High Dimensional Logistic Regression Model
using Adjusted Elastic Net Penalty

Zakariya Yahya Algamal
Department of Mathematical Sciences, Universiti Teknologi Malaysia
81310 Skudai, Johor, Malaysia
zak.sm_stat@yahoo.com

Muhammad Hisyam Lee
Department of Mathematical Sciences, Universiti Teknologi Malaysia
81310 Skudai, Johor, Malaysia
mhl@utm.my

Abstract

Reduction of the high dimensional binary classification data using penalized logistic regression is one of the challenges when the explanatory variables are correlated. To tackle both estimating the coefficients and performing the variable selection simultaneously, elastic net penalty was successfully applied in high dimensional binary classification. However, elastic net has two major limitations. First it does not encourage grouping effects when there is no high correlation. Second, it is not consistent in variable selection. To address these issues, an adjusted of the elastic net (AEN) and its adaptive adjusted elastic net (AAEM), are proposed to take into account the small and medium correlation between explanatory variables and to provide the consistency of the variable selection simultaneously. Our simulation and real data results show that AEN and AAEN have advantage with small, medium, and extremely correlated variables in terms of both prediction and variable selection consistency comparing with other existing penalized methods.

Keywords: High dimensional, Penalization, Logistic regression, LASSO, Elastic net.

1. Introduction

With the advancement of technologies, massive amount of data with increasing dimensions have been generated in many areas such as genetics, medical, economic and social sciences. The expansion of the data is in two dimensions: the number of variables and the number of observations. "High dimensional data" refers to the situation where the number of variables measured is greater than the number of observations in the data. This differs from traditional datasets for statistical analysis where we have many observations on a few variables. Such high dimensional data has posed new challenges to statistical analysis, because a lot of conventional statistical methods do not automatically apply into these datasets, for example, the curse of dimensionality makes many classical regression models, such as logistic regression, ineffective, because statistical issues associated with modeling high dimensional data include model overfitting, estimation instability, computational difficulty (Pourahmadi, 2013).

How to reduce the dimensionality has been an important research question in statistical applications. One way to handle the high dimensional data is to perform data reduction. To do this, various penalized methods have been proposed begin by ridge penalty (Hoerl

& Kennard, 1970). It estimates the regression coefficients through ℓ_2-norm penalty. It is well-known that ridge regression shrinks the coefficients of correlated predictor variables toward each other, allowing them to borrow strength from each other (Friedman, Hastie, & Tibshirani, 2010). The least absolute shrinkage and selection operator (LASSO) was proposed by Tibshirani (1996) to estimate the regression coefficients through ℓ_1-norm penalty. Zou and Hastie (2005) proposed the elastic net penalty which is based on a combined penalty of LASSO and ridge regression penalties in order to overcome the drawbacks of using the LASSO and ridge regression on their own.

Usually, in high dimensional data the explanatory variables are correlated. If there is a group of highly correlated variables, the LASSO will randomly select only one variable from this group and drop the rest whereas elastic net will select the whole group of the highly correla-ted explanatory variables (Zou & Hastie, 2005; Zhou, 2013). Analogously, Bondell and Reich (2008) proposed a penalty called OSCAR to encourage selection of a group of highly correlated explanatory variables. Elastic net often performs better than LASSO in terms of prediction error when there is correlation among variables, also OSCAR has a comparable performance similar to elastic net (Zeng & Xie, 2011). Tutz and Ulbricht (2009) proposed correlation-based penalty to deal with grouping effects. This penalty just makes variable shrinkage rather than variable selection. Elastic net penalty lacks consistent variable selection (oracle property), even though it outperforms LASSO. Zou and Zhang (2009) proposed adaptive elastic net to handle grouping effects and enjoying oracle property simultaneously. El Anbari and Mkhadri (2014) explained though experimental studies that elastic net seems to be slightly less reliable if the correlation between explanatory variables is not so extreme (i.e. $|\rho| \leq 0.95$). In this paper, an adjusted of the elastic net (AEN) and its adaptive adjusted elastic net (AAEM), are proposed to take into account the small and medium correlation between explanatory variables and to provide the consistency of the variable selection simultaneously. The remainder of this paper organizes as follows. Section 2 covers the penalized logistic regression methods. Description of the AEN and AAEM is explained in section 3. Sections 4 and 5 are devoted to simulation studies and results. While section 6 covered the real data analysis. We end this paper with a conclusion in section 7.

