Generalized Linear Model with Elastic Net Regularization for Gamma Distributed Response Variables (glmGammaNet)

Xin Chen, Aleksandr Y. Aravkin, and R. Douglas Martin
Department of Applied Mathematics
University of Washington

June 01, 2018
Limitations of Ordinary Least Squares (OLS)

\[ y = \beta x + \epsilon \]

- Assumes response variable is normally distributed
- Assumes the expectation of the response variable is equal to the linear predictor
- Cannot identify possible sparsity in the data set
- May encounter numerical problems when there is collinearity in the predictor variables
Limitations of Ordinary Least Squares (OLS)

\[ y = \beta x + \epsilon \]

- Assumes response variable is normally distributed
- Assumes the expectation of the response variable is equal to the linear predictor
Limitations of Ordinary Least Squares (OLS)

\[ y = \beta x + \epsilon \]

- Assumes response variable is normally distributed
- Assumes the expectation of the response variable is equal to the linear predictor
- Cannot identify possible sparsity in the data set
Limitations of Ordinary Least Squares (OLS)

\[ y = \beta x + \epsilon \]

- Assumes response variable is normally distributed
- Assumes the expectation of the response variable is equal to the linear predictor
- Cannot identify possible sparsity in the data set
- May encounter numerical problems when there is collinearity in the predictor variables
Generalized Linear Model with Elastic Net Regularization (glmNet)

- Allows the response variable to have distributions other than normal distribution
Generalized Linear Model with Elastic Net Regularization (glmNet)

- Allows the response variable to have distributions other than normal distribution
- Allows nonlinear relationship between the expectation of the response variables and the linear predictor
Generalized Linear Model with Elastic Net Regularization (glmNet)

- Allows the response variable to have distributions other than normal distribution
- Allows nonlinear relationship between the expectation of the response variables and the linear predictor
- Add $\ell_1$ regularization to discover sparsity
Generalized Linear Model with Elastic Net Regularization (glmNet)

- Allows the response variable to have distributions other than normal distribution
- Allows nonlinear relationship between the expectation of the response variables and the linear predictor
- Add $\ell_1$ regularization to discover sparsity
- Add $\ell_2$ regularization to deal with collinearity in the predictor variables

Chen and Martin (2018) used a glmNet model for exponentially-distributed spectral density estimate to compute the standard error of a variety of risk/performance measures
Generalized Linear Model with Elastic Net Regularization (glmNet)

- Allows the response variable to have distributions other than normal distribution
- Allows nonlinear relationship between the expectation of the response variables and the linear predictor
- Add $\ell_1$ regularization to discover sparsity
- Add $\ell_2$ regularization to deal with collinearity in the predictor variables
- Chen and Martin (2018) used a glmNet model for exponentially-distributed spectral density estimate to compute the standard error of a variety of risk/performance measures
# R packages related to glmNet

## Table 1: Comparison of R implementations for GLM

<table>
<thead>
<tr>
<th>Package</th>
<th>Function</th>
<th>Gamma Dist</th>
<th>Model Selection</th>
<th>Multicore</th>
</tr>
</thead>
<tbody>
<tr>
<td>glmnet</td>
<td>glmnet()</td>
<td>No</td>
<td>ElasticNet</td>
<td>Yes</td>
</tr>
<tr>
<td>h2o</td>
<td>h2o.glm()</td>
<td>No</td>
<td>ElasticNet</td>
<td>No</td>
</tr>
<tr>
<td>stats</td>
<td>glm()</td>
<td>Yes</td>
<td>AIC/BIC</td>
<td>Yes</td>
</tr>
<tr>
<td>bestglm</td>
<td>bestglm()</td>
<td>No</td>
<td>Subset AIC/BIC</td>
<td>Yes</td>
</tr>
<tr>
<td>glmGammaNet</td>
<td>glmGammaNet()</td>
<td>Yes</td>
<td>ElasticNet</td>
<td>Yes</td>
</tr>
</tbody>
</table>
glmGammaNet with log link function

- Objective Function with Elastic Regularization

\[ H((\beta; k, y, X, \alpha, \lambda)) = \text{NLL}(\beta; k, y, X) \]
\[ + \lambda \left( \alpha \| \beta \|_1 + \frac{1 - \alpha}{2} \| \beta \|_2^2 \right) \]
glmGammaNet with log link function

- **Objective Function with Elastic Regularization**

\[ H((\beta; k, y, X, \alpha, \lambda) = NLL(\beta; k, y, X) \]
\[ + \lambda \left( \alpha \|\beta\|_1 + \frac{1 - \alpha}{2} \|\beta\|^2 \right) \]

- For given \( \lambda \) and \( \alpha \), we use a modified version of FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) to minimize \( H \).
glmGammaNet with log link function

- Objective Function with Elastic Regularization

\[ H((\beta; k, y, X, \alpha, \lambda)) = NLL((\beta; k, y, X)) + \lambda \left( \alpha \| \beta \|_1 + \frac{1 - \alpha}{2} \| \beta \|_2^2 \right) \]

- For given \( \lambda \) and \( \alpha \), we use a modified version of FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) to minimize \( H \).
- Use Cross Validation to choose appropriate \( \lambda \) from a grid of possible values.
Objective Function with Elastic Regularization

\[ H((\beta; k, y, X, \alpha, \lambda)) = \text{NLL}(\beta; k, y, X) + \lambda \left( \alpha \|\beta\|_1 + \frac{1 - \alpha}{2} \|\beta\|_2^2 \right) \]

For given \(\lambda\) and \(\alpha\), we use a modified version of FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) to minimize \(H\).

Use Cross Validation to choose appropriate \(\lambda\) from a grid of possible values.

