BWM package

Sergi Baena-Miret, Ferran Reverter, Esteban Vegas

2024-05-21

Contents

Introduction **1**

Packages **1**

Data **1**

Block missing visualization **2**

Profiles **2**

Spliting in train and test sets **7**

Training **7**

Model estimation **8**
Parameter α 8
Parameter β 8
Degree of sparseness 8

Performance **9**

References **9**

Introduction

We present a guide to the `bwm` package.

Packages

The following R packages need to be loaded:

```
library(bwm)
library(visdat)
library(caret)
```

Data

We will load a **subset** from the dataset **Breast Cancer Data** gathered from Chierici et al. (2020). The dataset comprises three types of omics and clinical data obtained through the efforts of The Cancer Genome Atlas (TCGA) Research Network, accessible at https://www.cancer.gov/tcga. This dataset includes: gene

expression, protein abundance and gene copy number variation data, which were employed to predict the estrogen receptor status of breast invasive carcinoma (0: negative; 1: positive).

```
load(file="tutorial_bwm_small.RData")
```

The loaded `RData` object contains three elements:

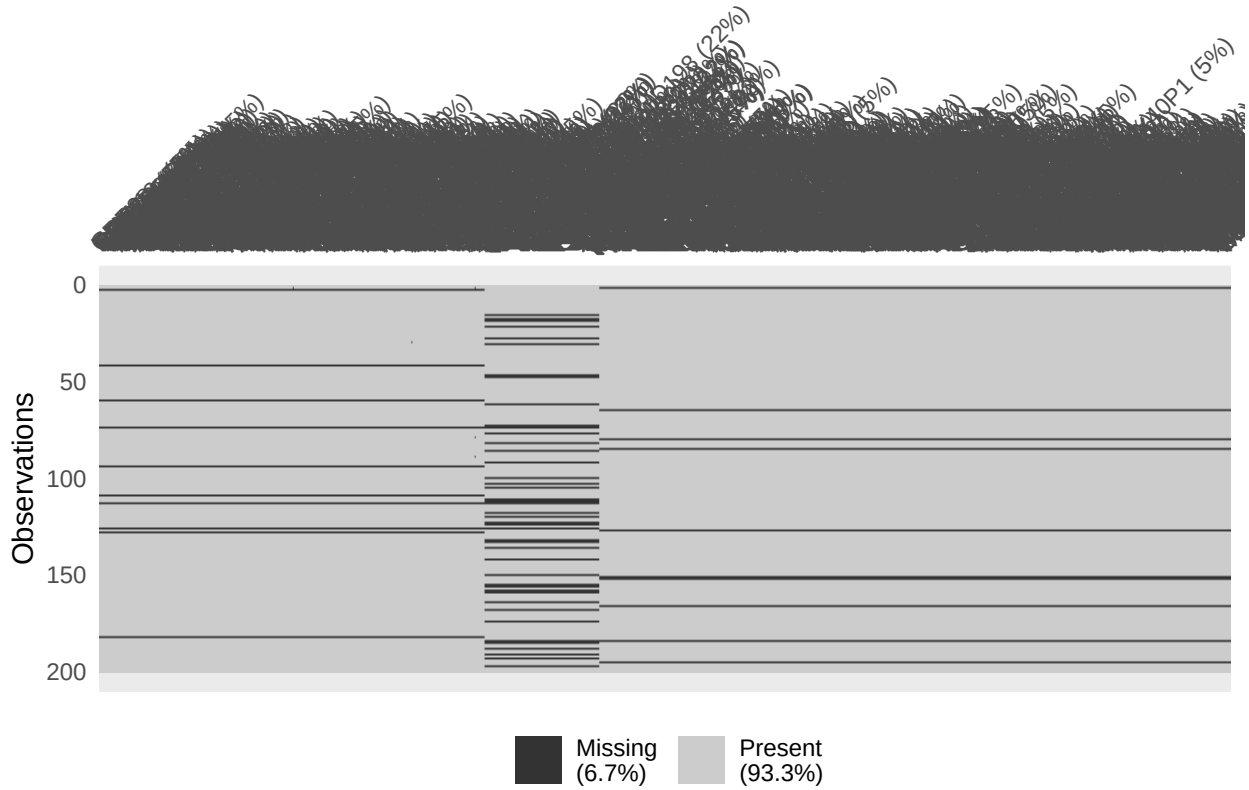
- the data frame `xdata_m` with data.
- the vector `y` with observation labels.
- the vector `p` that indicates how many features there are in each omic source.

