Ensembles and Combining Models

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

27 March 2019

GUIDEWIRE

1

Agenda

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

Page 2

2

Rationale and Effectiveness of Ensembles

3

What is the "best" model?

There isn't only one correct model.

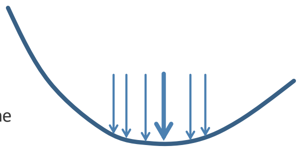
Consider credibility-weighting a statewide average with a countrywide average.

4

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."



5

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*.

6

AN UNREALISTIC ILLUSTRATION

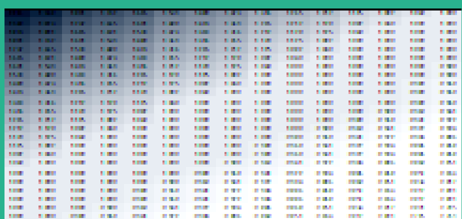
Ground Rules

1) We get to know reality & compare our models directly.
2) Assume the numbers are frequency relativities.
3) Volume is limited; we can only divide the data into three equally-sized groups.
4) Model predictions are just the average for each defined group.

Page7

7

AN UNREALISTIC ILLUSTRATION

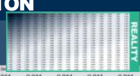


REALITY

Page8

8

AN UNREALISTIC ILLUSTRATION



MODEL 1

Group relatively homogeneous business together.

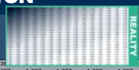
Sum of the squared error = 13.48

1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944

Page9

9

AN UNREALISTIC ILLUSTRATION



MODEL 2

A different way of splitting the data.

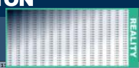
Sum of the squared error = 11.63

1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929

Page10

10

AN UNREALISTIC ILLUSTRATION



ENSEMBLE Models 1 & 2

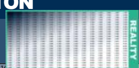
Sum of the squared error = 9.02

1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937

Page11

11

AN UNREALISTIC ILLUSTRATION



ENSEMBLE Models 1 - 5

Sum of the squared error = 8.47

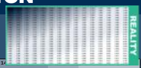
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.230	1.230	1.230	1.230	1.211	1.170	1.170	1.116	1.116	1.116	1.093	1.093	1.015	1.015	1.015
1.230	1.230	1.230	1.230	1.211	1.170	1.170	1.116	1.116	1.116	1.093	1.093	1.015	1.015	1.015
1.172	1.172	1.172	1.172	1.152	1.058	1.058	1.058	1.058	1.058	1.015	1.015	1.015	1.015	0.975
1.172	1.172	1.172	1.172	1.112	1.018	1.018	1.018	1.018	1.018	0.975	0.975	0.975	0.975	0.975
1.172	1.172	1.172	1.172	1.112	1.018	1.018	1.018	1.018	1.018	0.975	0.975	0.975	0.975	0.975
1.172	1.172	1.172	1.172	1.112	1.018	1.018	1.018	1.018	1.018	0.975	0.975	0.975	0.975	0.975
1.172	1.172	1.172	1.172	1.172	1.054	0.994	0.994	0.974	0.974	0.952	0.952	0.952	0.952	0.952
1.172	1.172	1.172	1.172	1.172	1.054	0.994	0.994	0.974	0.974	0.952	0.952	0.952	0.952	0.952
1.148	1.148	1.148	1.148	1.070	1.030	0.970	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.148	1.148	1.148	1.148	1.070	1.030	0.970	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.148	1.148	1.148	1.148	1.070	1.030	0.970	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.148	1.148	1.148	1.148	1.070	1.030	0.970	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.148	1.148	1.148	1.148	1.070	1.030	1.030	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.070	1.070	1.070	1.070	1.070	1.030	1.030	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.070	1.070	1.070	1.070	1.070	1.030	1.030	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.070	1.070	1.070	1.070	1.070	1.030	1.030	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928

