Supplementary Material for "Feature-weighted Ordinal Classification for Predicting Drug Response in Multiple Myeloma"

Ziyang Ma and Jeongyoun Ahn

March 2021

1 Simulation Studies

We carried out a simulation study to measure the empirical performance of the proposed FWOC and compared with BhGLM, PLDA and PCRM.

In the simulations, we simulated moderate high-dimensional datasets, with dimension of 500 and sample size of 150. The samples are assumed to come from four classes with natural ordering: $C_1 \prec C_2 \prec C_3 \prec C4$ and the sample sizes are distributed as $n_1 = 45, n_2 = 35, n_3 = 40, n_4 = 30$. The observations from class k are assumed to follow a multivariate normal distribution $N_p(\boldsymbol{\mu}_k, \Sigma_k)$, where $\boldsymbol{\mu}_k$ and Σ_k are the mean vector and covariance matrix for class k and $p = 500$. Among the 500 features, we assume that there are 20 signal variables which contribute to the separation of classes and the rest will be noise variables. We further divide the signal features into two categories, one consists of order-concordant variables, (i.e., ordinal variables), the other consists of order-discordant variables (i.e., nominal variables). In the following, based on how much ordinality is existing among the classes, we consider three different scenarios that are characterized by different mean structures of the signal variables:

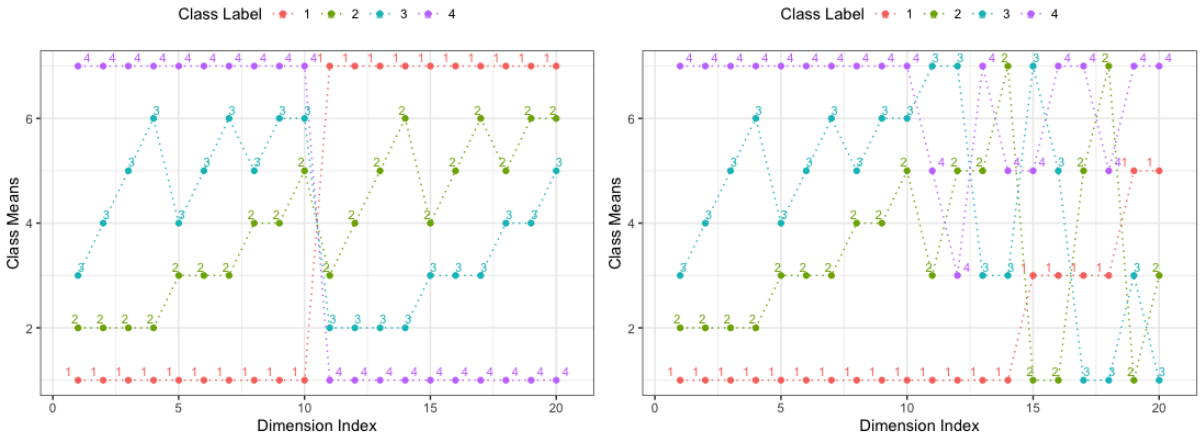
- linear ordinality
- nonlinear ordinality
- nominal situation

The signal variables in the scenario of 'linear ordinality' are all ordinal variables. There are ten ordinal variables and ten nominal variables in the scenario of 'nonlinear ordinality'. The signal variables in the nominal situations will all be nominal variables. Let the mean vector for class k be $\boldsymbol{\mu}_k = (e * \mathbf{m}_k, \mathbf{0})$, where e is the effect size, $\mathbf{m}_k$ is the mean structure for the signal variables, and $\mathbf{0}$ is the mean structure for the noise variables, whose elements are all zero. The values of $\mathbf{m}_k$ are shown in Figure 1. The effect size (e) over all the simulations is set to be 0.25. We fix the variance to be 1. For each mean structure, we consider three correlation structures: 1) identity matrix; 2) block auto-correlation matrix; 3) block compound symmetry matrix. The block auto-correlation matrix Σ_{auto} and block compound symmetry matrix Σ_{cs} are given in the following:

$$\Sigma_{\text{auto}} = \begin{pmatrix} A(\rho_1)_{ns\times ns} & \mathbf{0}_{(p-ns)\times(p-ns)} \\ \mathbf{0}_{(p-ns)\times(ns)} & I_{(p-ns)\times(p-ns)} \end{pmatrix}, \Sigma_{\text{cs}} = \begin{pmatrix} C(\rho_2)_{ns\times ns} & \mathbf{0}_{(p-ns)\times(p-ns)} \\ \mathbf{0}_{(p-ns)\times(ns)} & I_{(p-ns)\times(p-ns)} \end{pmatrix},$$