2. Penalized Logistic Regression Model

Logistic regression is a statistical method to model a binary classification problem. The regression function has a nonlinear relation with the linear combination of the explanatory variables. In binary classification, the response variable of the logistic regression has two values either 1 for the positive class or 0 for the negative class. Let $\mathbf{y}_i \in \{0,1\}$ be a vector of size $n \times 1$ of tissues, and let $\mathbf{x}_j$ be a $p \times 1$ vector of explanatory variables. The logistic transformation of the vector of probability estimates $\pi_i = p(y_i = 1 | \mathbf{x}_j)$ is modeled by a linear function, logit transformation:

$$\ln[\frac{\pi_i}{1-\pi_i}] = \beta_0 + \sum_{j=1}^{p} \mathbf{x}_j^T \beta_j, \ i = 1, 2, ..., n, \tag{1}$$

where β_0 is the intercept and β_j is a $p \times 1$ vector of unknown explanatory variable coefficients. The log-likelihood function of (1) is defined as

$$\ell(\beta_0, \beta) = \sum_{i=1}^{n} \left\{ \mathbf{y}_i \ln \pi(\mathbf{x}_{ij}) + (1 - \mathbf{y}_i) \ln(1 - \pi(\mathbf{x}_{ij})) \right\}, \ j = 1, 2, ...p. \tag{2}$$

Logistic regression offers the advantage of simultaneously estimating the probabilities $\pi(\mathbf{x}_{ij})$ and $1 - \pi(\mathbf{x}_{ij})$ for each class and classifying subjects. The probability of classifying the ith sample in class 1 is estimated by $\hat{\pi}_i = \exp(\beta_0 + \sum_{j=1}^{p} \mathbf{x}_j^T \beta_j) / 1 + \exp(\beta_0 + \sum_{j=1}^{p} \mathbf{x}_j^T \beta_j)$. The predicted class is then obtained by $I\{\hat{\pi}_i > 0.5\}$, where $I(\Box)$ is an indicator function.

Penalized logistic regression (PLR) adds a nonnegative penalty term to Eq. (1), such that the size of the explanatory variables coefficients in high dimension can be controlled. Several penalty terms have been discussed in the literature (Li, Jia, & Zhao, 2013; Tibshirani, 1996; Zhenqiu et al., 2007). The ℓ_1-norm penalty, proposed by Tibshirani (Tibshirani, 1996), is one of the popular penalization terms. The ℓ_1-norm penalty performs explanatory variables selection and estimation simultaneously by constraining the log-likelihood function of variables coefficients. Thus, the PLR is defined as:

$$PLR = \sum_{i=1}^{n} \left\{ \mathbf{y}_i \ln \pi(\mathbf{x}_{ij}) + (1 - \mathbf{y}_i) \ln(1 - \pi(\mathbf{x}_{ij})) \right\} + \lambda P(\beta). \tag{3}$$

The estimation of the vector β is obtained by minimizing Eq. (3)

$$\hat{\beta}_{PLR} = \arg\min_{\beta} \left[\sum_{i=1}^{n} \left\{ \mathbf{y}_i \ln \pi(\mathbf{x}_{ij}) + (1 - \mathbf{y}_i) \ln(1 - \pi(\mathbf{x}_{ij})) \right\} + \lambda P(\beta) \right],$$

(4)

where $\lambda P(\beta)$ is the penalty term that regularized the estimates. The penalty term depends on the positive tuning parameter, λ, which controls the tradeoff between fitting the data to the model and the effect of the penalization. In other words, it controls the amount of shrinkage. For the $\lambda = 0$, we obtain the maximum likelihood estimation (MLE) solution. In contrast, for large values of λ the influence of the penalization term on the coefficient estimates increases. Choosing the tuning parameter is an important part of the model fitting. If the focus is on classification, the tuning parameter should find the right balance between the bias and variance to minimize the misclassification error. Without loss of generality, it is assumed that the explanatory variables are standardized, $\sum_{i=1}^{n} x_{ij} = 0$ and $(n^{-1}) \sum_{i=1}^{n} x^2_{ij} = 1, \quad \forall j \in \{1, 2, ..., p\}$. As a result, the intercept β_0 is not penalized. The estimation of the vector β using the LASSO (ℓ_1-norm penalization) is defined as:

$$\hat{\beta}_{LASSO} = \arg\min_{\beta} \left[\sum_{i=1}^{n} \left\{ \mathbf{y}_i \ln \pi(\mathbf{x}_{ij}) + (1-\mathbf{y}_i)\ln(1-\pi(\mathbf{x}_{ij})) \right\} + \lambda \sum_{j=1}^{p} |\beta_j| \right], \qquad (5)$$