Page12

12

AN UNREALISTIC ILLUSTRATION

ENSEMBLE
Models 1-9



1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.191	1.191	1.152	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.191	1.152	1.140	1.140	1.096	1.053	1.053
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.232	1.152	1.140	1.140	1.096	1.053	1.053
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.232	1.193	1.140	1.140	1.096	1.053	1.053
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.196	1.148	1.136	1.095	1.009	1.009	1.009
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.196	1.148	1.136	1.095	1.009	1.009	1.009
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.196	1.148	1.136	1.101	1.009	1.009	1.009
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.196	1.148	1.136	1.101	1.009	1.009	1.009
1.269	1.269	1.269	1.269	1.258	1.236	1.236	1.169	1.169	1.122	1.076	1.042	0.992	0.992	0.982
1.269	1.269	1.269	1.269	1.258	1.236	1.236	1.169	1.130	1.122	1.076	1.042	0.992	0.992	0.982
1.237	1.237	1.237	1.237	1.226	1.140	1.140	1.103	1.064	1.056	0.998	0.998	0.992	0.992	0.960
1.237	1.237	1.237	1.237	1.203	1.118	1.118	1.042	1.042	1.034	0.976	0.976	0.970	0.970	0.960
1.166	1.166	1.166	1.166	1.133	1.047	1.008	1.008	1.008	1.000	0.976	0.976	0.970	0.970	0.960
1.124	1.166	1.166	1.166	1.133	1.047	1.005	1.005	1.008	1.000	0.966	0.966	0.960	0.960	0.960
1.091	1.091	1.133	1.133	1.133	1.026	0.992	0.992	0.984	0.976	0.953	0.947	0.947	0.947	0.947
1.091	1.091	1.091	1.133	1.091	1.026	0.992	0.992	0.981	0.976	0.953	0.947	0.947	0.947	0.947
1.078	1.078	1.078	1.078	1.035	1.012	0.979	0.979	0.968	0.960	0.938	0.931	0.931	0.933	0.933
1.078	1.078	1.078	1.078	1.035	1.012	0.979	0.979	0.968	0.960	0.931	0.931	0.931	0.933	0.933
1.078	1.078	1.078	1.078	1.035	1.012	0.979	0.979	0.968	0.960	0.931	0.931	0.931	0.933	0.933
1.078	1.078	1.078	1.078	1.035	1.012	0.979	0.979	0.968	0.954	0.931	0.931	0.931	0.933	0.933
1.078	1.078	1.078	1.078	1.035	1.003	1.003	0.969	0.952	0.944	0.931	0.931	0.931	0.933	0.933
1.035	1.035	1.035	1.035	1.035	1.003	1.003	0.963	0.952	0.944	0.931	0.931	0.931	0.933	0.933
1.035	1.035	1.035	1.035	1.035	1.003	0.996	0.963	0.952	0.944	0.931	0.931	0.931	0.933	0.933
1.035	1.035	1.035	1.035	1.035	0.996	0.996	0.963	0.952	0.944	0.931	0.931	0.931	0.933	0.933

Sum of the squared error
= 7.35

Page13

13

Ensembles

"Ensemble modeling has taken the [Predictive Analytics] industry by storm.

It's often considered the most important predictive modeling advancement of this century's first decade."

Siegel, E. (2013). *Predictive Analytics.*

Page14

14

Basic Approaches – Bagging and Boosting

15

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**

16

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).**

17

Basics of Ensembles

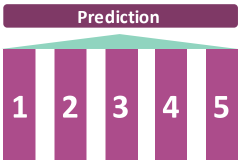
Bagging = Bootstrap aggregation

- 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."

18

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.

19

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.

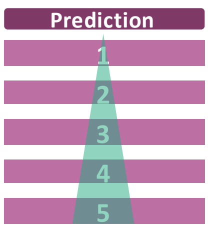
20

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.



21

Complexity and Reality

22

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)

Page23

23

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?**

Page24

24

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.

25

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?**

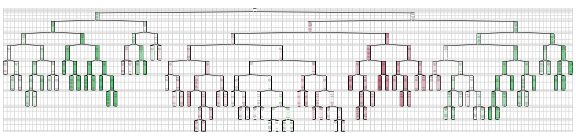
"GLMs…are simpler and easier to explain than advanced models." (p. 31)

"…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?**

26

Decision Trees are easy to explain



Decision Trees *are transparent.* Transparency does not equal simplicity.

