Supplementary Material

Code:

1.code for nonparametric Bayesian method：

```
proc discrim data=a method=npar kernel=normal r=0.5 crosslisterr;
class y;( y- default status or default loss rate)
priors proportional;
var x1-x13;(x- indicator)
run;
(The value of r should be adjusted continuously until the overall discrimination
accuracy is the highest)
```

2. Parameter Bayesian method code：

```
Proc discrim data=b crosslist;
Class y; ( y- default status or default loss rate)
Var x1-x9; (x- indicator)
Run;
```

3. Parameter clustering code：

```
proc cluster data=c method=ward;
var z;(z- default loss rate)
id xh;
proc tree;
run;
```

4. Nonparametric clustering code:

```
proc modeclus data=c method=6 k=2 test list;
var z; (z- default loss rate)
run;
```

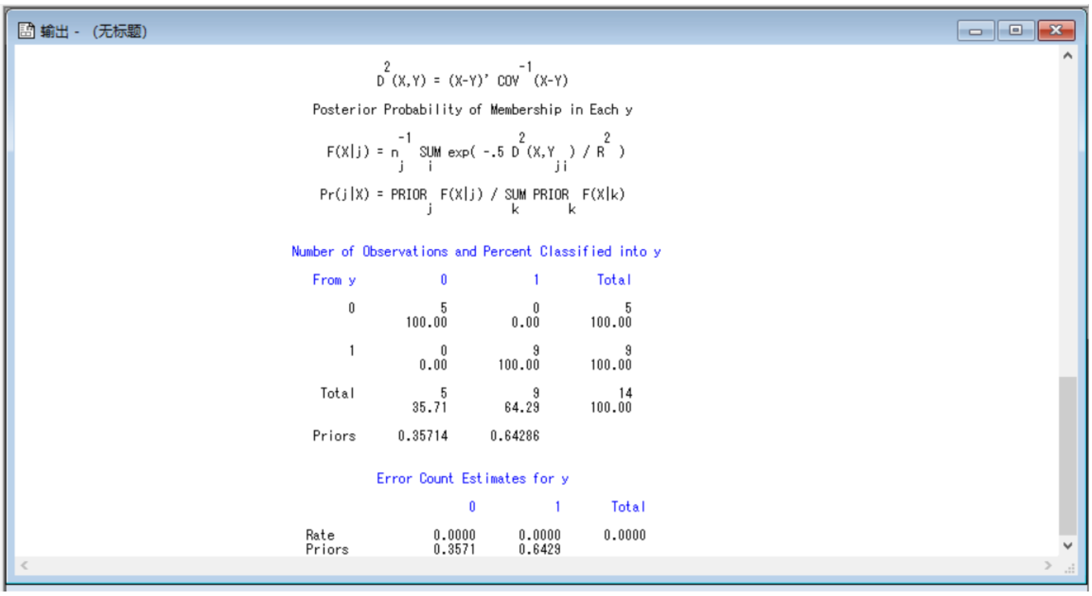
5.Logistic regression code：

```
proc logistic descending data=d;
model y = x1-x81 /selection = forward slentry = 0.05 details;
run;
(y- default status or default loss rate)
(x- indicator)
```

According to the reviewer’s comment, in order to verify the stability of the model, we randomly selected 80% of all samples as the training set and 20% of the samples as the test set, and carried out three simple cross-validation of the model. The verification results are as follows:

The results of the first simple cross-validation are as follows:

Nonparametric model results:



$$D^2(X,Y) = (X-Y)' \, COV^{-1} \, (X-Y)$$

Posterior Probability of Membership in Each y

$$F(X|j) = n_j^{-1} \; SUM_i \; \exp(-.5 \, D^2(X,Y_{ji}) / R^2)$$

$$Pr(j|X) = PRIOR_j \, F(X|j) \,/\, SUM_k \, PRIOR_k \, F(X|k)$$

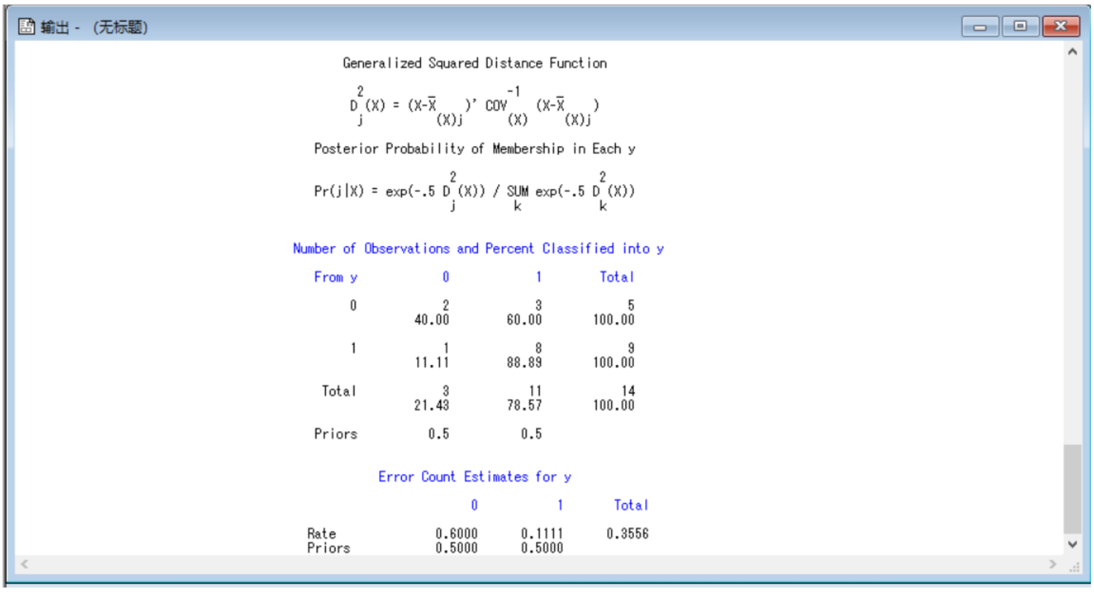
Number of Observations and Percent Classified into y

From y	0	1	Total
0	5 100.00	0 0.00	5 100.00
1	0 0.00	9 100.00	9 100.00
Total	5 35.71	9 64.29	14 100.00
Priors	0.35714	0.64286	

Error Count Estimates for y

	0	1	Total
Rate	0.0000	0.0000	0.0000
Priors	0.3571	0.6429	

Parametric model results:



Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} \, (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each y

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \,/\, SUM_k \, \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into y

From y	0	1	Total
0	2 40.00	3 60.00	5 100.00
1	1 11.11	8 88.89	9 100.00
Total	3 21.43	11 78.57	14 100.00
Priors	0.5	0.5	

Error Count Estimates for y

	0	1	Total
Rate	0.6000	0.1111	0.3556
Priors	0.5000	0.5000	

Logistic regression model results:

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} \, (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each y

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, \sum_k \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into y

From y	0	1	Total
0	9 56.25	7 43.75	16 100.00
1	10 24.39	31 75.61	41 100.00
Total	19 33.33	38 66.67	57 100.00
Priors	0.5	0.5	

Error Count Estimates for y

	0	1	Total
Rate	0.4375	0.2439	0.3407
Priors	0.5000	0.5000	

The results of the second simple cross-validation are as follows:

Nonparametric model results:

Posterior Probability of Membership in Each y

$$F(X|j) = n_j^{-1} \sum_i \exp(-.5 \, D^2(X,Y_{ji}) \, / \, R^2)$$

$$Pr(j|X) = PRIOR_j \, F(X|j) \, / \, \sum_k PRIOR_k \, F(X|k)$$

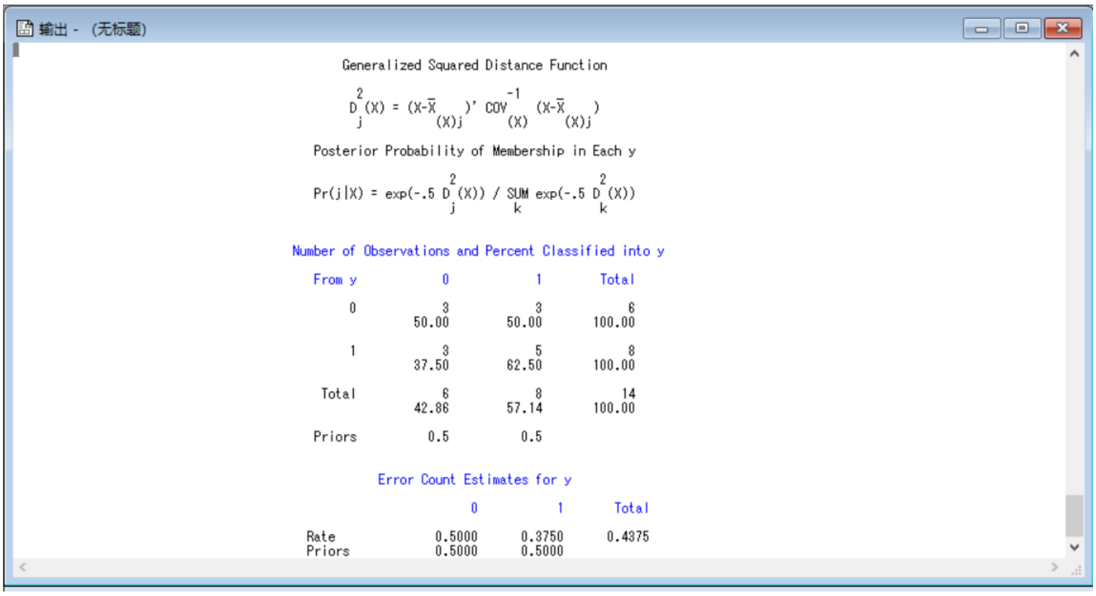
Number of Observations and Percent Classified into y

From y	0	1	Total
0	5 83.33	1 16.67	6 100.00
1	0 0.00	8 100.00	8 100.00
Total	5 35.71	9 64.29	14 100.00
Priors	0.42857	0.57143	

Error Count Estimates for y

	0	1	Total
Rate	0.1667	0.0000	0.0714
Priors	0.4286	0.5714	

Parametric model results:


输出 - (无标题)

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} \, (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each y

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, SUM_k \exp(-.5 \, D_k^2(X))$$

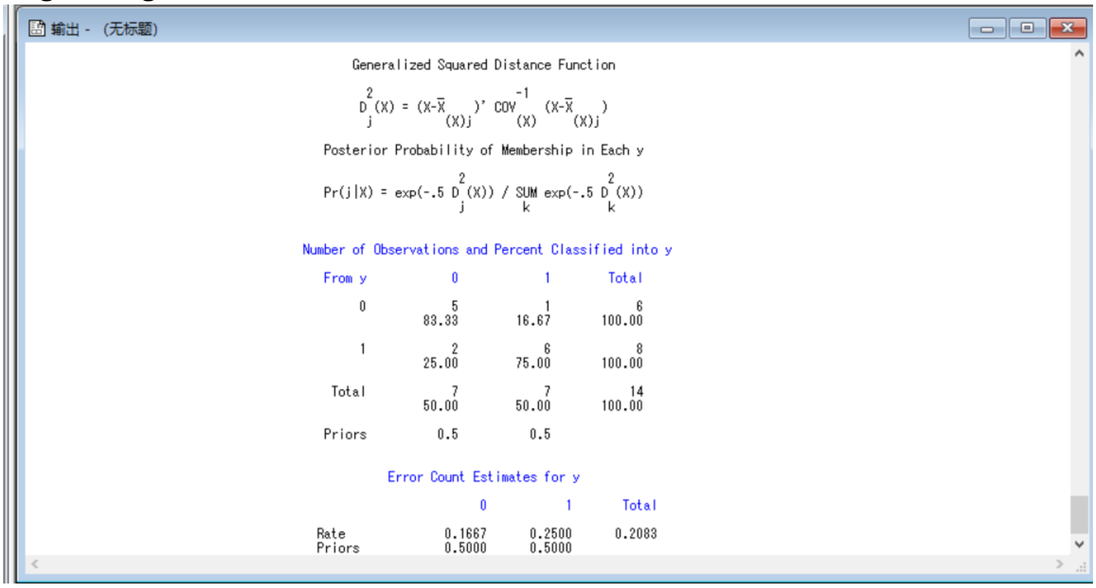
Number of Observations and Percent Classified into y

From y	0	1	Total
0	3 50.00	3 50.00	6 100.00
1	3 37.50	5 62.50	8 100.00
Total	6 42.86	8 57.14	14 100.00
Priors	0.5	0.5	

Error Count Estimates for y

	0	1	Total
Rate	0.5000	0.3750	0.4375
Priors	0.5000	0.5000	

Logistic regression model results:


输出 - (无标题)

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} \, (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each y

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, SUM_k \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into y

From y	0	1	Total
0	5 83.33	1 16.67	6 100.00
1	2 25.00	6 75.00	8 100.00
Total	7 50.00	7 50.00	14 100.00
Priors	0.5	0.5	

