TMJM
THE MINDANAWAN
Journal of Mathematics

The Mindanawan Journal of Mathematics
Official Journal of the Department of Mathematics and Statistics
Mindanao State University-Iligan Institute of Technology
ISSN: 2094-7380 (Print) | 2783-0136 (Online)

Vol. 4 (2022), no. 2, pp. 23–33

# On Bayesian Feature Selection Procedure Applied to Regression Problem with HDD

Aries P. Valeriano[1] and Bernadette F. Tubo[2,*]

[1,2]Department of Mathematics and Statistics
MSU-Iligan Institute of Technology, 9200 Iligan City, Philippines
aries.valeriano@g.msuiit.edu.ph, bernadette.tubo@g.msuiit.edu.ph

Received: 17th October 2022 Revised: 11th November 2022

**Abstract**

High-dimensional data (HDD) means that the number of features, $p$, are exceedingly high and only a few samples $n$, are available. Regression problem involves the understanding of how the response, $y$, depend simultaneously on some features $x$. Often, only a few $x$'s explain $y$, while the rest may only have a little or no influence at all to it. Moreover, most of the existing methodology on how the $x$'s are entered into a regression model is established on $p \leq n$.

This study investigates a recently introduced methodology called the Bayesian feature ranking (BFR) on its performance with respect to how well the data fit the regression model in the presence of HDD in the $x$'s with $y$ being continuous. The proposed methodology involves implementing a modified forward selection (MFS) procedure on the ranked features with different noise levels $\nu$ infused on $y$ via the BFR. MFS via BFR procedure allows the most top ranked features to be included in the model and addition of features to the model is done sequentially, with increment value $\Delta = 5$. For baseline comparison, MFS procedure on unranked features is conducted and evaluation of the derived models will be based on the derived values of $R^2$, a statistic for model fit. Results showed that in both simulated and real dataset, MFS via BFR consistently gave higher $R^2$ than the baseline MFS, implying that the model derived via BFR using ranked features of $x$ describe $y$ much better than the model using unranked features of $x$.

## 1 Introduction

Statistical modeling involves predicting or explaining response variable, $y$ from several explanatory variables $x_1, x_2, \ldots, x_p$, where $y$ can be continuous, indicating a regression problem, or categorical having two outcomes, or more than two, indicating a binary, or multi-class classification problem, respectively. Most classical methodologies such as linear and logistic regression are established on dataset with $p \leq n$, where $p$ is the number of $x_j$'s for $j = 1, 2, \ldots p$ and $n$ is the number of observations. When dealing with high-dimensional data (HDD), that is, $p >> n$, often, only a few $x_j$'s explain $y$, while the rest may have little or no influence on it at all. Thus, the challenge is on which $x_j$'s should be included in the final model as most feature ranking or variable selection methodologies are created for [6].

[*] Corresponding author

**2022 Mathematics Subject Classification**: MSC62, MSC65
**Keywords and Phrases**: High-dimensional data, Feature selection procedure, Regression model

-This research is supported by DOST-SEI ASTHRDP Grant

HDD are omnipresent in most field of study. To name a few: (1) in healthcare industry where features such as blood pressure, resting heart rate, immune system state, operation history, height, weight, existing conditions, and so on, frequently outnumber individuals engaged; (2) in financial industry with features such as PE Ratio, Market Cap, Trading Volume, Dividend Rate, and so on, frequently outnumber the given stocks, and (3) in genomics, where there are typically thousands of gene features and only a few hundred samples [11].

As a result, several statistical or machine learning methods are presented to deal with HDD, in which, the implementation of these procedures in the real world is made possible by recent advances in computer machines. After all, it will require an increase in processing power because these are typically extremely large, as seen in genomics, and large $p$ corresponds to higher dimensionality, complicating the model and computation [7, 9].

The primary focus of this study is in the investigation of a recently introduced feature selection methodology, the Bayesian feature ranking (BFR) where it is applied to high-dimensional regression problem. Other feature selection methodologies also exist, like the Random Forest (RF) with default parameters and Independent Screening by Generalized Correlation (HM) as presented by Enes Makalic and Daniel Scmidt (2011). An empirical investigation of the performance of BFR against RF with default parameters and HM which used TopX metric to compare performances showed that BFR outperformed the other two feature selection methods [8].

In this study, a different approach which is more specific than what Enes Makalic and Daniel Scmidt (2011) used for investigating the performance of BFR is implemented. This approach involves performing a modified forward selection (MFS) procedure on ranked features. Ranked features refers to features categorized by the BFR methodology with respect to the inclusion of some features to the regression model. Forward selection procedure is modified in such a way that the increment $\Delta = 5$ and the model fitted to the ranked features for each forward step is via the ridge regression confirmed using 5-fold cross validation. Ridge regression analysis is utilized because the model can deal with the multicollinearity and the high-dimensionality of the data without dropping any of the explanatory variable $x_j$'s [11]. Forward selection procedure is mainly employed as a variable selection procedure, which is why most applications of it automatically set $\Delta = 1$, allowing the model to explicitly evaluate the importance of each $x_j$ in the given dataset. However, in this study, forward selection procedure where $\Delta$ is set to be equal to 5 does not result in inefficiency of the procedure, but rather reduces computing time when performed with HDD in the R software.

