

POSTER PRESENTATION

Open Access

# Development of sparse Bayesian multinomial generalized linear model for multi-class prediction

Behrouz Madahian<sup>1</sup>, Lih Y Deng<sup>1</sup>, Ramin Homayouni<sup>2,3\*</sup>

From UT-KBRIN Bioinformatics Summit 2014  
Cadiz, KY, USA. 11-13 April 2014

## Background

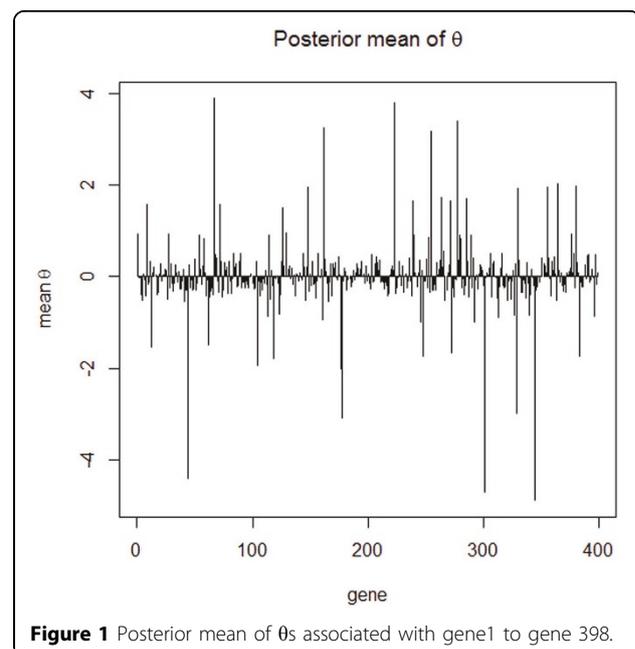
Gene expression profiling has been used for many years to classify samples and to gain insights into the molecular mechanisms of phenotypes and diseases. A major challenge in expression analysis is caused by the large number of variables assessed compared to relatively small sample sizes. In addition, identification of markers that accurately predict multiple classes of samples, such as those involved in the progression of cancer or other diseases, remains difficult.

## Materials and methods

In this study, we developed a multinomial Probit Bayesian model which utilized the double exponential prior to induce shrinkage and reduce the number of covariates in the model [1,2]. A fully Bayesian hierarchical model was developed in order to facilitate Gibbs sampling which takes into account the progressive nature of the response variable. Gibbs sampling was performed in R for 100k iterations and the first 20k were discarded as burn-in. The method was applied to a published dataset on prostate cancer progression downloaded from Gene Expression Omnibus at NCBI (GSE6099) [3]. The data set contained 99 prostate cancer cell types in four different progressive stages. The dataset was randomly divided into training (N=50) and test (N=49) groups such that each group contained an equal number of each cell type. Before applying our model, for each gene we performed ordinal logistic regression. Genes were ranked based on the p-value of association. Using a cut-off value of 0.05 after Benjamini and Hochberg FDR correction resulted in a final set of 398 genes.

## Results

Figure 1 shows the posterior mean of parameters associated with each gene. Using the top ten genes obtained from our model, we were able to achieve 86% classification accuracy in the training group and 82% accuracy in the test group. To test the robustness of the model, we switched the training and test groups and evaluated the classification accuracy. We obtained 88% classification accuracy on the new training group and 86% accuracy on the new test group. The classification accuracy by tumor type is shown in Table 1. Taken together, these results suggest that the Bayesian Multinomial Probit model applied to cancer progression data allows for reasonable subclass prediction.



**Figure 1** Posterior mean of  $\theta$ s associated with gene1 to gene 398.

\* Correspondence: rhomayon@memphis.edu

<sup>2</sup>Bioinformatics Program, University of Memphis, Memphis, TN, 38152, USA  
Full list of author information is available at the end of the article

**Table 1 Classification accuracy of prostate cancer subtypes in the train and test groups.**

Sample type	Train group	Test group	Train group Switched	Test group switched
Benign	100	94	100	100
PIN	29	33	50	71
PCA	86	86	100	100
MET	100	100	70	50

## Conclusion

Our future plan is to perform resampling on the selection of training and test groups in order to obtain more robust results and to compare the performance of the model to other popular classifiers such as Support Vector Machine and Random Forest.

## Acknowledgments

This work was supported by the University of Memphis Center for Translational Informatics and the Assisi Foundation of Memphis.

## Authors' details

<sup>1</sup>Department of Mathematical Sciences, University of Memphis, Memphis, TN, 38152, USA. <sup>2</sup>Bioinformatics Program, University of Memphis, Memphis, TN, 38152, USA. <sup>3</sup>Department of Biology, University of Memphis, Memphis, TN 38152, USA.

Published: 29 September 2014

## References

1. Park T, Casella G: **The Bayesian lasso.** *J Am Stat Assoc* 2008, **103**:681-686.
2. Albert J, Chib S: **Bayesian analysis of binary and polychotomous response data.** *J Am Stat Assoc* 1993, **88**:669-679.
3. Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, Kalyana-Sundaram S, Wei JT, Rubin MA, Pienta KJ, Shah RB, Chinnaiyan AM: **Integrative molecular concept modeling of prostate cancer progression.** *Nat Genet* 2007, **39**(1):41-51.

doi:10.1186/1471-2105-15-S10-P14

**Cite this article as:** Madahian et al.: Development of sparse Bayesian multinomial generalized linear model for multi-class prediction. *BMC Bioinformatics* 2014 **15**(Suppl 10):P14.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

