The Predictive Analytics & Futurism Section Presents

Practical Predictive Analytics Seminar

May 22, 2019 | Tampa Marriott Water Street | Tampa, FL

Presenters:

Talex Diede, MS

Jean-Marc Fix, FSA, MAAA

Brian D. Holland, FSA, MAAA

Ben Johnson, MS

Matthias Kullowatz, ASA, MAAA, SMS

Marshall Lagani, MA

Best practices

Jean-Marc Fix and Matthias Kullowatz

May 22, 2019

SOCIETY OF ACTUARIES®

SOA Antitrust Compliance Guidelines

Active participation in the Society of Actuaries is an important aspect of membership. While the positive contributions of professional societies and associations are well-recognized and encouraged, association activities are vulnerable to close antitrust scrutiny. By their very nature, associations bring together industry competitors and other market participants.

The United States antitrust laws aim to protect consumers by preserving the free economy and prohibiting anti-competitive business practices; they promote competition. There are both state and federal antitrust laws, although state antitrust laws closely follow federal law. The Sherman Act, is the primary U.S. antitrust law pertaining to association activities. The Sherman Act prohibits every contract, combination or conspiracy that places an unreasonable restraint on trade. There are, however, some activities that are illegal under all circumstances, such as price fixing, market allocation and collusive bidding.

There is no safe harbor under the antitrust law for professional association activities. Therefore, association meeting participants should refrain from discussing any activity that could potentially be construed as having an anti-competitive effect. Discussions relating to product or service pricing, market allocations, membership restrictions, product standardization or other conditions on trade could arguably be perceived as a restraint on trade and may expose the SOA and its members to antitrust enforcement procedures.

While participating in all SOA in person meetings, webinars, teleconferences or side discussions, you should avoid discussing competitively sensitive information with competitors and follow these guidelines:

- -Do not discuss prices for services or products or anything else that might affect prices
- -Do not discuss what you or other entities plan to do in a particular geographic or product markets or with particular customers.
- -Do not speak on behalf of the SOA or any of its committees unless specifically authorized to do so.
- -Do leave a meeting where any anticompetitive pricing or market allocation discussion occurs.
- -Do alert SOA staff and/or legal counsel to any concerning discussions
- -Do consult with legal counsel before raising any matter or making a statement that may involve competitively sensitive information.

Adherence to these guidelines involves not only avoidance of antitrust violations, but avoidance of behavior which might be so construed. These guidelines only provide an overview of prohibited activities. SOA legal counsel reviews meeting agenda and materials as deemed appropriate and any discussion that departs from the formal agenda should be scrutinized carefully. Antitrust compliance is everyone's responsibility; however, please seek legal counsel if you have any questions or concerns.

Presentation Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented. Attendees should note that the sessions are audio-recorded and may be published in various media, including print, audio and video formats without further notice.

Agenda

- CRISP-DM methodology
- Business understanding
- Data understanding
- Data quality
- ASOP 23
- Bias
- Documentation
  - Reproducibility
  - Version control

Cross Industry Process for Data Mining

- Business understanding
- Data understanding
- Data preparation
- Modeling
- Evaluation
- Deployment

Business Understanding

- What is the business problem you are trying to solve?
  - Business objective
  - Business success criteria
- Where are you now?
  - Assess current situation
  - Costs and benefits to move to the new structure
- Only then assess the technical goals (e.g. desired accuracy, speed...)

Data Understanding

- The data is key so over-document
  - Data collection
  - Data description
  - Data exploration
  - Data quality

Data Quality

- Emphasis on measures and acceptability threshold
- Completeness: % complete, relative to what
- Uniqueness: # real world/# in dataset
- Timeliness: time between occurrence and recording
- Validity: how much of the data is in the correct form/format
- Accuracy: data represents what it is supposed to be
- Consistency: with other data sets
- Implied: data is safe, legal and appropriately protected

Source: The Six Primary Dimensions for Data Quality Assessment
https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf

ASOP 23: Data

- Reliance
- Review of data for reasonableness and consistency
  - Definitions (of the data elements)
  - Identify questionable data values
  - Review of prior data (maybe less relevant in our context)
- Use of Data (as is, enhance, judgment, audit, inadequate)
- Professional judgment!

Bias

- Watch out for bias leading to unfair discrimination
- Bias can be due to not understanding correlations
- Harder to detect in machine learning approaches
- One way to guard against bias is to verify that explicitly adding the variable does not change your model results
- Generally due to problems in training set

Missing data

- Very common in “real world” data
- Missingness by columns (features) vs. rows (observations):
  - Guides how one might deal with the missing values
- A few solutions:
  - Remove observations/columns with many missing values
  - Impute all missing values in a particular column with a constant, and include a missing value indicator feature
  - Impute all missing values in a particular column using a function of other predictor variables

Reproducibility and version control

- Reproducibility: save workflows in scripts for easy replication
- R Projects help organize related workflows and data sources
- Version control is more than just saving workflows
  - Compare older and newer versions of documents to assess what changed, when it changed, and who messed everything up!
  - Split out current tasks into distinct versions that can be merged later; great for code testing and review (QA)
- Git/GitHub example later

Practical Predictive Analytics Seminar

Matthias Kullowatz
Ben Johnson

Session 3: Predictive Models in Life and Annuities

May 22, 2019

SOCIETY OF ACTUARIES

Theory

SOCIETY OF ACTUARIES

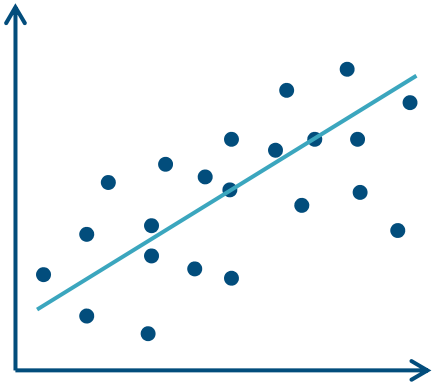
Agenda

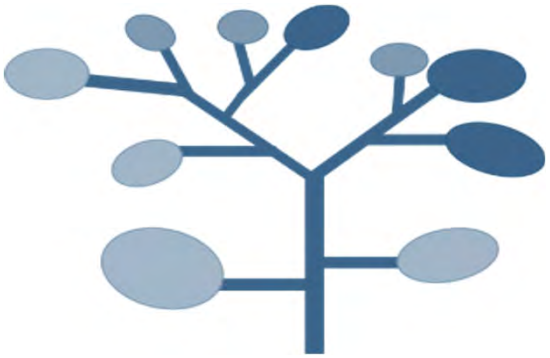
- Questions of interest for life and annuity products
- Logistic regression theory and application
- Associated theoretical concerns that may arise in the modeling process
- Model validation
- Hands-on time throughout!

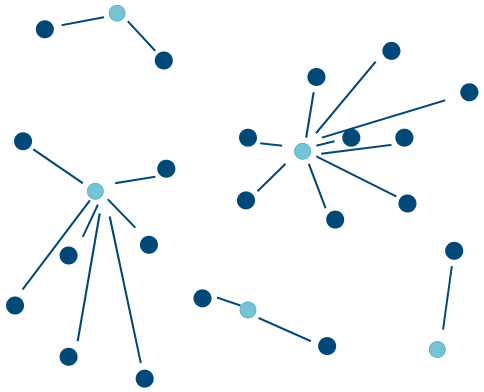
Questions of interest

- When will a policyholder…
  - Lapse?
  - Partially withdraw?
  - Die?
- How will a policyholder utilize the policy?
- What drives these “behaviors” and why?
- Are the findings implementable?