The motivation of the proposed methodology stems from the fact that the purpose of variable selection procedure is to find the best features to include in the model. Thus, the importance of ranked features via BFR can be determined by how well the model fits to it [10]. For baseline comparison of the MFS via BFR (on ranked features), MFS procedure (on unranked features) is also performed. This further demonstrates the power of employing a feature selection method using ranked features, particularly the BFR, before fitting a regression model to a dataset. In the simulation study of this research work, the model is formulated, such that, it has varying levels of correlation or degree of independence among the explanatory variables $x_j$ and infused noise $\nu$ in the response variable $y$.

## 2 Methodology

### 2.1 Data Description

In this paper, simulation study is done by generating 100 datasets with consideration of 3 linear regression functions described as follows by the authors in [2] to reflect varying scenarios, with sample size $n = 50$ and $p = 100$ features, in which each of it are converted to standardized values. In addition, noise $\nu$ was added to $y$ indicated by the variable $SNR \in \{1, 8\}$, that is, when $SNR = 1$, then the model has higher level of noise $\nu$ and if $SNR = 8$, then $\nu$ is in a lower level of noise. For baseline comparison, only these two extreme levels of noise (high and low) are considered in this paper following the work of Fan, J. et al. (2009) in [2]. The three linear regression functions are as follows:

Function I:

The generating regression coefficients were

$$\beta^* = (1.24, -1.34, -1.35, -1.80, -1.58, -1.60, 0'_{p-6})'$$

where $0_{p-6}$ is $(p-6)$-dimensional zero vector and $x_j \sim N_n(0, 1)$ for $j = 1, 2, ..., p$.

Function II:

The generating regression coefficients were

$$\beta^* = (4, 4, 4, -6\sqrt{2}, 0'_{p-4})'$$

where $x_j \sim N_n(0, 1)$ and the correlation between predictors was $\rho_{(x_j,x_4)} = 1/\sqrt{2} = 0.71$, $\forall j \neq 4$; $\rho_{(x_j,x_k)} = 0.5$, if $j$ and $k$ were distinct elements in $\{1, 2, ..., p\} \setminus \{4\}$.

Function III:

The generating regression coefficients were

$$\beta^* = (4, 4, 4, -6\sqrt{2}, 4/3, 0'_{p-5})'$$

where $x_j \sim N_n(0, 1)$. The correlation between predictors was $\rho_{(x_j,x_5)} = 0, \forall j \neq 5$, $\rho_{(x_j,x_4)} = 1/\sqrt{2} = 0.71$, $\forall j \notin \{4, 5\}$ and $\rho_{(x_j,x_k)} = 0.5$, if $j$ and $k$ were distinct elements in $\{1, 2, ..., p\} \setminus \{4, 5\}$.

Moreover, the proposed methodology is investigated using a secondary set of high-dimensional data. The "eyedata" which can be access through the R software by installing and loading "flare" package is explored. The dataset contains 120 samples with 200 explanatory variables. The data are obtained from rats with 200 different gene probes as explanatory variables and with 120 expression levels of the TRIM32 gene as the response variable [4].

## 2.2 The Regression Model

Considering the three linear regression functions described in subsection 2.1 and adding the different levels of noise $\nu$ to $y$, that is, $SNR \in \{1, 8\}$, the model building procedure is given as follows:

Let $\mu = E(y)$ be the expected value of an $n \times 1$ vector $y$. Then

$$\mu = \boldsymbol{x}\beta^*$$

where $\boldsymbol{x}$ is an $n \times p$ feature matrix and $\beta^*$ is a $p \times 1$ vector of regression coefficients which is defined previously. Then, the noise $\nu$ infused to $y$ is given as,

$$\nu = \sqrt{var(\mu)/SNR}$$

where $SNR \in \{1, 8\}$. So,

$$y = \mu + \nu \, \epsilon$$

where $\epsilon \sim N_n(0, 1)$ [8].

Now, in the evaluation of the performance of MFS with ranked features via BFR and baseline MFS with unranked features, the $R^2$ values for model fit are computed from the derived model.

## 2.3 Model Building and Evaluation

Figures 1 presents the procedure on how the model will be derived and how the model will be evaluated. Comparison of the performances of MFS via BFR (with ranked features) and baseline procedure MFS (with unranked features) will be done using simulated data.

