Manual for 'Mixture of Varying Coefficient Model with Component Selection'

July 17, 2018

1 Function

1.1 Mixture_VaryingCo_LongReg_Sp

Function for the EM algorithm for estimating the mixture of varying coefficient model without random effect.

1.1.1 Input

Y Y is a $n \times \max_i m_i$ matrix for the response matrix, where n is the number of subjects and m_i is the number observation of subject i. Code the missing value using `inf`.

X X is a $q \times (\max_i m_i) \times n$ matrix for covariate with varying coefficients. q is the number of components with varying coefficients. n is the number of subjects. m_i is the number observation of subject i.

U U is a $n \times \max_i m_i$ for modifier. n is the number of subjects. m_i is the number observation of subject i.

Z Z is a $p \times (\max_i m_i) \times n$ matrix for covariate with fixed coefficient. p is the number of components with fix coefficients. n is the number of subjects. m_i is the number observation of subject i.

Component The number of mixture components start with.

Knot A $q \times 1$ matrix for the number of knots used in splines. q is the number of components with varying coefficients.

Int The support of modifier. If the input for Int is empty (that means [] in matlab), use $[\min_{i,j} u_{ij}, \max_{i,j} u_{ij}]$ for the support of modifier.

lambda The penalty for mixture proportion.

penalty Indicates for whether penalized the mixture proportion. If penalty $= 0$: will not penalize the mixture proportion. If penalty $= 2$: use the proposed SCAD type penalty.

1.1.2 Output

A A $D \times C$ matrix for estimated coefficient for spline basis. D is the number of coefficient for spline. C is the (estimated) number of mixture components.

B A $p \times C$ matrix for estimated fixed coefficients. p is the number of components with fix coefficients. C is the (estimated) number of mixture components.

sigma A $1 \times C$ matrix for estimated variance of C components.C is the (estimated) number of mixture components.

C The estimated number of mixture components.

Loglikeli Log-likelihood of the estimated model.

1.2 Mixture_VaryingCo_LongReg_RanE_Sp

Main function for the nested EM algorithm for estimating the mixture of varying coefficient model with random effect.

1.2.1 Input

Y Y is a $n \times \max_i m_i$ matrix for the response matrix, where n is the number of subjects and m_i is the number observation of subject i. Code the missing value using `inf`.

X X is a $q \times (\max_i m_i) \times n$ matrix for covariate with varying coefficients. q is the number of components with varying coefficients. n is the number of subjects. m_i is the number observation of subject i.

U U is a $n \times \max_i m_i$ for modifier. n is the number of subjects. m_i is the number observation of subject i.

Z Z is a $p \times (\max_i m_i) \times n$ matrix for covariate with fixed coefficient. p is the number of components with fix coefficients. n is the number of subjects. m_i is the number observation of subject i.

V V is a $s \times (\max_i m_i) \times n$ matrix for covariate for random effect. s is the number of components for random effect. n is the number of subjects. m_i is the number observation of subject i.

Component The number of mixture components start with.

Knot A $q \times 1$ matrix for the number of knots used in splines. q is the number of components with varying coefficients.

Int The support of modifier. If the input for Int is empty (that means [] in matlab), use $[\min_{i,j} u_{ij}, \max_{i,j} u_{ij}]$ for the support of modifier.

lambda The penalty for mixture proportion. The actual penalty used will be $lambda \times n^{-0.25}$.

penalty Indicates for whether penalized the mixture proportion. If penalty $= 0$: will not penalize the mixture proportion. If penalty $= 2$: use the proposed SCAD type penalty.

isplot Indicates for whether plot the varying coefficient after the convergence of the algorithm. If 0: will not plot. If 1: will plot. The plot will come one by one. Press any button to get the next plot. The related information will appear in the title of the plot.

1.2.2 Output

ResA A $D \times C$ matrix for estimated coefficient for spline basis. D is the number of coefficient for spline. C is the (estimated) number of mixture components.

ResB A $p \times C$ matrix for estimated fixed coefficients. p is the number of components with fix coefficients. C is the (estimated) number of mixture components.

Ressigma A $1 \times C$ matrix for estimated variance of C components.C is the (estimated) number of mixture components.

