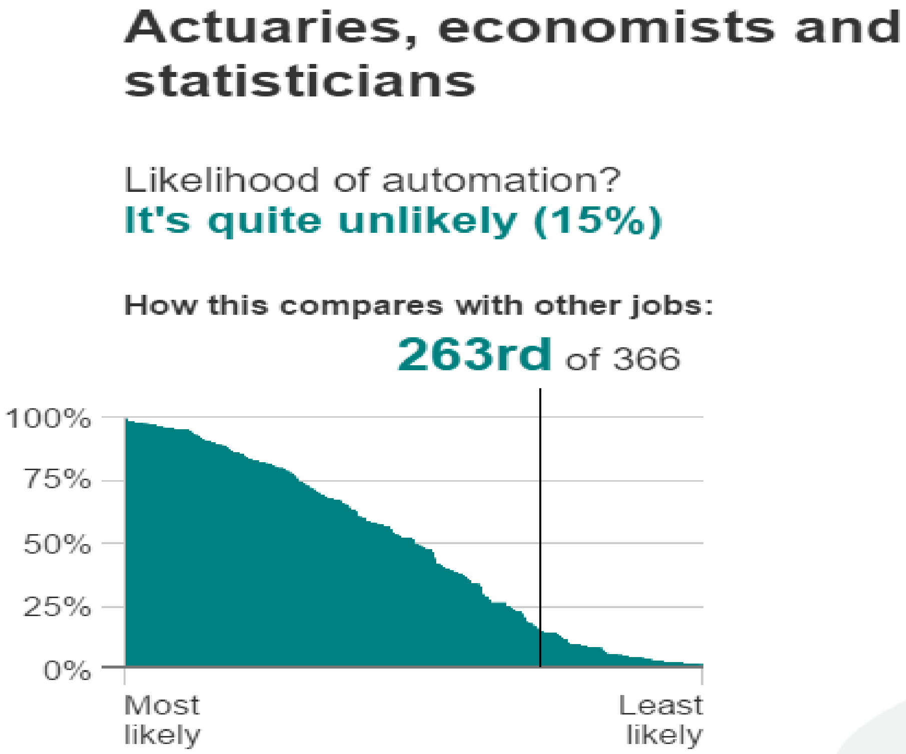
Institute
and Faculty
of Actuaries

Penalised Regression

Tony Ward
tony@statcore.co.uk

06 June 2016

Will a Robot Take your job in the next 20 years?



Will a Data Scientist take your job in the next 20 years?

UNSTRUCTURED DATA

MACHINE LEARNING

CLOUD COMPUTING

Today we will cover..

- How to improve the accuracy of predictive models using unstructured data

- How to replace an entire risk pricing modelling team with one Data Scientist

Institute
and Faculty
of Actuaries

How to improve the accuracy of predictive models using unstructured data

06 June 2016

Motivating Example: Salary Prediction

- Web Scraped 28,000 Jobs from www.theactuaryjobs.com
- Want to predict the salary based on the data provided in the advert.
- Data is both structured (sector, experience) and unstructured (job title, further info)

Job 1

Sign in or Create account | Recruiters

TheActuary Jobs

Institute and Faculty of Actuaries

Home | Find a job | Job alerts | Search recruiters | Careers advice | www.theactuary.com

Global insurance company with Lloyds syndicate is looking to hire a pricing actuary reporting directly into the Group pricing actuary you will have experience of pricing and be looking to lead a small team based in London. The business covers a variety of lines of business including marine, aviation, energy, property and liability. You will work closely with the underwriters, present to senior management and underwriters and own and manage rate adequacy and rate movement MI for international specialty lines of business. Above and beyond your pricing skills you must be able to demonstrate leadership qualities and have excellent communication skills

Head of International Pricing

Recruiter	[redacted]
Location	London
Salary	Circa £130K
Posted	17 Jun 2015
Closes	17 Jul 2015
Ref	ZB375
Sector	General insurance
Experience	Nearly qual (11+ exams), Qualified
Contract Type	Permanent
Hours	Full Time

Job 2

Our client, an investment consultancy, is looking to hire a research analyst.
You will work on investment manager research projects. Working closely with the investment consulting team, you will conduct research on multi-asset fund managers and equity fund managers. Candidates are expected to have experience in researching investment managers and in interest in financial markets and funds. You must be a good team player with good communication skills and be able to present to the investment committee.
Our client offers study support (e.g. CFA or actuarial qualifications)

Research Analyst

Recruiter	[redacted]
Location	London
Salary	£35k - 40k + bonus + benefits
Posted	18 May 2015
Closes	18 Jun 2015
Ref	Research Analyst
Sector	Investment, Management consultancy, Pensions
Experience	Graduate/post-graduate, Nearly qual (11+ exams), Other, Part qual (1-10 exams)
Contract Type	Permanent
Hours	Full Time

Some roles pay less than other!