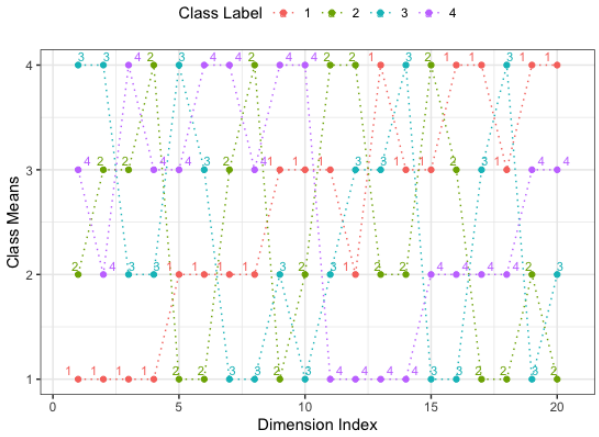
where $ns = 20$ is the number of signal variables, I is the identity matrix and $\mathbf{0}$ is the matrix whose elements are all zeros. $A(\rho_1)$ is the auto-correlation matrix with coefficient $\rho_1 = 0.9$, $C(\rho_2)$ is the compound symmetry matrix with coefficient $\rho_2 = 0.7$.

For each simulated dataset, we randomly split it to a training set with 70% observations and a test set with 30% observations. We used the training set for model building (parameter tuning) and the test set for model assessment. The test results are averaged over 100 repetitions, which are given in Figure 2, 3, 4 and 5.



(a) Linear ordinality

(b) Nonlinear ordinality



(c) Nominal situation

Figure 1: The class mean structures $\mathbf{m}_k$ for the three scenarios. The values are presented in the y-axis, the indexes of the signal variables are presented in the x-axis. Mean structures for the four classes are colored differently. Note that we allow different distances between means from adjacent classes for different dimensions, which are shown in the figure.

In general, the performance of FWOC is very competitive under the three scenarios. Especially in the scenario of 'nonlinear ordinality', the advantage of FWOC is the most obvious. Among the three scenarios, the performances of BhGLM and PCRM are only acceptable in the 'linear ordinality' case. The performance of BhGLM and PCRM is not well in other cases, which is not beyond our expectations. In terms of feature selection, PCRM performs well in achieving a sparse solution, but BhGLM fails with feature selection. In practice, data are complicated such that the classes have a high probability of being 'nonlinear' aligned, in which these model-based approaches might fail.