ResBsigma A $s \times s \times C$ matrix for estimated random effect of C components. C is the (estimated) number of mixture components.

ResPi A $1 \times C$ matrix for estimated mixture proportions.

ResLoglikeli Log-likelihood of the estimated model.

ResC The estimated number of mixture components.

1.3 Mix_VarCo_LLRR_PenBIC

Tuning hyper-parameters using information criterion.

1.3.1 Input

Y Y is a $n \times \max_i m_i$ matrix for the response matrix, where n is the number of subjects and m_i is the number observation of subject i. Code the missing value using `inf`.

X X is a $q \times (\max_i m_i) \times n$ matrix for covariate with varying coefficients. q is the number of components with varying coefficients. n is the number of subjects. m_i is the number observation of subject i.

U U is a $n \times \max_i m_i$ for modifier. n is the number of subjects. m_i is the number observation of subject i.

Z Z is a $p \times (\max_i m_i) \times n$ matrix for covariate with fixed coefficient. p is the number of components with fix coefficients. n is the number of subjects. m_i is the number observation of subject i.

V V is a $s \times (\max_i m_i) \times n$ matrix for covariate for random effect. s is the number of components for random effect. n is the number of subjects. m_i is the number observation of subject i.

Component The number of mixture components start with.

Knot A $q \times W_1$ matrix for the number of knots used in splines. q is the number of components with varying coefficients. W_1 is the number of candidate spline spaces.

Int The support of modifier. If the input for Int is empty (that means [] in matlab), use $[\min_{i,j} u_{ij}, \max_{i,j} u_{ij}]$ for the support of modifier.

lambda A $1 \times W_2$ matrix. The penalty for mixture proportion. W_2 is the number of candidate lambdas. The actual penalty used will be $lambda \times n^{-0.25}$.

penalty Indicates for whether penalized the mixture proportion. If penalty $= 0$: will not penalize the mixture proportion. If penalty $= 2$: use the proposed SCAD type penalty.

1.3.2 Output

AICResC Selected number of mixture components based on AIC. (Not proposed in the manuscript.)

BICResC Selected number of mixture components based on BIC.

AICResKnot Selected spline space based on AIC. (Not proposed in the manuscript.)

BICResKnot Selected spline space based on BIC.

1.4 SCAD

Function for calculating the SCAD penalty

1.4.1 Input

x The input value of the function

lambda The penalty value of SCAD. We set $\alpha = 3.7$ here.

deri Indicator for whether to return the derivative of SCAD. deri $= 1$: return the derivative. deri $= 0$: return the SCAD value.

1.4.2 Output

res The value of function.

1.5 Trun_PowB

Generate the truncated power basis.

1.6 Input

Int A 1×2 matrix for the interval of spline.

NumK Number of knots used in spline.

Degree Degree of spline.

x A point for calculating the value for spline function.

2 Examples

The Simu_Data1() and Simu_Data2() are functions for generating the simulation study case1 and case2, respectively. Sample code for simulation is as follows.

2.1 Simulation case 1

Res is the matrix storing the selected number of components. The algorithm will write an excel file when it finishes a simulation repetition. Sample code:

```
REP=100; %number of simulation repetitions
Res=[];
Knot=[0,1,2,3];
lambda=[0.2,0.3,0.35,0.4];
for rep=1:REP
   [Y,X,U,Z,V,~]=Simu_Data();
  [~,PenBICC,~,~]=Mix_VarCo_LLRR_PenBIC(Y,X,U,Z,V,10,Knot,[],lambda,2);
   Res=[Res,PenBICC]
   xlswrite('Res',Res);
end
```

2.2 Simulation case 2

Sample code:

```
REP=100;
Res=[];
Knot=[4 4 5 5;1 2 1 2];
lambda=[0.2 0.25 0.3 0.35];
for rep=1:REP
```

```
    [Y,X,U,Z,V,~]=Simu_Data2();
   [~,PenBICC,~,~]=Mix_VarCo_LLRR_PenBIC(Y,X,U,Z,V,10,Knot,[],lambda,2);
    Res=[Res,PenBICC]
    xlswrite('Res',Res);
end
```