where λ is a tuning parameter. It reduces to the MLE estimator when $\lambda = 0$. On the other hand, if $\lambda \to \infty$, the penalization term forces all the explanatory variables to be zeros. In practice, the value of λ is often chosen by a cross-validation procedure. To solve Eq. (5), the traditional numerical methods are through MLE or the Newton-Raphson's algorithm. However, the computation of these methods is prohibitive when the number of explanatory variables is large (Zhu & Hastie, 2004). Equation (5) can be efficiently solved by the coordinate descent algorithm (Friedman et al., 2010; Park & Hastie, 2008).

The LASSO has an advantage in that it is computationally feasible in high dimensional classification data. On the other hand, the LASSO has three main drawbacks. First of all, if $p > n$ (i.e. the explanatory variables are greater than the number of samples), the LASSO selects at most n variables because of the nature of the convex optimization problem. In addition, the LASSO cannot handle the effect of grouping. When the pairwise correlations among a group of explanatory variables are very high, then the LASSO tends to select only one variable from the whole group and does not take into account which one is selected (Zeny, Xiaojian, Sanjeena, & Paul, 2012). Lastly, the LASSO lacks the oracle properties, as stated in Fan and Li (Fan & Li, 2001).

Elastic net is a penalization method for explanatory variables selection, which is introduced by Zou and Hastie (2005) to deal with the first two drawbacks of LASSO. Elastic net tries to merge the ℓ_2-norm and the ℓ_1-norm penalizations, by using ridge regression penalty to deal with high correlation problem while taking advantage of LASSO penalization in variable selection property. The PLR using elastic net penalty is defined by

$$\hat{\beta}_{Elastic} = \arg\min_{\beta} \left[\sum_{i=1}^{n} \left\{ \mathbf{y}_i \ln \pi(\mathbf{x}_{ij}) + (1-\mathbf{y}_i)\ln(1-\pi(\mathbf{x}_{ij})) \right\} + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right]. \quad (6)$$

As we observe from Eq. (6), elastic net estimator depends on two non-negative tuning parameters λ_1 and λ_2 which lead to penalized logistic regression solution. However, elastic net performs well when the pairwise correlations between variables are very high. El Anbari and Mkhadri (2014) stated that if the absolute correlation between genes is less than 0.95, elastic net may be slightly less reliable. Moreover, elastic net does not take into account the correlation structure among genes (Bühlmann, Rütimann, van de Geer, & Zhang, 2013). Additionally, it was pointed out by Zou and Zhang (2009) that the elastic net fails in terms of achieving oracle property, although the grouping effect problem for elastic net remains. As a result, adaptive elastic net was introduced by Zou and Zhang (2009) and Ghosh (2011), which it combines the ℓ_2-norm penalization with the adaptive LASSO.

3. Adjusted Elastic Net Penalty

In this section, we present our proposed adjusted method, AEN and AAEN, in logistic regression model. The main idea behind AEN is to take into account the information about the empirical correlation of the data matrix in the ℓ_2-norm term because elastic net does not. Suppose without loss of generality that the explanatory variables are scaled, we define the AEN estimator as

$$\hat{\beta}_{AEN} = \arg\min_{\beta} \left\{ -\ell(\beta) + \lambda_1 \sum_{j=1}^{k} |\beta_j| + \lambda_2 \left[\sum_{j=1}^{k-1} \sum_{p>j} (\beta_j - r_{j,p}\beta_p)^2 + \beta_k^2 \right] \right\} \qquad (7)$$

where λ_1 and λ_2 are non-negative tuning parameters. $r_{j,p}$ is the correlation between j and p explanatory variables where $p > j$. The quantity $(\beta_j - r_{j,p}\beta_p)^2$ is helpful to make AEN reliable if the correlation between explanatory variables is not so extreme. The last term from Eq. (7) is greater than zero for any vector β. Therefore, $(r_{j,p})'(r_{j,p})$ represents a Cholesky's decomposition. After suitable data argumentation, Eq. (7) is equivalent to a LASSO. The AEN was solved using coordinate descent optimization (Friedman et al., 2010) which a computationally efficient method for solving this type of convex optimization problem. The optimal AEN model was found by a grid search over the parameters λ_1 and λ_2.

