Evaluating models built in Emblem

Liam McGrath, ACAS, CPCU
WTW
February 22, 2022
Agenda

- Introduction to Emblem
- Correlations
- Residuals
- Summary Graphs
- Questions
Emblem
Supports each stage of the analytics process

Explore your data
A variety of visual summary tools to explore and understand data, including correlations, distributions and average response by factor.

Feature engineering
Feature engineering is often described as the biggest driver of model success and is at the core of how Emblem works.

Modeling and validation
Fit highly flexible and interpretable GLMs. No black box, see and control the complexity. Automation capabilities for interaction detection and feature selection.

Model review and maintenance
A framework that provides an efficient and consistent approach to reviewing and assessing models.
Correlations

- Correlation statistics
  - Measure association between explanatory variables
  - Identify potentially aliased variables (i.e. perfectly/near-perfectly correlated variables)
  - Visualization for easier interpretation
Residuals

Overview

Examining residuals

- Assess the adequacy of the variance function
  - Normality
  - Heteroskedasticity
- Identify mis-fit variables
Residuals

Overview

Residuals measure the error of our individual predictions

- Pearson – basic residual, difference between predicted and observed
  \[ r_p = \frac{y - \mu}{\sqrt{V(\mu)}} \]

- Deviance – accounts for non-normal data distributions
  \[ r_D = \text{sgn}(y - \mu) \sqrt{d} \]

- Standardized Pearson/Deviance - accounts for correlation between predicted and observed
  \[ r_{SP} = \frac{r_p}{\sqrt{(1 - h)}} \quad r_{SD} = \frac{r_D}{\sqrt{(1 - h)}} \]

- Studentized Standardized Pearson/Deviance - accounts for inclusion of scale parameter
  \[ r_{SSP} = \frac{r_{SP}}{\sqrt{\phi}} \quad r_{SSD} = \frac{r_{SD}}{\sqrt{\phi}} \]

- Anscombe – similar to Deviance, accounts for non-normal distributions
Residuals
Impact of error/link

Error Structure: Normal
Link: Identity
Residuals
Impact of error/link

Error Structure: Gamma
Link: Log
Residuals
Without vehicle value in the model…

![Scatter plot showing residuals](image-url)
Residuals

Without vehicle value in the model…
Residuals
With vehicle value added to the model...
Residuals
Discrete distributions
Residuals

Crunched residuals

- How can we make discrete data less discrete? Aggregate!
- Crunched residuals
  - Group observations with similar predicted values into buckets (much less likely that all observations have the same observed value)
  - Calculated residuals at the group level
    \[ r_i = \frac{\sum (Actual - Expected)}{\sqrt{\sum Expected}} \]
Summary Graphs

Overview

[Graph showing predicted values for the number of drivers]
Summary Graphs

Line definitions

- Observed Average

- Fitted Average

- Model Prediction at Base Levels:
  - Represents the effect in the model of this variable, holding all other variables constant
  - Similar to coefficients, can be converted to relativities

- Model Approximate Unsimplified:
  - Representation of the “true” effect of the variable, including both what is modeled and the leftover effect
  - The modeled component (Model Prediction at Base Levels) is multivariate
  - The leftover component is one-way

\[
\text{Model Approximate Unsimplified} = \frac{\text{Model Prediction at Base Levels} \times \text{Observed Average}}{\text{Fitted Average}}
\]

Residual/Leftover