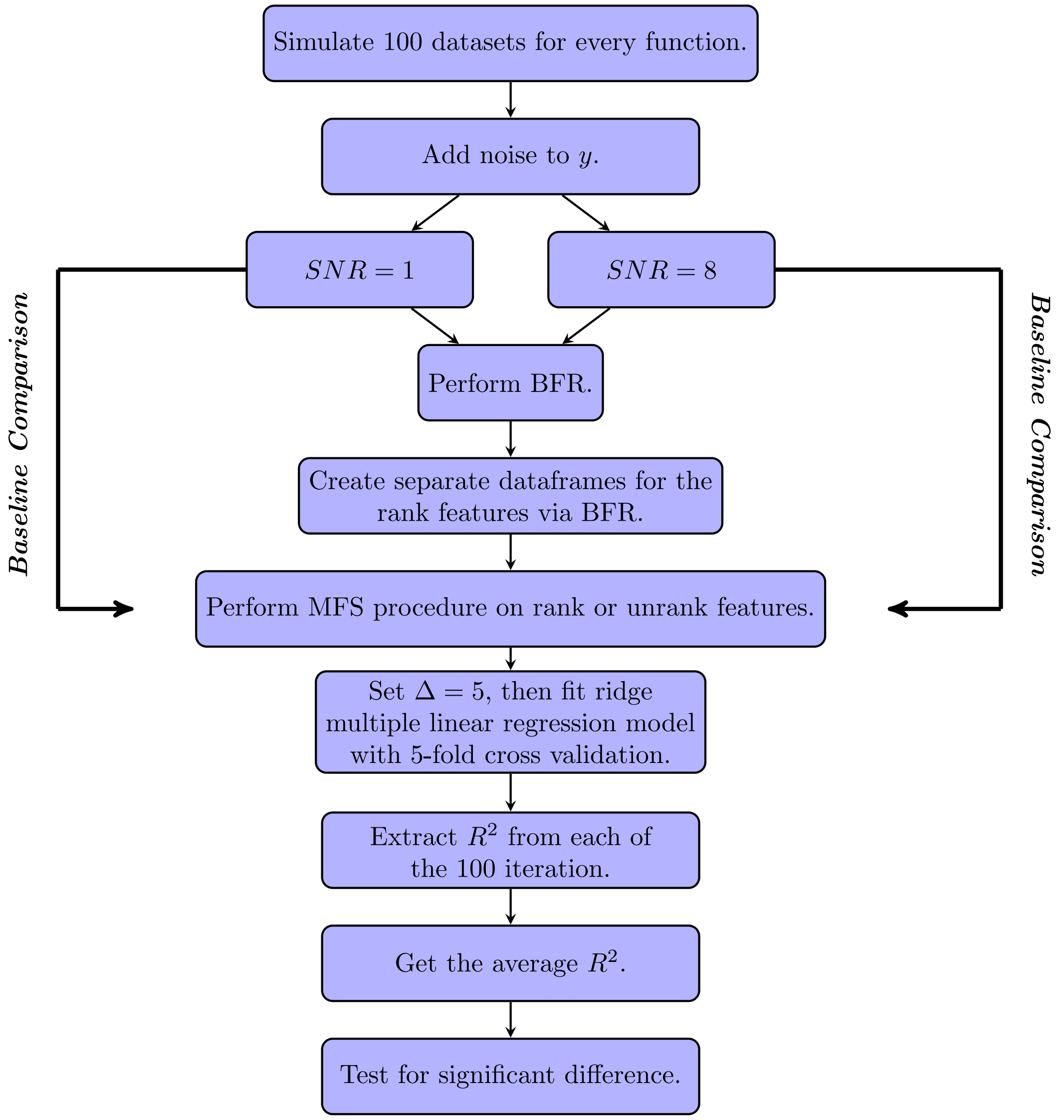

Simulate 100 datasets for every function.
Add noise to $y$.
$SNR = 1$
$SNR = 8$
Perform BFR.
Create separate dataframes for the rank features via BFR.
Perform MFS procedure on rank or unrank features.
Set $\Delta = 5$, then fit ridge multiple linear regression model with 5-fold cross validation.
Extract $R^2$ from each of the 100 iteration.
Get the average $R^2$.
Test for significant difference.
Baseline Comparison
Baseline Comparison

Figure 1: *Schematic Diagram for Simulated Dataset*

Figure 2 shows the procedure in assessing the performance of the methodologies MFS via BFR and the baseline MFS using the the dataset ”eyedata” as described in subsection 2.1.

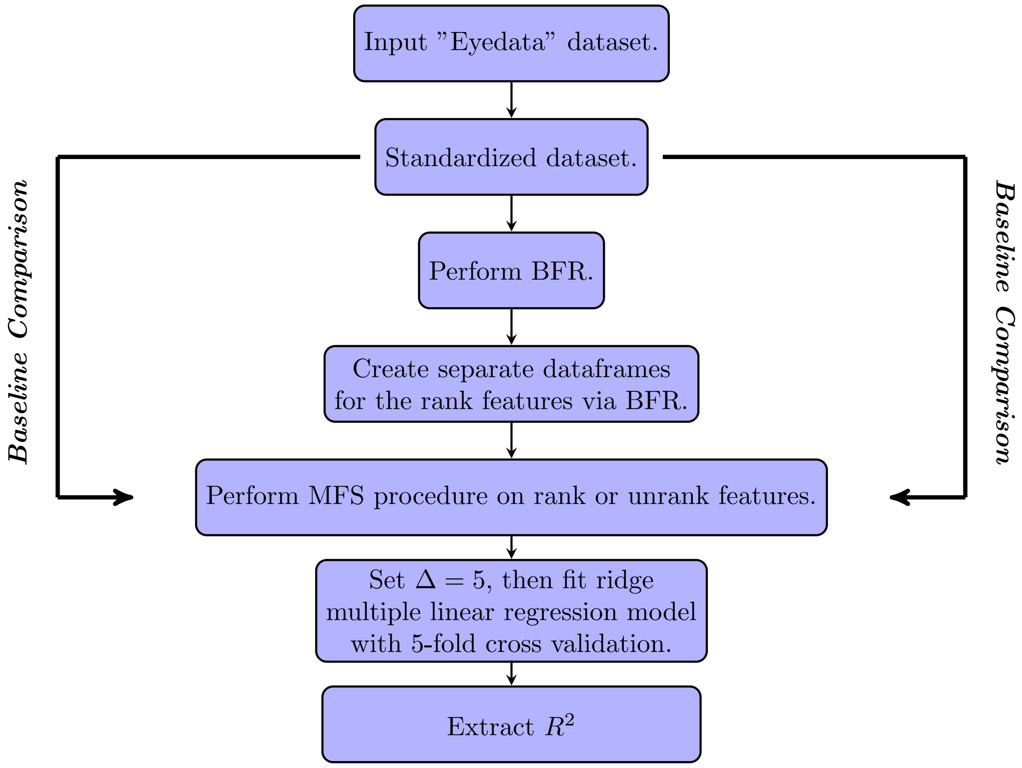

Input ”Eyedata” dataset.
Standardized dataset.
Perform BFR.
Create separate dataframes for the rank features via BFR.
Perform MFS procedure on rank or unrank features.
Set $\Delta = 5$, then fit ridge multiple linear regression model with 5-fold cross validation.
Extract $R^2$
Baseline Comparison
Baseline Comparison