They want volunteers?

Pricing Actuary - Retail

Recruiter	[redacted]
Location	London
Salary	£0 per annum
Posted	23 Dec 2013
Closes	20 Jan 2014
Ref	JK10399
Contact	[redacted]
Sector	General insurance
Experience	Nearly qual (11+ exams), Qualified
Contract Type	Permanent
Hours	Full Time

Setting up the Problem

- **TASK:** To predict the probability of a job paying more than £70,000
- **INPUT VARIABLES**
  - Sector: *General Insurance, Pensions,...*
  - Experience: *Graduate/post-graduate, Part qual (1-10 exams),..*
  - City
  - County
  - Job title e.g. *Head of International Pricing*
  - Further Info e.g. *Global insurance company with Lloyds syndicate is looking to hire a pricing actuary reporting directly…*

How to handle unstructured data

- Turn the unstructured data into a structured tabular form (a Document Term Matrix).

Job Title	Analyst	Actuari	Head	International	Pricing	Research	Retail
Head of International Pricing			1	1	1		
Research Analyst	1					1	
Pricing Actuary - Retail		1			1		1

- The presence/absence of each word are the potential predictors in our model
- Do this for both “Job Title” and “Further Info” text items
- Now our modelling dataset is very wide!!

Modelling Dataset

Rating Factor	Levels
Experience	7
Sector	14
Country	89
City	99
Job Title	189
Further Info	691

- Over 1000 different categorical levels to consider
- Fit a GLM with binomial error term (Logistic Regression).

Statistical Modelling - process

- One way tables
- Correlated Variables
- Forward / Backward Selection
- Statistical Tests of Coefficients
- Likelihood Ratio Test of Nested Models
- Interaction with Time
- Interaction with Random Factors
- Check Residuals

Time consuming!

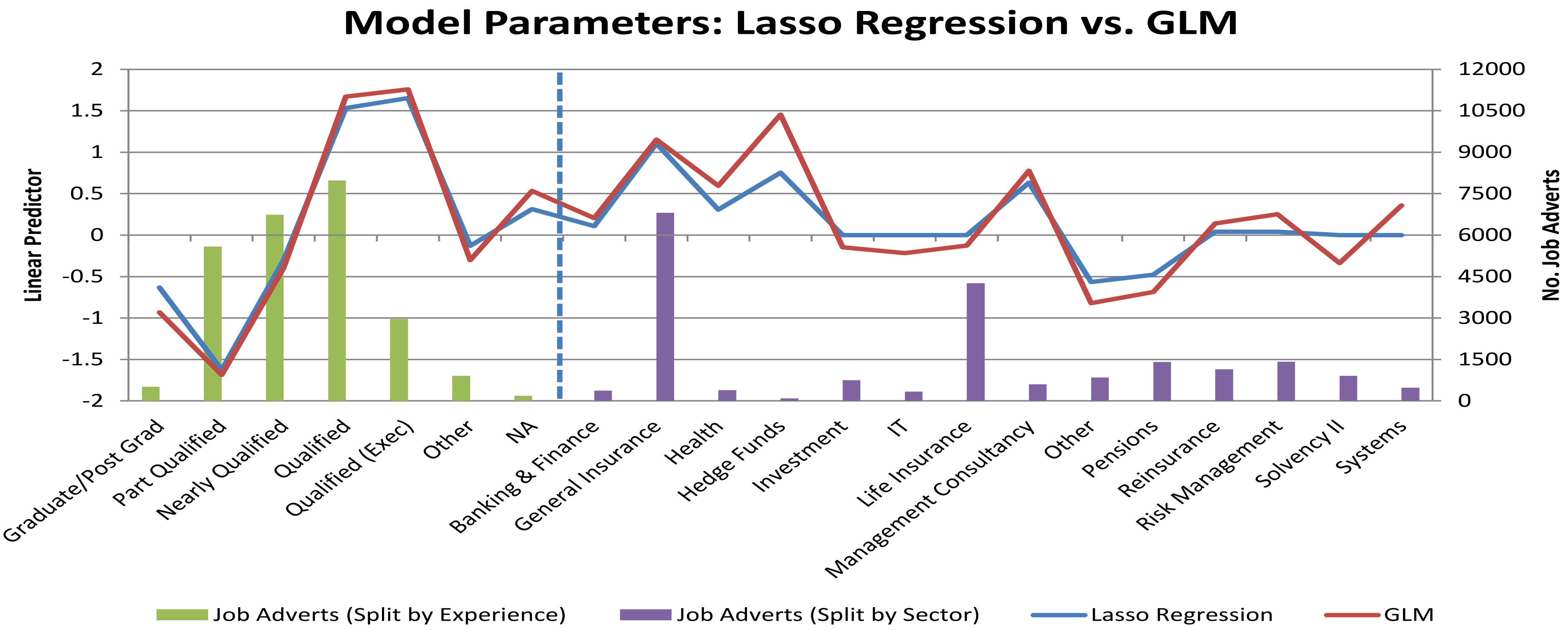
Penalised Regression

- Penalised Regression refers to the family of methods for building Generalised Linear Models using Regularisation.
- Also known as “Shrinkage Methods” since they shrink model coefficients towards zero.
- Methods include Lasso Regression, Ridge Regression and Elastic Net.
- Widely used in the machine learning community
- Used in insurance?

