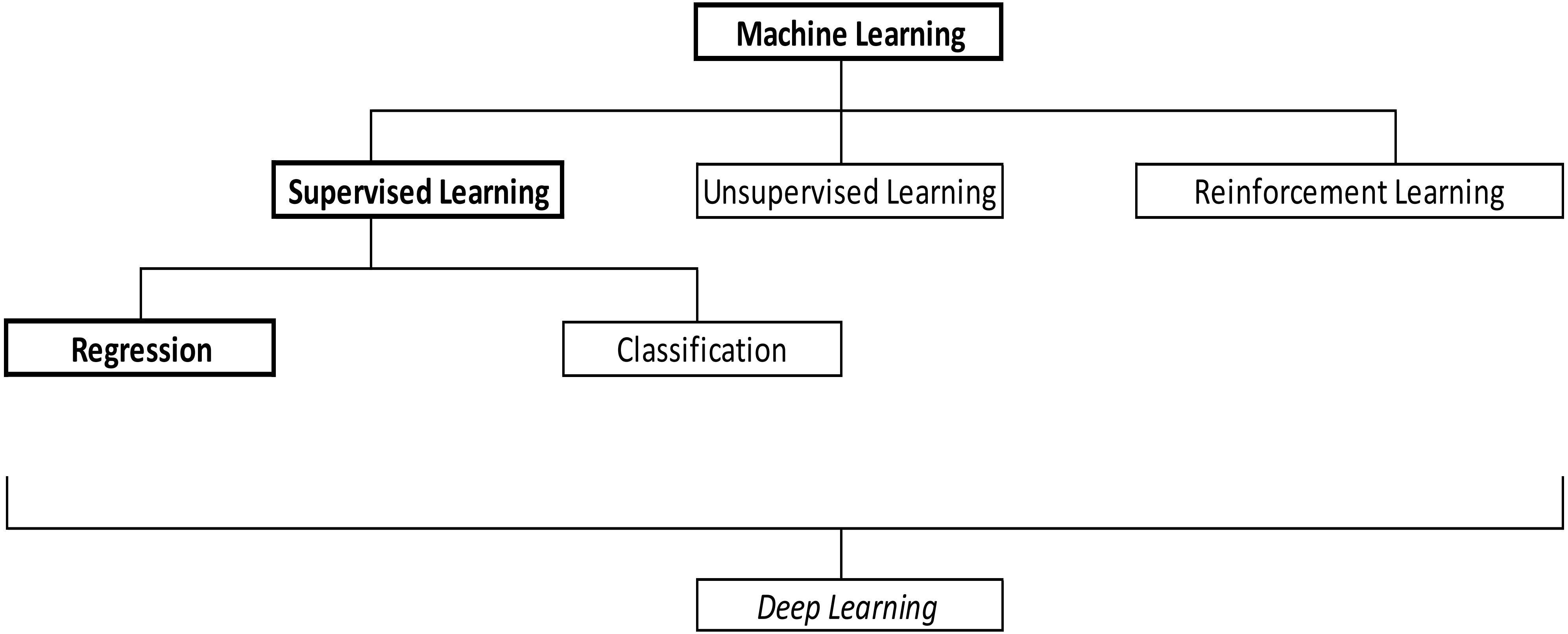
Agenda

- Introduction
- *Machine Learning*
- Deep Learning
- Applications in Actuarial Science
- Discussion and Conclusion

Machine Learning

- Machine Learning is concerned with "the study of algorithms that allow computer programs to automatically improve through experience" (Mitchell 1997)
  - Machine learning approach to AI - systems trained to recognize patterns within data to acquire knowledge (Goodfellow, Bengio and Courville 2016).
- Earlier attempts to build AI systems = hard code knowledge into knowledge bases
- But doesn't work for highly complex tasks e.g. image recognition, scene understanding and inferring semantic concepts (Bengio 2009)
- ML Paradigm – feed data to the machine and let it figure it out!

Map of Machine Learning



Supervised Learning

- Supervised learning = application of machine learning to datasets that contain features and outputs with the goal of predicting the outputs from the features (Friedman, Hastie and Tibshirani 2009).

y (*outputs*) **X (*features*)**

```
> freMTPL2freq
          IDpol ClaimNb    Exposure Area VehPower VehAge DrivAge BonusMalus VehBrand  VehGas Density Region
     1:       1       1 0.100000000    D        5      0      55         50      B12 Regular    1217    R82
     2:       3       1 0.770000000    D        5      0      55         50      B12 Regular    1217    R82
     3:       5       1 0.750000000    B        6      2      52         50      B12  Diesel      54    R22
     4:      10       1 0.090000000    B        7      0      46         50      B12  Diesel      76    R72
     5:      11       1 0.840000000    B        7      0      46         50      B12  Diesel      76    R72
   ---
678009: 6114326       0 0.002739726    E        4      0      54         50      B12 Regular    3317    R93
678010: 6114327       0 0.002739726    E        4      0      41         95      B12 Regular    9850    R11
678011: 6114328       0 0.002739726    D        6      2      45         50      B12  Diesel    1323    R82
678012: 6114329       0 0.002739726    B        4      0      60         50      B12 Regular      95    R26
678013: 6114330       0 0.002739726    B        7      6      29         54      B12  Diesel      65    R72
```

So, ML is just regression, right?