The data frame `xdatas_m` contains 200 observations (49 negative and 151 positive), having the expression of 390 genes, the abundance of 116 proteins and the copy number variation of 639 genes.

Block missing visualization

In the dataset there are 10 observations for which gene expression is not available, also 10 observations for which genetic variation is not available and 89 for which protein abundance is not available. This results in an overall percentage of missing data of 6.7%. We can visualize the block-wise missing structure with the `vis_miss()` function

```
vis_miss(xdatas_m,large_data_size=10^9)
```



Profiles

We can extract the profiles using the function `get_profile`.

For each observation, the profile indicates whether a certain omic is present. The observation profile is given by

$$I(\mathsf{gene}) \cdot 2^2 + I(\mathsf{protein}) \cdot 2^1 + I(\mathsf{copy}) \cdot 2^0$$

where $I(\mathsf{omic}) = 1$, if omic is present on that observation or 0 else case. Then, profile 1 corresponds to observations only having copy number variation, profile 2 corresponds to only have protein, profile 3 corresponds to have protein and copy, profile 4 to only have gene expression, profile 5 to have gene expression and copy, profile 6 to have gene expression and protein, and profile 7 to have all three omics.

In accordance with the order of the sources in vector `p`, we get the profile of each observation.

```
get_profile(p , xdatas_m, showTable = T, F)
```

```
##     Source 1 Source 2 Source 3 Profile
## 1          1        1        1       7
## 2          0        1        0       2
## 3          0        1        1       3
## 4          1        1        1       7
## 5          1        1        1       7
## 6          1        1        1       7
## 7          1        1        1       7
## 8          1        1        1       7
## 9          1        1        1       7
## 10         1        1        1       7
## 11         1        1        1       7
## 12         1        1        1       7
## 13         1        1        1       7
## 14         1        1        1       7
## 15         1        1        1       7
## 16         1        0        1       5
## 17         1        1        1       7
## 18         1        0        1       5
## 19         1        0        1       5
## 20         1        1        1       7
## 21         1        1        1       7
## 22         1        0        1       5
## 23         1        1        1       7
## 24         1        1        1       7
## 25         1        1        1       7
## 26         1        1        1       7
## 27         1        1        1       7
## 28         1        0        1       5
## 29         1        1        1       7
## 30         0        1        1       3
## 31         1        0        1       5
## 32         1        1        1       7
## 33         1        1        1       7
## 34         1        1        1       7
## 35         1        1        1       7
## 36         1        1        1       7
## 37         1        1        1       7
## 38         1        1        1       7
## 39         1        1        1       7
## 40         1        1        1       7
```

```
## 41          1         1         1         7
## 42          0         1         1         3
## 43          1         1         1         7
## 44          1         1         1         7
## 45          1         1         1         7
## 46          1         1         1         7
## 47          1         0         1         5
## 48          1         0         1         5
## 49          1         1         1         7
## 50          1         1         1         7
## 51          1         1         1         7
## 52          1         1         1         7
## 53          1         1         1         7
## 54          1         1         1         7
## 55          1         1         1         7
## 56          1         1         1         7
## 57          1         1         1         7
## 58          1         1         1         7
## 59          1         1         1         7
## 60          0         1         1         3
## 61          1         1         1         7
## 62          1         0         1         5
## 63          1         1         1         7
## 64          1         1         1         7
## 65          1         1         0         6
## 66          1         1         1         7
## 67          1         1         1         7
## 68          1         1         1         7
## 69          1         1         1         7
## 70          1         1         1         7
## 71          1         1         1         7
## 72          1         1         1         7
## 73          1         0         1         5
## 74          0         0         1         1
## 75          1         1         1         7
## 76          1         1         1         7
## 77          1         0         1         5
## 78          1         1         1         7
## 79          0         1         1         3
## 80          1         1         0         6
## 81          1         1         1         7
## 82          1         0         1         5
## 83          1         1         1         7
## 84          1         1         1         7
## 85          1         1         0         6
## 86          1         0         1         5
## 87          1         1         1         7
## 88          1         1         1         7
## 89          0         1         1         3
## 90          1         1         1         7
## 91          1         1         1         7
## 92          1         0         1         5
## 93          1         1         1         7
## 94          0         1         1         3
```