Penalised Regression

- Same model form as traditional GLM.
  - Error Term (Poisson, Gamma, Binomial)
  - link function (log, logit)
  - Offsets
  - Relativities
- Key difference is how we build the model
- Same process as with other machine learning algorithms: k-fold cross validation to optimise the model hyper-parameters

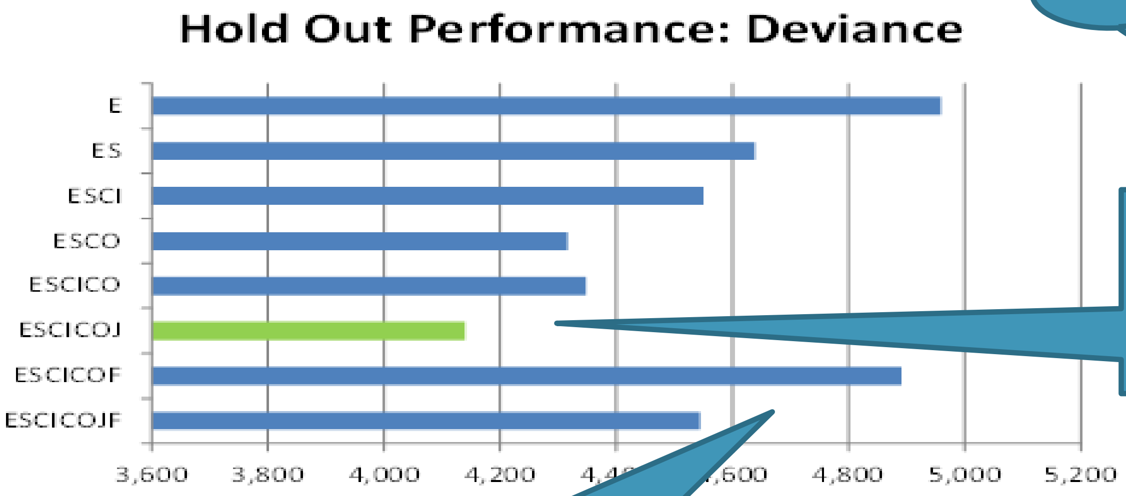
Lasso Example: Salary Prediction



Candidate Models

Model Name	Factors Included
S	Sector
E	Experience
ES	Experience, Sector
ESCI	Experience, Sector, City
ESCICO	Experience, Sector, City, Country
ESCICOJ	Experience, Sector, City, Country, **Job Title**
ESCICOF	Experience, Sector, City, Country, **Further Info**
ESCICOJF	Experience, Sector, City, County, **Job Title, Further Info**

Results



Jobs containing “chief” “head” “expert” “director” got higher predictions

Model ESCICOJ which included “Job Title” performed the best

Jobs containing “junior” “trainee” “student” received lower predictions

Adding “Further Info” damaged performance

Learnings

- Unstructured data can boost predictive performance
- Penalised Regression isn’t a silver bullet – throwing all your data at it doesn’t work.
- I’m glad I work in General Insurance and not Pensions!

Institute
and Faculty
of Actuaries

How does Lasso Regression work?

06 June 2016

Traditional GLM: Maximum Likelihood

- The GLM relativities are found by maximum likelihood.
- In practice we minimise the negative of the log likelihood