Predictive model forms



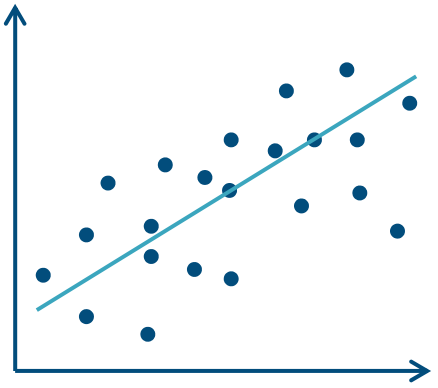




Icon made by Freepik from www.flaticon.com

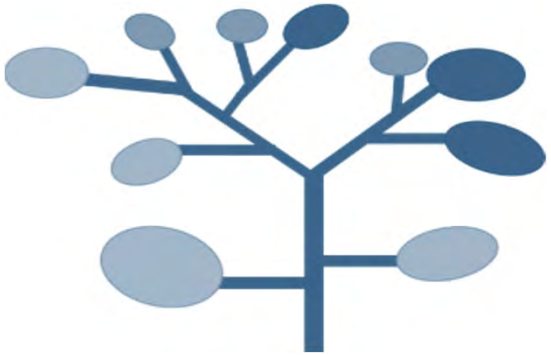
Regression

- OLS, GLM, regularization (ridge, lasso, elastic net)
- Pros
  - Quick fitters
  - Interpretable coefficients and output
  - Harder to overfit
  - Widely used
- Cons
  - Constrained by parametric, functional form
  - Multicollinearity issues



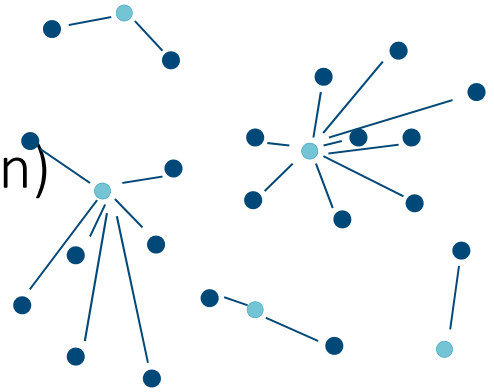
Tree-based models

- Decision trees, random forest, GBM
- Pros
  - Inherently models interactions between drivers
  - Models relationships non-parametrically
- Cons
  - Black-box formula
  - Doesn't interpolate or extrapolate well



Clustering, et. al.

- Supervised: k-nearest neighbors
- Unsupervised: K-means, hierarchical
- Pros
  - Reduces dimensionality (ease of interpretability)
  - Easy to explain predictions (k-nearest neighbors)
- Cons
  - Sensitive to outliers
  - Reduces dimensionality (loss of information)



Neural networks

- Pros
  - Inherent interaction effects/non-parametric
  - Well-suited for problems with many predictor variables
    - Image recognition and text analysis-type problems
- Cons
  - Black-box formula (even more opaque than GBM/RF)
  - Estimate uncertainty harder to measure
  - Computationally intensive

Icon made by Freepik from www.flaticon.com

Other modeling methods/techniques

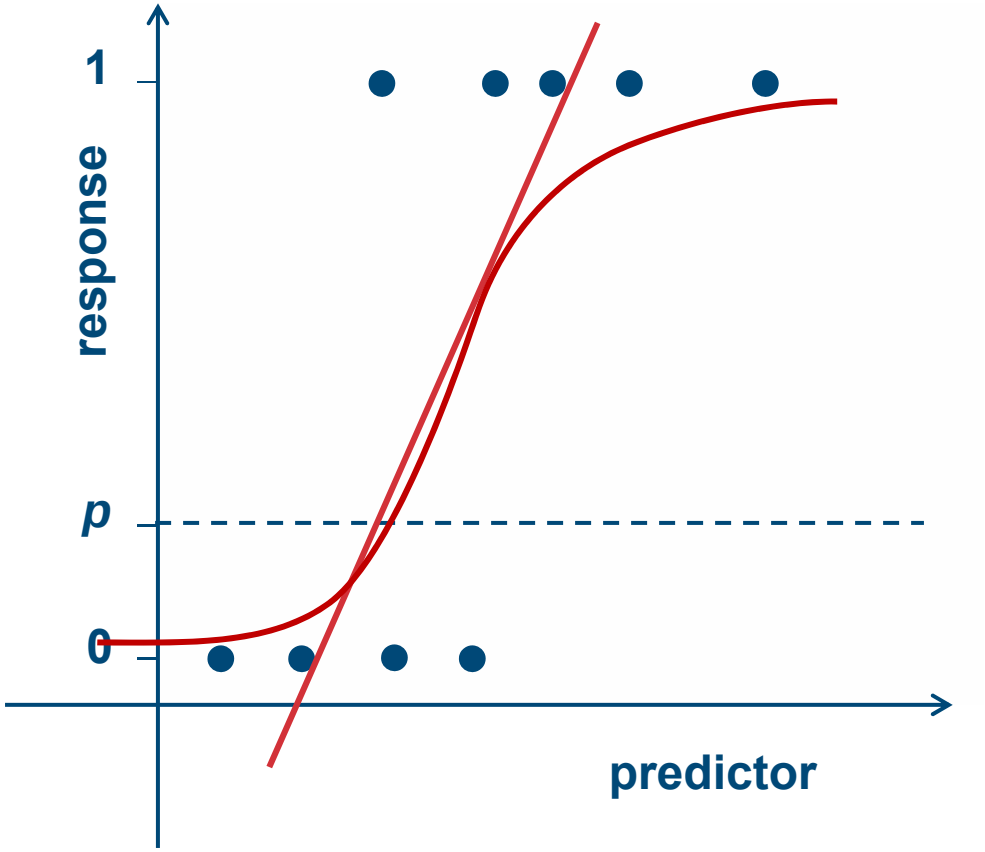
- Survival models
  - Cox proportional hazards
  - Accelerated failure time
- Support vector machines
- Agent-based modeling
- Splines (with regularization)

Icon made by Freepik from www.flaticon.com

Logistic GLM

- For predicting probabilities of binary outcomes
- Link function provides much needed flexibility
- Predictor variables can be quantitative or qualitative

Why a link function?