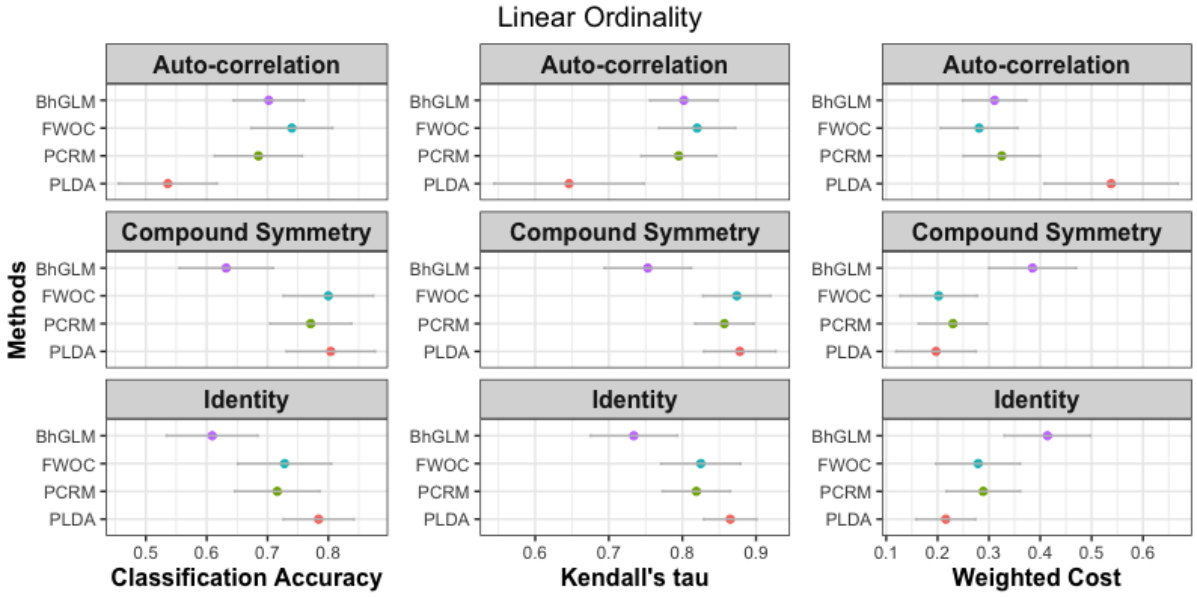


Figure 2: The average classification accuracy, Kendall's τ and weighted cost (when $d = 1$) over 100 simulated data sets under the scenario of linear ordinality. Standard deviations are represented by error bars. The three columns show the three metrics, whose values are displayed on the x axis. Different correlation structures under the scenario are presented in the rows.

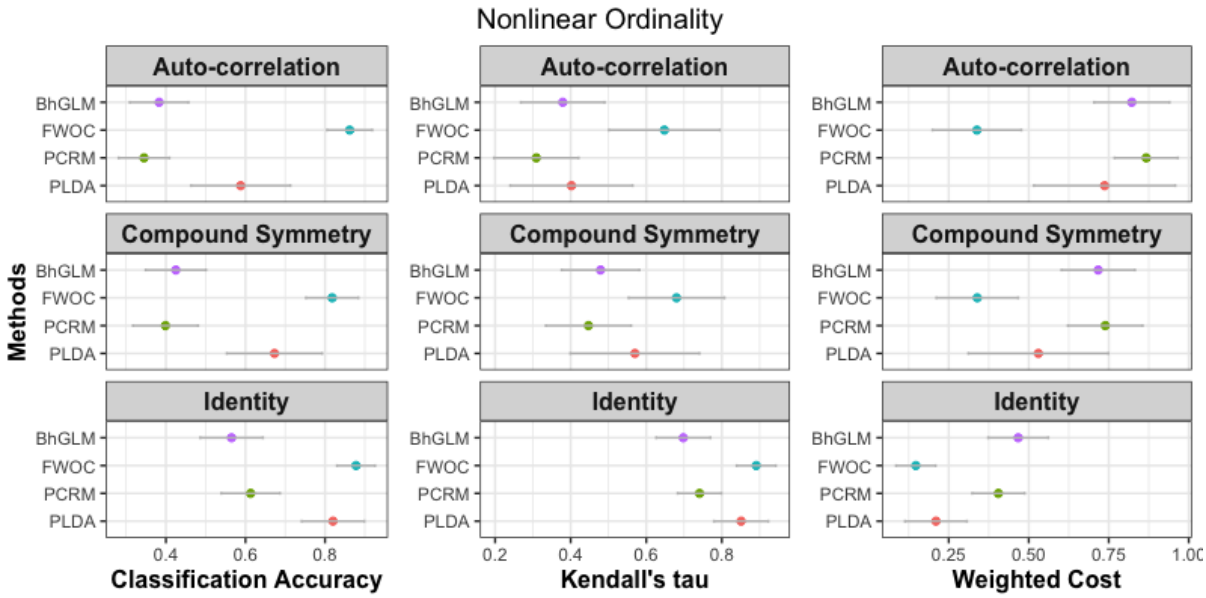


Figure 3: The average classification accuracy, Kendall's τ and weighted cost (when $d = 1$) over 100 simulated data sets under the scenario of nonlinear ordinality. Standard deviations are represented by error bars. The three columns show the three metrics, whose values are displayed on the x axis. Different correlation structures under the scenario are presented in the rows.