- Not exactly. ML relies on a different approach to building, parameterizing and testing statistical models, based on statistical learning theory. For other ideas – see Richman (2018)
- Distinction between tasks of predicting and explaining, see Shmueli (2010). Focus on predictive performance leads to:
  - Building algorithms to predict responses instead of specifying a stochastic data generating model (Breiman 2001)…
  - … favouring models with good predictive performance at expense of intepretablity.
  - Accepting bias in models if this is expected to reduce the overall prediction error.
  - Quantifying predictive error (i.e. out-of-sample error)

Unsupervised learning

- Unsupervised learning = application of machine learning to datasets containing only features to find structure within these datasets (Sutton and Barto 2018).
- Task of unsupervised learning is to find meaningful patterns using only the features.
- Recent examples:
  - modelling yield curves using Principal Components Analysis (PCA) for the Interest Rate SCR in SII
  - mortality modelling – Lee-Carter model uses PCA to reconstruct mortality curves

The ML Actuary

- Actuarial problems are often supervised regressions =>
- If an actuarial problem can be expressed as a regression, then machine and deep learning techniques can be applied:
  - P&C pricing
  - IBNR reserving
  - Experience analysis
  - Mortality modelling
  - Lite valuation models
- But don't forget about unsupervised learning either!

Agenda

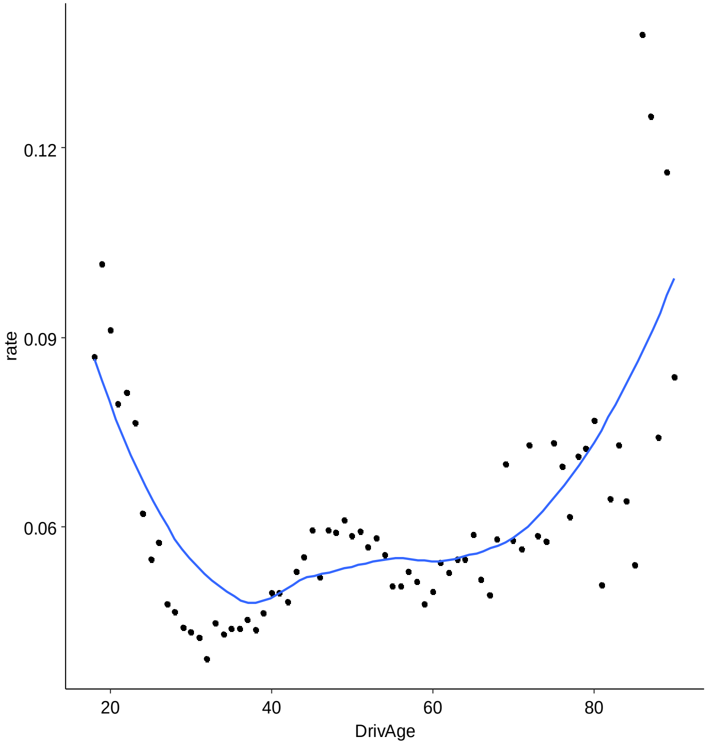
- Introduction
- Machine Learning
- *Deep Learning*
- Applications in Actuarial Science
- Discussion and Conclusion

Feature Engineering (Model Specification)

- Suppose we realize that Claims depends on Age^2 => enlarge feature space by adding Age^2 to data. Other options – add interactions/basis functions e.g. splines

y (*outputs*) **X (*features*)**

```
> freMTPL2freq
         IDpol ClaimNb    Exposure Area VehPower VehAge DrivAge BonusMalus VehBrand  VehGas Density Region DrivAge_2
     1:      1       1 0.100000000    D        5      0      55         50      B12 Regular    1217    R82      3025
     2:      3       1 0.770000000    D        5      0      55         50      B12 Regular    1217    R82      3025
     3:      5       1 0.750000000    B        6      2      52         50      B12  Diesel      54    R22      2704
     4:     10       1 0.090000000    B        7      0      46         50      B12  Diesel      76    R72      2116
     5:     11       1 0.840000000    B        7      0      46         50      B12  Diesel      76    R72      2116
    ---
678009: 6114326       0 0.002739726    E        4      0      54         50      B12 Regular    3317    R93      2916
678010: 6114327       0 0.002739726    E        4      0      41         95      B12 Regular    9850    R11      1681
678011: 6114328       0 0.002739726    D        6      2      45         50      B12  Diesel    1323    R82      2025
678012: 6114329       0 0.002739726    B        4      0      60         50      B12 Regular      95    R26      3600
678013: 6114330       0 0.002739726    B        7      6      29         54      B12  Diesel      65    R72       841
>
```



