Analysis

OSL

First, let's look at our benchmark - OSL regression. We use full set of data that is mean centering and standardizing to fit the OSL model. OSL model will be served as our benchmark for comparison with later four different methods. So we calculate OSL model's MSE, which is equal to 0.0447862

Here is more information about OSL regression:

Table 1: Summary Table of OSL Regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0000	0.0107	0.00	1.0000
Income	-0.5982	0.0180	-33.31	0.0000
Limit	0.9584	0.1646	5.82	0.0000
Rating	0.3825	0.1652	2.32	0.0211
Cards	0.0529	0.0129	4.08	0.0001
Age	-0.0230	0.0110	-2.09	0.0374
Education	-0.0075	0.0109	-0.69	0.4921
GenderFemale	-0.0116	0.0108	-1.07	0.2832
StudentYes	0.2782	0.0109	25.46	0.0000
MarriedYes	-0.0091	0.0110	-0.82	0.4107
EthnicityAsian	0.0160	0.0134	1.19	0.2347
EthnicityCaucasian	0.0110	0.0133	0.83	0.4083

Table 1 exhibits estimates of all coefficients and their corresponding t-test and p-value.

From above information, we can conclude that predictors such as *Income*, *Limit*, *Rating*, *Cards*, and *Student* are extremely significant.

Ridge Regression

Next, let's look at ridge regression. We load mean centerized and standardized data before the analysis. Here are the steps: 1. Check mission value in data for both train and test set 2. Initialize lambda and use random seeds (set.seed()) for cross-validation 3. Use train set to conduct 10-fold cross-validation to find out the best tuning parameter

Let's look at plot of the cross-validation errors in terms of the tunning parameter