$$\min_{\beta_0,\beta}\left(-l(\beta|X,Y)\right)$$

β_0	Intercept
β	Relativities
$l(\beta\|X,Y)$	Log Likelihood

Lasso Regression

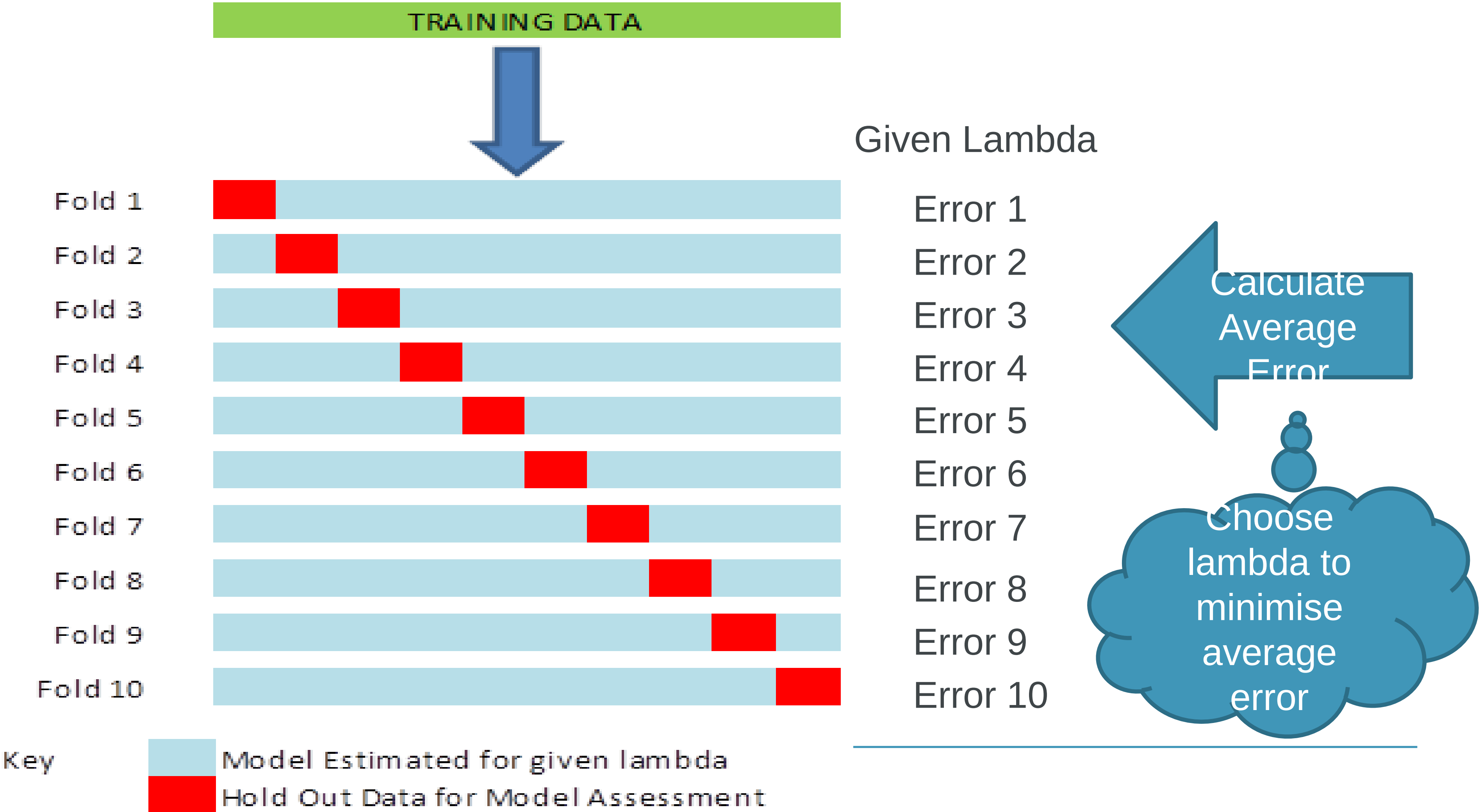
- Lasso Regression fits the GLM subject to the following constraint

$$\sum_{j=1}^{p} |\beta_j| \leq t$$

- Models ar $\min_{\beta_0,\beta}\left(-\frac{1}{N} l(\beta|X,Y) + \lambda \sum_{i=1}^{N} |\beta_i|\right)$ lihood