Choose \(\lambda\) that corresponds to smallest CV error
glmGammaNet with log link function

- Objective Function with Elastic Regularization

\[
H(\beta; k, y, X, \alpha, \lambda) = NLL(\beta; k, y, X) + \lambda \left( \alpha \| \beta \|_1 + \frac{1 - \alpha}{2} \| \beta \|_2^2 \right)
\]

- For given \( \lambda \) and \( \alpha \), we use a modified version of FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) to minimize \( H \).
- Use Cross Validation to choose appropriate \( \lambda \) from a grid of possible values.
  - Choose \( \lambda \) that corresponds to smallest CV error
  - Choose \( \lambda \) that corresponds to the \( p \)th percentile of CV errors
glmGammaNet with log link function

- Objective Function with Elastic Regularization

\[ H((\beta; k, y, X, \alpha, \lambda) = \text{NLL}(\beta; k, y, X) + \lambda \left( \alpha \| \beta \|_1 + \frac{1 - \alpha}{2} \| \beta \|_2^2 \right) \]

- For given \( \lambda \) and \( \alpha \), we use a modified version of FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) to minimize \( H \).
- Use Cross Validation to choose appropriate \( \lambda \) from a grid of possible values.
  - Choose \( \lambda \) that corresponds to smallest CV error
  - Choose \( \lambda \) that corresponds to the \( p \)th percentile of CV errors
  - Choose largest \( \lambda \) with CV error that is smaller than the sum of the smallest CV error and its standard deviation
glmGammaNet with log link function (Cont’d)

- Gamma distribution is commonly used to model non-negative, positively-skewed, continuous variables

\[ f(y; k, \theta) = \frac{1}{\Gamma(k)\theta^k} y^{k-1} e^{-\frac{y}{\theta}}, \quad k > 0, \theta > 0 \]
Gamma distribution is commonly used to model non-negative, positively-skewed, continuous variables

\[ f(y; k, \theta) = \frac{1}{\Gamma(k)\theta^k} y^{k-1} e^{-\frac{y}{\theta}}, \quad k > 0, \theta > 0 \]

Log link function is an effective way of ensuring positive predicted values

\[ \log E(Y_i) = \log(k\theta_i) = X_i \beta \]
glmGammaNet with log link function (Cont’d)

- Gamma distribution is commonly used to model non-negative, positively-skewed, continuous variables

\[
f(y; k, \theta) = \frac{1}{\Gamma(k)\theta^k} y^{k-1} e^{-\frac{y}{\theta}}, \quad k > 0, \theta > 0
\]

- Log link function is an effective way of ensuring positive predicted values

\[
\log E(Y_i) = \log (k\theta_i) = X_i.\beta
\]

- Negative log-likelihood given observations \( y \) and \( X \) is

\[
NLL(\beta; k, y, X) = \sum_{i=1}^{N} \log \Gamma(k) + k \cdot X_i.\beta
\]

\[
- k \cdot \log k - (k - 1) \log y_i + k \cdot y_i e^{-X_i.\beta}
\]
Numerical Experiment: L1 Error of Fitted Coefficients

We run 1000 Monte Carlo Simulations to demonstrate glmGammaNet performance. 10 out of 15 true coefficients are 0.

<table>
<thead>
<tr>
<th></th>
<th>error.L1</th>
<th>% error.L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>glmGamma</td>
<td>0.41</td>
<td>10.4</td>
</tr>
<tr>
<td>glmGammaNet</td>
<td>0.37</td>
<td>9.3</td>
</tr>
<tr>
<td>glmGammaNet.percentile</td>
<td>0.36</td>
<td>9.2</td>
</tr>
<tr>
<td>glmGammaNet.1sd</td>
<td>0.55</td>
<td>14.0</td>
</tr>
<tr>
<td>glmGammaNet.percentile.nonzero</td>
<td>0.33</td>
<td>8.3</td>
</tr>
<tr>
<td>glmGammaNet.1sd.nonzero</td>
<td>0.23</td>
<td>5.8</td>
</tr>
</tbody>
</table>

Table 2: L1 Error of Fitted Coefficients for Different GLM methods
Numerical Experiment: Variable Selection

Figure 1: Histogram of number of zero coefficients selected over 1000 simulations
### Table 3: Fitted glmGammaNet Coefficients for Hedge Fund Returns

<table>
<thead>
<tr>
<th></th>
<th>beta0</th>
<th>beta1</th>
<th>beta2</th>
<th>beta3</th>
<th>beta9</th>
<th>beta10</th>
<th>beta11</th>
<th>beta12</th>
<th>zero.coeffs</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTAG</td>
<td>-5.929</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>DIS</td>
<td>-2.982</td>
<td>-2.718</td>
<td>-3.925</td>
<td>2.087</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>EM</td>
<td>-2.055</td>
<td>-2.916</td>
<td>8.393</td>
<td>-6.364</td>
<td>-0.059</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>ED</td>
<td>-3.448</td>
<td>-1.903</td>
<td>-0.265</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>FIA</td>
<td>-2.43</td>
<td>-3.372</td>
<td>-3.619</td>
<td>-0.019</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>GM</td>
<td>-6.171</td>
<td>-4.017</td>
<td>5.607</td>
<td>4.989</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>SS</td>
<td>-2.979</td>
<td>-9.025</td>
<td>21.119</td>
<td>-4.339</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>FoF</td>
<td>-3.761</td>
<td>-2.204</td>
<td>-1.917</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>
References

- R package available at https://github.com/chenx26/glmGammaNet
References

- R package available at https://github.com/chenx26/glmGammaNet
R package available at
https://github.com/chenx26/glmGammaNet