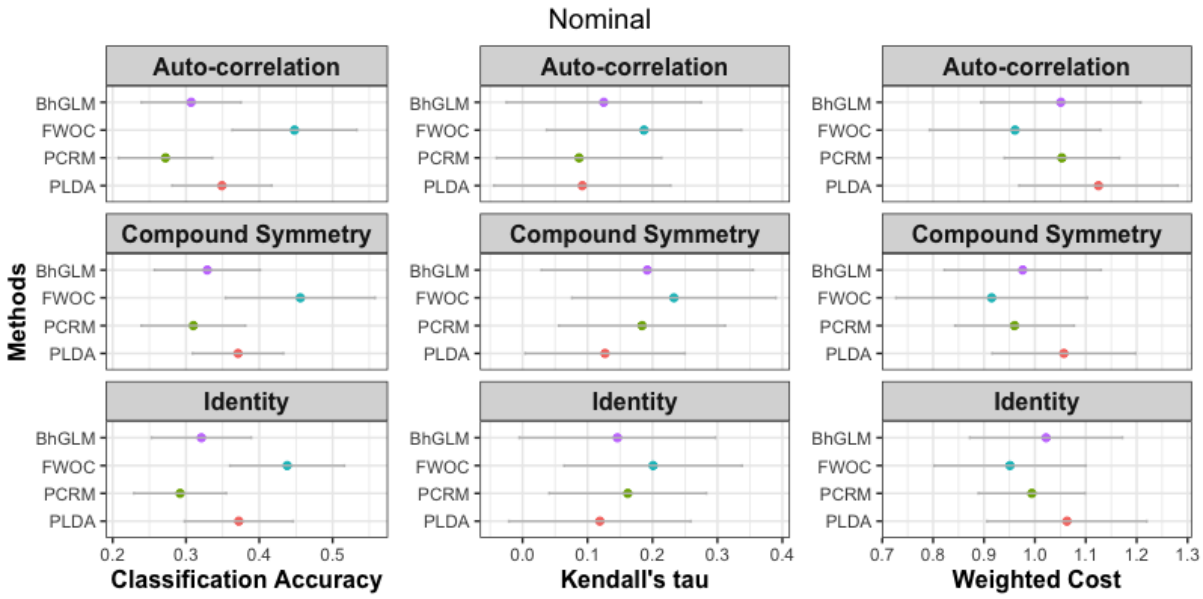
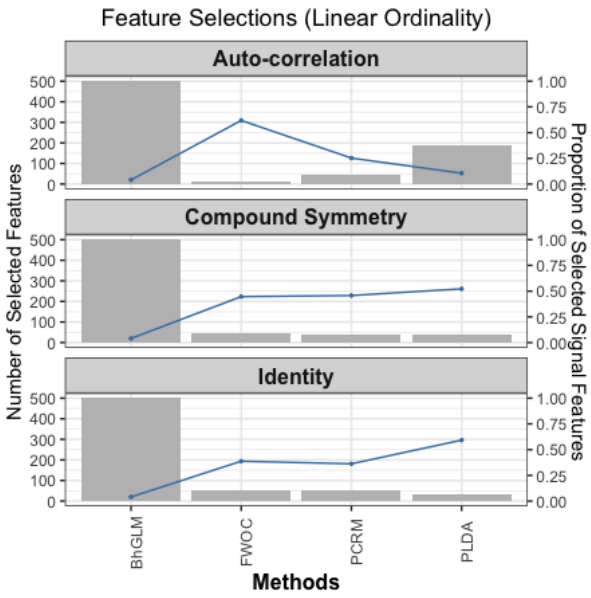
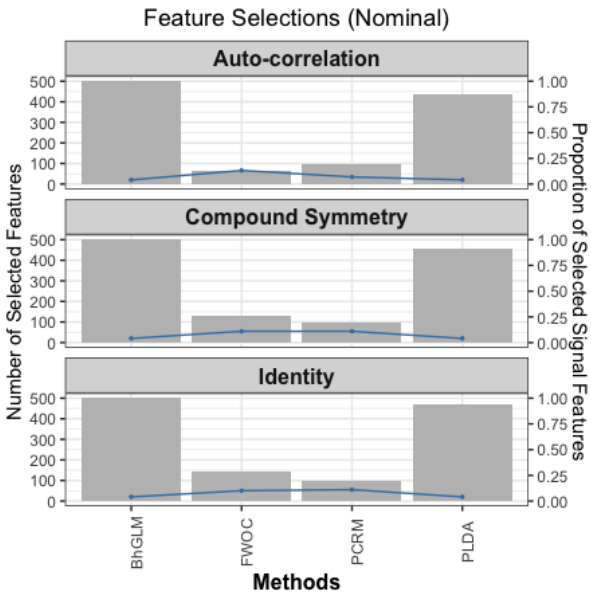