Penalty Term Lambda

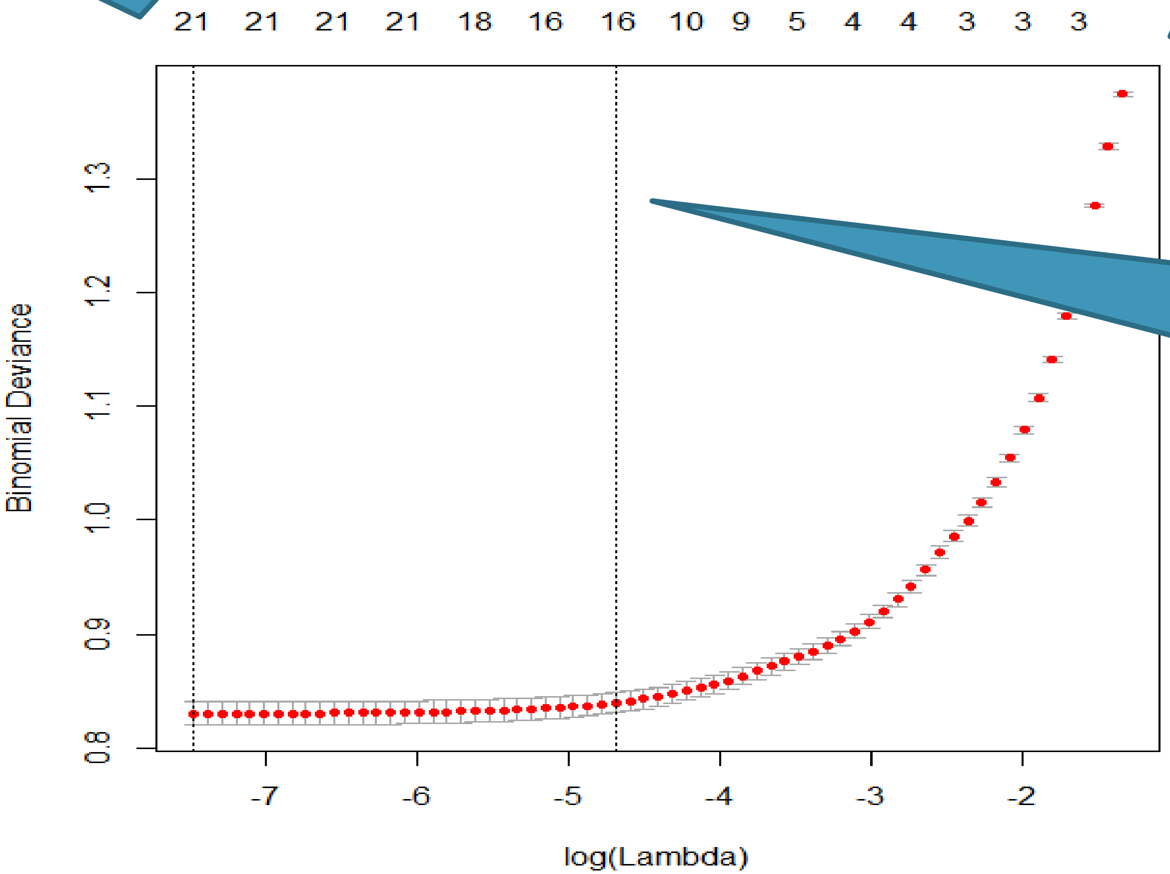
- We need to find the best λ to use.
  - Big lambda = Strong Penalty >> Most coefficients set to zero
  - Small lambda = Weak Penalty >> Approaching Traditional GLM
- To select λ we split the data into training and test. Using the training data we test a range of λ values using 10-Fold Cross Validation.
- Lambda is chosen to minimise the cross validation error.


TRAINING DATA
Given Lambda
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Fold 6
Fold 7
Fold 8
Fold 9
Fold 10
Error 1
Error 2
Error 3
Error 4
Error 5
Error 6
Error 7
Error 8
Error 9
Error 10
Calculate Average Error
Choose lambda to minimise average error
Key
Model Estimated for given lambda
Hold Out Data for Model Assessment

Small Lambda / Weak Penalty

Coefficients enter model

Big Lambda / Strong Penalty



At this penalty we have 16 coefficients in the model – the rest are zero

Institute
and Faculty
of Actuaries

How to replace an entire risk pricing modelling team with one Data Scientist

06 June 2016

Statistical Modelling :

- Three or four analysts per product.
- Each person can only build one model at once

Home Risk Pricing Team

FLOOD FIRE THEFT AD

Motor Risk Pricing Team

TPD TPI AD THEFT

- The only way to speed this up is to hire more staff
- This doesn’t remove the endless debate around borderline decisions.

You are the bottleneck!

Machine Learning

FLOOD

FIRE

THEFT

AD

TPI

AD

TPD

THEFT

- Model building is automated therefore many models can be built in parallel
- No more debate about borderline decisions – remember the hedge fund example
- Quicker Speed to Market
- Models can be continuously updated as new data arrives.
- **Now the bottleneck is IT!**

Success criteria

- Speed to Market
- Interpretation
- Implementation
- Accuracy

Challenging!