Even "simpler" modeling techniques can make complex models.

27

Explaining GLM Results

Consider a Loss Cost GLM with 25 predictors, some of them being two-way interactions.

One predictor is Age of Roof with the following relativities:

Age of Roof	Rel
0-7	0.90
8-12	1.00
13+	1.10

What can be said of the group of customers with roofs aged 7 years or less?

How many of us said that the predicted loss cost of the Age of Roof 0-7 group is 10% less than the base customer?

How many wondered if Age of Roof was part of any interaction terms?

28

Explaining GLM Results

Relativities are how GLMs *model* the target. Relativities are how GLMs parse the predictable variation in the target (i.e. the "signal") to multiple different predictors.

Age of Roof	# Expos	Rel
0-7	7,000	0.90
8-12	7,600	1.00
13+	5,400	1.10

Terr	# Expos	Rel
1	4,000	1.25
2	4,600	1.10
3	3,000	1.00
4	4,200	0.98
5	4,200	0.85

When the exposures across two predictors are correlated, the single-predictor relativity doesn't reflect the entirety of the model prediction.

Consider historical hail storms in Territory 1 that were not removed from the data.

29

Explaining GLM Results

Here there is *no* correlation between the exposure distributions of Age of Roof and Territory.

Age of Roof	# Expos	Rel
0-7	7,000	0.90
8-12	7,600	1.00
13+	5,400	1.10

Terr	# Expos	Rel
1	4,000	1.25
2	4,600	1.10
3	3,000	1.00
4	4,200	0.98
5	4,200	0.85

Exposures	Territory					
Age of Roof	1	2	3	4	5	Total
0-7	1,400	1,610	1,050	1,470	1,470	7,000
8-12	1,520	1,748	1,140	1,596	1,596	7,600
13+	1,080	1,242	810	1,134	1,134	5,400
Total	4,000	4,600	3,000	4,200	4,200	20,000

% of Exposures	Territory					
Age of Roof	1	2	3	4	5	Total
0-7	20%	23%	15%	21%	21%	100%
8-12	20%	23%	15%	21%	21%	100%
13+	20%	23%	15%	21%	21%	100%
Total	20%	23%	15%	21%	21%	100%

Relativities	Territory					
Age of Roof	1	2	3	4	5	Ave Rel
0-7	1.125	0.990	0.900	0.882	0.765	0.9336
8-12	1.250	1.100	1.000	0.980	0.850	1.0373
13+	1.375	1.210	1.100	1.078	0.935	1.1410

In this case, the predicted loss cost for the Age of Roof 0-7 group of customers *is* 10% lower than the base group of customers.

30

Explaining GLM Results

Here there *is* a correlation, though not enough to be a convergence problem.

Age of Roof	# Expos	Rel
0-7	7,000	0.90
8-12	7,600	1.00
13+	5,400	1.10

Terr	# Expos	Rel
1	4,000	1.25
2	4,600	1.10
3	3,000	1.00
4	4,200	0.98
5	4,200	0.85

Exposures	Territory					
Age of Roof	1	2	3	4	5	Total
0-7	2,500	2,300	1,050	1,000	950	7,800
8-12	910	1,378	1,140	2,111	2,061	7,600
13+	590	1,122	810	1,089	1,189	4,800
Total	4,000	4,800	3,000	4,200	4,200	20,200

% of Exposures	Territory					
Age of Roof	1	2	3	4	5	Total
0-7	32%	29%	13%	13%	12%	100%
8-12	12%	18%	15%	28%	27%	100%
13+	12%	23%	17%	23%	25%	100%
Total	20%	23%	15%	21%	21%	100%

Relativities	Territory					
Age of Roof	1	2	3	4	5	Ave Rel
0-7	1.125	0.990	0.900	0.882	0.765	0.9799
8-12	1.250	1.100	1.000	0.980	0.850	1.0018
13+	1.375	1.210	1.100	1.078	0.935	1.1137

In this case, the predicted loss cost for the Age of Roof 0-7 group of customers is only 2% lower than the base group of customers.

And this is only 2 of 25 predictors!
How easy is this to explain?

