

Ensembles and Combining Models



Christopher Cooksey, FCAS, MAAA, CSPA
Head Actuary, Data & Analytics

21 March 2018



Agenda

Rationale and Effectiveness of Ensembles
Basic Approaches – Bagging and Boosting
Complexity and Reality
Combining Linear Regression and Ensembles

Rationale and Effectiveness of Ensembles

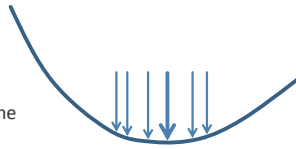
What is the “best” model?

There isn't only one correct model.

Consider credibility-weighting a statewide average with a countrywide average.

What is the “best” model?

If you have two models, each of which perform similarly from a statistical perspective, which do you choose?



Normally we work with some function to define “best.”

Multiplicity of Models

“...there is often a multitude of different descriptions [equations $f(x)$] in a class of functions giving about the same minimum error rate.”

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, Vol. 16, No. 3.

“Data will often point with almost equal emphasis on several possible models, and it is important that the statistician recognize and accept this.”

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*.

Basics of Ensembles

How do you take one set of data and one modeling method and get multiple models?!

1. Data
2. Modeling technique(s)
3. Method for combining models

Page 11

Basics of Ensembles

Remember our credibility-weighting of statewide and countrywide averages?

1. We get variety from using different data.
2. Only one technique is used (averaging).
3. We combine through $n/(n+k)$.

Page 12

Basics of Ensembles

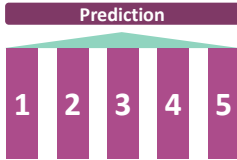
Bagging = Bootstrap aggregation

- One modeling technique is used on several randomly sampled versions of the data.
- Bootstrapped datasets are built by sampling with replacement to build several equal size datasets.

Component models within an ensemble are "learners."

Page 13

Basics of Ensembles



Individual learners stand side-by-side. Weighting can be applied to the average.

Bagging

With learners built on different versions of the data, bagging averages predicted estimates together, thereby reducing the variance of the prediction.

Page 11

Basics of Ensembles

Adaboost (short for adaptive boosting) is one of the original versions of boosting.

Predictions from the first learner are compared to actuals. Misclassified instances are given more weight ("boosted") in subsequent learners. Later learners have a chance to explicitly correct errors from previous ones.

Letting subsequent models focus on the residuals of prior models is the essence of a boosting approach.

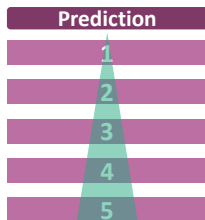
Page 12

Basics of Ensembles

Boosting

- Approach to the data is modified, not the data itself.
- Boosting is effective at reducing the bias of the prediction.

Learners layer on top of each other. Subsequent learners take into account the results of prior learners.



Page 13

Complexity and Reality

Horizontal lines for notes

Actuarial Review – Jan/Feb 2017

Distinguished between GLM and Decision Trees versus Advanced Analytics and Machine Learning.

“For advanced analytics, the product team needs to weigh the benefit of the added lift compared to the need for transparency.” (p. 31)

“GLMs...are simpler and easier to explain than advanced models.” (p. 31)

“...greater sophistication also makes the reasons behind the results less transparent and harder to explain.” (p. 32)

Page 31

Horizontal lines for notes

Actuarial Review – Jan/Feb 2017

“For advanced analytics, the product team needs to weigh the benefit of the added lift compared to the need for transparency.” (p. 31)

- Well stated – benefit versus need.
• In our conventional wisdom, do we put these as co-equal?

Page 31

Horizontal lines for notes

Accuracy and Interpretability

“Framing the question as the choice between accuracy and interpretability is an incorrect interpretation of what the goal of a statistical analysis is.

The point of a model is to get useful information about the relation between the response and predictor variables. Interpretability is a way of getting information.”