Figure 2: *Schematic Diagram for Real Dataset*

# 3 Result and Discussions

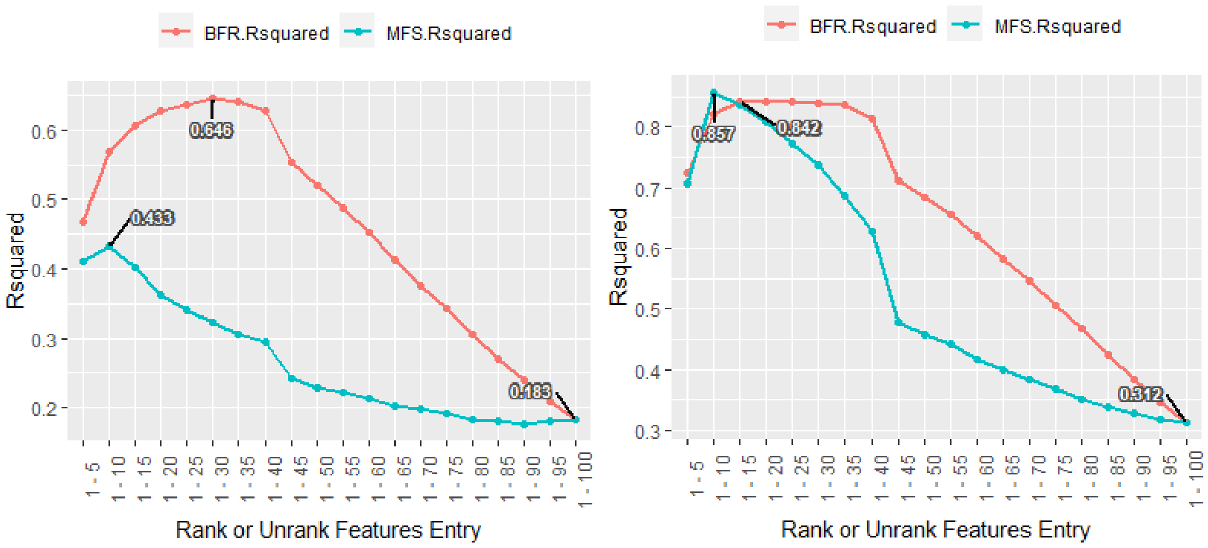
## 3.1 Using the Simulated Data

Every linear function in the simulation study reflect a unique scenario. It is to note that Function I depict a scenario where the $x_j$'s are independent to each other. Function II presents a scenario where the $x_j$'s have varying levels of correlation. Lastly, Function III reflect a scenario where some $x_j$'s are independent to each other, while the rest have different levels of correlation. Thus, for both Functions II and III, the data is generated to exhibit multicollinearity among the $x_j$'s. Moreover, two levels of noise $\nu$ is infused to the simulated response variable $y$, that is, if $SNR = 1$, then $y$ has higher level of noise and if $SNR = 8$, then $y$ has lower level of noise in $y$.

### 3.1.1 With Function I and $SNR \in \{1, 8\}$

Figure 3 displays the performance of MFS via BFR and baseline MFS for different values of noise $\nu$ added to $y$ using Function I. The labeled datapoints at the starting features entry in Figure 3(a) and Figure 3(b) correspond to the optimum performance of each procedure. When all of the features (1-100) are entered into the model, the $R^2$ value of both feature selection procedures coincide, so that it would not matter if features are ranked or unranked.

It can be observed in Figure 3(a) that BFR consistently gave higher $R^2$, which may indicate that MFS via BFR outperforms baseline MFS for higher level of $\nu$. The highlighted values represents the maximum performance of each procedure. Moreover, Table 1 presents the list of the values of $R^2$ for both BFR and MFS when features 1 to 50 are entered in the model with increment of $\Delta = 5$. Only the first 50 features are considered since an abrupt decrease in the model fit is detected when at most 50% of the features are included in the model (Figures 3(a) and 3(b)).



(a) $SNR = 1$ (b) $SNR = 8$

Figure 3: *Model Fit Performance of MFS via BFR and baseline MFS for Function I*

Baseline MFS attains its maximum $R^2_{MFS} = 43.3\% < R^2_{BFR} = 56.9\%$ when 1-10 features are included in the model, afterwards, the model fit performance decreases as more features are included in the model. On the other hand, $R^2_{BFR} = 64.6\%$ is maximum when 1-30 features are included in the regression model as compared with $R^2_{MFS} = 32.3\%$. For lower level of $\nu$, baseline MFS attains maximum performance when unranked features 1-10 are included in the model with $R^2 = 85.7\%$ in contrast with $R^2_{BFR} = 82\%$. Thereafter, BFS outperforms MFS in terms of its $R^2$ values.