31

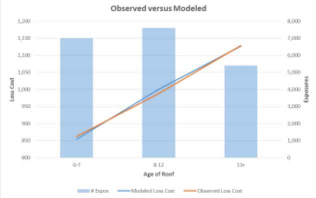
Observed versus Modeled

If we multiply the relativities through all 25 predictors (and the constant), we get the model's predicted loss cost.

A common exhibit for evaluating GLMs is this Observed versus Modeled graph.

(Similar to Monograph 5's Simple Quantile Plot, but with the x-axis being a given predictor's levels.)

Used to check the balance of the model.



32

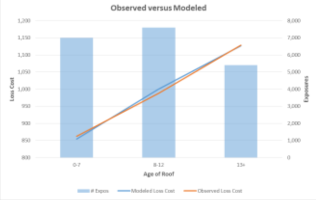
Observed versus Modeled

Observed versus Modeled graphs, plotted across individual predictors, take the whole model into account.

Note that this exhibit keys off of the model's prediction.

Nowhere does it rely on that prediction coming from a GLM, or any other method.

If the model makes a prediction, now matter how complicated or sophisticated it is, this graph can be made.



33

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.

34

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?

Easy to explain what the model is doing ≠ Easy to explain reality

35

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.

36

Combining Linear Regression and Ensembles

37

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

38

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.

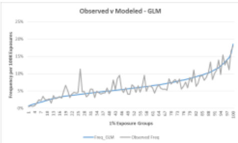
In both cases, modeler discretion was limited to the number of iterations. The assumption here is that both techniques could be improved by human intervention.

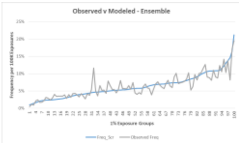
39

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





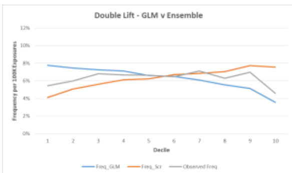
40

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?



41

Combining Linear Regression and Ensembles

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

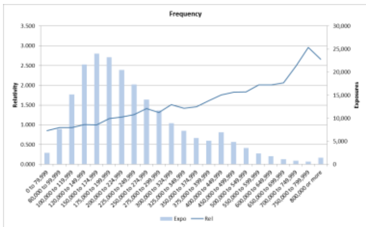
	(log) Linear	Non-linear, Combinatorial
GLM	Efficient representation	Possible (to a degree) to represent, but cumbersome to explore
Ensembles of Trees	Inefficient representation	Natural representation and exploration

42

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…

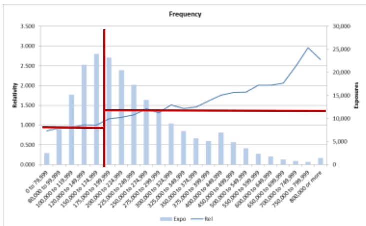


43

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…

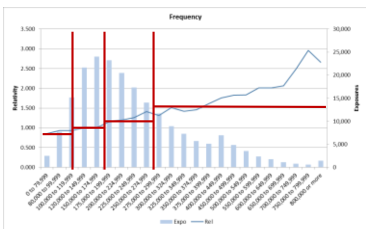


44

Combining Linear Regression and Ensembles

A tree-based approach would have to go several layers deep to even approximate the information in the GLM for this linear relationship.

This is inefficient.

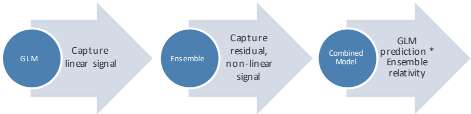


45

Combining Linear Regression and Ensembles

This isn't a competition. We should combine methods in ways that enhance their strengths and limit their weaknesses.

The first approach we'll try is to build a GLM and then model the residuals using the Ensemble.

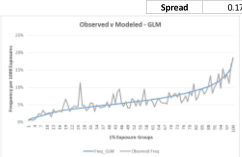


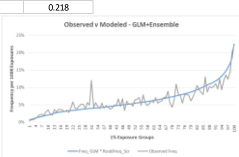
46

Case Study – GLM versus GLM+Ensemble

The predictions from the Ensemble add noticeable and consistent lift to the model. Ensemble relativities ranged from +64% to -39%.