```
## 95          1         1         1         7
## 96          1         1         1         7
## 97          1         1         1         7
## 98          1         1         1         7
## 99          1         1         1         7
## 100         1         0         1         5
## 101         1         1         1         7
## 102         1         1         1         7
## 103         1         0         1         5
## 104         1         1         1         7
## 105         1         0         1         5
## 106         1         1         1         7
## 107         1         1         1         7
## 108         1         1         1         7
## 109         0         1         1         3
## 110         1         1         1         7
## 111         1         0         1         5
## 112         1         0         1         5
## 113         0         0         1         1
## 114         1         1         1         7
## 115         1         1         1         7
## 116         1         1         1         7
## 117         1         1         1         7
## 118         1         0         1         5
## 119         1         1         1         7
## 120         1         0         1         5
## 121         1         1         1         7
## 122         1         1         1         7
## 123         1         0         1         5
## 124         1         0         1         5
## 125         1         1         1         7
## 126         0         0         1         1
## 127         1         1         0         6
## 128         0         1         1         3
## 129         1         1         1         7
## 130         1         1         1         7
## 131         1         1         1         7
## 132         1         0         1         5
## 133         1         0         1         5
## 134         1         1         1         7
## 135         1         1         1         7
## 136         1         0         1         5
## 137         1         1         1         7
## 138         1         1         1         7
## 139         1         1         1         7
## 140         1         1         1         7
## 141         1         1         1         7
## 142         1         0         1         5
## 143         1         1         1         7
## 144         1         1         1         7
## 145         1         1         1         7
## 146         1         1         1         7
## 147         1         1         1         7
## 148         1         1         1         7
```

```
## 149        1        1        1       7
## 150        1        0        1       5
## 151        1        1        0       6
## 152        1        1        0       6
## 153        1        1        1       7
## 154        1        1        1       7
## 155        1        0        1       5
## 156        1        0        1       5
## 157        1        1        1       7
## 158        1        0        1       5
## 159        1        0        1       5
## 160        1        1        1       7
## 161        1        1        1       7
## 162        1        1        1       7
## 163        1        1        1       7
## 164        1        0        1       5
## 165        1        1        1       7
## 166        1        1        0       6
## 167        1        1        1       7
## 168        1        0        1       5
## 169        1        1        1       7
## 170        1        1        1       7
## 171        1        1        1       7
## 172        1        1        1       7
## 173        1        1        1       7
## 174        1        0        1       5
## 175        1        1        1       7
## 176        1        1        1       7
## 177        1        1        1       7
## 178        1        1        1       7
## 179        1        1        1       7
## 180        1        1        1       7
## 181        1        1        1       7
## 182        0        1        1       3
## 183        1        1        1       7
## 184        1        0        0       4
## 185        1        0        1       5
## 186        1        1        1       7
## 187        1        1        1       7
## 188        1        0        1       5
## 189        1        1        1       7
## 190        1        1        1       7
## 191        1        0        1       5
## 192        1        1        1       7
## 193        1        0        1       5
## 194        1        1        1       7
## 195        1        1        0       6
## 196        1        1        1       7
## 197        1        0        1       5
## 198        1        1        1       7
## 199        1        1        1       7
## 200        1        1        1       7

##   [1] 7 2 3 7 7 7 7 7 7 7 7 7 7 7 7 5 7 5 5 7 7 5 7 7 7 7 7 5 7 3 5 7 7 7 7 7 7
```

```
##  [38] 7 7 7 7 3 7 7 7 7 5 5 7 7 7 7 7 7 7 7 7 7 7 3 7 5 7 7 6 7 7 7 7 7 7 7 5 1
##  [75] 7 7 5 7 3 6 7 5 7 7 6 5 7 7 3 7 7 5 7 3 7 7 7 7 7 5 7 7 5 7 5 7 7 7 3 7 5
## [112] 5 1 7 7 7 7 5 7 5 7 7 5 5 7 1 6 3 7 7 7 5 5 7 7 5 7 7 7 7 7 5 7 7 7 7 7 7
## [149] 7 5 6 6 7 7 5 5 7 5 5 7 7 7 7 5 7 6 7 5 7 7 7 7 7 5 7 7 7 7 7 7 7 3 7 4 5
## [186] 7 7 5 7 7 5 7 5 7 6 7 5 7 7 7
## Levels: 1 2 3 4 5 6 7
```

We can compute the profiles size

```
table(get_profile(p , xdatas_m, showTable = F, F))
```

```
##
##   1   2   3   4   5   6   7
##   3   1  10   1  40   8 137
```

In summary, we note that there are 3 observations in profile 1, 1 in profile 2, and so up to 137 in profile 7.

