The Journal of Open Source Software

GENRE (GPU Elastic-Net REgression): A CUDA-Accelerated Package for Massively Parallel Linear Regression with Elastic-Net Regularization

Christopher Khan[1] **and Brett Byram**[1]

1 Vanderbilt University

DOI: 10.21105/joss.02475

Software

- Review
- Repository
- Archive

Editor: Pending Editor

Submitted: 10 July 2020
Published: 12 July 2020

Summary

GENRE (GPU Elastic-Net REgression) is a package that allows for many instances of linear regression with elastic-net regularization to be processed in parallel on a GPU by using the C programming language and NVIDIA's (NVIDIA Corporation, Santa Clara, CA) Compute Unified Device Architecture (CUDA) parallel programming framework. Linear regression with elastic-net regularization (Zou & Hastie, 2005) is a widely utilized tool when performing model-based analyses. The basis of this method is that it allows for a combination of L1-regularization and L2-regularization to be applied to a given regression problem. Therefore, feature selection and coefficient shrinkage are performed while still allowing for the presence of groups of correlated features. The process of performing these model fits can be computationally expensive, and one of the fastest packages that is currently available is glmnet (Friedman, Hastie, & Tibshirani, 2010), (Qian, Hastie, Tibshirani, & Simon, 2013), (Hastie & Qian, 2014). This package provides highly efficient Fortran implementations of several different types of regression. In the case of its implementation of linear regression with elastic-net regularization, the objective function shown in (eq. 1) is minimized.

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{2N} \sum_{i=1}^{N} \left(\boldsymbol{y}_i - \sum_{j=1}^{P} \boldsymbol{X}_{ij} \boldsymbol{\beta}_j \right)^2 + \lambda \left(\alpha \left\| \boldsymbol{\beta} \right\|_1 + \frac{(1-\alpha) \left\| \boldsymbol{\beta} \right\|_2^2}{2} \right) \tag{1}$$

To minimize this objective function, cyclic coordinate descent is utilized as the optimization algorithm. This algorithm consists of minimizing the objective function with respect to one model coefficient at a time. Cycling through all of the coefficients results in one iteration, and this process continues until specified convergence criteria are satisfied. As previously stated, glmnet is highly efficient for single model fits, but performing thousands of these fits will still require significant computational time due to each one being executed in a serial fashion on a CPU. However, by using GENRE, massively parallel processing can be performed in order to achieve significant speedup. This is due to the fact that modern GPUs consist of thousands of computational cores that can be utilized. Moreover, although the processing in GENRE is performed using the C programming language and CUDA, a MEX-interface is included to allow for this code to be called within the MATLAB (The MathWorks, Inc., Natick, MA) programming language for convenience. This also means that with modification, the MEX-interface can be replaced with another interface if it is desired to call the C/CUDA code in another language, or the C/CUDA code can be utilized without an interface.

Statement of Need

The core motivation for developing GENRE was that many of the available packages for performing linear regression with elastic-net regularization focus on achieving high performance in terms of computational time or resource consumption for single model fits. However, they often do not address the case in which there is a need to perform many model fits in parallel. For example, the research project that laid the foundation for GENRE involved performing ultrasound image reconstruction using an algorithm called Aperture Domain Model Image REconstruction (Byram & Jakovljevic, 2014), (Byram, Dei, Tierney, & Dumont, 2015), (Dei & Byram, 2017). This algorithm is computationally expensive due to the fact that in one stage, it requires thousands of instances of linear regression with elastic-net regularization to be performed in order to fit models of ultrasound data. Originally, this algorithm was implemented on a CPU, and it typically required several minutes to reconstruct one ultrasound image. The primary bottleneck was performing all of the required model fits due to the fact that glmnet was used to compute each fit serially. However, a GPU implementation of the algorithm was developed, and this implementation provided a speedup of over two orders of magnitude, which allowed for multiple ultrasound images to be reconstructed per second. The main contributor to this speedup was the fact that the model fits were performed in parallel on the GPU.

Aside from this application, there are a number of other applications that can potentially benefit from having the ability to perform model fits in a massively parallel fashion, which is why the code was developed into a package. For example, linear regression with elastic-net regularization has been applied to the field of genomics in order to develop predictive models that utilize genetic markers (Ogutu, Schulz-Streeck, & Piepho, 2012), (Waldmann, Mészáros, Gredler, Fuerst, & Sölkner, 2013). In addition, like ADMIRE, there are a variety of other signal processing applications. For example, this regression method has been used to create models of fMRI data in order to predict the mental states of subjects and provide insight into neural activity (Carroll, Cecchi, Rish, Garg, & Rao, 2009). Moreover, another signal processing example is that linear regression models with elastic-net regularization have been used in combination with hidden Markov random field segmentation to perform CT estimation for the purposes of MRI-based attenuation correction for PET/MRI (Chen et al., 2014). Now, by using GENRE, the models in each of the aforementioned examples can be computed in parallel in order to reduce the amount of processing time that is required.

Acknowledgements

This work was supported by NIH grants R01EB020040 and S10OD016216-01 and NAVSEA grant N0002419C4302.

References

Byram, B., Dei, K., Tierney, J., & Dumont, D. (2015). A model and regularization scheme for ultrasonic beamforming clutter reduction. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, *62*(11), 1913–1927. doi:10.1109/tuffc.2015.007004

Byram, B., & Jakovljevic, M. (2014). Ultrasonic multipath and beamforming clutter reduction: A chirp model approach. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, *61*(3), 428–440. doi:10.1109/tuffc.2014.2928

Carroll, M., Cecchi, G., Rish, I., Garg, R., & Rao, A. (2009). Prediction and interpretation of distributed neural activity with sparse models. *NeuroImage*, *44*(1), 112–122. doi:10.1016/j.neuroimage.2008.08.020

Chen, Y., Juttukonda, M., Lee, Y., Su, Y., Espinoza, F., Lin, W., Shen, D., et al. (2014). MRI based attenuation correction for pet/mri via mrf segmentation and sparse regression estimated ct. *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. doi:10.1109/isbi.2014.6868131

Dei, K., & Byram, B. (2017). The impact of model-based clutter suppression on cluttered, aberrated wavefronts. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, *64*(10), 1450–1464. doi:10.1109/tuffc.2017.2729944

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1). doi:10.18637/jss.v033.i01

Hastie, T., & Qian, J. (2014). Glmnet vignette. Retrieved from http://www.web.stanford.edu/~hastie/Papers/Glmnet_Vignette.pdf

Ogutu, J. O., Schulz-Streeck, T., & Piepho, H. P. (2012). Genomic selection using regularized linear regression models: Ridge regression, lasso, elastic net and their extensions. *BMC Proceedings*, *6*(S2). doi:10.1186/1753-6561-6-s2-s10

Qian, J., Hastie, T., Tibshirani, R., & Simon, N. (2013). Glmnet for matlab. Retrieved from http://www.stanford.edu/~hastie/glmnet_matlab/

Waldmann, P., Mészáros, G., Gredler, B., Fuerst, C., & Sölkner, J. (2013). Evaluation of the lasso and the elastic net in genome-wide association studies. *Frontiers in Genetics*, *4*. doi:10.3389/fgene.2013.00270

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320. doi:10.1111/j.1467-9868.2005.00503.x