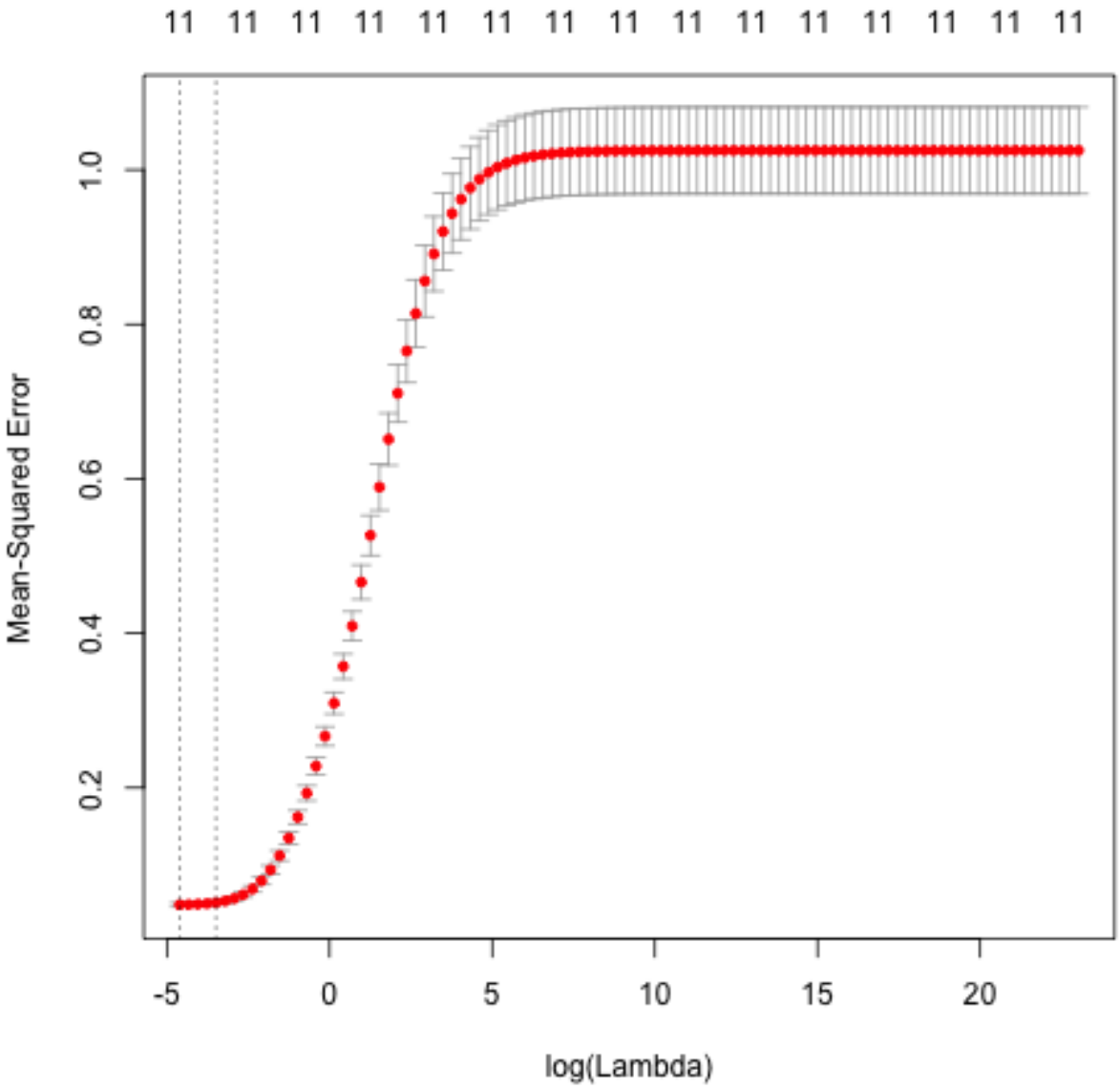

Figure 1: The cross-validation errors in terms of the tunning parameter

4. Find the λ for the best model: best $\lambda = 0.01$
5. Use the test set to compute the test Mean Square Error: test MSE $= 0.0454419$
6. Refit the model on the full data set using the best lambda and get official coefficients.

Here is the official coefficient table:

```
## Loading required package: Matrix
```

Table 2: Coefficient Table of Ridge Regression

	s0
(Intercept)	0.00
Income	-0.57
Limit	0.72
Rating	0.59
Cards	0.04
Age	-0.03
Education	-0.01
GenderFemale	-0.01
StudentYes	0.27
MarriedYes	-0.01
EthnicityAsian	0.02
EthnicityCaucasian	0.01

Lasso Regression

Next, let's look at lasso regression. We load mean centerized and standardized data before the analysis. Here are the steps: 1. Check mission value in data for both train and test set 2. Initialize lambda and use random seeds (set.seed()) for cross-validation 3. Use train set to conduct 10-fold cross-validation to find out the best tuning parameter

Let's look at plot of the cross-validation errors in terms of the tunning parameter