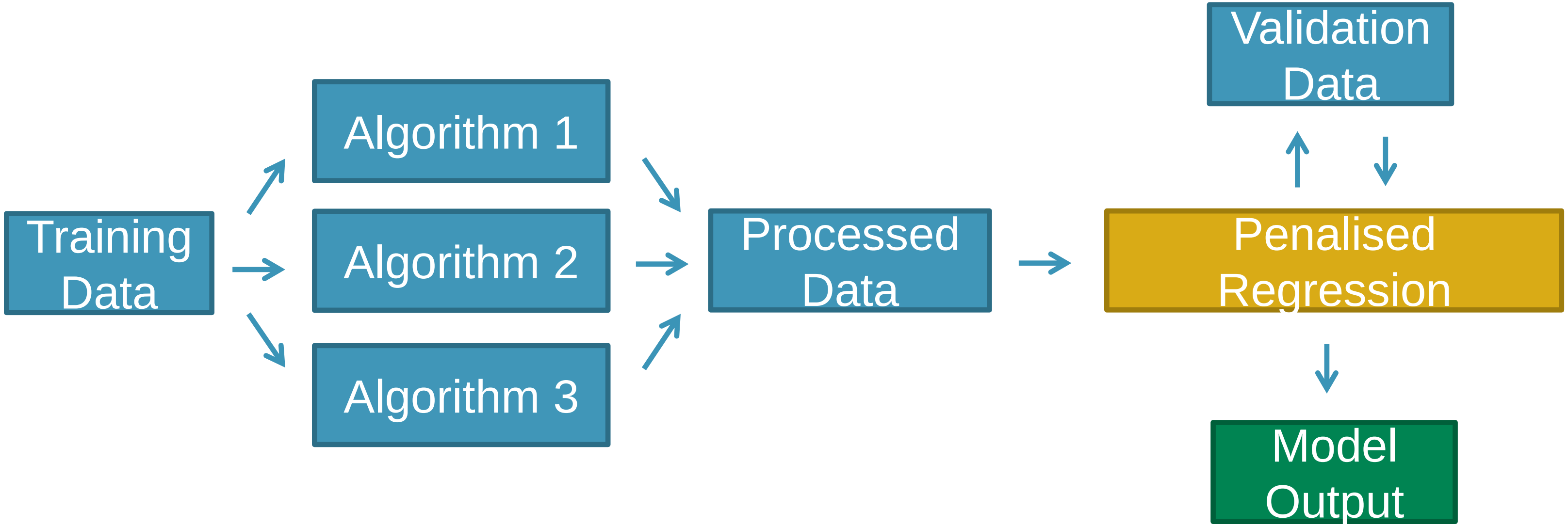
Challenges

- Throwing all data at Penalised Regression doesn’t work.
- Data sets are big!
- Numerical Inputs
- Interactions

Advantages

- How much data should we include?
- 2 Accident Years?
- 3 Accident Years?
- Now we can make this decision scientifically – test them all!

My Solution – a machine learning pipeline



Success criteria

- Speed to Market
- Interpretation
- Implementation
- Accuracy

Summary

- How unstructured data can improve models
- Introduction to Lasso Regression
- How to replace a risk pricing team with one Data Scientist

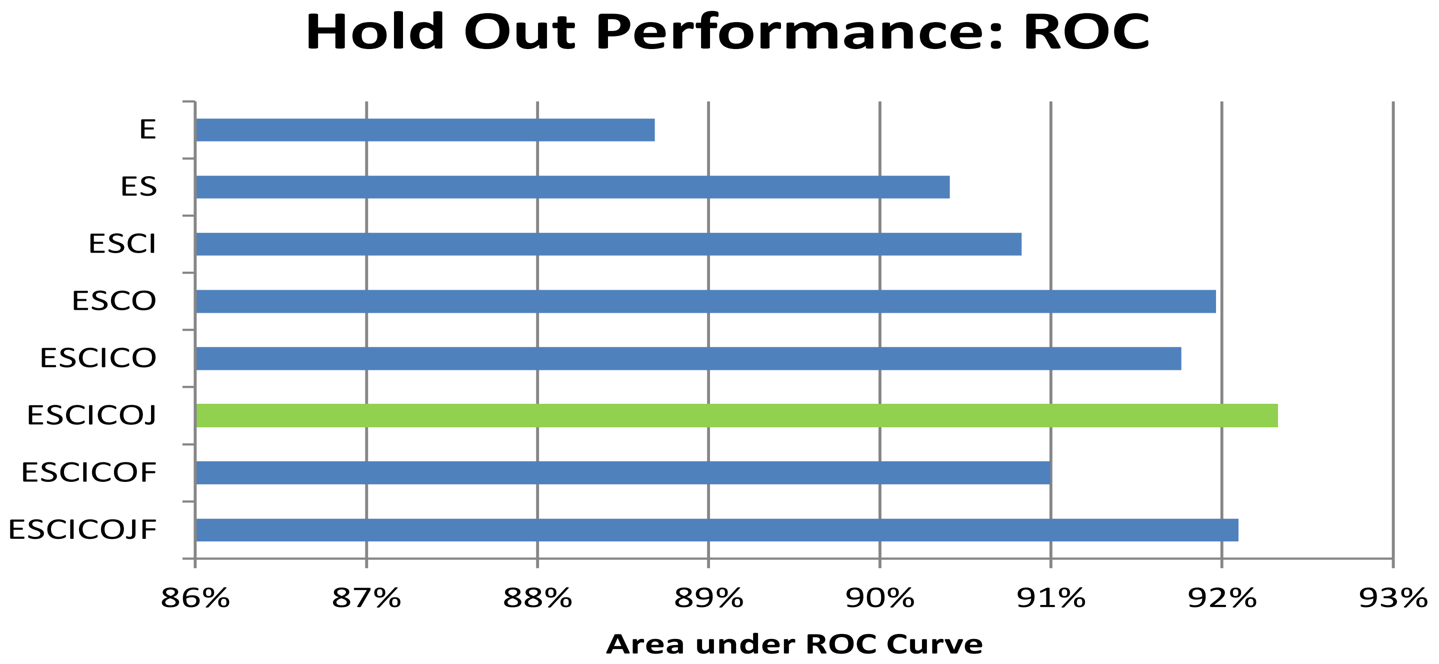
Thank you!

