Short introduction to package MAVE with example

Hang Weiqiang(E0010758@u.nus.edu)

February 1, 2017

1 Introduction

Package MAVE provides several methods, including MAVE and OPG methods proposed by [4, 5, 6], to find the central space (CS) and the central mean space (CMS). It also implements sliced inverse regression of a kernel version; see [1, 4]. Formal definition of the central space and the central mean space can be found in [2, 3]. For comparison, a package dr in CRAN also contains other sufficient dimension reduction methods[7].

The main part of package MAVE is written in C++ based on RcppArmadillo package. If there is any problem during installation, please update your Rcpp package and install RcppArmadillo package and try again.

2 Usage

The primary function in this package is MAVE(). The input arguments include an $n \times p$ covariate matrix X, an $n \times 1$ respond matrix Y, and the method argument for dimension reduction. The options for the method argument are 'csopg', 'csmave','meanopg','meanmave' and 'ksir', and the default is 'csopg'. 'csopg' and 'csmave' are methods of finding CS by OPG and MAVE respectively, 'meanopg' and 'meanmave' are methods of finding CMS by OPG and MAVE, 'ksir' is the sliced inverse regression of kernel version. MAVE() is very easy to use. The following is a simple example.

```
> set.seed(12345)
> library(MAVE)
> X <- matrix(rnorm(200*5),200,5)
> b1 <- matrix(c(1,1,0,0,0),5,1)
```

```
> b2 <- matrix(c(0,0,1,1,0),5,1)
> eps <- matrix(rnorm(200),200,1)
> y <- X%*%b1 + (X%*%b2)*eps
> rd <- MAVE(X,y)
```

In the example above, 'csopg' is actually used. In many cases, the result of OPG is as good as MAVE but its computation is much more time-efficient, hence OPG method is recommended.

The class of object created by MAVE function is a S3 class variable containing the matrix of CS or CMS of different dimensions. By typing the name of the returned object, the output can be viewed.

```
> rd

Call:
MAVE(x = X, y = y)

            [,1]
[1,]  0.73036278
[2,]  0.66689551
[3,]  0.07898421
[4,]  0.12409445
[5,] -0.01351479
            [,1]        [,2]
[1,]  0.73087051  0.02846618
[2,]  0.66709161  0.17485426
[3,]  0.08910854 -0.78195783
[4,]  0.11294015 -0.59225540
[5,] -0.01101216  0.07994467
            [,1]       [,2]        [,3]
[1,]  0.72527299 -0.1283312 -0.62517073
[2,]  0.68482867  0.2307634  0.64981465
[3,]  0.03807079 -0.7397801  0.41193890
[4,]  0.05828901 -0.6104505 -0.11030679
[5,] -0.01190724  0.1017591  0.07100167
(only the first 3  central space  are displayed,
        to display the space of dimension k, call object$dir[[k]])
```

The output will display the first three matrices of CSs or CMSs estimated by MAVE function. To select dimension of the space by cross-validation method, DIM method can be used.

```
> dim <- DIM(rd)
> dim

Call:
DIM(rd = rd)

Dimension        1            2            3            4            5
CV-value         0.19         0.18         0.2          0.21         0.23

The selected dimension is  2
```

The output shows the corresponding cv-value for different dimensions. More information is stored in the object. These includes:

```
> names(dim)

[1] "ky"       "x"        "call"     "method"   "dir"      "cv"        "best_dir"
```

In order to access other information such as CS or CMS all dimensions, just type

```
> dim$dir

[[1]]
            [,1]
[1,]  0.73036278
[2,]  0.66689551
[3,]  0.07898421
[4,]  0.12409445
[5,] -0.01351479

[[2]]
            [,1]        [,2]
[1,]  0.73087051  0.02846618
[2,]  0.66709161  0.17485426
[3,]  0.08910854 -0.78195783
[4,]  0.11294015 -0.59225540
[5,] -0.01101216  0.07994467

[[3]]
            [,1]       [,2]        [,3]
[1,]  0.72527299 -0.1283312 -0.62517073
```

```
[2,]  0.68482867  0.2307634  0.64981465
[3,]  0.03807079 -0.7397801  0.41193890
[4,]  0.05828901 -0.6104505 -0.11030679
[5,] -0.01190724  0.1017591  0.07100167

[[4]]
            [,1]       [,2]       [,3]       [,4]
[1,] 0.726903040 -0.1225114  0.5761828 -0.2586219
[2,] 0.681825923  0.2425726 -0.5749784  0.2682844
[3,] 0.066267461 -0.6961773 -0.5080248 -0.4816879
[4,] 0.048194359 -0.6393452  0.1588111  0.7493841
[5,] 0.003362716  0.1808991 -0.2326006  0.2598818

[[5]]
            [,1]        [,2]        [,3]       [,4]        [,5]
[1,] 0.751665455 -0.03301067 -0.55273257  0.2905645  0.20968626
[2,] 0.651380673 -0.10344711  0.66448326 -0.2946907 -0.19136693
[3,] 0.057159109  0.82967878  0.29982542  0.4609763  0.07727573
[4,] 0.085927398  0.50295631 -0.40325111 -0.5693021 -0.50292654
[5,] 0.007182062  0.21651111  0.02102295 -0.5404865  0.81271380
```

In order to access the best CS or CMS selected by DIM method, just type

```
> dim$best_dir

            [,1]        [,2]
[1,]  0.73087051  0.02846618
[2,]  0.66709161  0.17485426
[3,]  0.08910854 -0.78195783
[4,]  0.11294015 -0.59225540
[5,] -0.01101216  0.07994467
```

References

[1] Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414), 316-327.

[2] Cook, R.D.(1998),*Regression Graphics*. New York: Wiley

[3] Cook, R. D., and Li, B. (2002). Dimension reduction for conditional mean in regression. *Annals of Statistics*, 455-474.

[4] Xia, Y., Tong, H., Li, W. K., and Zhu, L. X. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3), 363-410.

[5] Xia, Y. (2007) A constructive approach to the estimation of dimension reduction directions. *Annals of Statistics*, 35(3), 2654-2690

[6] Wang, H., and Xia, Y. (2008). Sliced regression for dimension reduction. *Journal of the American Statistical Association*, 103(482), 811-821.

[7] Weisberg, S. (2002). Dimension reduction regression in R. Journal of Statistical Software, 7(1), 1-22.