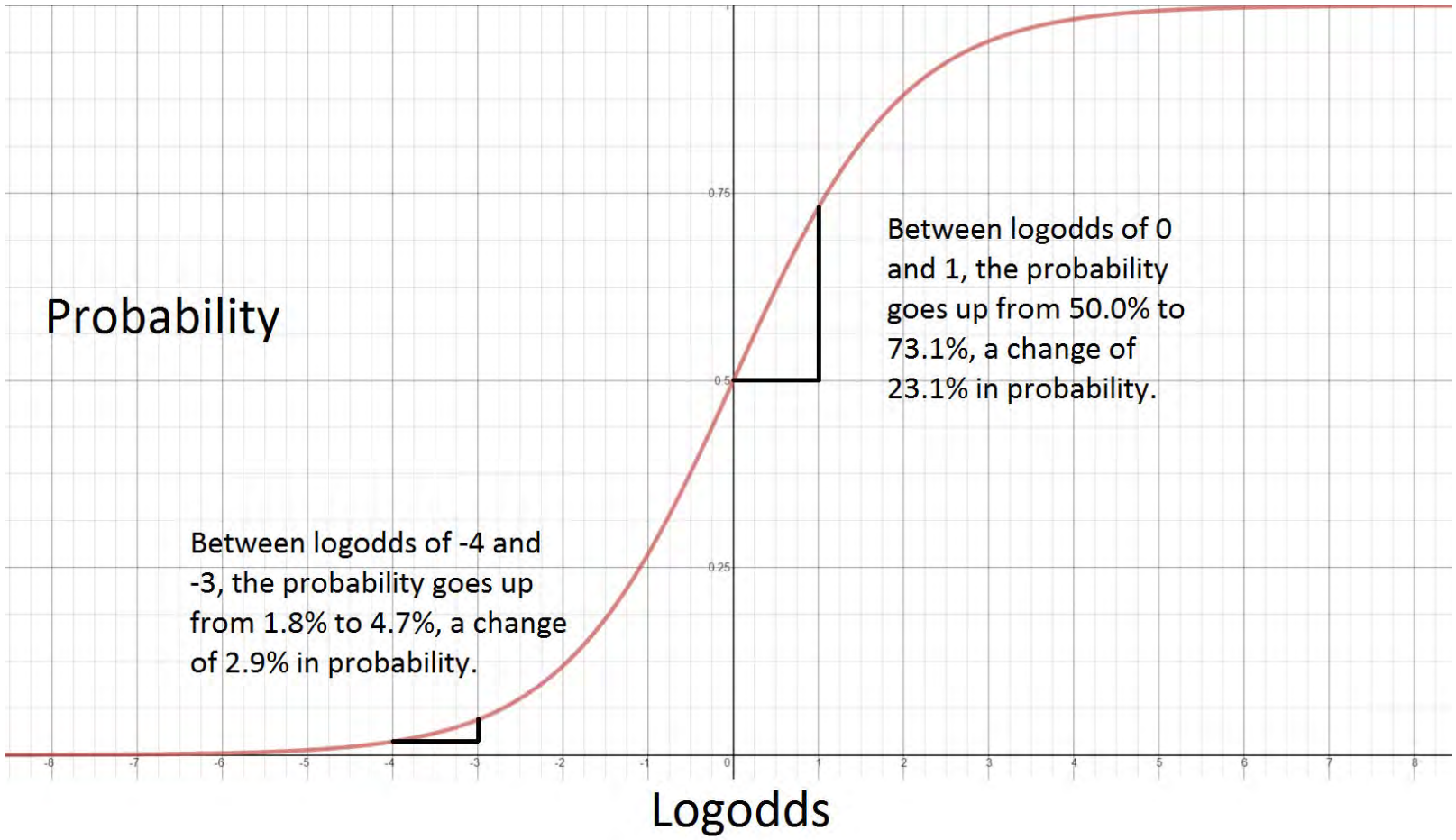


The logistic function

- $\hat{y} = g(L) = \frac{e^L}{1+e^L}$
  - $L = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \widehat{\beta_2} x_2 + \cdots + \widehat{\beta_p} x_p$
  - $\lim_{L \to \infty} g(L) = 1 \; and \; \lim_{L \to -\infty} g(L) = 0$

- $g^{-1}(\hat{y}) = \ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = L$
  - Logit function ("logodds")

Consequences of logit link



Interpretation of coefficients

- $\ln\left(\frac{\hat{y}(x)}{1-\hat{y}(x)}\right) = \widehat{\beta_0} + \widehat{\beta_1}x \Rightarrow \frac{\hat{y}(x)}{1-\hat{y}(x)} = e^{\widehat{\beta_0}+\widehat{\beta_1}x}$
- Continuous x-value:
  - $\frac{\hat{y}(x+1)}{1-\hat{y}(x+1)} \div \frac{\hat{y}(x)}{1-\hat{y}(x)} = \frac{e^{\widehat{\beta_0}+\widehat{\beta_1}(x+1)}}{e^{\widehat{\beta_0}+\widehat{\beta_1}x}}$

$$= e^{\widehat{\beta_1}}$$

- Odds ratio

Theoretical extras

- Independent observations
- The model is fit by maximizing the following:

$$loglikelihood = \sum [Y_i \ln(\hat{y}_i) + (1 - Y_i)\ln(1 - \hat{y}_i)]$$

  - $AIC = -2 \times loglikelihood + 2 \times parameters$
  - $BIC = -2 \times loglikelihood + \ln(N) \times parameters$

Hands-on: Fit logistic GLM in R!