Representation learning

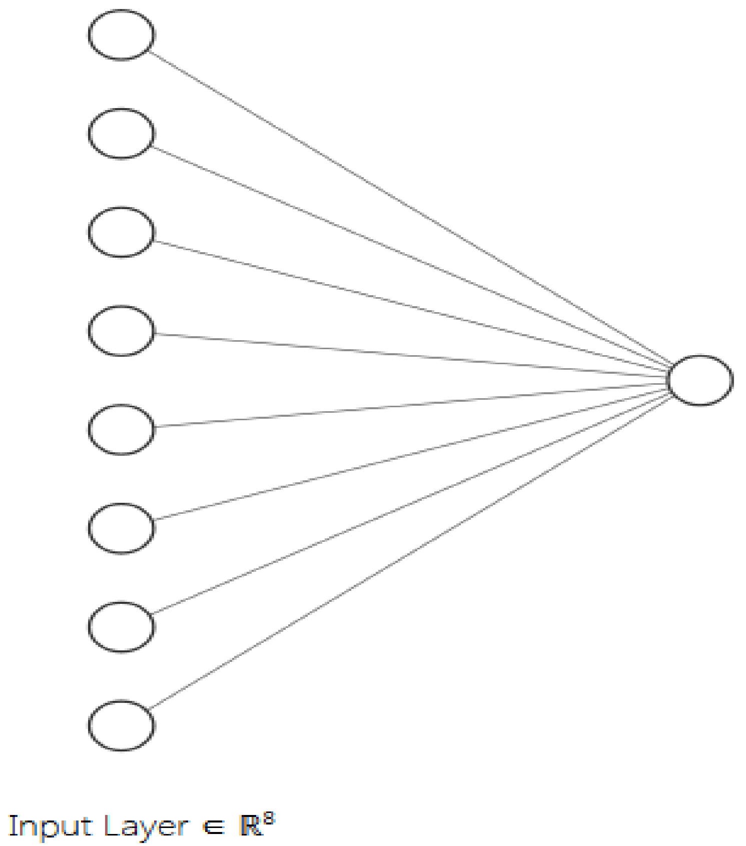
- In many domains, including actuarial science, traditional approach to designing machine learning systems relies on humans for feature engineering. But:
  - designing features is time consuming/tedious
  - relies on expert knowledge that may not be transferable to a new domain
  - becomes difficult with very high dimensional data
- Representation Learning = ML technique where algorithms automatically design features that are optimal for a particular task. Traditional examples are PCA (unsupervised) and PLS (supervised)
- Simple/naive RL approaches often fail when applied to high dimensional data

Deep Learning

- Deep Learning = representation learning technique that <u>automatically constructs hierarchies of complex features</u>
- Modern example of deep learning is feed-forward neural networks, which are multi-layered machine learning models, where each layer learns a new representation of the features.
- The principle: Provide data to the network and let it figure out what and how to learn.
- Desiderata for AI by Bengio (2009):
  - *"Ability to learn with little human input the low-level, intermediate, and high-level abstractions that would be useful to represent the kind of complex functions needed for AI tasks."*

