

Predictive Analytics Workshop

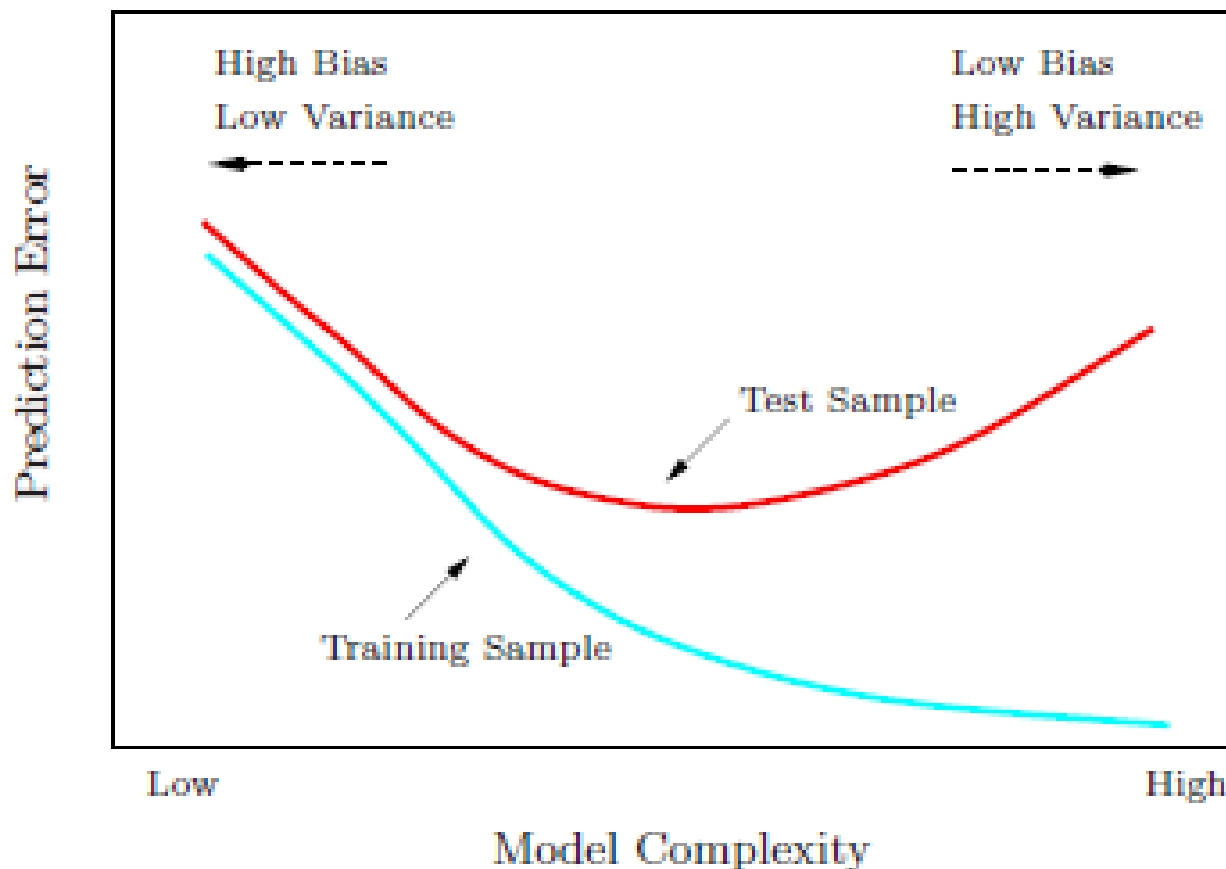
Model Testing

Brian Hartman, PhD ASA
Assistant Professor of Statistics
Actuarial Program Director
Brigham Young University



18th Annual Intercompany Long Term Care Insurance Conference

Traversing the bias-variance trade-off



Hastie et al. 2009



- Within a model, the major concern is whether the data violates any model assumption.
 - Examples of important assumptions
- We can also make statistical comparisons across models to see if the bias-variance trade-off is favorable.
 - Discussion of testing metrics
 - Walk through R examples of testing this
 - Test the models we built above can compare them
 - Show issues with non-penalized GLM



- Beware of R^2
- Measures like AIC and BIC give a combined picture of bias and variance, giving credit for a good fit but penalizing for additional parameters.
- Deviance can be used to determine how the proposed model fits relative to a saturated one.
- Other methods like step-wise variable selection or LASSO



K-fold cross-validation

- Use subset of data to develop coefficients
- Calculate error of predicted values on holdout data
- Average error across the k tests

3-Fold	Test 1	Test 2	Test 3	MSE on holdout data	Average
1 33%	1 Holdout	1 Use	1 Use		Test 1
2 33%	2 Use	2 Holdout	2 Use		Test 2
3 33%	3 Use	3 Use	3 Holdout		Test 3
Calibration data	100%	100%	100%		



- The Poisson GLM can be tested by observing the Deviance and Pearson Chi-Square ratios. Customarily, anything above 1 is considered to be a rejection of the chosen distribution.
- The above issue is related to the phenomenon of over-dispersion. Potential remedies:
 - Use an over-dispersed Poisson model.
 - The Negative Binomial can also be explored.
- Other tests for influential observations, residuals that have patterns, and promoting parsimony are useful techniques.

[Run R 0300: 3.0.9 – 3.0.26]