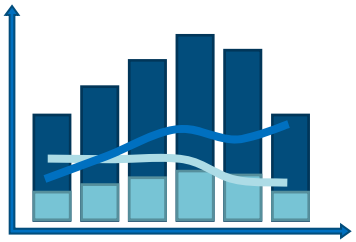
SOCIETY OF ACTUARIES

Practical concerns

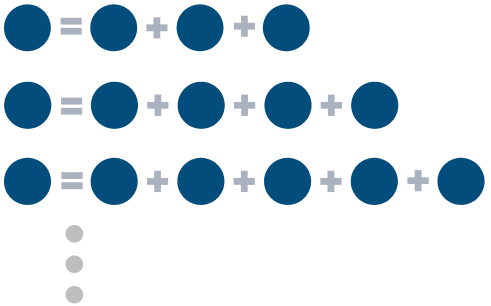
SOCIETY OF ACTUARIES

Predictive analytics process

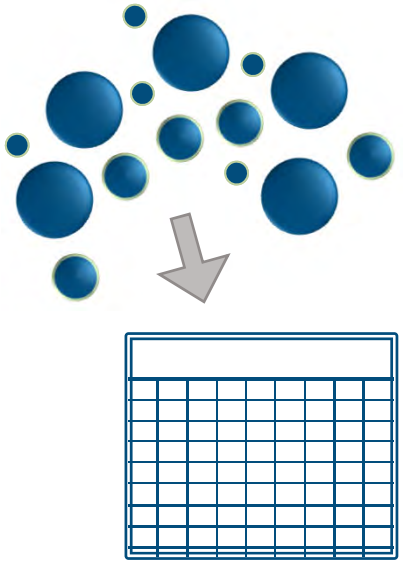
Exploratory Analysis

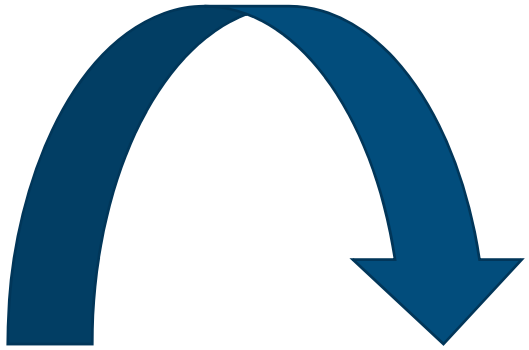


Modeling

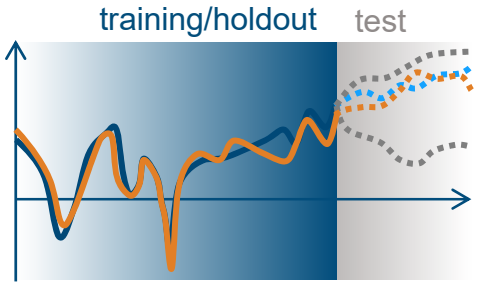


Data Prep





Validation



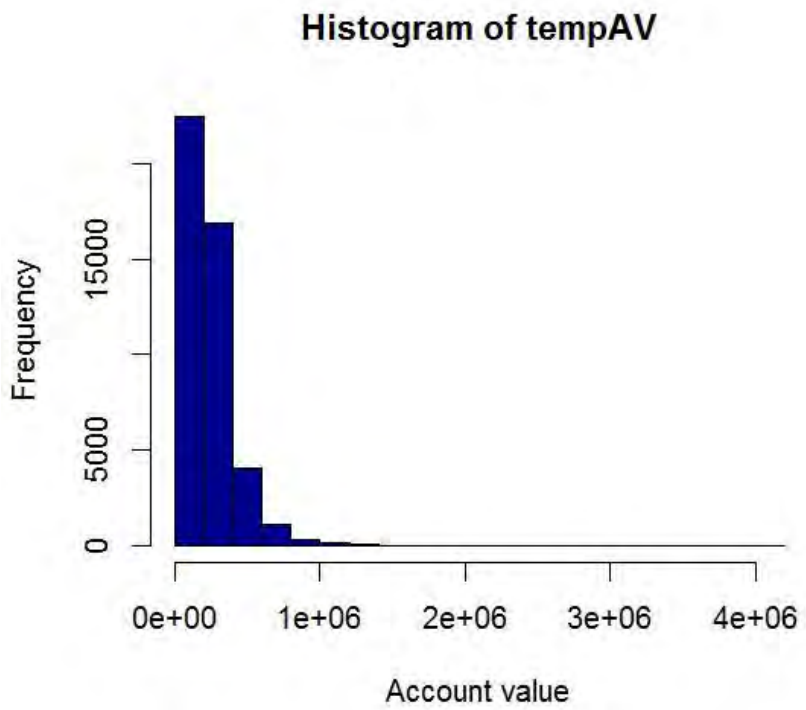
Practical concerns: Data

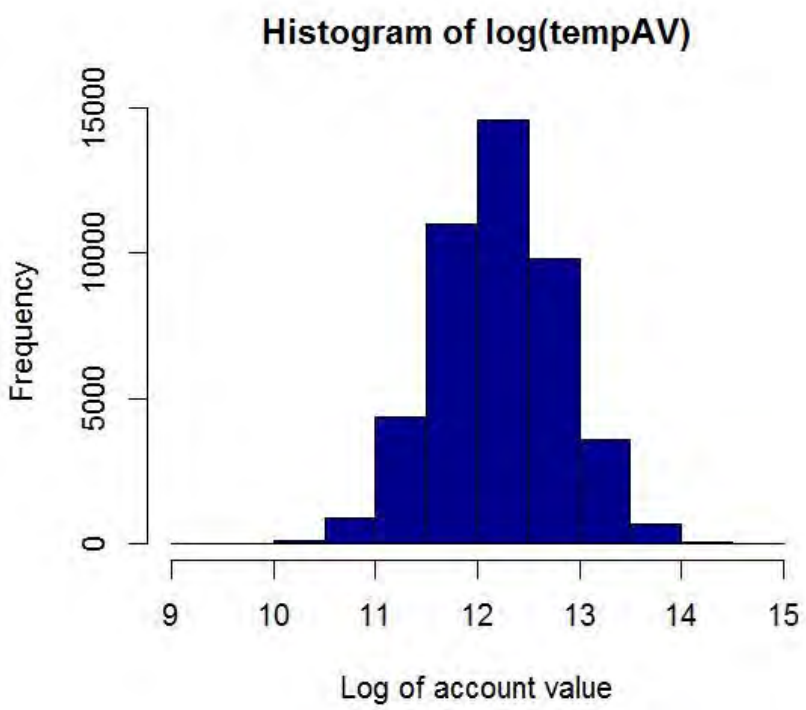
- Formatting variables (1)
- Identifying and dealing with outlier data values (2)
- Accounting for missing data (2)
- Derive new variables for modeling (3)
- Compile dataset into appropriate format (4)

Practical concerns: Modeling

- Holdout dataset (2A)
- Fitting a model (2C)
- Multicollinearity concerns (2E)
- Setting reference levels for factors (DataPrep 2)
- Piecewise terms (2F)
- Undersampling (3)

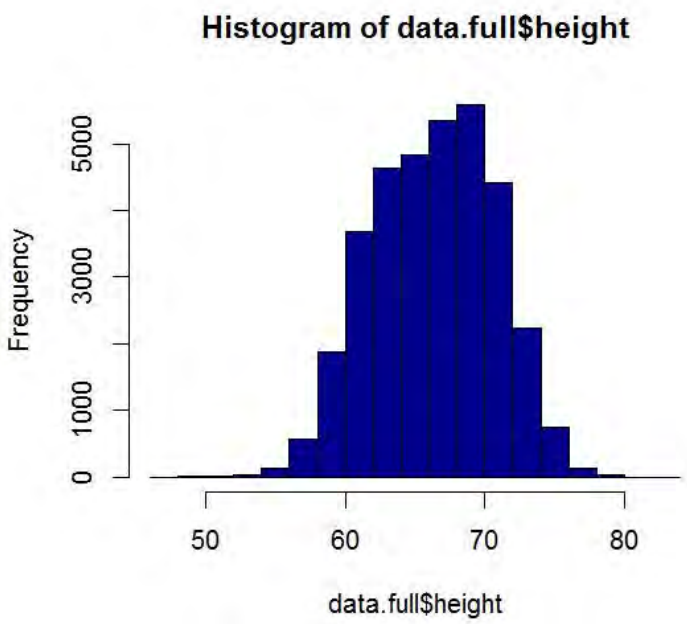
Data outliers

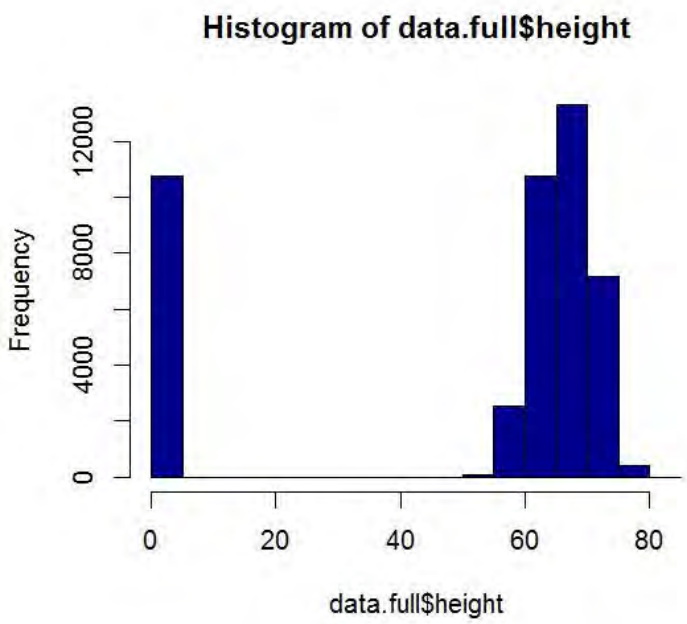




Missing values

```
> summary(data.full$height)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  47.00   64.00   67.00   66.82   70.00   83.00   10739
```



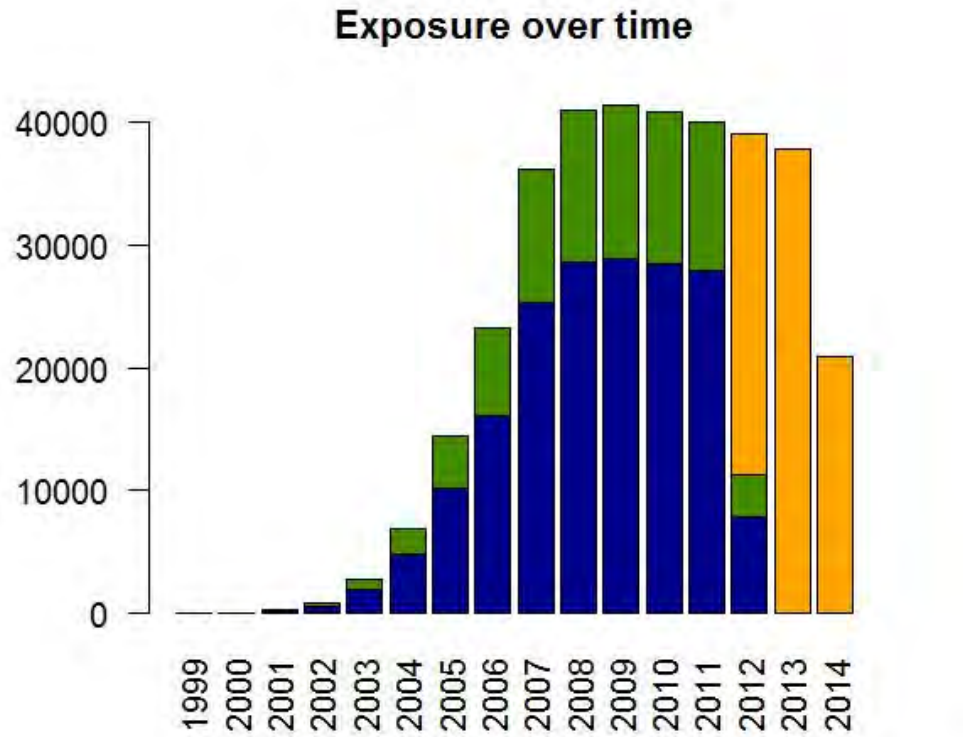


Missing values

Model	NA treatment	Intercept	Height coefficient	Flag coefficient
Death ~ height	Removed	-4.418	0.0100	N/A
Death ~ height + Ind	Set to 0	-3.580	0.0100	-0.838
Death ~ height + Ind	Set to mean	-4.245	0.0100	-0.173
Death ~ height	Set to 0	-3.589	-0.0024	N/A
Death ~ height	Set to mean	-4.343	0.0095	N/A