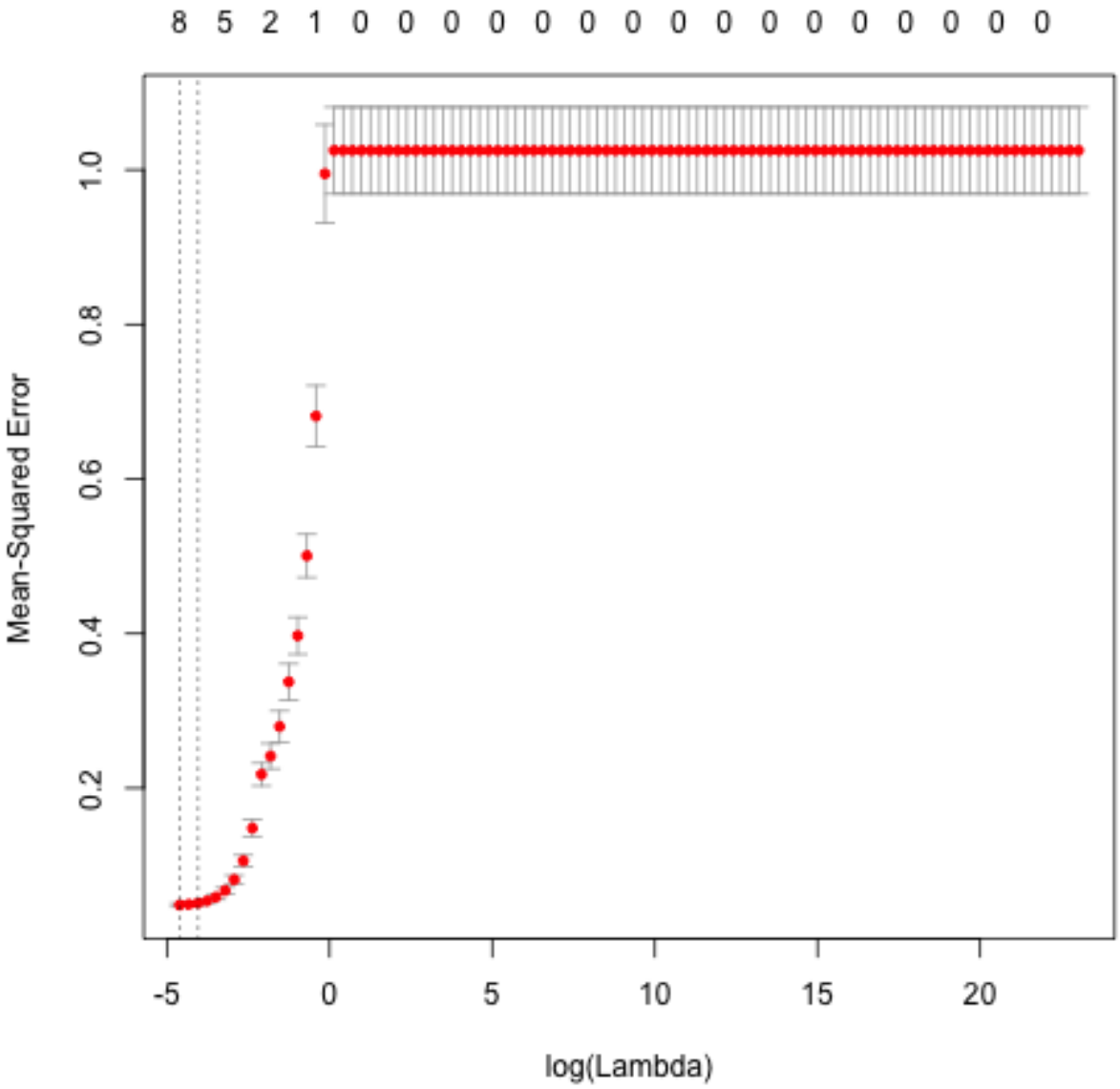

Figure 2: The cross-validation errors in terms of the tunning parameter

4. Find the λ for the best model: best $\lambda = 0.01$
5. Use the test set to compute the test Mean Square Error: test MSE $= 0.0467553$
6. Refit the model on the full data set using the best lambda and get official coefficients.

Here is the official coefficient table:

Table 3: Coefficient Table of Lasso Regression

	s0
(Intercept)	0.00
Income	-0.55
Limit	0.93
Rating	0.37
Cards	0.04
Age	-0.02
Education	0.00
GenderFemale	0.00
StudentYes	0.27
MarriedYes	0.00
EthnicityAsian	0.00
EthnicityCaucasian	0.00

Principal Component Regression (PCR)

Next, let's look at PCR. We load mean centerized and standardized data before the analysis. Here are the steps: 1. Check mission value in data for both train and test set 2. Use random seeds (set.seed()) for cross-validation 3. Use train set to conduct 10-fold cross-validation to find out the best number of principal components used

Let's look at plot of the cross-validation errors in terms of the number of principal components used