	GLM	GLM+Ensemble
Min	0.7%	0.7%
Max	18.5%	22.5%
Lift	26.3	33.3
Spread	0.178	0.218





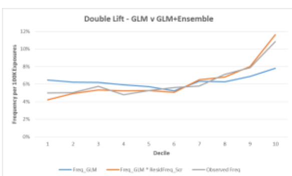
47

Case Study – GLM versus GLM+Ensemble

A double lift chart shows a clearly better result as well.

Specifically in the cases where the combined model and the GLM disagree, the combined models is consistently and dramatically more accurate.

Remember that these results are on a pure Test dataset.

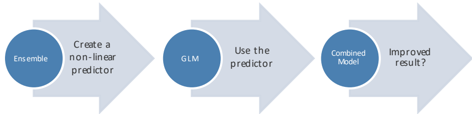


48

Combining Linear Regression and Ensembles

What if we let the Ensemble go first instead?

Part of the Ensemble output for the approach we used presents the model prediction as a 3-digit score. This Score was attached to the data and considered as an additional predictor representing the non-linear signal in the data.

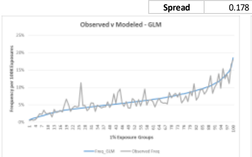


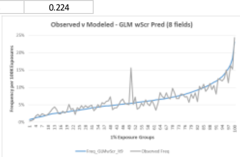
49

Case Study – GLM versus GLM with non-linear predictor

Like the other combined approach, the lift of the model is noticeably improved.

	GLM	GLM wScr Pred
Min	0.7%	0.8%
Max	18.5%	23.1%
Lift	26.3	30.8
Spread	0.178	0.224



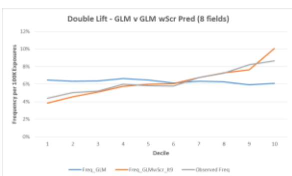


50

Case Study – GLM versus GLM with non-linear predictor

And again, a double lift chart shows a clearly better result as well.

Specifically in the cases where the combined model and the GLM disagree, the combined models is consistently and dramatically more accurate.



51

Case Study – GLM versus GLM with non-linear predictor

It is interesting to examine the output of the forward stepwise procedure for the base GLM and the GLM with the non-linear predictor.

Baseline GLM	
Variable(s) Added	Deviance
NULL MODEL	18,402
Field1	17,830
Field2	17,548
Field3	17,148
Field4	17,019
Field5	16,763
Field6	16,670
Field7	16,640
Field8	16,584

GLM with non-linear predictor	
Variable(s) Added	Deviance
NULL MODEL	18,402
Scr_Freq_f6bdf	16,648
Field1	16,486
Field9	16,466
Field7	16,439
Field3	16,407
Field5	16,373
Field2	16,370
Field10	16,357

52

Case Study – Combined versus Combined

Is there a performance difference in the two combined model approaches? Not on the basis of lift.

It is notable that the creation of a non-linear predictor serves to simplify the entire model. The same lift is achieved with the loss of fewer degrees of freedom.

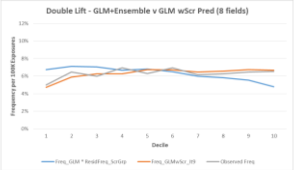
	GLM+Ensemble	GLM wScr Pred
Min	0.7%	0.8%
Max	22.5%	23.1%
Lift	33.3	30.8
Spread	0.218	0.224
# Levels	76	70
dF	67	62
Price Points	27,417,600	5,140,800

53

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.



54

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.

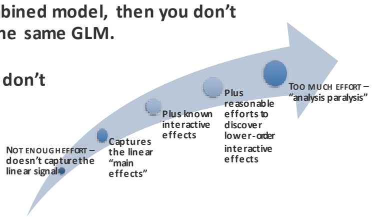
Page55

55

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.



Page56

56

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.**

Page57

57

Questions?

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

Guidewire Software

ccooksey@guidewire.com

Page58

58