Tony Ward

www.statcore.co.uk

tony@statcore.co.uk

Appendix: ROC chart



Appendix: Ridge Regression

- Ridge Regression fits the GLM subject to the following constraint

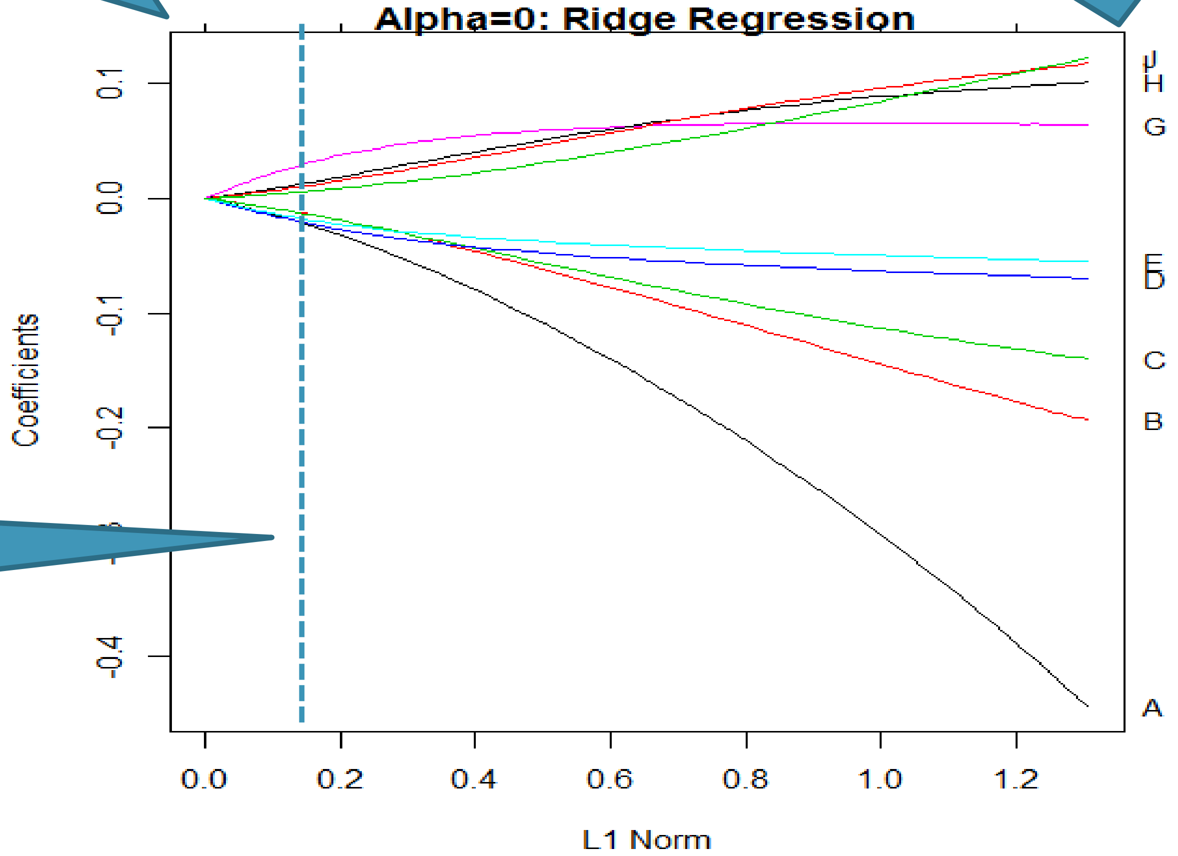
$$\sum_{j=1}^{p} \beta_j^2 \leq t$$

- Models are $\min_{\beta_0,\beta} \left(-\frac{1}{N} l(\beta|X,Y) + \lambda \sum_{i=1}^{N} \frac{\beta_i^2}{2} \right)$ hood