Error Count Estimates for y

	0	1	Total
Rate	0.1667	0.2500	0.2083
Priors	0.5000	0.5000	

The results of the third simple cross-validation are as follows:

Nonparametric model results:

输出 - (无标题)

$$D^2(X,Y) = (X-Y)' \, COV^{-1} \, (X-Y)$$

Posterior Probability of Membership in Each y

$$F(X|j) = n_j^{-1} \, SUM_i \, \exp(-.5 \, D^2(X,Y_{ji}) / R^2)$$

$$Pr(j|X) = PRIOR_j \, F(X|j) \, / \, SUM_k \, PRIOR_k \, F(X|k)$$

Number of Observations and Percent Classified into y

From y	0	1	Total
0	7 100.00	0 0.00	7 100.00
1	0 0.00	7 100.00	7 100.00
Total	7 50.00	7 50.00	14 100.00
Priors	0.5	0.5	

Error Count Estimates for y

	0	1	Total
Rate	0.0000	0.0000	0.0000
Priors	0.5000	0.5000	

Parametric model results:

输出 - (无标题)

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} \, (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each y

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, SUM_k \, \exp(-.5 \, D_k^2(X))$$

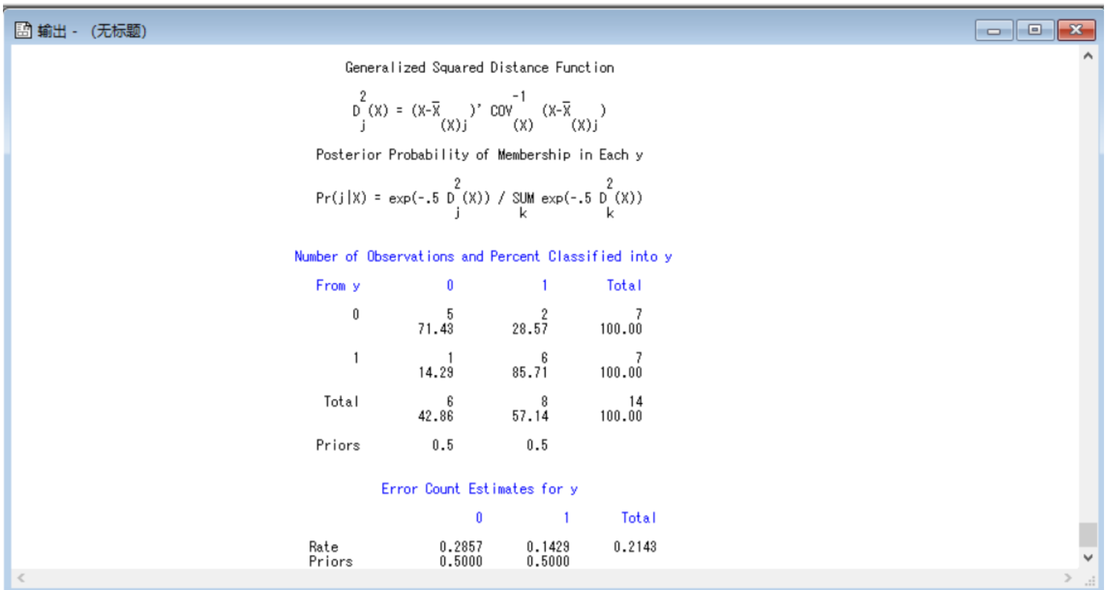
Number of Observations and Percent Classified into y

From y	0	1	Total
0	4 57.14	3 42.86	7 100.00
1	4 57.14	3 42.86	7 100.00
Total	8 57.14	6 42.86	14 100.00
Priors	0.5	0.5	

Error Count Estimates for y

	0	1	Total
Rate	0.4286	0.5714	0.5000
Priors	0.5000	0.5000	

Logistic regression model results:



输出 - (无标题)

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} \, (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each y

$$Pr(j|X) = \exp(-.5\, D_j^2(X)) \,/\, SUM_k \exp(-.5\, D_k^2(X))$$

Number of Observations and Percent Classified into y

From y	0	1	Total
0	5 71.43	2 28.57	7 100.00
1	1 14.29	6 85.71	7 100.00
Total	6 42.86	8 57.14	14 100.00
Priors	0.5	0.5	

Error Count Estimates for y

	0	1	Total
Rate	0.2857	0.1429	0.2143
Priors	0.5000	0.5000	