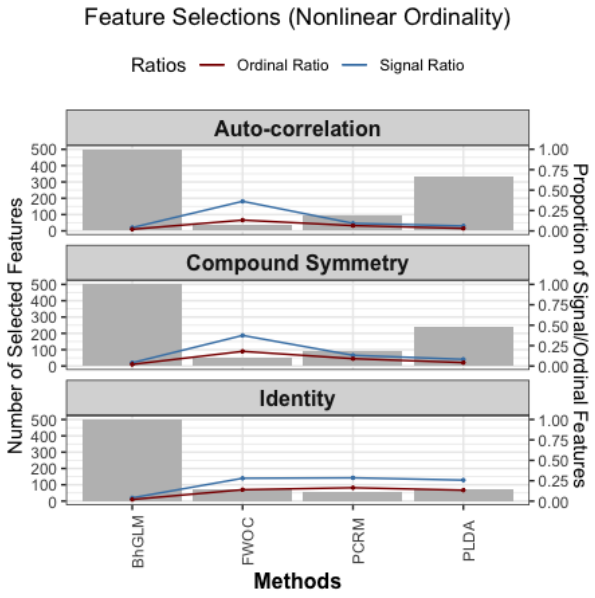
Figure 4: The average classification accuracy, Kendall's τ and weighted cost (when $d = 1$) over 100 simulated data sets under the scenario of nonlinear ordinality. Standard deviations are represented by error bars. The three columns show the three metrics, whose values are displayed on the x axis. Different correlation structures under the scenario are presented in the rows.



(a) Linear ordinality



(b) Nonlinear ordinality



(c) Nominal situation

Figure 5: The bargraph shows the number of selected features by each method, which is scaled on the left y axis. The line plot shows the ratio of selected signal features over all selected features, which is scaled on the right y axis. The blue line shows the signal ratio and the red line shows the ordinal ratio (ordinal ratio is only available under the scenario of nonlinear ordinality). The three rows show the three correlation structures

2 Comparisons Between 'Rank-correlation' and 'Equal weights'

Here, we discuss the difference between using $\bar{W}$ and identity matrix I in the L_2 penalty. In $\bar{W}$, Kendall's τ was used to measure the weights. We can expect that $\bar{W}$ will favor ordinal signal variables than nominal signal variables when they both exist. We used a simple setting to illustrate our idea. Similar with the simulation studies, we simulated toy datasets under the scenario of 'linear ordinality', 'nonlinear ordinality' and 'nominal situation'. We set $n = 100$, $p = 30$ and the samples are assumed to come from four ordinal classes with sizes: $n_1 = 20$, $n_2 = 30$, $n_3 = 30$, $n_4 = 20$. Data are normally distributed with identity covaraince matrix and the effect size is set to be 1.5. There are 10 signal variables in each scenario. Linear scenario contains 10 ordinal signal variables, nominal scenario contains 10 nominal variables and nonlinear scenario contains 5 ordinal variables and 5 nominal variables. We generated 30 datasets in each scenario and varied $r \in (0, 0.1, 0.3)$. The average absolute values of the coefficients of the first two discriminating vectors are given in Figure 6, 7 and 8. Note that for linear and nominal case, there is no obvious difference between the coefficients, as both $\bar{W}$ and I selected all the signal variables. However, when the signal variables are different, i.e., in the nonlinear case, as shown in Figure 7, $\bar{W}$ obtained higher weights in ordinal variables and less weights in nominal variables compared with I. The difference is larger as r gets smaller. This confirms that the proposed method indeed encourages more ordinally interpretable results while controlling the class separability in L_2 sense.