Furthermore, the adaptive version of AEN, AAEN, is defined by

$$\hat{\beta}_{AAEN} = \arg\min_{\beta} \left\{ -\ell(\beta) + \lambda_1 \sum_{j=1}^{k} w_j |\beta_j| + \lambda_2 \left[\sum_{j=1}^{k-1} \sum_{p>j} (\beta_j - r_{j,p}\beta_p)^2 + \beta_k^2 \right] \right\} \quad (8)$$

where

$$w_j = (1/|\hat{\beta}_{j(AEN)}|)^\gamma, \quad j = 1, 2, ..., p \qquad (9)$$

where $\gamma > 0$. For simplicity, $\gamma = 1$ was used for both simulation study and real data application.

4. Simulation Study

In this section, simulation studies are used to investigate the performance of the proposed AEN and AAEN. Furthermore, we compare AEN and AAEN with elastic net. In all simulations the response variable was generated as $y \sim B[\exp(\sum_{j=1}^{p} \mathbf{x}_j^T \beta_{true})/1 + \exp(\sum_{j=1}^{p} \mathbf{x}_j^T \beta_{true})]$. All simulation cases are replicate 50 times.

For every simulation case and in each replication we generate training, validation, and testing data. The training data were used for model fitting. The validation data were used to determine the tuning parameters. The testing data were used to evaluate the penalization methods. For each case, the observation numbers of the corresponding data sets are denoted by training/validation/testing. Based on the simulated data, we used three metrics to evaluate all penalization methods which were studied in this paper, miss-classification errors for the test data (ME_t), *hits* which stands for the number of correctly identified true variables, and *false positive* (FP) which denotes to the number of zero variables which are wrongly considered as true variables.

Since we investigate a penalization method with both variable selection and grouping property, we use simulation scenario with different values of the correlation and different numbers of training, validation, and testing observations.

Simulation Scenario: In this setting, we generate data sets with sample sizes 50/50/100 and 1000 explanatory variables. Four cases are studied. The grouping effects were generated as follows

$$Group\ 1:\ x_i = w_1 + \varepsilon_i, w_1 \sim N(0,1),\quad i = 1,2,3;$$
$$Group\ 2:\ x_i = w_2 + \varepsilon_i, w_2 \sim N(0,1),\quad i = 4,5,6;$$
$$Group\ 3:\ x_i = w_3 + \varepsilon_i, w_3 \sim N(0,1),\quad i = 7,8,9.$$

Furthermore, the noisy explanatory variables were generated as $x_i \sim N(0,1),\ i = 10,11,....,1000$.

Case A: In this case we set $\varepsilon_i \sim N(0,0.01),\ i = 1,2,...,9$. The correlations among variables within each group are 0.98. The true variables' parameters were $\beta = (\underbrace{0.3,....,0.3}_{9}, \underbrace{0,....,0}_{991})$.

Case B: This simulation is like case A except that $\varepsilon_i \sim N(0,0.6),\ i = 1,2,...,9$. Thus, there are correlations within each group around 0.7.

Case C: Similar to case A, we set $\varepsilon_i \sim N(0,0.8),\ i = 1,2,...,9$ in order to get correlations within each group equal 0.5.

Case D: Similar to previous cases, in order to get correlations within each group equal 0.3. We assume that $\varepsilon_i \sim N(0,1.5),\ i = 1,2,...,9$.

5. Simulation Results

To examine the performance of the AEN and AAEN penalties we compare it with elastic net. For the tuning parameters of elastic net, AEN, and AAEN, a prior value of λ_2 is required to transform the original training data set to the new augmented training data set. A sequence of values for λ_2 is given, where $0 \le \lambda_2 \le 100$. The miss-classification error for the training set (ME_t) is computed as the criterion of evaluation. Figure 1 displays the corresponding boxplots of the ME_t for the three used methods for the four cases. It is clearly seen that AEN and AAEN has less variability comparing with elastic net. Also, it can be seen that AEN and AAEN are slightly similar.

Table 1 summarizes the median ME_t and the standard deviation (Std. Dev.) of the median ME_t which estimated by using bootstrap with $B = 100$ resampling on 50 ME_t values. In addition, the median number of *hits* and FP are reported too. In each case, the bold font indicates the best method on ME_t, Std. Dev., *hits*, and FP. Table 1 reveals that the AAEN method produces considerably smaller median ME_t and standard deviation among all methods in all cases. For example, in case A the median ME_t of AAEN is 5.782 with standard deviation equals to 1.561 which is smaller than 5.824 (1.766) and 6.348 (2.010) for AEN and elastic net methods respectively. Furthermore, the reduction of ME_t is

usually substantial compared to elastic net. For example, the reduction in case A, case B, case C, and case D is 0.67%, 1.60%, 5.01%, and 6.60% respectively. Moreover, in case A, there is high collinearity among variables. Elastic net is supposed to have the best performance then AEN because elastic net deals with extremely highly correlations. In addition, our method performs well in terms of ME_t when the correlation is small and medium. Besides, from the simulation results we can observe that elastic net came the last method.