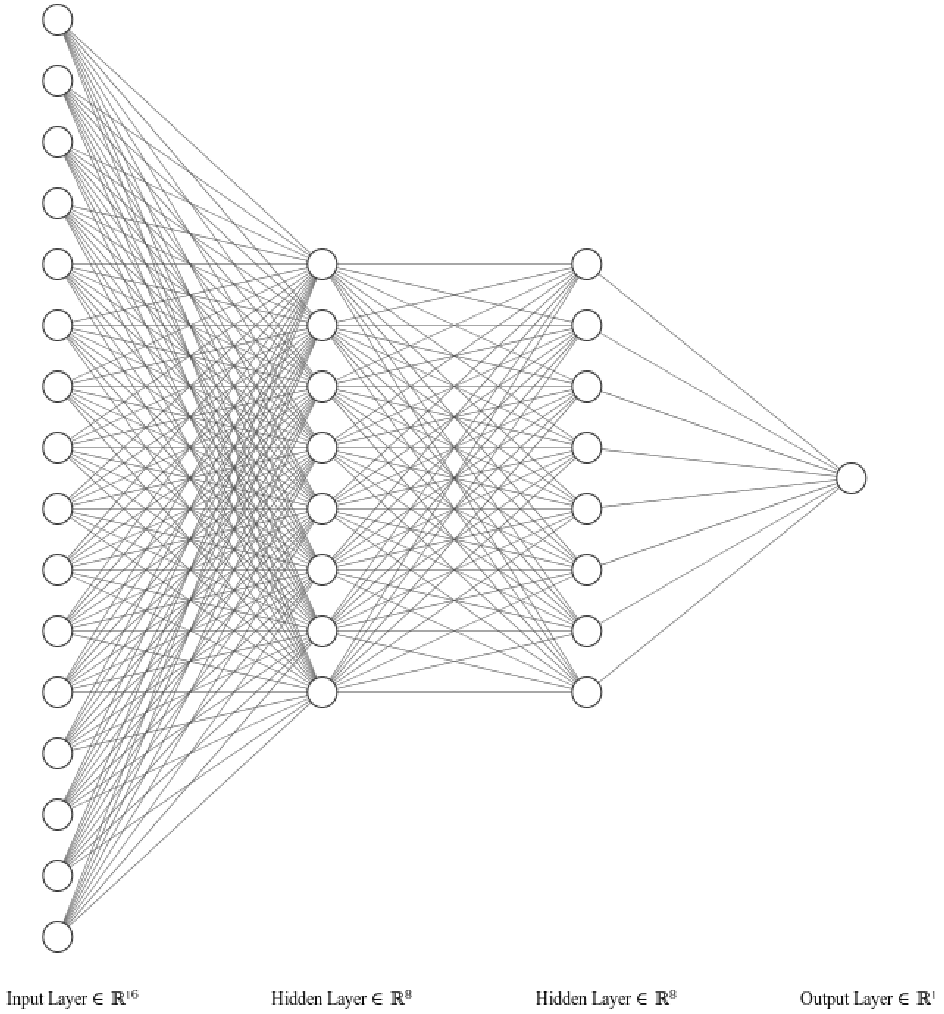
Single Layer NN = Linear Regression

- Single layer neural network
- Circles = variables
- Lines = connections between inputs and outputs
- Input layer holds the variables that are input to the network…
- … multiplied by weights (coefficients) to get to result
- Single layer neural network is a linear regression!



Deep Feedforward Net

- Deep = multiple layers
- Feedforward = data travels from left to right
- Fully connected network = all neurons in layer connected to all neurons in previous layer
- More complicated representations of input data learned in hidden layers
- Subsequent layers represent regressions on the variables in hidden layers



Embedding layers

- Several specialized types of neural networks depending on purpose
- Embedding layer learns dense vector transformation of sparse input vectors and clusters similar categories together; see Section 3.3 in Richman (2018)
- Embeddings often capture actuarially meaningful relationships in categorical data – can be interpreted as relativities

	Actuary	Accountant	Quant	Statistician	Economist	Underwriter
Actuary	1	0	0	0	0	0
Accountant	0	1	0	0	0	0
Quant	0	0	1	0	0	0
Statistician	0	0	0	1	0	0
Economist	0	0	0	0	1	0
Underwriter	0	0	0	0	0	1

	Finance	Math	Stastistics	Liabilities
Actuary	0.5	0.25	0.5	0.5
Accountant	0.5	0	0	0
Quant	0.75	0.25	0.25	0
Statistician	0	0.5	0.85	0
Economist	0.5	0.25	0.5	0
Underwriter	0	0.1	0.05	0.75

Summary of architectures

- Key principle - Use architecture that expresses useful priors about the data => major performance gains:
- Deep feedforward network – structured (tabular) data
- Embedding layers – categorical data (or real values restructured as categorical data)
- Deep autoencoder (non-linear PCA) – unsupervised learning
- Convolutional neural network – data with spatial/temporal dimension e.g. images and time series
- Recurrent neural network – data with temporal structure

Agenda

- Introduction
- Machine Learning
- Deep Learning
- *Applications in Actuarial Science*
- Discussion and Conclusion

Recent Papers applying DL

- Searches within actuarial literature confined to articles written after 2006, when current resurgence of interest in neural networks began (Goodfellow, Bengio and Courville 2016).
  - Pricing of non-life insurance (Noll, Salzmann and Wüthrich 2018; Wüthrich and Buser 2018) **X**
  - IBNR Reserving (Kuo 2018b; Wüthrich 2018b; Zarkadoulas 2017) **X**
  - Analysis of telematics data (Gao, Meng and Wüthrich 2018; Gao and Wüthrich 2017; Wüthrich and Buser 2018; Wüthrich 2017)
  - Mortality forecasting (Hainaut 2018; Richman and Wüthrich 2018) **X**
  - Approximating nested stochastic simulations (Hejazi and Jackson 2016, 2017)
  - Forecasting financial markets (Smith, Beyers and De Villiers 2016)