Spliting in train and test sets

We will separate at random the observations, 67% for training and the rest for testing. Preserving the ratio of positives and negatives in both parts.

```
set.seed(123456)
prob_train <- 0.67

# We select the indices that we will use for training
indexes_partition <- sample(1:dim(xdatas_m)[1],dim(xdatas_m)[1]*prob_train)
# Data matrix numeric variables
xdatas_train <- xdatas_m[indexes_partition, ]
xdatas_test <- xdatas_m[-indexes_partition, ]


y_train <- y[indexes_partition]
y_test <- y[-indexes_partition]
```

```
table(y_train)
```

```
## y_train
## Negative Positive
##       35       99
```

```
table(y_test)
```

```
## y_test
## Negative Positive
##       14       52
```

Training

We will train the model with the `bwm.train()` function. We highlight the following arguments:

- `p` ... Numeric vector that indicates how many features there are in each omic source. In this example we have `p=c(390, 116, 639)`.
- `X` ... Data matrix for training. In rows, the observations, in columns, the features grouped in same order as indicated by the vector `p`.
- `y` ... Response vector for training.

More information about the function arguments can be found in `bwm` package manual.

Spended time to find the solution is close to 5 minutes.

```
mod<-bwm.train(p=p, X=xdatas_train, y=y_train,lambda = 0.00000005, L.step = 10, maxIter=300,
               maxIter.beta = 50, to.normalize = T)
```

```
##   |                                                                      |
## For the parameter lambda = 5e-08 the algorithm has converged before reaching the maximum number of i
```

Function `bwm.train` returns a list with 8 information slots. Main slots are:

- $\boldsymbol{\alpha}_i$ estimations $i = 1, ..., 7$
- $\boldsymbol{\beta}_j$ estimations $j = 1, 2, 3$

Model estimation

Parameter α

```
mod$alpha
```

```
## [[1]]
## [1] 0.000000000 0.000000000 0.008064516
##
## [[2]]
## [1] 0.000000000 0.009433962 0.000000000
##
## [[3]]
## [1] 0.00000000 0.01030928 0.01030928
##
## [[4]]
## [1] 0.008264463 0.000000000 0.000000000
##
## [[5]]
## [1] 0.008928571 0.000000000 0.008928571
##
## [[6]]
## [1] 0.01041667 0.01041667 0.00000000
##
## [[7]]
## [1] 0.01136364 0.01136364 0.01136364
```

Parameter β

```
betageneexpres<-mod$beta[1:p[1]]
betaprot<-mod$beta[(p[1]+1):(p[1]+p[2])]
betacnv<-mod$beta[(p[2]+1):(p[2]+p[3])]
```

Degree of sparseness

```
length(which(betageneexpres==0))
```

```
## [1] 159
```

```
length(which(betaprot==0))
```

```
## [1] 8
```

```
length(which(betacnv==0))
```

```
## [1] 115
```

Performance

We will evaluate the performace on test set using the function `bwm.predict()`.

```
y_pred_test <- bwm.predict(mod, xdatas_test, p)
confusionMatrix(as.factor(y_pred_test),as.factor(y_test))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction Negative Positive
##   Negative       11        7
##   Positive        3       45
##
##                Accuracy : 0.8485
##                  95% CI : (0.739, 0.9249)
##     No Information Rate : 0.7879
##     P-Value [Acc > NIR] : 0.1449
##
##                   Kappa : 0.5896
##
##  Mcnemar's Test P-Value : 0.3428
##
##             Sensitivity : 0.7857
##             Specificity : 0.8654
##          Pos Pred Value : 0.6111
##          Neg Pred Value : 0.9375
##              Prevalence : 0.2121
##          Detection Rate : 0.1667
##    Detection Prevalence : 0.2727
##       Balanced Accuracy : 0.8255
##
##        'Positive' Class : Negative
##
```

References

Chierici, Marco, Nicole Bussola, Alessia Marcolini, Margherita Francescatto, Alessandro Zandonà, Lucia Trastulla, Claudio Agostinelli, Giuseppe Jurman, and Cesare Furlanello. 2020. "Integrative Network Fusion: A Multi-Omics Approach in Molecular Profiling." *Frontiers in Oncology* 10: 1065.