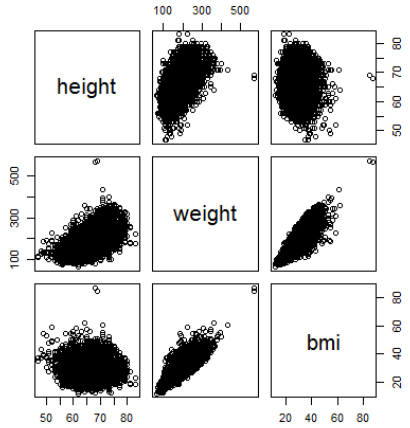
- The first three models are mathematically equivalent
- The second two are biased

Training versus holdout data



Multicollinearity

- pairs()

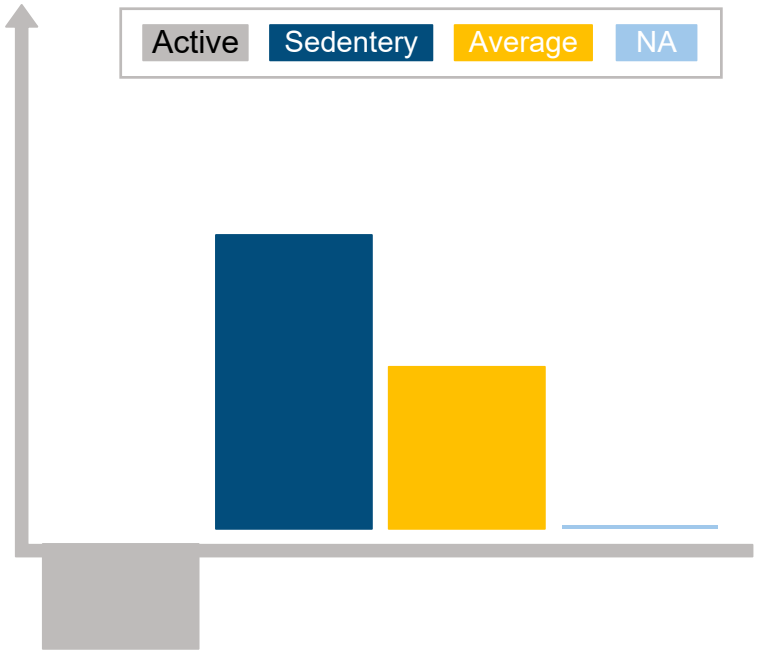


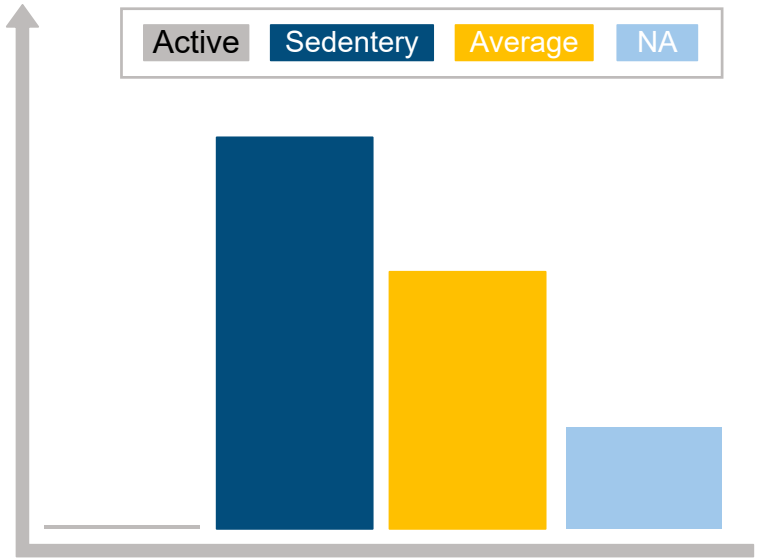
- cor()

```
          height    weight     bmi
height   1.000000 0.637640 0.052578
weight   0.637640 1.000000 0.795710
bmi      0.052578 0.795710 1.000000
```

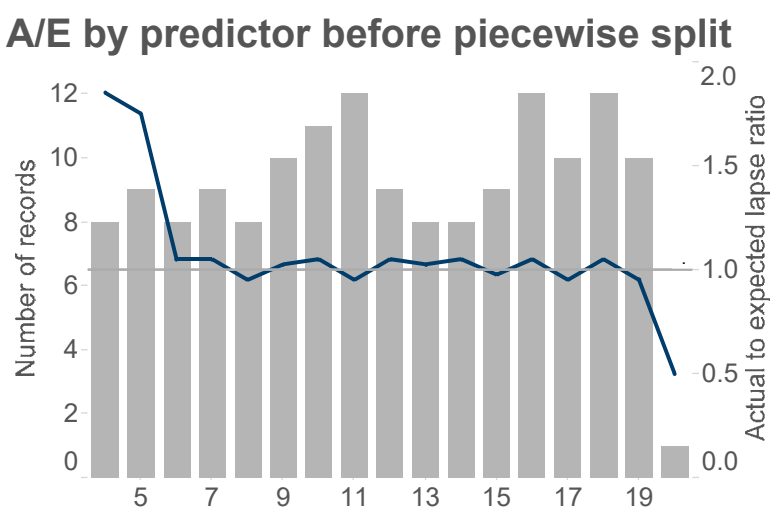
- vif()

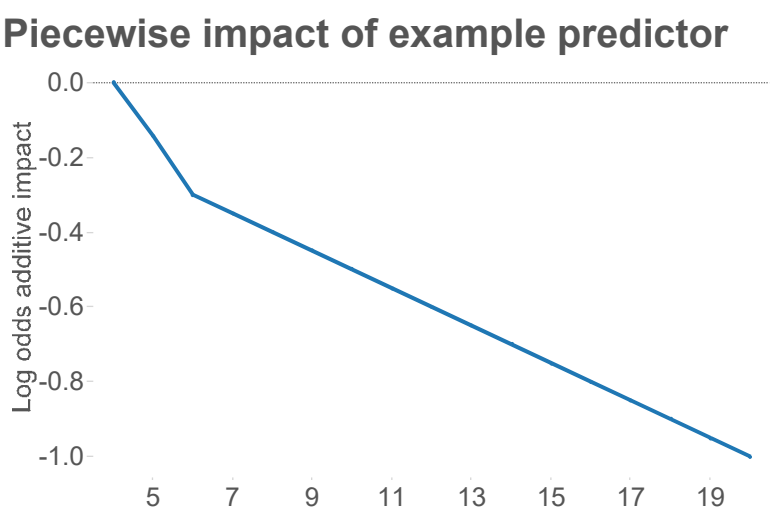
Reference levels





Piecewise linear effects





Undersampling

- For logistic regression, undersampling can help improve runtimes:
  - All deaths (n) +
  - Randomly selected non-deaths (3n)
- Fitting the model Death ~ AttAge

Dataset	Records	Runtime	Intercept	AttAge coefficient
Full	259,284	2.15	-14.13	0.129
Undersampled	25,152	0.12	-10.99	0.123

Hands-on: Practical concerns in R!