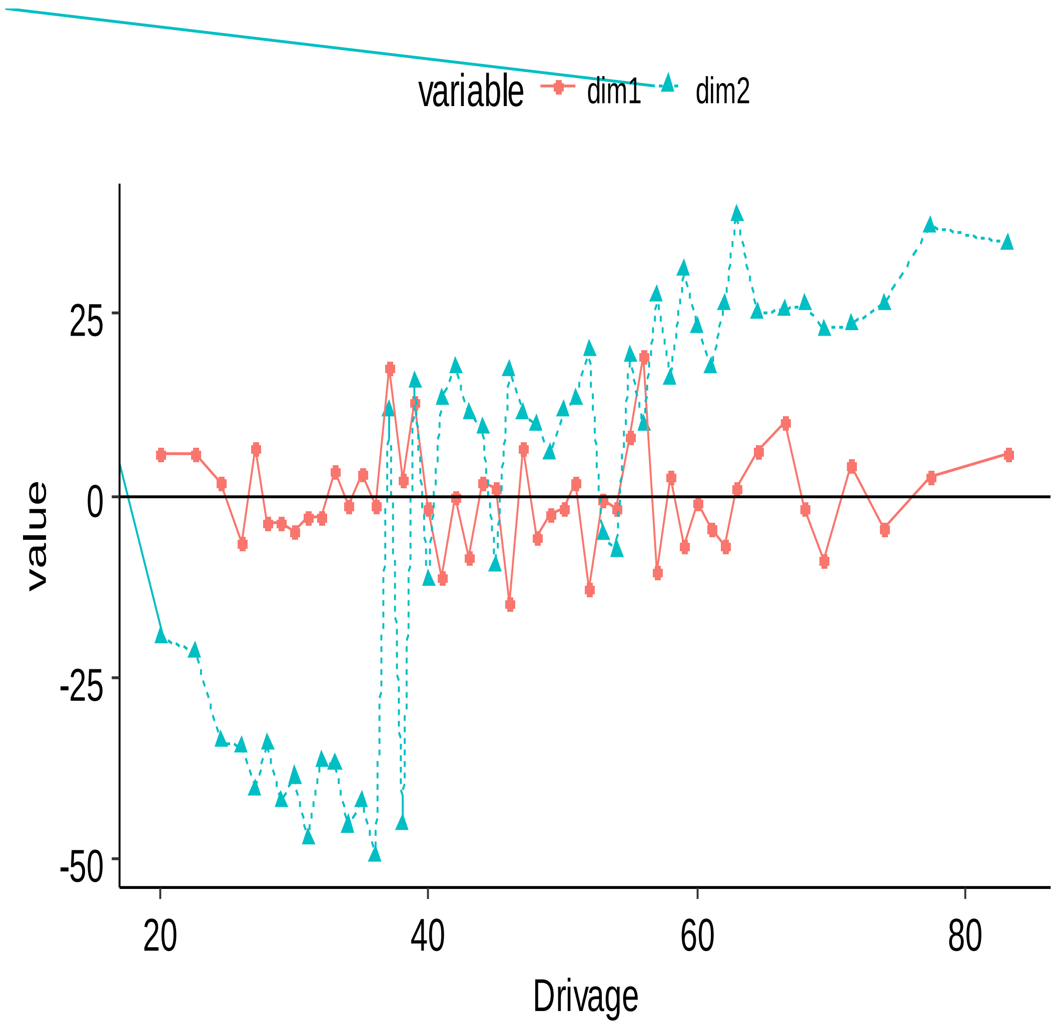
Non-life pricing (1)

- Non-life Pricing (tabular data fit with GLMs) seems like obvious application of ML/DL
- Noll, Salzmann and Wüthrich (2018) is tutorial paper (with code) in which apply GLMs, regression trees, boosting and (shallow) neural networks to French TPL dataset to model frequency
  - ML approaches outperform GLM
  - Boosted tree performs about as well as neural network…
  - ….mainly because ML approaches capture some interactions automatically
  - In own analysis, found that surprisingly, off the shelf approaches do not perform particularly well on frequency models.
  - These include XGBoost and 'vanilla' deep networks

Non-life pricing (2)

- Deep neural network applied to raw data (i.e. no feature engineering) did not perform well
- Embedding layers provide significant gain in performance over GLM and other NN architectures
- Layers learn a (multi-dimensional) schedule of relativities at each age (shown after applying t-SNE)
- Transfer learning – can boost performance of GLM

Model	OutOfSample
GLM	0.3217
GLM_Keras	0.3217
NN_shallow	0.3150
NN_no_FE	0.3258
NN_embed	0.3068
GLM_embed	0.3194
NN_learned_embed	0.2925



IBNR Reserving

- IBNR Reserving boils down to regression of future reported claim amounts on past => good potential for ML/DL approaches
  - Granular reserving for claim type/property damaged/region/age etc difficult with normal chain-ladder approach as too much data to derive LDFs judgementally
  - Wüthrich (2018b) (who provides code + data) extends chain-ladder as a regression model to incorporate features into derivation of LDF