For variable selection accuracy, the penalization methods should include all important variables (non-zero variables), *hits* and FP were used to measure the performance of AAEN, AEN, and elastic net in term of selecting the non-zero variables. From Table 1 both AAEN and AEN succeed in selecting the true non-zero variables in most of the cases in term of *hits*. For example, AEN selects the all nine non-zero variables. Moreover, when the correlation coefficient varies from small, medium, to extremely high correlation elastic net selects less non-zero variables comparing to AAEN and AEN. We can expect such a result because elastic has its limitation in biased selection. In terms of FP, AAEN and AEN methods usually select less ineffective variables than elastic net in most cases. To this end, it is obvious from our simulation results that the AAEN and AEN methods perform better in term of ME_t by obtaining smaller values, *hits*, and FP followed by elastic net for small, medium, and extremely high correlation and has greater advantage of variable selection with grouping effects in logistic regression model.

6. Real Data Results

To evaluate our proposed method in the field of binary classification, a publicly well-known binary cancer classification dataset was used, which is the prostate cancer dataset published by (Singh et al., 2002). It consists of 102 samples of 52 prostate tumor samples and 50 non-tumor tissues, where each sample has 12600 genes. A subset of 5966 genes was adapted in the classification. In order to enable a fair comparison, typically, the dataset was randomly partitioned into a training dataset, which comprised 70% of the samples, and a test dataset, which consisted of 30% of the samples. The partition repeated 50 times. In order to get the best value of the pair (λ_1, λ_2), the 10-fold CV was employed using the training dataset. All the applications were conducted in R using the *glmnet* package.

Table 2 shows the median number of explanatory variables selected by each of the AAEN, AEN, and elastic net in the training data set, and the corresponding median ME_t. It can be seen that AAEN performs best in term of prediction error where the ME_t of the AAEN is approximately 3.73% lower than AEN and 6.39% lower than elastic net. Moreover, AAEN selects less explanatory variables than the other two methods.

7. Conclusion

A study of adjusted elastic net was proposed by applying on logistic regression model. AAEN and AEN with elastic net were compared by using simulation studies and real data application. Both the simulation and real data results show that the AAEN and AEN are outperforming the elastic net in terms of ME_t of test data and variable selection accuracy. We can conclude that AAEN and AEN more reliable for grouping effects when there are

broader ranges of correlation between variables in applying penalized logistic regression model.

References

1. Anbari, M., & Mkhadri, A. (2014). Penalized regression combining the L 1 norm and a correlation based penalty. *Sankhya B, 76*(1), 82-102. doi:10.1007/s13571-013-0065-4.

2. Bondell, H. D., & Reich, B. J. (2008). Simultaneous regression shrinkage, variable Selection, and supervised clustering of predictors with OSCAR. *Biometrics, 64*(1), 115-123. doi:10.1111/j.1541-0420.2007.00843.x

3. Bühlmann, P., Rütimann, P., van de Geer, S., & Zhang, C.-H. (2013). Correlated variables in regression: Clustering and sparse estimation. *Journal of Statistical Planning and Inference, 143*(11), 1835-1858. doi:10.1016/j.jspi.2013.05.019.

4. El Anbari, M., & Mkhadri, A. (2014). Penalized regression combining the L1 norm and a correlation based penalty. *Sankhya B, 76*(1), 82-102. doi:10.1007/s13571-013-0065-4.

5. Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association, 96*(456), 1348-1360.

6. Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software, 33*(1), 1-22.

7. Ghosh, S. (2011). On the grouped selection and model complexity of the adaptive elastic net. *Statistics and Computing, 21*(3), 451-462. doi:10.1007/s11222-010-9181-4.

8. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55-67.

9. Li, J., Jia, Y., & Zhao, Z. (2013). Partly adaptive elastic net and its application to microarray classification. *Neural Computing and Applications, 22*(6), 1193-1200. doi:10.1007/s00521-012-0885-6.

10. Park, M. Y., & Hastie, T. (2008). Penalized logistic regression for detecting gene interactions. *Biostatistics, 9*(1), 30-50. doi:10.1093/biostatistics/kxm010.