SOCIETY OF ACTUARIES

Validation

SOCIETY OF ACTUARIES

Validation and comparison

- Overall model fit (4A)
  - Bias-variance tradeoff
- Comparison between two candidate models (4B)

Model fit

- R^2
- Log-likelihood/AIC/BIC
- Actual-to-expected plots (4A-i)
- Confusion matrix (4A-ii)
- AUC (4A-iii)

Confusion matrix

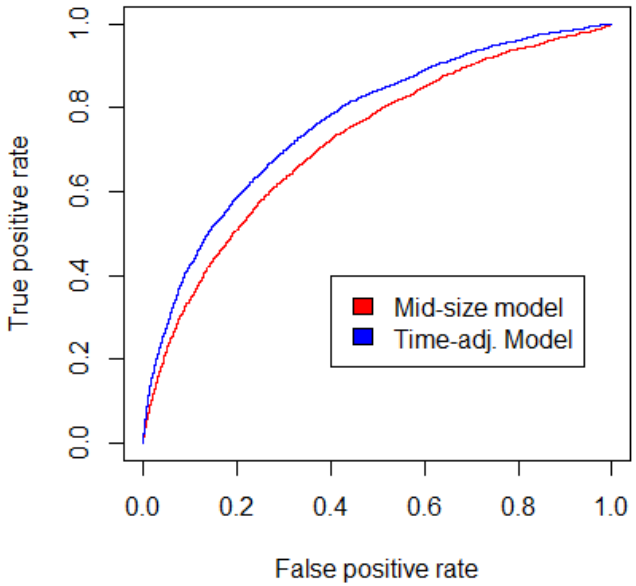
- Select a threshold for predicting the outcome
- Build a 2x2 contingency table

Prediction	Death		
	0	1	Total
0	65,815	835	66,650
1	18,500	1,313	19,813
Total	84,315	2,148	86,463

True positive rate = 1,313/2,148 = 0.658 (1 – Type-II error)
False positive rate = 18,500/84,315 = 0.301 (Type-I error)

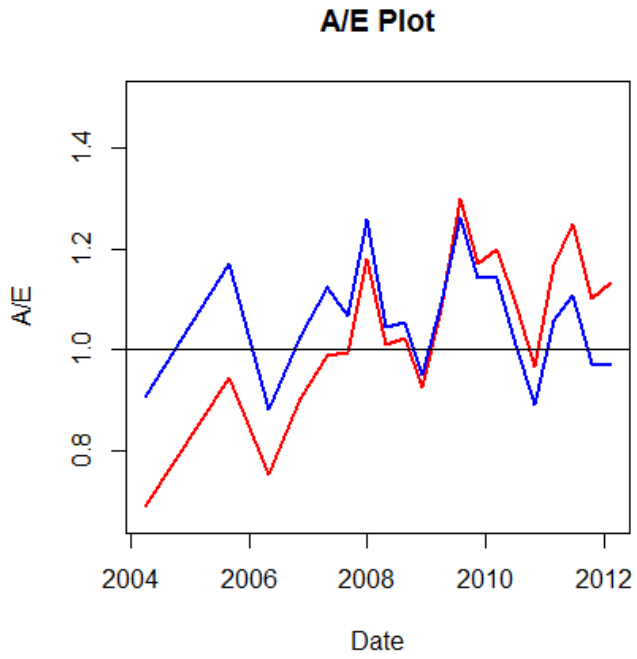
Area under the curve (AUC)

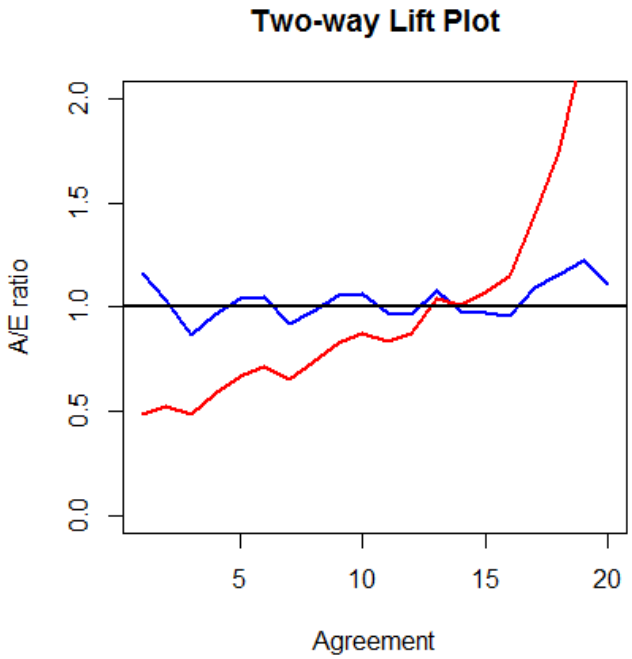
- The curve here is the relationship of the true positive rate and false positive rate as the threshold moves from 0 to 1



Model comparison: Lift charts

- Actual to expected (4B)
- Two-way lift (4B)





Hands-on: Validation in R!

SOCIETY OF ACTUARIES

Thank you!

SOCIETY OF ACTUARIES

Practical Predictive Analytics Seminar

Talex Diede

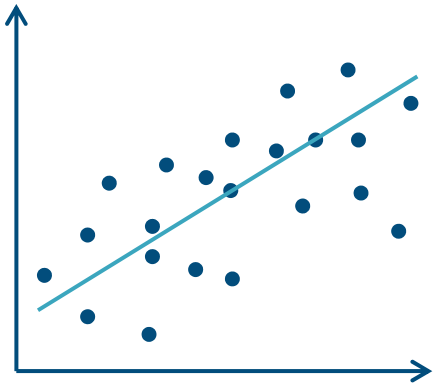
Session 4: Machine Learning Topics

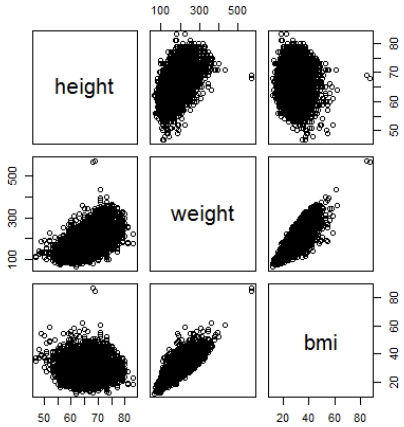
May 22, 2019

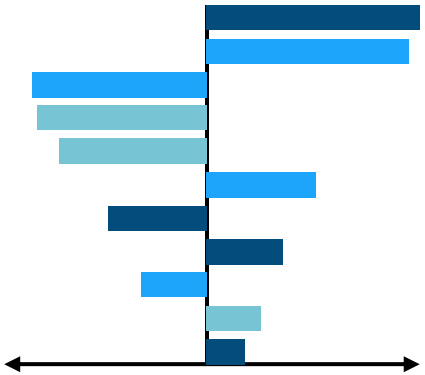
SOCIETY OF ACTUARIES

GLM review

- Linear model
- Interpretable
- Issues:
  - Multicollinearity
  - Variable selection
  - Variable importance
  - Interactions





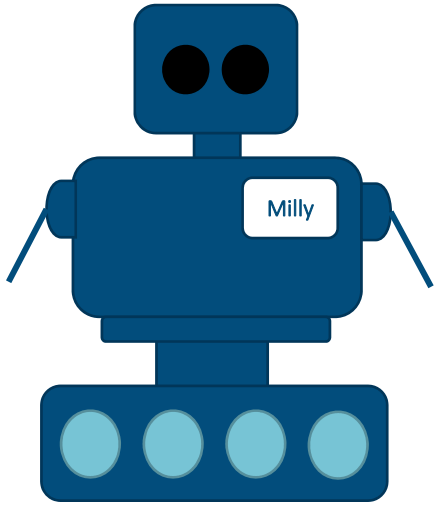


Why machine learning?

