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

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 ℓ_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 ℓ_1 regularization to discover sparsity
- Add ℓ_2 regularization to deal with collinearity in the predictor variables

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 ℓ_1 regularization to discover sparsity
- Add ℓ_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

Package	Function	Gamma Dist	Model Selection	Multicore
glmnet	glmnet()	No	ElasticNet	Yes
h2o	h2o.glm()	No	ElasticNet	No
stats	glm()	Yes	AIC/BIC	Yes
bestglm	bestglm()	No	Subset AIC/BIC	Yes
glmGammaNet	**glmGammaNet()**	**Yes**	**ElasticNet**	**Yes**

Table 1: Comparison of R implementations for GLM

glmGammaNet with log link function

- Objective Function with Elastic Regularization

$$
\begin{aligned}
H((\beta; k, y, X, \alpha, \lambda) =& NLL(\beta; k, y, X) \\
&+ \lambda \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right)
\end{aligned}
$$

glmGammaNet with log link function

- Objective Function with Elastic Regularization

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

- For given λ and α, 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 λ and α, we use a modified version of FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) to minimize H.
- Use Cross Validation to choose appropriate λ from a grid of possible values.

glmGammaNet with log link function

- Objective Function with Elastic Regularization

$$
\begin{aligned}
H((\beta; k, y, X, \alpha, \lambda) = & NLL(\beta; k, y, X) \\
& + \lambda \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right)
\end{aligned}
$$

- For given λ and α, we use a modified version of FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) to minimize H.
- Use Cross Validation to choose appropriate λ from a grid of possible values.
  - Choose λ 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 λ and α, we use a modified version of FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) to minimize H.
- Use Cross Validation to choose appropriate λ from a grid of possible values.
  - Choose λ that corresponds to smallest CV error
  - Choose λ that corresponds to the pth percentile of CV errors

glmGammaNet with log link function

- Objective Function with Elastic Regularization

$$
\begin{aligned}
H((\beta; k, y, X, \alpha, \lambda) =& NLL(\beta; k, y, X) \\
&+ \lambda \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right)
\end{aligned}
$$

- For given λ and α, we use a modified version of FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) to minimize H.
- Use Cross Validation to choose appropriate λ from a grid of possible values.
  - Choose λ that corresponds to smallest CV error
  - Choose λ that corresponds to the pth percentile of CV errors
  - Choose largest λ 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

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

glmGammaNet with log link function (Cont'd)

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

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

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

$$\log \mathrm{E}(Y_i) = \log(k\theta_i) = X_{i\cdot}\beta$$

glmGammaNet with log link function (Cont'd)

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

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

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

$$\log \mathrm{E}(Y_i) = \log(k\theta_i) = X_{i\cdot}\beta$$

- Negative log-likelihood given observations y and X is

$$\begin{aligned} NLL(\beta; k, y, X) = \sum_{i=1}^{N} & \log\Gamma(k) + k \cdot X_{i\cdot}\beta \\ & - k \cdot \log k - (k-1)\log y_i + k \cdot y_i e^{-X_{i\cdot}\beta} \end{aligned}$$

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.

	error.L1	% error.L1
glmGamma	0.41	10.4
glmGammaNet	0.37	9.3
glmGammaNet.percentile	0.36	9.2
glmGammaNet.1sd	0.55	14.0
glmGammaNet.percentile.nonzero	0.33	8.3
glmGammaNet.1sd.nonzero	0.23	5.8

Table 2: L1 Error of Fitted Coefficients for Different GLM methods

Numerical Experiment: Variable Selection

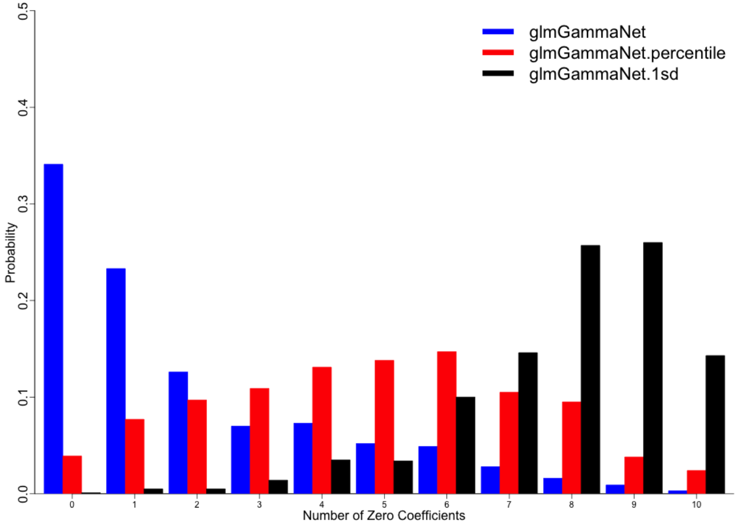

Figure 1: Histogram of number of zero coefficients selected over 1000 simulations

Expected Shortfall Standard Error using glmGammaNet

	beta0	beta1	beta2	beta3	beta9	beta10	beta11	beta12	zero.coeffs
CTAG	-5.929	0	0	0	0	0	0	0	12
DIS	-2.982	-2.718	-3.925	2.087	0	0	0	0	5
EM	-2.055	-2.916	8.393	-6.364	-0.059	0	0	0	3
EMN	-3.759	-9.592	-21.114	140.738	-20.777	-10.717	-5.338	-2.578	0
ED	-3.448	-1.903	-0.265	0	0	0	0	0	10
FIA	-2.43	-3.372	-3.619	-0.019	0	0	0	0	7
GM	-6.171	-4.017	5.607	4.989	0	0	0	0	5
SS	-2.979	-9.025	21.119	-4.339	0	0	0	0	4
FoF	-3.761	-2.204	-1.917	0	0	0	0	0	8

Table 3: Fitted glmGammaNet Coefficients for Hedge Fund Returns

References

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

References

- R package available at https://github.com/chenx26/glmGammaNet
- Chen, X., A. Y. Aravkin and R. D. Martin (2018). "Generalized Linear Model for Gamma Distributed Variables via Elastic Net Regularization". arXiv preprint arXiv:1804.07780

References

- R package available at https://github.com/chenx26/glmGammaNet
- Chen, X., A. Y. Aravkin and R. D. Martin (2018). "Generalized Linear Model for Gamma Distributed Variables via Elastic Net Regularization". arXiv preprint arXiv:1804.07780
- Chen, X. and R. D. Martin, Standard Errors of Risk and Performance Estimators with Serially Correlated Returns (January 30, 2018). Available at SSRN: https://ssrn.com/abstract=3085672 or http://dx.doi.org/10.2139/ssrn.3085672