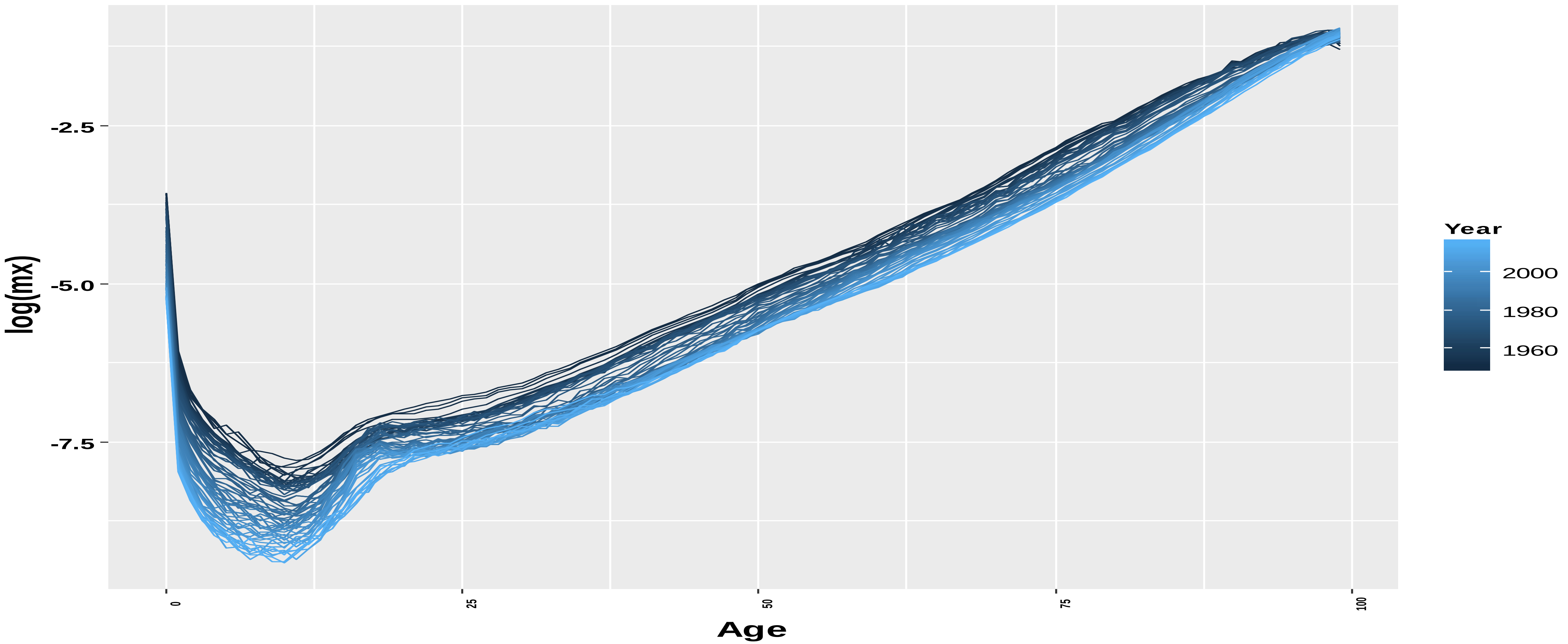
$$\hat{C}_{i,j+1} = f(X).C_{i,j}$$

  - DeepTriangle of Kuo (2018b) is less traditional approach. Joint prediction of Paid + Outstanding claims using Recurrent Neural Networks and Embedding Layers
  - Better performance than CL/GLM/Bayesian techniques on Schedule P data from USA