- Data continues to grow
- Powerful
- Flexible
- Computational enhancements
  - Cheaper
  - More available
- It's sexy

Machine learning techniques

- Regularization methods
- Classification and regression trees
- Ensemble models
- Others:
  - Clustering
  - Bayesian
  - Neural network
  - Deep learning

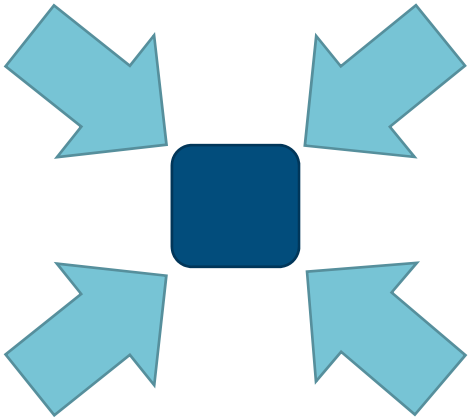


Regularization Methods

SOCIETY OF ACTUARIES

What is “regularization”?

- Regularization is a technique used to avoid the problem of overfitting. The idea is to add a complexity term to the loss function to penalize more complex models.



Regularization methods

- Ridge regression
- LASSO
- ElasticNet

- In R:
  - Packages: **glmnet**, MASS, ridge, lars, elasticnet, …

Ridge regression

- weight decay
- L2-norm penalty
- $loglikelihood =$

$$-\sum[Y_i \ln(\hat{y}_i) + (1 - Y_i)\ln(1 - \hat{y}_i)] + \boldsymbol{\lambda}\sum \boldsymbol{\beta^2}$$

LASSO

- Least absolute shrinkage and selection operator
- L1-norm penalty
- $loglikelihood =$

$$-\sum[Y_i \ln(\hat{y}_i) + (1 - Y_i)\ln(1 - \hat{y}_i)] + \boldsymbol{\lambda}\sum|\boldsymbol{\beta}|$$

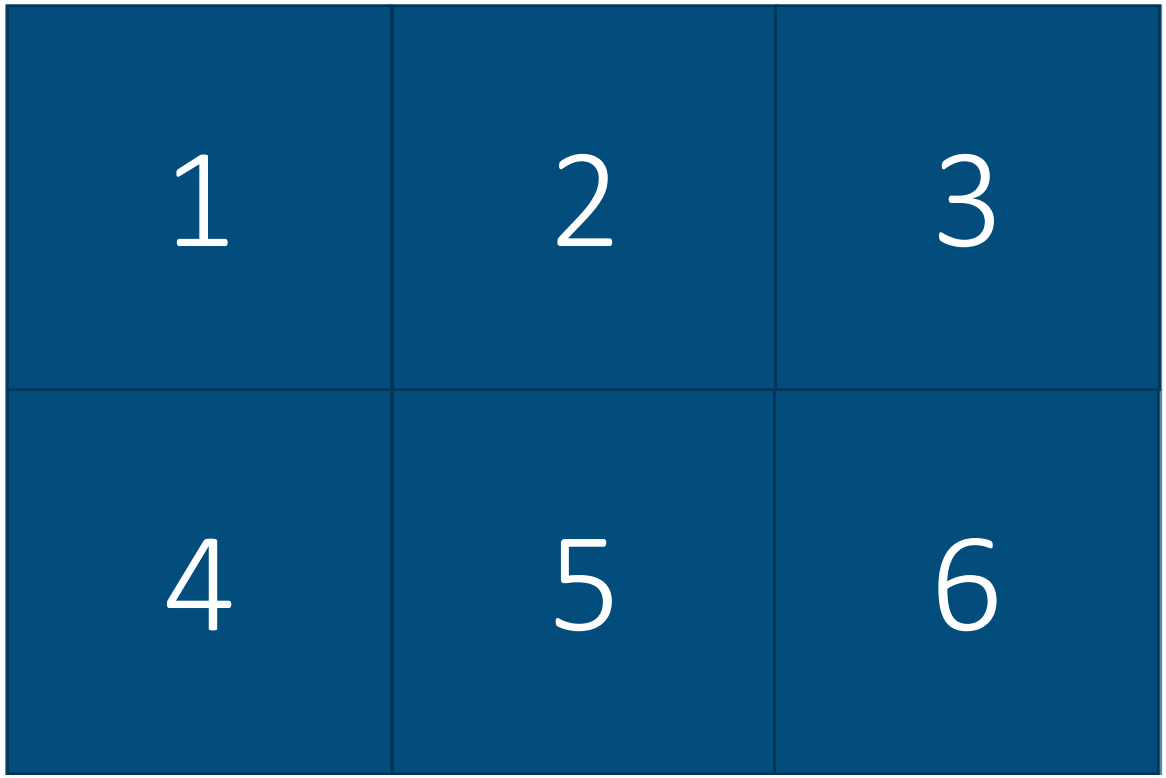
ElasticNet

- Convex combination of ridge and LASSO
- L2 & L1-norm penalties
- $loglikelihood =$

$$-\sum[Y_i \ln(\hat{y}_i) + (1 - Y_i)\ln(1 - \hat{y}_i)] + \boldsymbol{\lambda}\left((\mathbf{1} - \boldsymbol{\alpha})\sum \boldsymbol{\beta}^{\mathbf{2}} + \boldsymbol{\alpha}\sum|\boldsymbol{\beta}|\right)$$

Aside: Cross-Validation

- Useful for smaller datasets



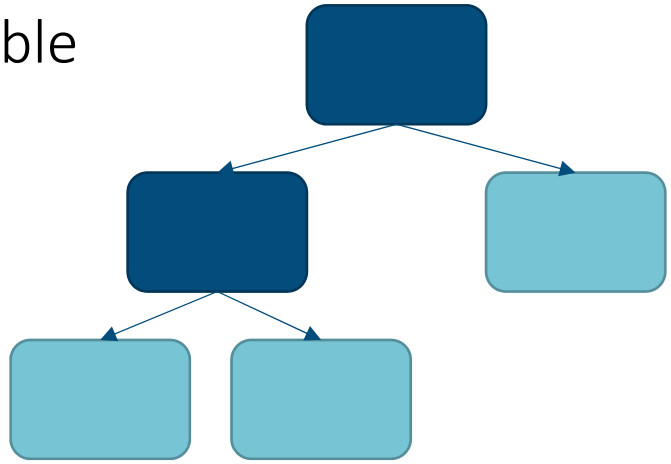
Classification and Regression Trees (CART)

SOCIETY OF ACTUARIES

Trees

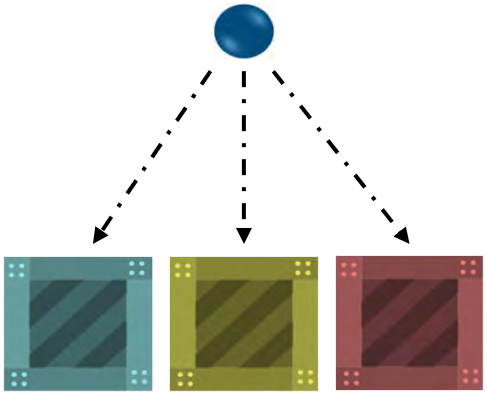
- Sequence of questions/rules for splitting the data
- Elements of CART algorithms
  - Rules for splitting data at each node
  - Stopping criteria
  - Prediction for the target variable