Table 1: $R^2$ *values of MFS via BFR and baseline MFS procedures for Function I*

| (a) $SNR = 1$ | | | |
|---|---|---|---|
| RUFE | $R^2_{BFR}$ | $R^2_{MFS}$ | $t$-test ($p$-value) |
| 1 - 5 | 0.469 | 0.411 | <0.001* |
| 1 - 10 | 0.569 | 0.433 | <0.001* |
| 1 - 15 | 0.607 | 0.403 | <0.001* |
| 1 - 20 | 0.629 | 0.362 | <0.001* |
| 1 - 25 | 0.636 | 0.342 | <0.001* |
| 1 - 30 | 0.646 | 0.323 | <0.001* |
| 1 - 35 | 0.641 | 0.306 | <0.001* |
| 1 - 40 | 0.627 | 0.295 | <0.001* |
| 1 - 45 | 0.554 | 0.243 | <0.001* |
| 1 - 50 | 0.521 | 0.229 | <0.001* |

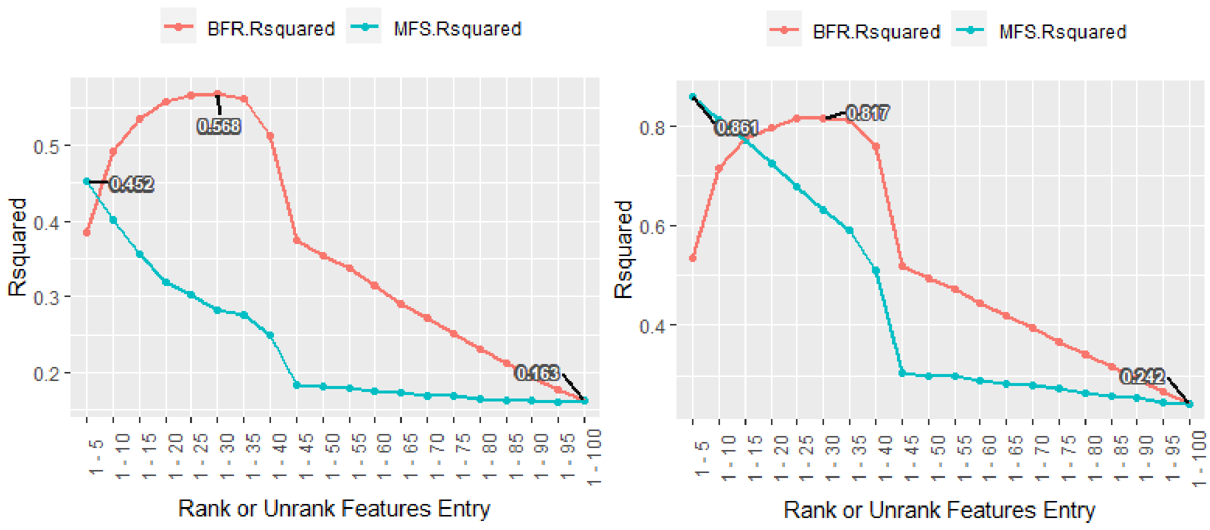
| (b) $SNR = 8$ | | | |
|---|---|---|---|
| RUFE | $R^2_{BFR}$ | $R^2_{MFS}$ | $t$-test ($p$-value) |
| 1 - 5 | 0.724 | 0.706 | 0.114 |
| 1 - 10 | 0.820 | 0.857 | <0.001* |
| 1 - 15 | 0.842 | 0.836 | 0.480 |
| 1 - 20 | 0.842 | 0.808 | <0.001* |
| 1 - 25 | 0.841 | 0.773 | <0.001* |
| 1 - 30 | 0.838 | 0.736 | <0.001* |
| 1 - 35 | 0.836 | 0.687 | <0.001* |
| 1 - 40 | 0.813 | 0.629 | <0.001* |
| 1 - 45 | 0.711 | 0.478 | <0.001* |
| 1 - 50 | 0.684 | 0.458 | <0.001* |

*RUFE - Ranked or Unranked Features Entry

Employing the $t$-test analysis to verify if there is a significant difference between the $R^2$ values for both procedures, the derived $p$-values for significance are summarized in Table 1. For higher level of $\nu$, all the $p$-values all close to 0 indicating that $H_0$: pairwise $R^2$ values are equal is rejected favoring the hypothesis $H_1$: pairwise $R^2$ values are significantly different. Meaning, each pairwise values of $R^2$ are statistically different. This implies that the performance of MFS via BFR outperforms baseline MFS for all feature entries when $SNR = 1$. On the other hand, when $SNR = 8$, that is, for lower level of $\nu$, the same result is obtained except for feature entries 1-5 and 1-15, where both methodologies performed more or less the same.

### 3.1.2 With Function II and $SNR \in \{1, 8\}$

Figure 4 presents the performance of both BFR and MFS feature selection procedures using Function II. When $SNR = 1$, BFR has the highest performance power over MFS after the inclusion of 30 features in the model with $R^2_{BFR} = 56.8\%$ versus $R^2_{MFS} = 28.2\%$ (see Table 2). $R^2_{MFS} = 45.2\%$ is maximum when 1-5 features are included in the model in contrast to $R^2_{BFR} = 38.6\%$ and these values are statistically different favoring baseline MFS for better model fit with 1-5 features, but eventually its performance declines as more features are included in the model. When $SNR = 8$, BFR outperforms MFS only after the inclusion of the first 20 features with $R^2_{BFR} = 79.9\%$ compared to $R^2_{MFS} = 72.5\%$. BFR attains its maximum $R^2 = 81.7\%$ when 30 features are entered into the model. Similarly, MFS attains maximum $R^2_{MFS} = 86.1\%$ for 1-5 features but the model fit performance decreases afterwards. Results may indicate that BFR attains a better model fit that MFS when the level of noise in $y$ is low.