Mortality Forecasting Project

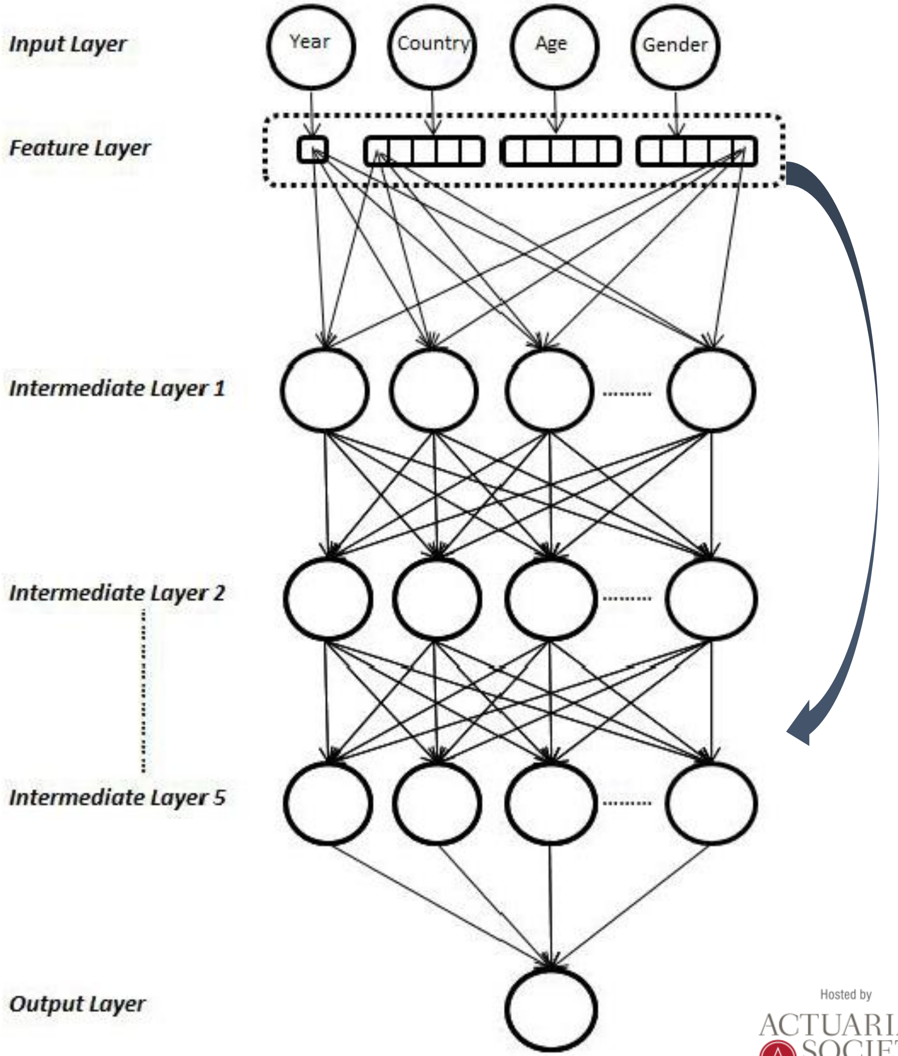
- Richman and Wüthrich (2018)
- Mortality data sourced from Human Mortality Database (HMD) covering mortality rates for ~41 countries, for both genders, from 1950-2016
- Divide data into training and test sets:
  - Training set = observations at ages 0-99 occurring in the years before 2000
  - Test set = observations in the years 2000-2016
- Countries in the HMD that have at least ten years of data before year 2000 (excludes Germany, Croatia and Chile)
- 38 of the 41 countries used = <u>aim to forecast 76 distinct sets of mortality rates</u>

Female Mortality, USA, 1950-2016



Deep neural network

- Categorical inputs to network defined using <u>embedding layers</u> = vector valued step functions of parameters calibrated from input data
- Year input is numerical
- Intermediate layers combine the inputs into new features (128 nodes per layer) using non-linear transformations
- Deep networks hard to optimize => add a skip connection (He, Zhang, Ren *et al.* 2016)



Results

- Results of comparing the models
- Best performing model is deep neural network…
- …produces the best out-of-time forecasts 51 out of 76 times
- for purposes of large scale mortality forecasting, deep neural networks dramatically outperform traditional single and multi-population forecasting models

	Model	Average MSE	Median MSE	Best Performance
1	LC_SVD	5.50	2.48	7
2	LC_ACF_region	3.46	2.50	10
3	ACF_BP	6.12	3.00	4
4	CAE_BP	5.59	3.46	4
5	DEEP	2.68	1.38	51

Agenda

- Introduction
- Machine Learning
- Deep Learning
- Applications in Actuarial Science
- *Discussion and Conclusion*

Discussion

- Emphasis on predictive performance and potential gains of moving from traditional actuarial and statistical methods to machine and deep learning approaches.
- Measurement framework utilized within machine learning – focus on testing predictive performance => focus on measurable improvements in predictive performance led to refinements and enhancements of deep learning architectures
- Learned representations from deep neural networks often have readily interpretable meaning
- Very useful for high-frequency and high-dimensional data

Conclusion

- Deep learning can enhance the predictive power of models built by actuaries
- Application of deep learning techniques to actuarial problems seems to be rapidly emerging field within actuarial science => appears reasonable to predict more advances in the near-term.
- Deep learning is not a panacea for all modelling issues - applied to the wrong domain, deep learning will not produce better or more useful results than other techniques.
- Winter might be coming – if actuaries do not take the lead in applying deep learning, someone else will.

References (1)

Bengio, Y. 2009. "Learning deep architectures for AI", Foundations and trends® in Machine Learning 2(1):1-127.

De Brébisson, A., É. Simon, A. Auvolat, P. Vincent et al. 2015. "Artificial neural networks applied to taxi destination prediction", arXiv arXiv:1508.00021

Dong, W., J. Li, R. Yao, C. Li et al. 2016. "Characterizing driving styles with deep learning", arXiv arXiv:1607.03611

Dong, W., T. Yuan, K. Yang, C. Li et al. 2017. "Autoencoder regularized network for driving style representation learning", arXiv arXiv:1701.01272

Friedman, J., T. Hastie and R. Tibshirani. 2009. The Elements of Statistical Learning : Data Mining, Inference, and Prediction. New York: Springer-Verlag. http://dx.doi.org/10.1007/978-0-387-84858-7.

Gao, G., S. Meng and M. Wüthrich. 2018. Claims Frequency Modeling Using Telematics Car Driving Data. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3102371. Accessed: 29 June 2018.