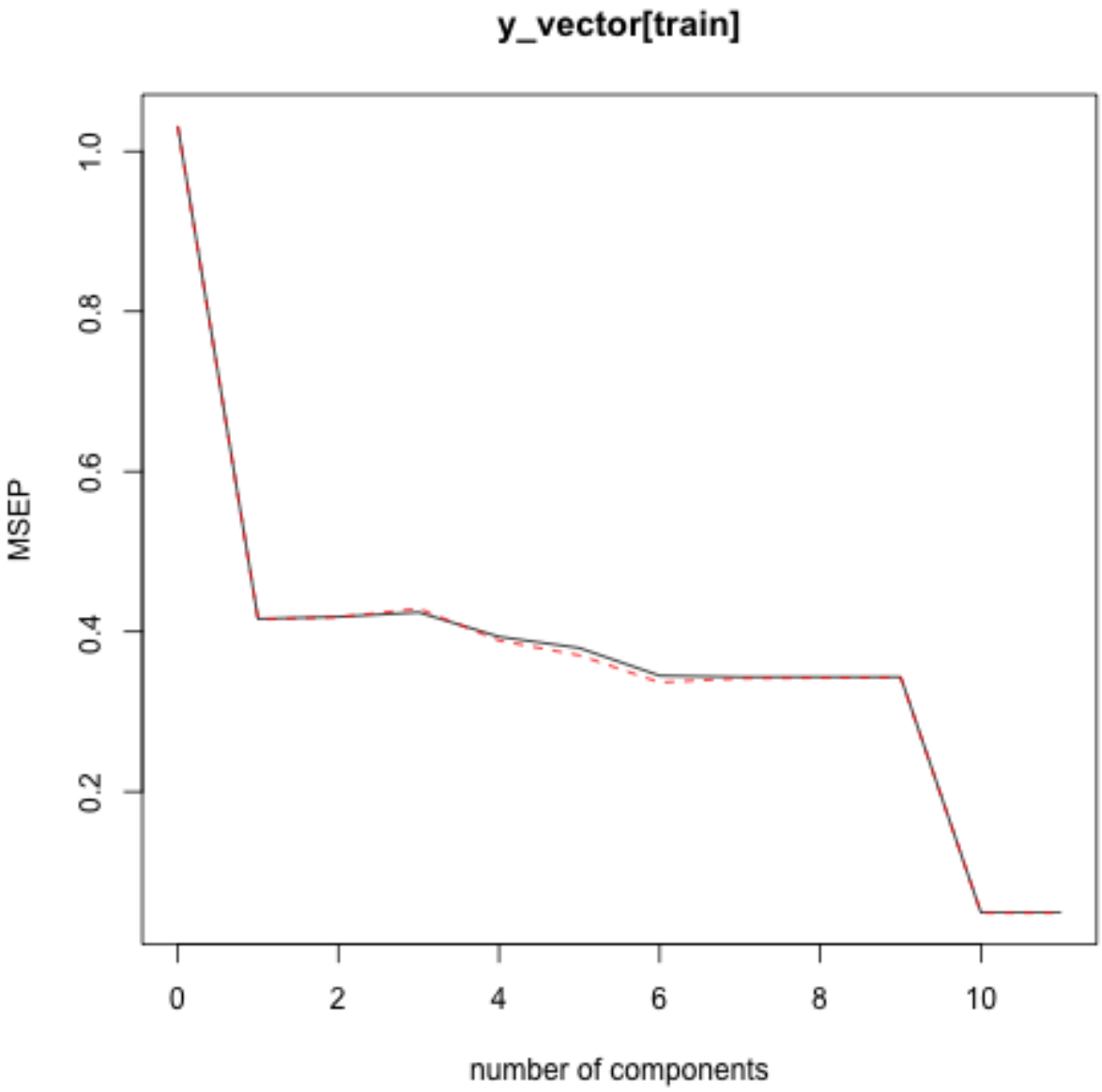

Figure 3: The cross-validation errors in terms of the number of principal components used

4. Find the number of principal components considered for the best model: M = 10
5. Use the test set to compute the test Mean Square Error: test MSE = 0.0460076
6. Refit the model on the full data set using M = 10and get official coefficients.

Here is the official coefficient table:

Table 4: Coefficient Table of PCR

	y_vector.10 comps
Income	-0.60
Limit	0.67
Rating	0.67
Cards	0.04
Age	-0.02
Education	-0.01
GenderFemale	-0.01
StudentYes	0.28
MarriedYes	-0.01
EthnicityAsian	0.02
EthnicityCaucasian	0.01

Partial Least Squares Regression (PLSR)

Next, let's look at PLSR. We load mean centerized and standardized data before the analysis. Here are the steps: 1. Check mission value in data for both train and test set 2. Use random seeds (set.seed()) for cross-validation 3. Use train set to conduct 10-fold cross-validation to find out the best number of partial least squares directions

Let's look at plot of the cross-validation errors in terms of the number of partial least squares directions