(a) $SNR = 1$ (b) $SNR = 8$

Figure 4: *Model Fit Performance of MFS via BFR and baseline MFS for Function II*

Tables 2(a) and 2(b) gives the $R^2$ values of BFR and MFS procedures when at most 50% of features are included into the model at two different noise level $\nu$. The highlighted values indicate the higher $R^2$ values between the two feature selection methodologies. It can be deduced that baseline MFS requires less number of features (only 5 unranked features) to attain optimal performance but the BFR gave mostly higher values of $R^2$. Moreover, all $p$-values points to the the rejection of the hypothesis of pairwise equality of the $R^2$ values favoring MFS via BFR over baseline MFS with respect to a better model fit at higher level of $\nu$ infused to $y$. For lower level of $\nu$, the same conclusion is derived except at features entries 1-15 where the $R^2$ value of both procedures are not significant different.

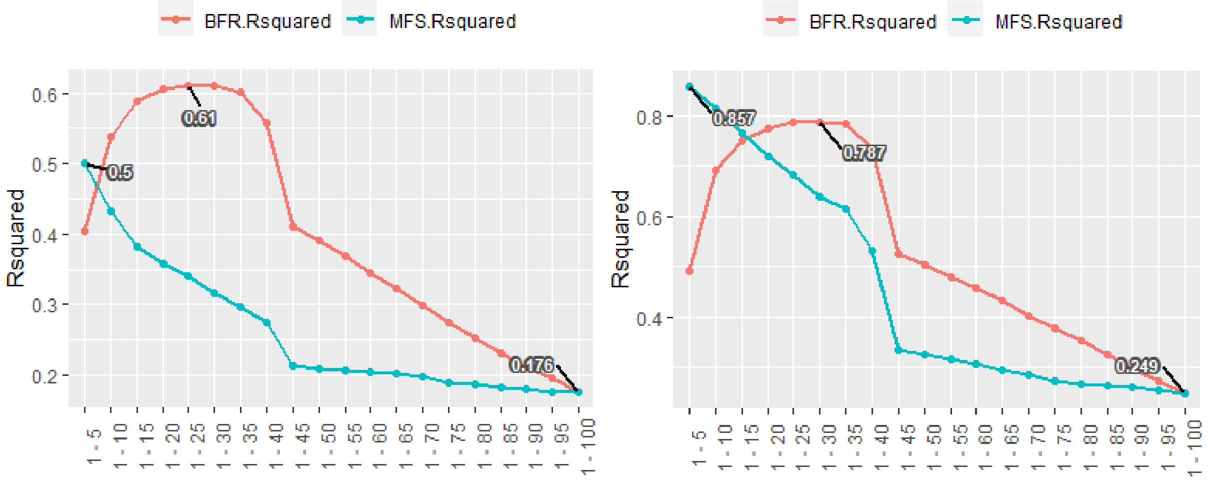
Table 2: $R^2$ *values of MFS via BFR and baseline MFS for Function II*

| (a) $SNR = 1$ | | | |
|---|---|---|---|
| RUFE | $R^2_{BFR}$ | $R^2_{MFS}$ | $T$-test ($p$-value) |
| 1 - 5 | 0.386 | 0.452 | <0.001* |
| 1 - 10 | 0.492 | 0.401 | <0.001* |
| 1 - 15 | 0.535 | 0.357 | <0.001* |
| 1 - 20 | 0.557 | 0.319 | <0.001* |
| 1 - 25 | 0.566 | 0.302 | <0.001* |
| 1 - 30 | 0.568 | 0.282 | <0.001* |
| 1 - 35 | 0.561 | 0.276 | <0.001* |
| 1 - 40 | 0.512 | 0.250 | <0.001* |
| 1 - 45 | 0.375 | 0.183 | <0.001* |
| 1 - 50 | 0.354 | 0.181 | <0.001* |

| (b) $SNR = 8$ | | | |
|---|---|---|---|
| RUFE | $R^2_{BFR}$ | $R^2_{MFS}$ | $T$-test ($p$-value) |
| 1 - 5 | 0.535 | 0.861 | <0.001* |
| 1 - 10 | 0.717 | 0.815 | <0.001* |
| 1 - 15 | 0.775 | 0.773 | 0.829 |
| 1 - 20 | 0.799 | 0.725 | <0.001* |
| 1 - 25 | 0.816 | 0.679 | <0.001* |
| 1 - 30 | 0.817 | 0.633 | <0.001* |
| 1 - 35 | 0.813 | 0.591 | <0.001* |
| 1 - 40 | 0.760 | 0.510 | <0.001* |
| 1 - 45 | 0.520 | 0.306 | <0.001* |
| 1 - 50 | 0.496 | 0.300 | <0.001* |

### 3.1.3 With Function III and $SNR \in \{1, 8\}$

Figure 5(a) and Figure 5(b) shows a result where BFR outperforms MFS after the inclusion of 10 features in the model at higher level of $\nu$ ($R^2_{BFR} = 53.8\% > R^2_{MFS} = 43.3\%$). However, at lower level of $\nu$, that is, when $SNR = 8$, BFR will perform better than MFS only after the inclusion of 20 features. Again, doing pairwise comparison (see Table 3) on the $R^2$ values for both procedures shows that BFR gave higher values of $R^2$ (except when at most 10 features are initially entered into the model). However, the $R^2$ values of MFS decreases as more features are included in the regression model. Moreover, BFR consistently performs better than MFS with respect to model fit performance when $SNR = 8$ except for feature entries 1-15, where both procedures performed similarly.