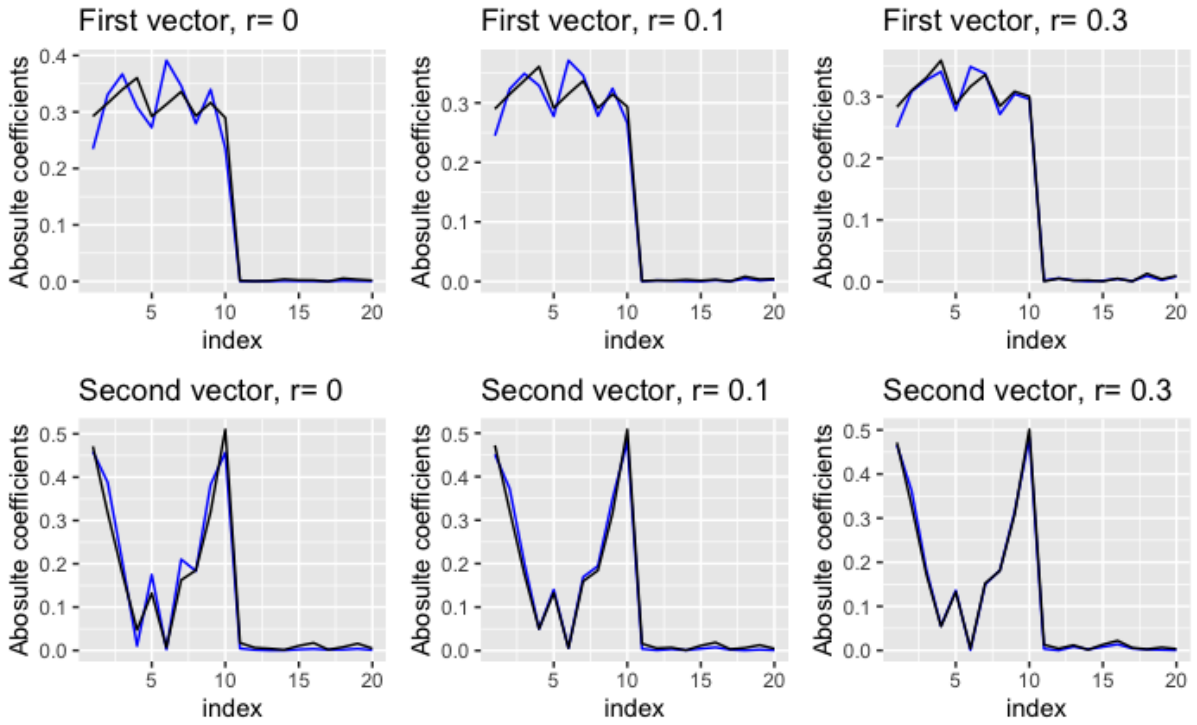


Figure 6: The average absolute values of the first 20 coefficients of the first two discriminant vectors (linear scenario). Blue lines represent results from $\bar{W}$ and black lines represent results form I. Indices of 1-10 indicate ordinal signal variables.

