Supplementary Materials

1 Multi-Step Estimate

As neither $M_1(\beta)$ nor $M_2(\beta)$ provides a good approximation to $1 - AUC(\beta)$, we proposed an iterative algorithm to approximately minimize $1 - AUC(\beta)$. More specifically, we suggest the following adaptive algorithm

1. Set the initial
$$\beta \leftarrow \text{argmin}_{\beta} M_2(\beta)$$

2. Update β as
$$\beta \leftarrow \text{argmin}_{\gamma} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\{1 - \gamma'(X_i - Y_j)\}_+}{1 + |1 - \hat{\beta}'_{k-1}(X_i - Y_j)|},$$
where the minimization can be solved via linear programming technique.

3. Repeat step (2) until convergence or the number of iteration reaches a pre-specified number.

In this subsection, we conduct a small simulation study to examine the operational characteristics of the above algorithm. Especially we investigate whether the iteration always converges to a fixed point and whether the iteration improves the performance of the constructed score in maximizing $AUC(\beta)$. To this end, we generate the covariates $\{X_1, \cdots, X_n\}$ and $\{Y_1, \cdots, Y_m\}$ from the following models:

1. (multivariate normal) $X_i \overset{iid}{\sim} N(\mu_1, \Sigma_1)$ and $Y_j \overset{iid}{\sim} N(\mu_2, \Sigma_2)$, where X_i and Y_j are 3-dimensional random vectors.

2. (log-normal mixture) $\log(X_i) \overset{iid}{\sim} 0.8N(\mu_1, \Sigma_1) + 0.2N(\mu_3, \Sigma_3)$ and $\log(Y_j) \overset{iid}{\sim} 0.8N(\mu_2, \Sigma_2) + 0.2N(\mu_3, \Sigma_3)$.

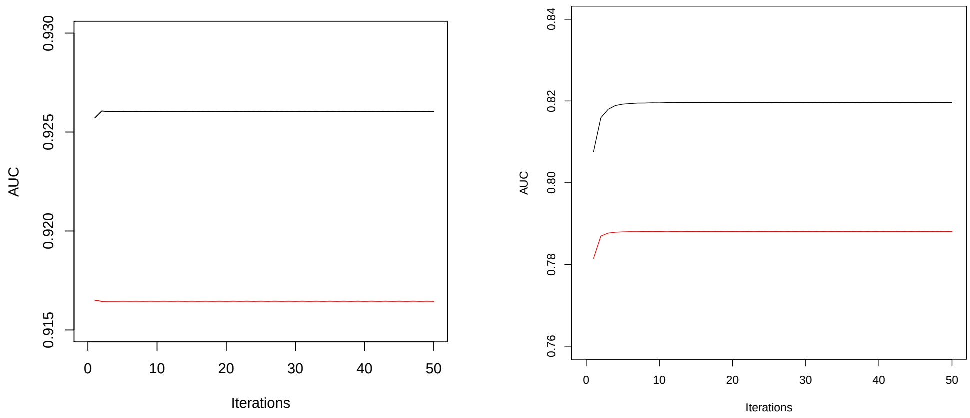
In both settings, we let $\mu_1 = (1, 0, 0)'$, $\mu_2 = (0, 1, 0)'$, $\mu_3 = (0, 0, 0)'$,

$$\Sigma_1 = \Sigma_2 = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix},$$

$\Sigma_3 = 5I_3$, I_3 is the 3 by 3 identity matrix. For each generated data set, we construct the combined score at each iteration and obtain the corresponding AUC in both training and validation set. The validation set is generated from the same distribution as the training set with 2000 observations in each group. Based on results from 500 generated data sets, we calculate the empirical average of resulting AUC based on scores from each iteration and examine whether the estimated weight converges after 50 iteration. The relative difference between two $\hat{\beta}$ in consecutive iterations needs to less than 0.01% in order to be considered as convergence. Figure 1 plots the empirical average of AUC against iteration steps. In the multivariate normal case, the iteration does not improve the AUC in the validation set. On the other hand, the AUC in the validation set increases from 0.781 to 0.787 after one iteration and 0.788 after two iterations. Further iterations does not improve the AUC anymore. AUC in the training set follows the similar pattern. 92% and 96% of the iterations converges in multivariate and log-normal mixture cases, respectively. For those failed to converge, the estimated $\hat{\beta}$ may oscillate between few values closed to each other. This result is not surprising as most of gain in prediction performance of adaptive lasso and SCAD regularization, if exists, is also realized in first one or two steps (Zou and Li, 2008). In the light of the numerical study, we suggest that iteration may improve the AUC when there are potential outliers and one or two iterations suffice to harvest the gain in maximizing $AUC(\beta)$.

References

[1] Zou, H., Li, R. (2008) One-step sparse estimates in nonconcave penalized likelihood models, *Annals of Statistics*, **36**, 1509-1533.



(a) multivariate normal

(b) log-normal mixture

Figure 1: AUC in training and validation sets (black: training set; red: validation set)