Breiman, L. (2001). Statistical Modeling: The Two Cultures. Statistical Science, Vol. 16, No. 3.

Page 11

Horizontal lines for notes

Actuarial Review – Jan/Feb 2017

“For advanced analytics, the product team needs to weigh the benefit of the added lift compared to the need for transparency.” (p. 31)

- Well stated – benefit versus need.
• In our conventional wisdom, do we put these as co-equal?

“GLMs...are simpler and easier to explain than advanced models.” (p. 31)

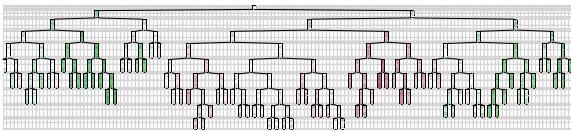
“...greater sophistication also makes the reasons behind the results less transparent and harder to explain.” (p. 32)

- Is this as true as we think it is?

Page 12

Horizontal lines for notes

Decision Trees are easy to explain



Decision Trees are transparent. Transparency does not equal simplicity.

Even “simpler” modeling techniques can make complex models.

Page 13

Horizontal lines for notes

Focus on Reality, not the Model

“...when a model is fit to data to draw quantitative conclusions...the conclusions are about the model’s mechanism, not about nature’s mechanism.”

“These truisms have often been ignored...It is a strange phenomenon – once a model is made, then it becomes truth and the conclusions from it are infallible.”

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, Vol. 16, No. 3.

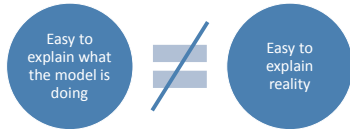
Page 11

Focus on Reality, not the Model

Which should we care about more?

- Explaining how the model works, or
- Explaining what the model says about reality, about our risks, about our customers?

Which do our business partners care about more?



Page 12

Dealing with Complexity

These tasks can be done for simple and complex models alike:

Tasks	Methods
Does the model work?	Show how it predicts hold-out data.
Does the model effectively differentiate?	Lift curves, Gini coefficients, etc.
Which predictors are more important?	Run models with and without predictors.
How does the data relate to specific predictors?	Observed versus modeled graphs.
What are the reasons for a given prediction?	Approximate the model with a simpler model.
Is the model stable over time?	Divide data by time and test.

This is not an exhaustive list. The point is that most of what is required from a predictive model doesn't relate to its inner workings.

Page 13

Actuarial Review – Jan/Feb 2017

Distinguished between GLM and Decision Trees versus Advanced Analytics and Machine Learning.

“For advanced analytics, the product team needs to weigh the benefit of the added lift compared to the need for transparency.” (p. 31)

“GLMs...are simpler and easier to explain than advanced models.” (p. 31)

“...greater sophistication also makes the reasons behind the results less transparent and harder to explain.” (p. 32)

Page 11

Horizontal lines for notes

Horizontal lines for notes

Combining Linear Regression and Ensembles

Case Study

Worker’s Compensation data
Exposures represent \$100,000 in payroll

Frequency target

Training Data: 70% of pre-2014 data, selected at random
Validation Data: 30% of pre-2014 data, the balance of this group
Test Data: 2014 and 2015 data

All results here are shown on the Test data

Page 12

Horizontal lines for notes

Case Study

Two modeling methodologies are used.

- A forward stepwise GLM targeting a collection of 30 possible predictors.
- A boosted ensemble of trees using the same collection of 30 possible predictors. Analogous to the forward stepwise GLM, an automated process was used to select the primary model parameters of learning rate and tree depth.

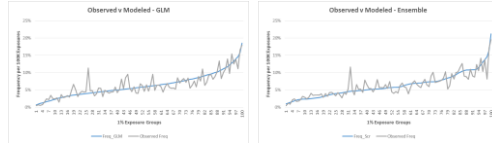
In both cases, modeler discretion was limited to the number of iterations. The assumption here is that both techniques could be improved by human intervention.