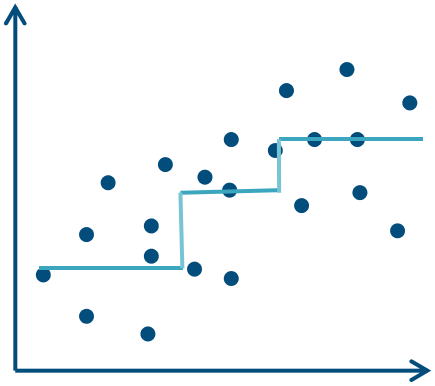
N = 350
0 = 200/350
1 = 150/350



Classification vs regression

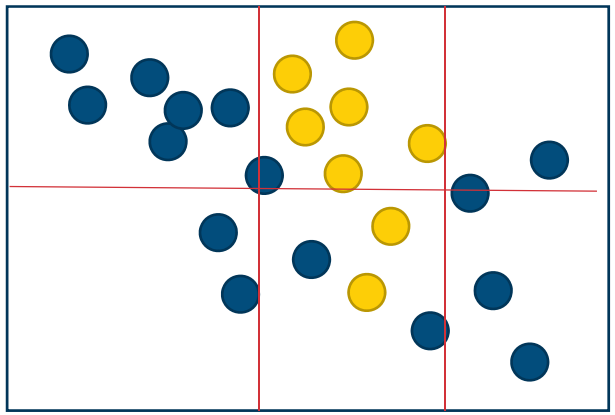
- Classification trees: used for categorical or binary target variables
  - Predict the category a policy will fall into
- Regression trees: continuous target variable
  - Predict the value of the continuous target





Splitting nodes

- Goal: choose the split that results in nodes with maximum homogeneity
- Classification: “Impurity” function
  - Entropy
  - Misclassification rate
  - Gini index
  - Twoing
- Regression: Squared residuals minimization



Stopping rules

- Depth
- Size
- Number of nodes
- Complexity parameter

Ensemble Models

SOCIETY OF ACTUARIES

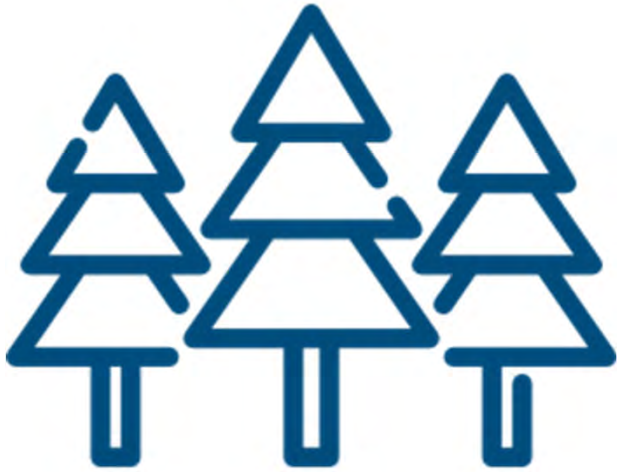
Overview

- What:
  - An ensemble model is the aggregation of two or more related but different models, averaged into a single prediction.

- Why:
  - Improve accuracy of predictions
  - Improve stability of the model

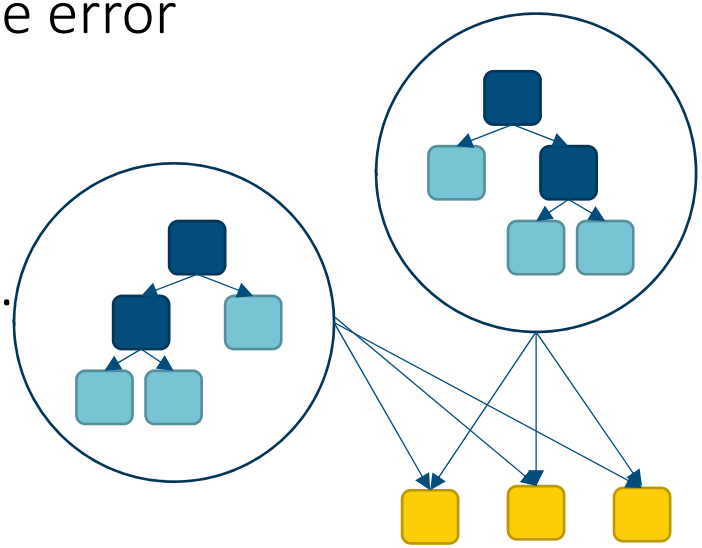
Ensemble methods

- Bagging
- Boosting
- Stacking



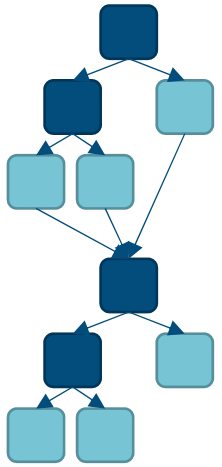
Bagging

- What is it:
  - Building multiple models from different subsamples of the training dataset, results are then combined for the final prediction.
  - Helps to reduce the variance error
- Example:
  - Random Forest
  - R package: **randomForest**, …



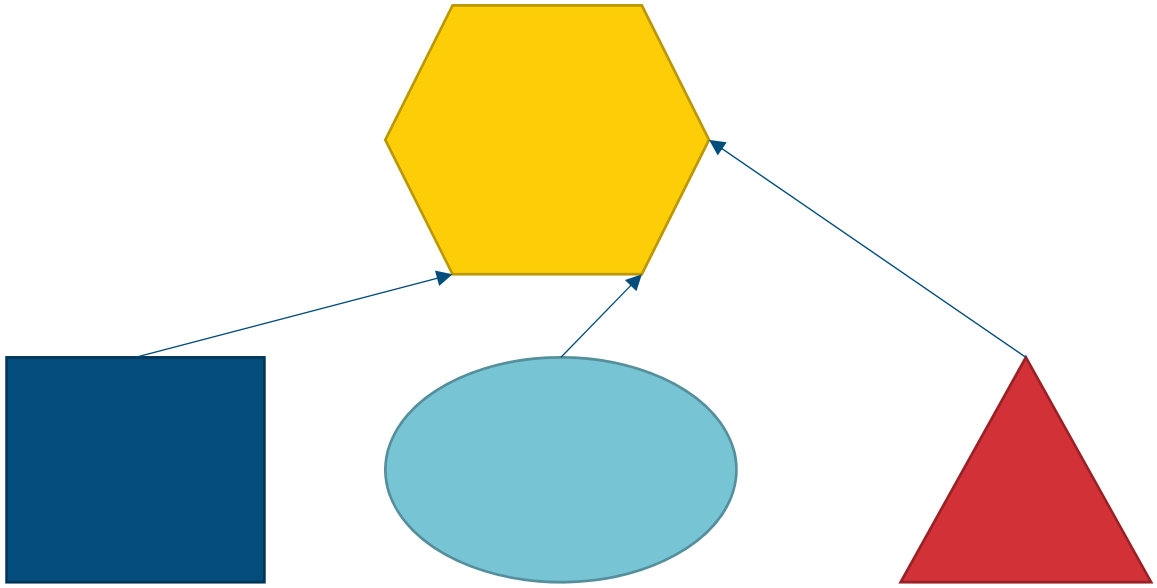
Boosting

- What is it:
  - Building multiple models, each of which is built to improve the prediction errors of a prior model
  - Has shown better predictive accuracy than bagging, but more likely to overfit
- Example:
  - Gradient Boosted Machines (GBM)
  - R packages: **gbm**, xgboost, …



Stacking

- What is it:
  - Building multiple models, typically different types of models, then having a supervisor model that determines how to best combine those results



Back to R!

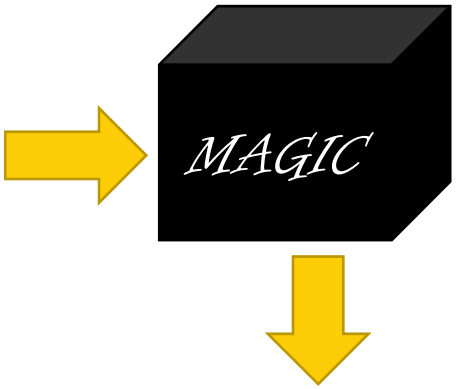
SOCIETY OF ACTUARIES

Final Thoughts

SOCIETY OF ACTUARIES

Weighing your options

- Implementation
- Explanation
- Cost



$$\text{Log Odds} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \bullet\bullet\bullet$$

Other considerations

- Actuarial judgment
- Model selection
- Data issues
- Hardware/Software

Now you're on your way!