Gao, G. and M. Wüthrich. 2017. Feature Extraction from Telematics Car Driving Heatmaps. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3070069. Accessed: June 29 2018.

Goodfellow, I., Y. Bengio and A. Courville. 2016. Deep Learning. MIT Press.

Guo, C. and F. Berkhahn. 2016. "Entity embeddings of categorical variables", arXiv arXiv:1604.06737

Hainaut, D. 2018. "A neural-network analyzer for mortality forecast", Astin Bulletin 48(2):481-508.

Hannun, A., C. Case, J. Casper, B. Catanzaro et al. 2014. "Deep speech: Scaling up end-to-end speech recognition", arXiv arXiv:1412.5567

Hejazi, S. and K. Jackson. 2016. "A neural network approach to efficient valuation of large portfolios of variable annuities", Insurance: Mathematics and Economics 70:169-181.

Hejazi, S. and K. Jackson. 2017. "Efficient valuation of SCR via a neural network approach", Journal of Computational and Applied Mathematics 313:427-439.

Krizhevsky, A., I. Sutskever and G. Hinton. 2012. "Imagenet classification with deep convolutional neural networks," Paper presented at Advances in Neural Information Processing Systems. 1097-1105.

Kuo, K. 2018. "DeepTriangle: A Deep Learning Approach to Loss Reserving", arXiv arXiv:1804.09253

References (2)

Makridakis, S., E. Spiliotis and V. Assimakopoulos. 2018. "The M4 Competition: Results, findings, conclusion and way forward", International Journal of Forecasting

Mitchell, T. 1997. Machine learning. McGraw-Hill Boston, MA.

Noll, A., R. Salzmann and M. Wüthrich. 2018. Case Study: French Motor Third-Party Liability Claims. SSRN. https://ssrn.com/abstract=3164764 Accessed: 17 June 2018.

Richman, R. 2018. Thoughts on the M4 Conference. http://ronaldrichman.co.za/2018/12/17/thoughts-on-the-m4-conference/. Accessed: 20 March 2019.

Richman, R. and M. Wüthrich. 2018. "A Neural Network Extension of the Lee-Carter Model to Multiple Populations", Available at SSRN: https://ssrn.com/abstract=3270877

Schreiber, D. 2017. The Future of Insurance. DIA Munich 2017: https://www.youtube.com/watch?time_continue=1&v=LDOhFHJqKqI. Accessed: 17 June 2018.

Shmueli, G. 2010. "To explain or to predict?", Statistical Science:289-310.

Smith, M., F. Beyers and J. De Villiers. 2016. "A method of parameterising a feed forward multi-layered perceptron artificial neural network, with reference to South African financial markets", South African Actuarial Journal 16(1):35-67.

Sutton, R. and A. Barto. 2018. Reinforcement learning: An introduction, Second Edition. MIT Press.

Wijnands, J., J. Thompson, G. Aschwanden and M. Stevenson. 2018. "Identifying behavioural change among drivers using Long Short-Term Memory recurrent neural networks", Transportation Research Part F: Traffic Psychology and Behaviour 53:34-49.

Wu, Y., M. Schuster, Z. Chen, Q. Le et al. 2016. "Google's neural machine translation system: Bridging the gap between human and machine translation", arXiv arXiv:1609.08144

Wüthrich, M. 2018a. "Neural networks applied to chain–ladder reserving", European Actuarial Journal 8(2):407-436.

Wüthrich, M. 2018b. v-a Heatmap Simulation Machine. https://people.math.ethz.ch/~wueth/simulation.html. Accessed: 1 July 2018.

Wüthrich, M. and C. Buser. 2018. Data analytics for non-life insurance pricing. Swiss Finance Institute Research Paper. https://ssrn.com/abstract=2870308. Accessed: 17 June 2018.

Wüthrich, M.V. 2017. "Covariate selection from telematics car driving data", European Actuarial Journal 7(1):89-108.

Zarkadoulas, A. 2017. "Neural network algorithms for the development of individual losses." Unpublished thesis, Lausanne: University of Lausanne.

Get involved

- Insurance Data Science conference - 14 June 2019
- ETH Zurich
- https://insurancedatascience.org/
- Amazing line-up of papers, presentations and speakers!

- Kasa.ai – launching soon, led by Kevin Kuo of Rstudio
- An open research group encouraging innovation in insurance analytics
- Some interesting projects planned

Thanks for listening - Any questions?

Paper: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3218082

Contact: ron@ronaldrichman.co.za

IAA AAI SECTION COLLOQUIUM 2019

THE MODERN ACTUARY - CHALLENGE • INFLUENCE • LEAD

ASTIN • IAAHS • IAALS • IACA • PBSS

2 - 5 April
2019
Cape Town
South Africa
CTICC

www.colloquium2019.org.za

Hosted by

ACTUARIAL SOCIETY OF SOUTH AFRICA