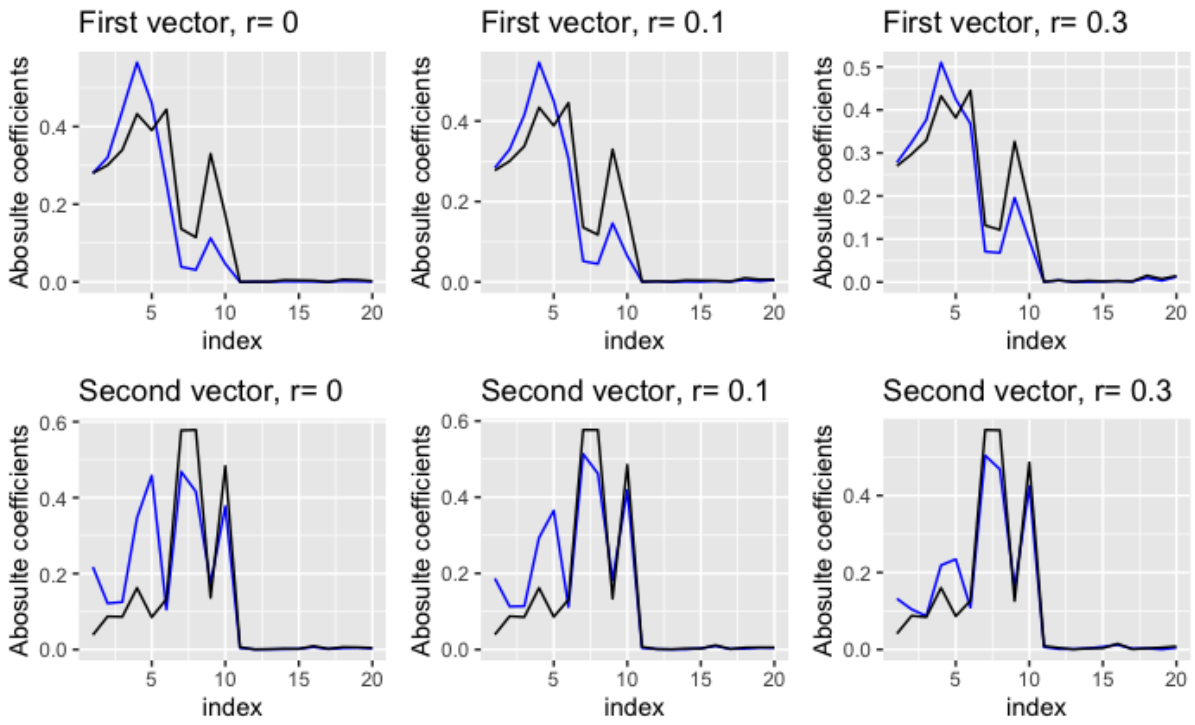


Figure 7: The average absolute values of the first 20 coefficients of the first two discriminant vectors (nonlinear scenario). Blue lines represent results from $\bar{W}$ and black lines represent results form I. Indices of 1-5 indicate ordinal signal variables and 6-10 indicate nominal signal variables.

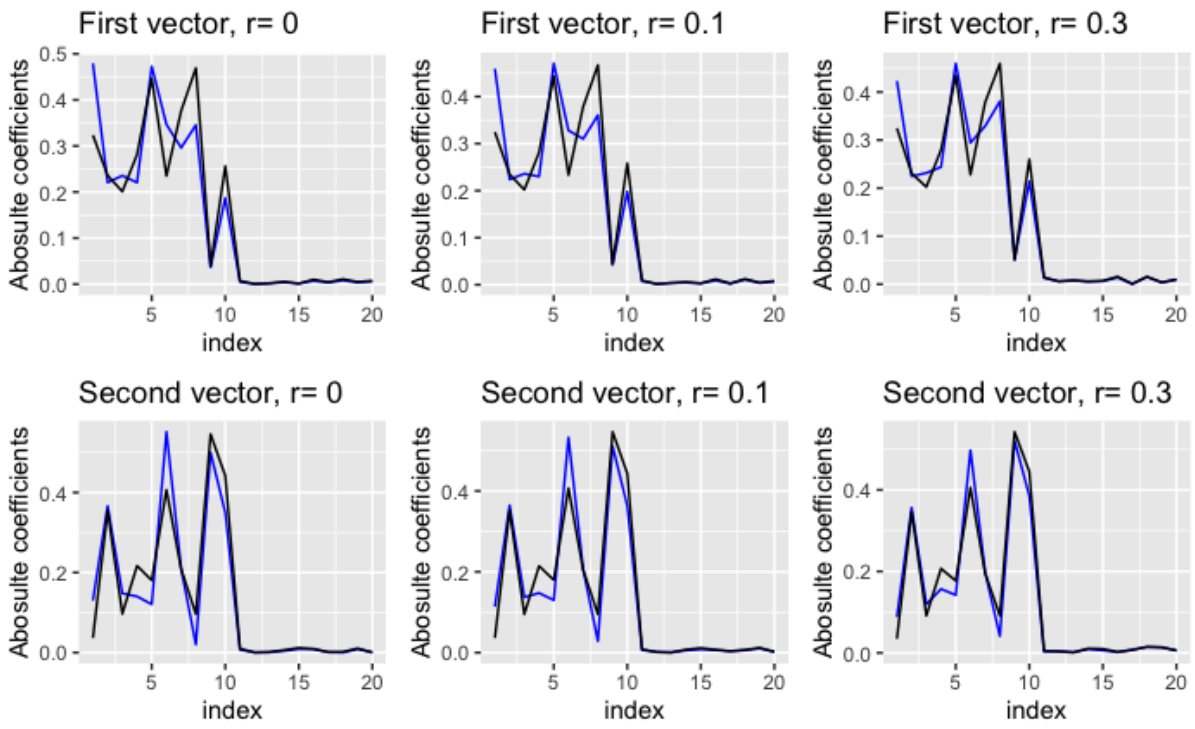


Figure 8: The average absolute values of the first 20 coefficients of the first two discriminant vectors (nominal scenario). Blue lines represent results from $\bar{W}$ and black lines represent results form I. Indices of 1-10 indicate nominal signal variables.