(a) $SNR = 1$ (b) $SNR = 8$

Figure 5: *Model Fit Performance of MFS via BFR and baseline MFS for Function III*

Table 3: *$R^2$ values of MFS via BFR and baseline MFS for Function III*

| (a) $SNR = 1$ | | | |
|---|---|---|---|
| RUFE | $R^2_{BFR}$ | $R^2_{MFS}$ | $T$-test ($p$-value) |
| 1 - 5 | 0.403 | 0.500 | <0.001* |
| 1 - 10 | 0.538 | 0.433 | <0.001* |
| 1 - 15 | 0.587 | 0.382 | <0.001* |
| 1 - 20 | 0.605 | 0.359 | <0.001* |
| 1 - 25 | 0.610 | 0.340 | <0.001* |
| 1 - 30 | 0.610 | 0.316 | <0.001* |
| 1 - 35 | 0.601 | 0.297 | <0.001* |
| 1 - 40 | 0.557 | 0.274 | <0.001* |
| 1 - 45 | 0.411 | 0.214 | <0.001* |
| 1 - 50 | 0.391 | 0.209 | <0.001* |

| (b) $SNR = 8$ | | | |
|---|---|---|---|
| RUFE | $R^2_{BFR}$ | $R^2_{MFS}$ | $T$-test ($p$-value) |
| 1 - 5 | 0.492 | 0.857 | <0.001* |
| 1 - 10 | 0.690 | 0.813 | <0.001* |
| 1 - 15 | 0.749 | 0.764 | 0.168 |
| 1 - 20 | 0.775 | 0.718 | <0.001* |
| 1 - 25 | 0.786 | 0.681 | <0.001* |
| 1 - 30 | 0.787 | 0.638 | <0.001* |
| 1 - 35 | 0.784 | 0.613 | <0.001* |
| 1 - 40 | 0.735 | 0.531 | <0.001* |
| 1 - 45 | 0.524 | 0.334 | <0.001* |
| 1 - 50 | 0.503 | 0.327 | <0.001* |

## 3.2 Using Real Data

The two methodologies are also processed using the real dataset described in subsection 2.1. The derived $R^2$ values are plotted in Figure 6. It can be observed that MFS via BFR consistently obtained higher $R^2$ values than the baseline MFS, where BFR exhibits an increasing trend of values from ranked gene probes entry 1-5 up to 1-35. Moreover, it can be observed that there is a sudden decline of the $R^2$ values when at most 50% of the features (gene probes) are already included in the model. This may imply that inclusion of at most 50% of the ranked features via BFR is enough to derived a regression model with HDD explanatory variables to obtain a better fit. This proposition was also observed in the simulation study. In this dataset, the MFS via BFR outperforms the baseline MFS in all cases of feature entries.

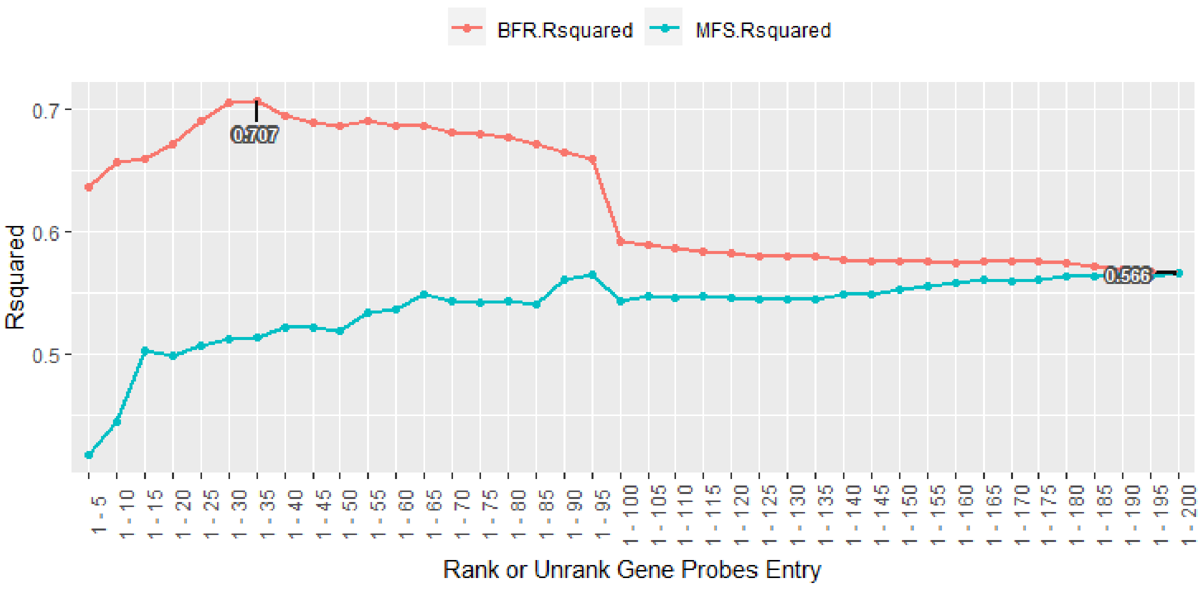


Figure 6: *Model Fit Performance of MFS via BFR and baseline MFS for Real Data*

It would be interesting to do further analysis of the first 35 gene probes included in the derived regression model for it gave an increasing or upward trend of $R^2$ values. The names of these gene probe are listed in Table 4. It can be observed that 3 gene probes are identified by both procedures that are included in the model, namely $X11$, $X13$ and $X19$ (these are highlighted in Table 4).

