Casualty Actuarial Society

2015 Ratemaking & Product Management Seminar

What's Next – Ensembles

Christopher Cooksey, FCAS, MAAA
Chief Actuary, EagleEye Analytics

Agenda

1. Multiplicity of models
2. Introducing the idea of ensembles
3. Basics of ensembles
4. Objections to ensembles

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.*

Multiplicity of Models

Actuaries should be comfortable with this notion – we leverage the idea that there isn’t only one correct model through credibility:

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

Here we have two estimates derived through different means. Instead of choosing one, we blend the information from each.

An unrealistic illustration

Ground rules…

- In this example, we get to know reality and how closely our models approximate it.
- For convenience, assume the numbers are frequency relativities.
- Volume is limited. To have a credible model we can only divide the data into three equally sized groups.
- Model predictions are just the average for each group defined.

An unrealistic illustration

Consider the unrealistic scenario where we know reality.

Reality

2.026	1.948	1.801	1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000
1.948	1.873	1.732	1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000
1.873	1.801	1.665	1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000
1.801	1.732	1.601	1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000
1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000
1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000
1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	0.980
1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980
1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941
1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922
1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904
1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886
1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851
1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834
1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785
1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785	0.769

An unrealistic illustration

Models group relatively homogeneous business together to find useful average performance.

Reality

2.026	1.948	1.801	1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000
1.948	1.873	1.732	1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000
1.873	1.801	1.665	1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000
1.801	1.732	1.601	1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000
1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000
1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000
1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	0.980
1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980
1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941
1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922
1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904
1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886
1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851
1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834
1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785
1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785	0.769

Model 1

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
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

The sum of the squared error of Model 1 is 13.48.

An unrealistic illustration

The first way of splitting the data is not the only way of splitting the data.

Reality

2.026	1.948	1.801	1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000
1.948	1.873	1.732	1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000
1.873	1.801	1.665	1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000
1.801	1.732	1.601	1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000
1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000
1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000
1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	0.980
1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980
1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941
1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922
1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904
1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886
1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851
1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834
1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785
1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785	0.769

Model 2

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
1.287	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

The sum of the squared error of Model 2 is 11.63.

An unrealistic illustration

Here we are not dividing the data nine ways. We are combining the *information* from each model.

Reality

2.026	1.948	1.801	1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000
1.948	1.873	1.732	1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000
1.873	1.801	1.665	1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000
1.801	1.732	1.601	1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000
1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000
1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000
1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	0.980
1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980
1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941
1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922
1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904
1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886
1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851
1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834
1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785
1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785	0.769

Ensemble – Models 1-2

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
1.274	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

The sum of the squared error of this ensemble is 9.02.

An unrealistic illustration

And we can keep going…

Reality

2.026	1.948	1.801	1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000
1.948	1.873	1.732	1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000
1.873	1.801	1.665	1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000
1.801	1.732	1.601	1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000
1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000
1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000
1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	0.980
1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980
1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941
1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922
1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904
1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886
1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851
1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834
1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785
1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785	0.769

Ensemble – Models 1-5

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
1.278	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

The sum of the squared error of this ensemble is 8.47.

An unrealistic illustration

And we can keep going…and going…

Reality

2.026	1.948	1.801	1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000
1.948	1.873	1.732	1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000
1.873	1.801	1.665	1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000
1.801	1.732	1.601	1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000
1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000
1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000
1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	0.980
1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980
1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941
1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922
1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904
1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886
1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851
1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834
1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785
1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785	0.769

Ensemble – Models 1-9

1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.191	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	1.191	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

The sum of the squared error of this ensemble is 7.35.

A realistic effect

Ensembles are remarkable for allowing complexity without over fitting the data.

"Ensembles remain robust even as they become increasingly complex. They seem to be immune to this limitation, as if soaked in a magic potion against overlearning."

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

This is *not* to say they can't over fit and don't need to be tested!

Worth the price

"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.*

Adoption of ensemble methods and ideas by actuaries would be transformative, not incremental.

Basics of Ensembles

You can't take one set of data and one modeling approach and get multiple models!

An ensemble approach can be understood in three pieces.

1. **Data being used in the modeling**
2. **Modeling technique(s) being used**
3. **Method for combining models**

Basics of Ensembles

Consider again our simple 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).**

The two most common ways to build ensembles are boosting and bagging.

Basics of Ensembles

Bagging is short for bootstrap aggregation. One modeling technique is used on several randomly sampled versions of the data.

Bootstrap datasets are built by simply sampling with replacement to build several equal size datasets.

Typically, the component models within an ensemble are referred to as "learners".

Basics of Ensembles

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

Individual learners stand side-by-side.

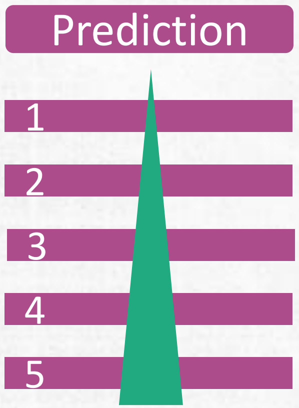
Weighting can applied to the average.

Random forests uses this bagging approach.

Basics of Ensembles

Boosting uses a different approach. The data is not modified, rather our approach to the data.

Conceptually, learners can be seen as layering on top of each other rather than standing side-by-side. Subsequent learners take into account the results of prior learners.



Boosting is effective at reducing the bias of the prediction.

Basics of Ensembles

Adaboost (short for adaptive boosting) is probably the most well-known version 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.

Note: *there are many variants of boosting.*

Basics of Ensembles

One important additional concept is the need to de-couple the learners within an ensemble.

The whole point here is to explore a larger version of the solution space. If the learners combined are all very similar, then the power of ensembles is reduced.

Typically, randomness is introduced into the building of each component model.

Objections to Ensembles

Resistance to ensembles usually centers around the complexity of the models.

Fans of *Occam's Razor* claim that simpler models are better.

Ironically, this is an over-simplification of *Occam's Razor*. Simpler is preferred in the absence of certainty, when multiple models perform equally well. *If an ensemble performs better, then it is simply the better model.*

Objections to Ensembles

"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.

Objections to Ensembles

Still, complexity can be a barrier acceptance. Several assumptions are fairly common.

1. *"All modeling techniques are equally difficult to explain." (Consider neural nets vs. trees.)*
2. *"Because it can't be explained in simple terms, there is no opportunity for insight."*
3. *"Departments of insurance won't accept them."*

In general, people often give up without putting effort into solving the problems.

Objections to Ensembles

Remember too that there are a variety of contexts and needs for predictive analytics in insurance.

Don’t think a complex model will be accepted for pricing in your underwriting-driven culture?

- *How about marketing?*
- *How about claims management?*
- *How about agent placement?*
- *How about internal monitoring?*

Conclusions

- The issue of multiple models exists whether you leverage the concept or not.
- Ensembles extract more information from data without paying the expected pricing in over fitting.
- Ensembles can be understood in three parts – how they approach the data, the modeling technique(s) used, and the method of combining.
- Results from ensembles approaches are transforming other industries and are worth the effort for actuaries to explore.

Questions?

Christopher Cooksey
Chief Actuary, EagleEye Analytics
EEAnalytics.com

ccooksey@eeanalytics.com
Cell: 740-398-2629