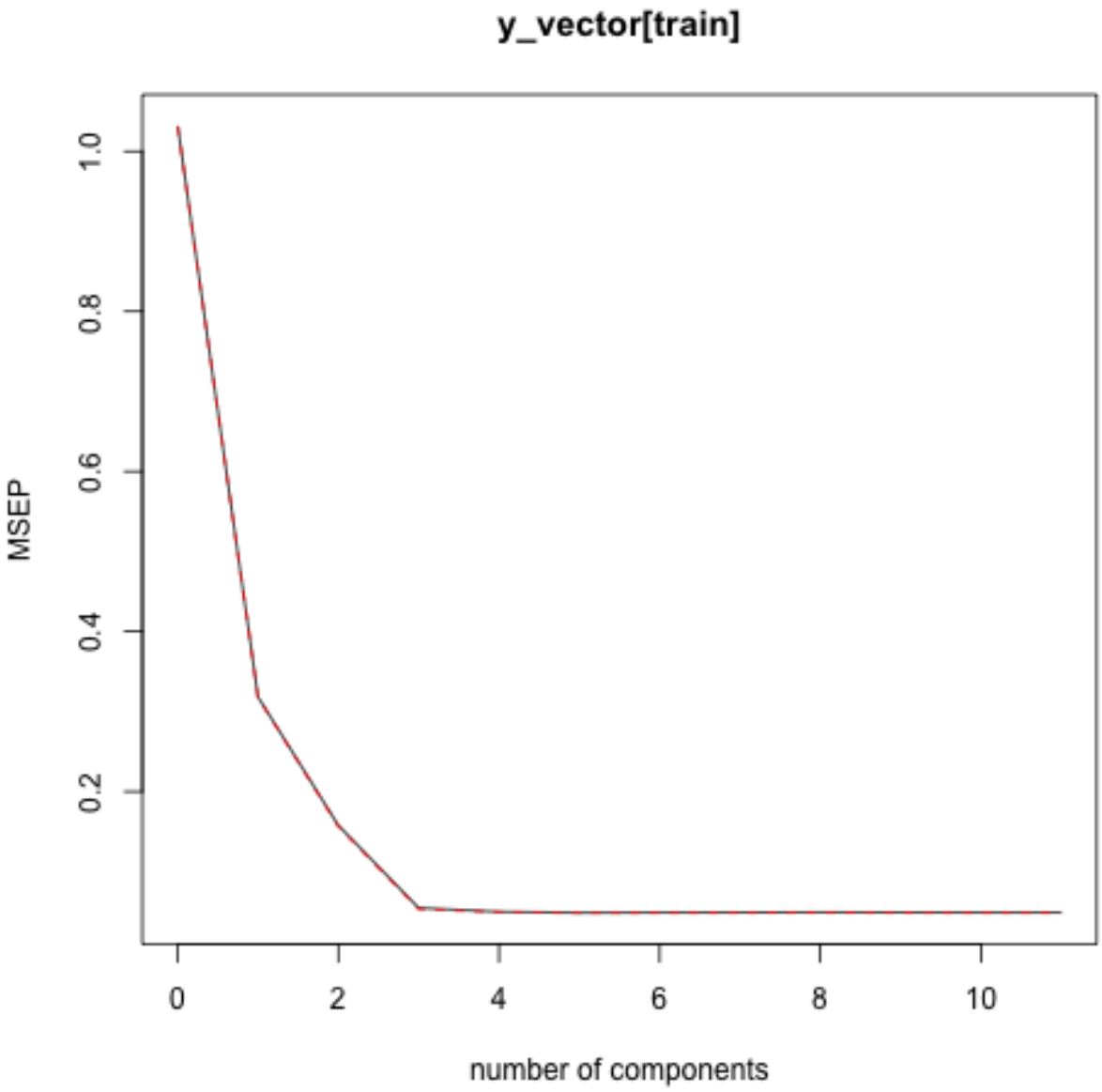

Figure 4: The cross-validation errors in terms of the number of partial least squares directions

4. Find the number of partial least squares directions for the best model: M = 5
5. Use the test set to compute the test Mean Square Error: test MSE = 0.0465753
6. Refit the model on the full data set using M = 5and get official coefficients.

Here is the official coefficient table:

Table 5: Coefficient Table of PLSR

	y_vector.5 comps
Income	-0.60
Limit	0.68
Rating	0.67
Cards	0.04
Age	-0.02
Education	-0.01
GenderFemale	-0.01
StudentYes	0.28
MarriedYes	-0.01
EthnicityAsian	0.02
EthnicityCaucasian	0.01