Page 11

Case Study – GLM versus Ensemble

How do these methods compare when simply building a “ground-up” frequency model? On the surface, similar lift and fit.

	GLM	Ensemble
Min	0.7%	0.9%
Max	18.5%	21.2%
Lift	26.3	22.5
Spread	0.178	0.203



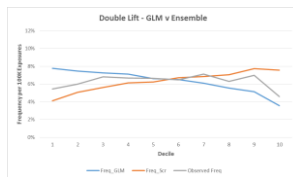
Page 12

Case Study – GLM versus Ensemble

A double lift chart shows mixed results as well.

However, is this comparison valid?

Is this the proper way to take advantage of the particular strengths and weaknesses of each approach?



Page 13

Combining Linear Regression and Ensembles

We often think about the linear and non-linear signal in the data.

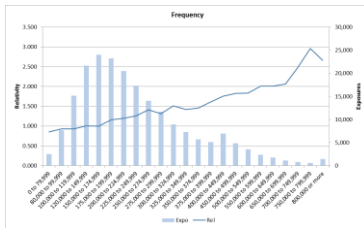
	(log) Linear	Non-linear, Combinatorial
GLM	Efficient representation	Possible (to a degree) to represent, but cumbersome to explore
Ensembles of Trees	Inefficient representation	Natural representation and exploration

Page 11

Combining Linear Regression and Ensembles

When there is linear signal, a GLM represents this in a straight-forward manner.

Imagine what it would take for a tree to represent this same information...

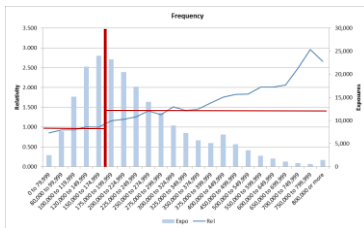


Page 11

Combining Linear Regression and Ensembles

When there is linear signal, a GLM represents this in a straight-forward manner.

Imagine what it would take for a tree to represent this same information...

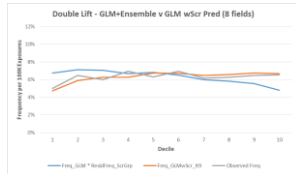


Page 11

Case Study – Combined versus Combined

The double lift chart in this case shows a clear winner.

Despite being a simpler model, when the two approaches disagree the GLM which uses a non-linear predictor is consistently more accurate than a GLM plus a refinement based on a residual Ensemble model.



Page 11

Case Study – Combined versus Combined

Is there really a clear winner?

In the case of Pricing, there are distinct advantages to modeling the residuals of a baseline GLM.

- By taking the GLM results as a given, the “complicated” model produces a single rate adjustment factor.
- The combined model still looks like a traditional rating plan.
- The Ensemble-based adjustment factor can be considered on its own terms – acceptability to agents, customers, regulators, etc.

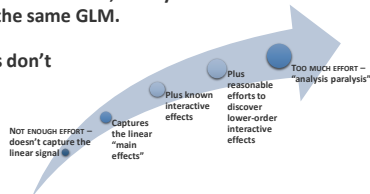
Also, we should note this is one result for one target on one dataset for one line of business.

Page 12

GLM within a combined approach

It is important to note that if you know from the beginning you are building a combined model, then you don't necessarily build the same GLM.

Combined models don't necessarily take more time.



Page 13

Summary

- Ensembles work by combining information from multiple models.
- Bagging averages predictions; boosting focuses on residuals.
- GLMs parse effects to individual fields. The question of who has a high or low prediction is different.
- Observed versus Modeled graphs are independent of modeling method. They can be used to explain complex models.
- Reality, with its simple trends *and* complexity exists without regard to our modeling method.
- There is great potential to combine modeling methods.

Page 11

Questions?

Christopher Cooksey, FCAS, MAAA, CSPA
Head Actuary, Data and Analytics

Guidewire Software

ccooksey@guidewire.com

Page 12