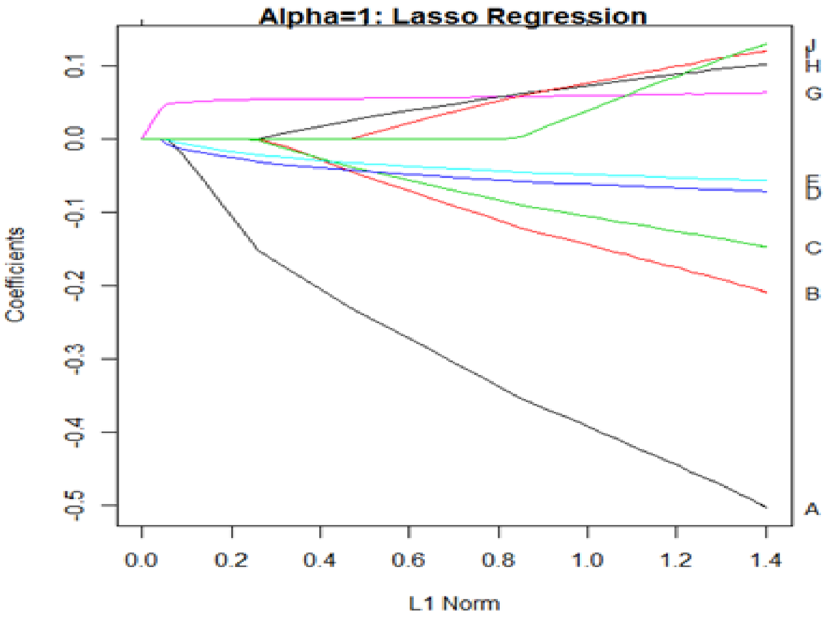
Strong Penalty > All coefficients are zero

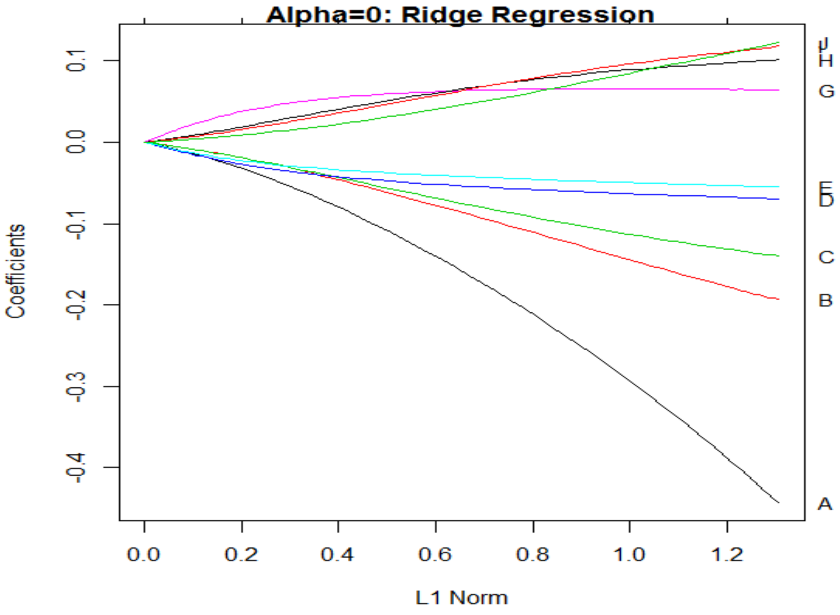
Weak Penalty > All coefficients non - zero

All coefficients quickly enter the model



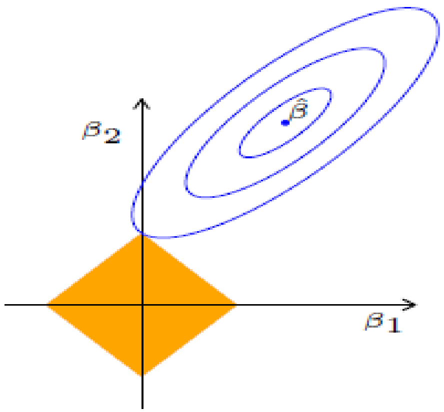
Lasso vs Ridge

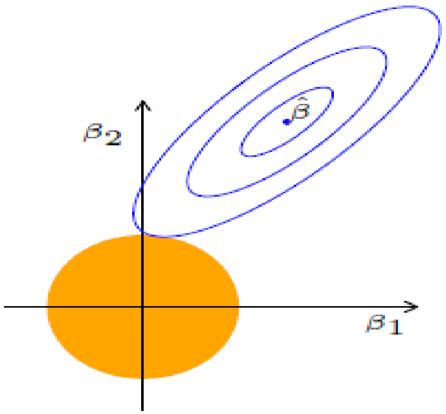




Appendix: Lasso vs Ridge – Part 1

- To understand why the lasso shrinks coefficients to zero and the ridge does not consider the following. The chart depicts the lasso (left) and ridge regression (right) when there are only two parameters





Appendix: Elastic Net

- The elastic-net penalty is controlled by α, and bridges the gap between Lasso (α=1) and Ridge (α=0). The tuning parameter λ controls the overall strength of the penalty

- Models are estimated by penalised maximum likelihood

$$\min_{\beta_0,\beta}\left(-\frac{1}{N}l(\beta|X,Y)+\lambda\left((1-\alpha)\sum_{i=1}^{N}\frac{\beta_i^2}{2}+\alpha\sum_{i=1}^{N}|\beta_i|\right)\right)$$

Appendix: Lasso v Ridge v Elastic Net