11. Pourahmadi, M. (2013). *High-Dimensional Covariance Estimation: With High-Dimensional Data*. NJ: Wiley.

12. Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., . . . Sellers, W. R. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell, 1*(2), 203-209. doi:http://dx.doi.org/10.1016/S1535-6108(02)00030-2.

13. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*(1), 267-288. Retrieved from http://www.jstor.org/stable/2346178.

14. Tutz, G., & Ulbricht, J. (2009). Penalized regression with correlation-based penalty. *Statistics and Computing, 19*(3), 239-253. Retrieved from <Go to ISI>://WOS:000267111500003.

15. Zeng, L., & Xie, J. (2011). Group variable selection for data with dependent structures. *Journal of Statistical Computation and Simulation, 82*(1), 95-106. doi:10.1080/00949655.2010.529812.

16. Zeny, Z. F., Xiaojian, Y., Sanjeena, S., & Paul, D. M. (2012). The LASSO and sparse least squares regression methods for SNP selection in predicting quantitative traits. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 9*(2), 629-636. Retrieved from http://doi.ieeecomputersociety.org/10.1109/TCBB.2011.139.

17. Zhenqiu, L., Feng, J., Guoliang, T., Suna, W., Fumiaki, S., & Ming, T. (2007). Sparse logistic regression with Lp penalty for biomarker identification. *Statistical Applications in Genetics and Molecular Biology, 6*(1), 1-22.

18. Zhu, J., & Hastie, T. (2004). Classification of gene microarrays by penalized logistic regression. *Biostatistics, 5*(3), 427-443. doi:10.1093/biostatistics/kxg046.

19. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B-Statistical Methodology, 67*, 301-320. Retrieved from <Go to ISI>://WOS:000227498200007.

20. 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

21. Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics, 37*(4), 1733-1751. doi:10.1214/08-AOS625.
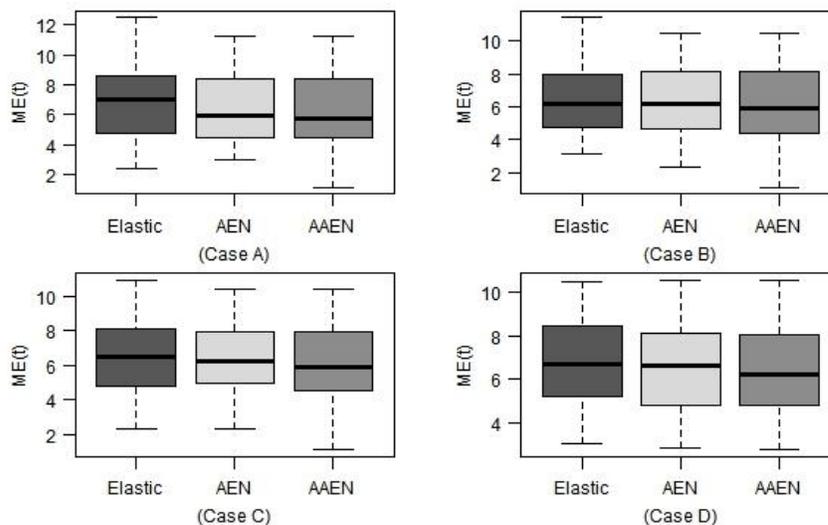
Figure 1: Comparison of median ME_t for three methods

Table 1: **Comparison among methods over 50 replications for the simulation scenario**

	ME$_t$ (Std. Dev.)	*hits*	FP
$\rho = 0.98$			
Elastic net	6.348 (2.010)	5	26.5
AEN	5.824 (1.766)	9	24.5
AAEN	**5.785 (1.561)**	**8.5**	**23**
$\rho = 0.7$			
Elastic net	6.129 (1.856)	4	23
AEN	6.055 (1.735)	9	21.5
AAEN	**5.953 (1.572)**	**9**	**21**
$\rho = 0.5$			
Elastic net	6.369 (1.891)	4	26
AEN	6.209 (1.831)	9	24.5
AAEN	**5.899 (1.604)**	**9**	**24**
$\rho = 0.3$			
Elastic net	6.671 (2.146)	4	25
AEN	6.640 (1.896)	8	21.5
AAEN	**6.202 (1.739)**	**9**	**20**

Table 2: **Comparison among methods for the real dataset**

	ME$_t$	No. of selected variables
Elastic net	10.713	48
AEN	10.417	44
AAEN	**10.028**	**42**