IJARIIT INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 4)
Available online at: www.ijariit.com

Improved sparse logistic regression for efficient feature selection

Shreyal Gajare
shreyalgajare@gmail.com
Maharashtra Institute of Technology, Pune, Maharashtra

Shilpa Sonawani
shilpasonwani@mit.edu.in
Maharashtra Institute of Technology, Pune, Maharashtra

ABSTRACT

Variable and feature selection have become the focus of most of the research areas of application for which datasets with hundreds of thousands of parameters are available. These areas may include health risk prediction, text processing of internet documents, gene array analysis etc. Features gathered for analysis may not be completely informative; some may contain noise, some may require normalization or most of them can be irrelevant. The goal of feature selection is to improve predictions of predictors, provide faster & cost effective predictors and provide a better understanding of the underlying process and the data. In feature selection, sparse logistic regression method serves best for countably selecting the important attributes. The logistic loss function is included for sparsity purpose. The regularization parameter is responsible for controlling the sparsity factor. It provides the most effective and relevant variable selection for a specific model. Most of the applications where prediction is required, this method serves its best purpose.

Keywords*: Feature selection, Regression Methods, the Loss function*

1. INTRODUCTION

Prediction of a specific task using the neural network is one of the most widely performed operations. In particular, the high dimensional nature of many modeling tasks in bioinformatics, going from sequence analysis over microarray analysis to spectral analyses and literature mining has given rise to a wealth of feature selection techniques being presented in the field.

Hence, feature selection methods if applied prior to the prediction model will definitely help in improving the model performance along with accuracy. Feature selection is one part from dimensionality reduction which helps in reducing the attributes which are not required for further processing. Feature selection can be applied on both supervised (classification) as well as unsupervised (clustering) data. But here the focus is on supervised data where the class labels are known beforehand.

2. FEATURE SELECTION

In machine learning, feature selection is also known as variable selection or attribute selection. It is the process of selecting a subset of relevant features for prediction model construction. It removes various irrelevant features from the set. Feature selection is classified into three main categories as shown in the figure.

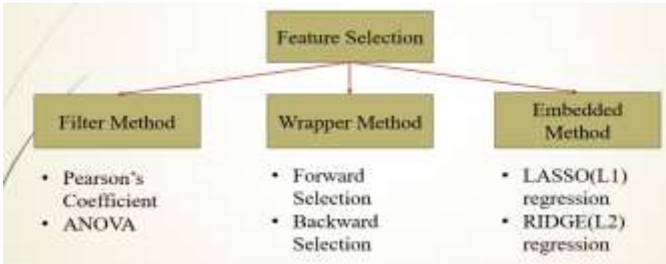

Fig. 1: Feature selection classification

A. **Filter Method:** Filter method selects the features regardless of the model used. It selects the features based on discriminating criteria that are relatively dependent on classification. Features with higher variance may get selected assuming they may contain useful data. Thus, feature variable and target variable relationship are not maintained. Examples of this type are information gain, chi-square test, Fisher score, correlation coefficient and variance threshold.
B. **Wrapper Method:** Wrapper methods select a subset of features and train the model using them. Based on the inferences, the features should be added or removed is decided. The problem is reduced to a search problem as every time new addition of features leads to training and cross-validating the data repetitively. These methods are computationally very expensive. Examples of wrapper methods are forward selection, backward selection, and recursive feature elimination.
C. **Embedded Method:** Embedded methods combine the advantages of both filter and wrapper methods. These are implemented with the algorithms possessing their own built-in feature selection strategy. Examples of these methods are L1 regularization (LASSO), L2 regression (RIDGE) and elastic regression along with regularized trees, random multinomial logit.

So, among all the methods above embedded methods serves the best purpose for selecting the features. Filter and wrapper methods are prone to overfitting and are not applicable on a higher scale. Embedded methods possess the regression techniques which contains the loss function or regularization parameter which avoids overfitting of attributes and takes care of sparsity.

3. PROPOSED FEATURE SELECTION APPROACH

For selecting only required features regression techniques are widely used for high dimensional data. L1 sparse logistic regression is used for generalized linear models and can be understood as adding penalty against complexity to reduce the degree of overfitting or variance of the model by adding more bias.

In this process first, the estimator is trained on the initial dataset. The estimator used is a logistic regression,

$$\mathrm{P}\ (\mathrm{y}|\mathrm{x}) = \frac{1}{1+\exp(-y\,(w^T x + b))}$$

Where, x is the input features, y is the class label. The training procedure is responsible for minimizing the loss function.

$$\mathrm{LL}(\mathrm{w},\mathrm{b}) = \frac{1}{n}\sum_{i=1}^{n} \log(1+\exp(-y_i\,(w^T x_i + b)))$$

Then logistic loss function is used in this procedure to avoid overfitting of parameters and hold sparsity.