Table 4: *List of the first 35 gene probes included in the derived model*

| RU | BFR | MFS |
|---|---|---|
| 1 | X87 | X1 |
| 2 | X62 | X2 |
| 3 | X180 | X3 |
| 4 | X153 | X4 |
| 5 | X140 | X5 |
| 6 | X76 | X6 |
| 7 | X134 | X7 |
| 8 | X200 | X8 |
| 9 | X187 | X9 |
| 10 | X155 | X10 |
| 11 | X71 | X11 |
| 12 | X102 | X12 |
| 13 | X50 | X13 |
| 14 | X54 | X14 |
| 15 | X184 | X15 |
| 16 | X13 | X16 |
| 17 | X96 | X17 |
| 18 | X185 | X18 |
| 19 | X19 | X19 |
| 20 | X92 | X20 |
| 21 | X174 | X21 |
| 22 | X90 | X22 |
| 23 | X41 | X23 |
| 24 | X146 | X24 |
| 25 | X66 | X25 |
| 26 | X11 | X26 |
| 27 | X48 | X27 |
| 28 | X99 | X28 |
| 29 | X164 | X29 |
| 30 | X172 | X30 |
| 31 | X42 | X31 |
| 32 | X101 | X32 |
| 33 | X147 | X33 |
| 34 | X157 | X34 |
| 35 | X114 | X35 |

*RU - Rank or Unrank

## 4 Conclusion and Recommendation

This study considers three Functions I, II, and III and two levels of $\nu$ infused to $y$ in a high-dimensional regression problem. It considers and compares the performance of MFS via BFR (with ranked feature) versus baseline MFS (with unranked feature) selection procedures. Simulation study on models with varied degree of correlation among the explanatory variables and infused noise $\nu$ in the response variable shows that MFS via BFR gave better model fit than the baseline MFS as shown in a higher $R^2$ values, on the average, consequently, higher performance power. In other words, ranking the features via a bayesian approach prior to fitting a regression model is better than inclusion of unranked features directly into the model. Moreover, when BFR and MFS were employed to a real dataset, it shows that BFR outperformed MFS since it consistently gave higher values of $R^2$. It was also observed that in both the simulation study and real data application that the $R^2$ values gradually decreases after the inclusion of about 50% of the given ranked features. This may indicate that inclusion of at most 50% of the ranked features via BFR is sufficient to have a final regression model with explanatory variables that are HDD. Also, since this study focused only on model fitting, it is recommended that further investigation be done with the proposed methodology in the area of predictive modeling or forecasting.

## Acknowledgements

Aries P. Valeriano is grateful to DOST-SEI ASTHRDP for the financial support throughout his graduate study as MS Statistics student in MSU-IIT.

# References

[1] *An Introduction to t-tests — Definitions, Formula and Examples.* Retrieved May 2, 2022, from https://www.scribbr.com/statistics/t-test/

[2] Fan, J., Samworth, R., & Wu, Y. (2009). *Ultrahigh dimensional feature selection: Beyond the linear model.* Journal of Machine Learning and Research, (10), 2013-2038. Retrieved from https://fan.princeton.edu/papers/08/GLIMSIS.pdf

[3] Glen, S. *Forward Selection: Definition.* Retrieved March 21, 2022, from https://www.statisticshowto.com/forward-selection/

[4] Győrffy, B., & Schäfer, R. (2008). *Meta-analysis of gene expression profiles related to relapse-free survival in 1,079 breast cancer patients.* Breast Cancer Research and Treatment, 118(3), 433-441. https://doi.org/10.1007/s10549-008-0242-8

[5] *High Dimensional Data.* Retrieved from https://www.sciencedirect.com/topics/computer-science/high-dimensional-data

[6] Johnstone, I. M., & Titterington, D. M. (2009). *Statistical challenges of high-dimensional data.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 367(1906), 4237-4253. https://doi.org/10.1098/rsta.2009.0159

[7] Kumar, A. (2021, February 21). *R-squared, R2 in Linear Regression: Concepts, Examples - Data* . Retrieved July 26, 2022, from https://vitalflux.com/r-squared-explained-machine-learning/

[8] Makalic, E., & Schmidt, D. F. (2011). *A Simple Bayesian Algorithm for Feature Ranking in High Dimensional Regression Problems.* AI 2011: Advances in Artificial Intelligence, 223-230. https://doi.org/10.1007/978-3-642-25832-9_23

[9] Shetty, B. (2019, July 19). *What is Curse of Dimensionality? A Complete Guide | Built In.* Retrieved March 22, 2022, from https://builtin.com/data-science/curse-dimensionality

[10] *Variable Selection in Multiple Regression.* Retrieved March 22, 2022, from https://advstats.psychstat.org/book/mregression/selection.php

[11] Zach,. (2021, February 10). *What is High Dimensional Data? (Definition & Examples).* Retrieved March 22, 2022, from https://www.statology.org/high-dimensional-data/