Minimize $\quad \sum_{xy} \log(\,1+\exp(-w^T x\,.\,y)) + \lambda w^T w$

Sparsity refers that only a few entries in a matrix remain non zero. L1 regression has the property of producing many coefficients with zero values or very small values with few large coefficients. Sparser solutions are good for feature selection in high dimensional spaces as well as for prediction speed.

λ is the regularization parameter which controls the sparsity of the model. Higher the value of λ, large the number of attributes has zero coefficients. Then ultimately feature score is calculated as,

$$S_k = \frac{1}{n}\sum_{i}^{n} w_i^k$$

Hence, this methodology is applied to select relevant features before undergoing a prediction task.

3.1 System Evaluation

Based on the embedded methods of feature selection the analysis is done. In regression L1, L2, as well as elastic net regularization methods, are considered for comparing the performance results. The two factors most concerned are the loss function and the regularization parameter.

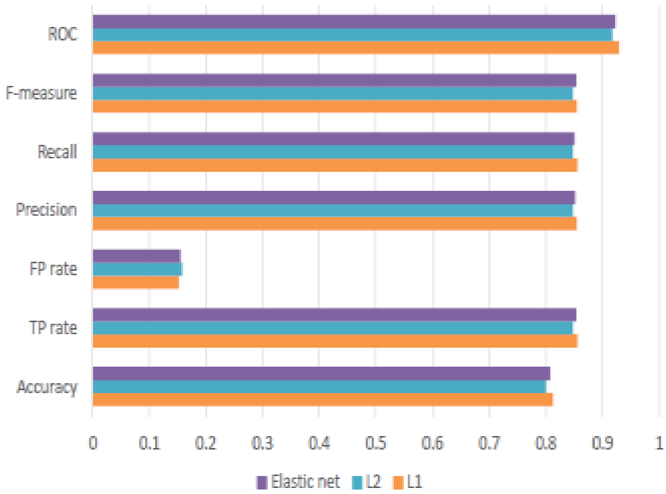

Fig. 2: Evaluation results

So, according to the table, L1 is more robust as compared to L2 and elastic regression. L2 always produces only one solution and nonsparse outputs. Elastic net regularization and L1 regularization shows approximately similar results, but as elastic is the combination of L1 and L2, L1 is better than elastic regularization. L2 is not used for feature selection purpose.

Thus, the evaluation denotes that sparse logistic regression provides good prediction accuracy since shrinking and removing the coefficients can reduce the variance. This increases model interpretability by eliminating irrelevant variables also reducing overfitting.

4. CONCLUSION

Although various feature selection methods are available for selection of relevant attributes, it should be verified with correct measures. The goal of using feature selection initially on the dataset reduces the repetitive training on data and increase the accuracy of the model. Thus L1 sparse logistic regression is suited to sparser solutions.

In future work, methodology combining improved logistic regression along with elastic net regularization can be worked upon. This can prove effective for grouped selection of parameters.

5. REFERENCES

[1] Milad Nezhada, Dongxiao Zhu, “SAFS: A Deep Feature Selection Approach for Precision Medicine”, IEEE International Journal of Bioinformatics and Biomedicine (BIBM), 2016.

[2] Tapio Manninen, Heikki Huttunen, “Leukemia Prediction Using Sparse Logistic Regression”, PLoS ONE 8(8): e72932. Journal.pone.0072932, 2013.

[3] Roman Zakharov, Pierre Dupont, “Ensemble Logistic Regression for Feature Selection”, IAPR International Conference on Pattern Recognition in Bioinformatics, pp 133-144, Springer, 2011.

[4] Zena M. Hira, Duncan F. Gillies, “ A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data”, Advances in Bioinformatics, Hindawi Publishing Corporation, vol. 10, pp.13, 2015.

[5] Felix Abramovich, Vadim Grinshtein, “High – dimensional classification by sparse logistic regression”, Published by Oxford University Press on Bioinformatics, vol. 34, pp. 485 – 493, 2018.

[6] Jiayu Zhou, Fei Wang, “From micro to macro: Data driven Phenotyping by Densification of Longitudinal Electronic medical records”, KDD '14 Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 135 – 144.

[7] Ales Zamuda, Christine Zarges, Gregor Stiglic, “Stability Selection using Genetic Algorithm and Logistis Linear Regression on Healthcare Records”, Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 143 – 144, 2017.

[8] Fei Wang, Ping Zhang, “Clinical Risk Prediction with Multilinear Sparse Logistic Regression”, KDD '14 Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 145 – 154.

