Home Assignment Tao Xiaoying

Task 1:

The minimized expressions (as function of λ):

LASSO:

$\min_{\beta} \Sigma(y_i - \beta_i)^2$ s.t. $\sum|\beta_i| \le s$

function of λ: $F(\beta_i, \lambda) = \sum(y_i - \beta_i)^2 + \lambda(|\beta_i| - s)$

Ridge:

$\min_{\beta} \Sigma(y_i - \beta_i)^2$ s.t. $\sum {\beta_i}^2 \le s$

function of λ: $F(\beta_i, \lambda) = \sum(y_i - \beta_i)^2 + \lambda({\beta_i}^2 - s^2)$

Solution and how the estimated $\boldsymbol{\beta_i}$-s depend on $\boldsymbol{\lambda}$:

LASSO:

$$\widehat{\beta_\iota} = \arg\min_{\beta} \sum (y_i - \beta_i)^2 + \lambda(|\beta_i| - s)$$

$$= \max_{\beta}(|y_i| - \lambda, 0) \sin y_i$$

$$\widehat{\beta_\iota} = \begin{cases} y_i - \lambda\,,\ y_i > 0 \\ \ 0\,,\quad y_i = 0 \\ y_i + \lambda\,,\ y_i < 0 \end{cases}$$

$$\lambda \to 0 : OLS$$

$\lambda \to \infty :\ y_i = \overline{y_\iota}$, all coefficients to zero

Ridge:

$$\widehat{\beta_\iota} = \frac{1}{1+\lambda} y_i$$

$$\lambda \to 0 : \widehat{\beta_\iota} = y_i$$

$$\lambda \to \infty :\ \widehat{\beta_\iota} = 0$$

Difference:

In Ridge regression, the penalty term sets coefficients close to zero($\widehat{\beta_\iota} \to 0$) ; in LASSO regression, the penalty term sets coefficients exactly to zero($\widehat{\beta_\iota} = 0$).

Note: $\widehat{\beta_\iota} = \beta_i\ hat$

Task 2:

We assume there are n times of bootstrapping (resampling).

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n=\frac{1}{e}$$

Task 3:

Part a:

wage - regression

	Model Comparison					
	wage					
	(1)	(2)	(3)	(4)	(5)	(6)
abil	0.46***	0.45***	0.43***	0.42***	0.42***	0.42***
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
educ	1.48***	0.83	1.15	1.57***	1.05	1.11
	(0.18)	(1.01)	(1.00)	(0.18)	(1.01)	(1.01)
educsq		0.03	0.02		0.02	0.02
		(0.04)	(0.04)		(0.04)	(0.04)
exper	0.64***	1.80***	1.72***	1.62***	1.74***	1.77***
	(0.12)	(0.53)	(0.53)	(0.47)	(0.53)	(0.53)
expersq		-0.05**	-0.04*	-0.04*	-0.04*	-0.05*
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
ne					1.62**	0.37
					(0.82)	(1.95)
nc					0.05	-0.83
					(0.71)	(1.45)
west					0.59	-3.22
					(0.86)	(2.63)
south						
ne:urban						1.61
						(2.14)
urban:nc						1.19
						(1.65)
urban:west						4.34
						(2.78)
urban:south						
urban			3.39***	3.41***	3.24***	2.13*
			(0.73)	(0.72)	(0.73)	(1.24)
Constant	-14.21***	-16.04**	-21.43***	-23.66***	-21.16***	-20.89***
	(3.21)	(6.63)	(6.65)	(4.14)	(6.64)	(6.67)
Observations	788	788	788	788	788	788
R^2	0.14	0.15	0.17	0.17	0.18	0.18
Note:				*p<0.1; **p<0.05; ***p<0.01		

```
  rmse_test rmse_train model_name model_pretty_name nvars        r2      BIC
1  8.302615   7.706826  modellev1               (1)     3 0.1425066 5487.954
2  8.246951   7.681825  modellev2               (2)     5 0.1480611 5496.172
3  8.209718   7.576952  modellev3               (3)     6 0.1711639 5481.178
4  8.229096   7.577844  modellev4               (4)     5 0.1709686 5474.694
5  8.205926   7.551626  modellev5               (5)     9 0.1766954 5495.910
6  8.228620   7.539081  modellev6               (6)    12 0.1794286 5513.298
```

Model 4 is the best, because BIC is smallest, variables are not too much, and RMSE of test sample is relatively small.

Part b:

Models are the same as part a, except the dependent variable.

```
  rmse_test rmse_train model_name model_pretty_name nvars        r2      BIC
1 0.5243854  0.5308307  modellev1               (1)     3 0.1749542 1271.495
2 0.5259355  0.5278201  modellev2               (2)     5 0.1842864 1275.870
3 0.5225591  0.5187933  modellev3               (3)     6 0.2119485 1255.353
4 0.5187993  0.5207489  modellev4               (4)     5 0.2059961 1254.614
5 0.5194135  0.5166744  modellev5               (5)     9 0.2183727 1268.912
6 0.5219456  0.5145111  modellev6               (6)    12 0.2249041 1282.308
```

Model 4 is the best, because BIC is smallest, and RMSE of test sample is the smallest, and variables are not too much.

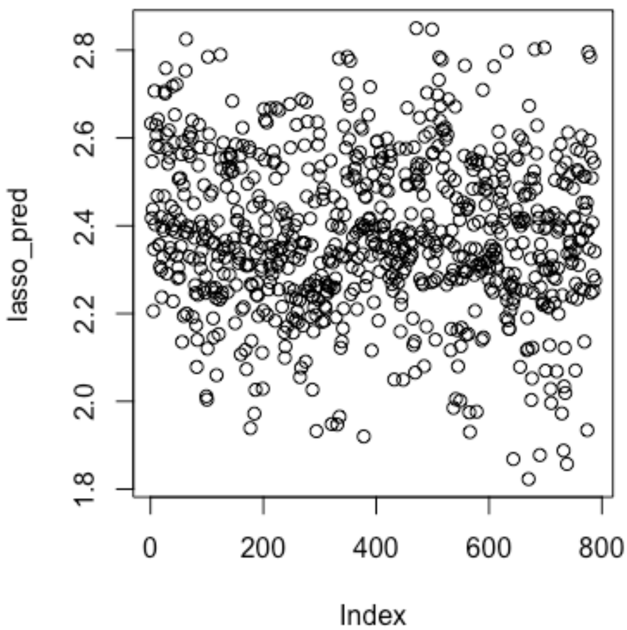
RMSE and Bias without correction: 14.54694, 11.31387
RMSE and Bias with correction (adjustment: 0.5*sigma(mb)^2): 8.172926, 0.9456576

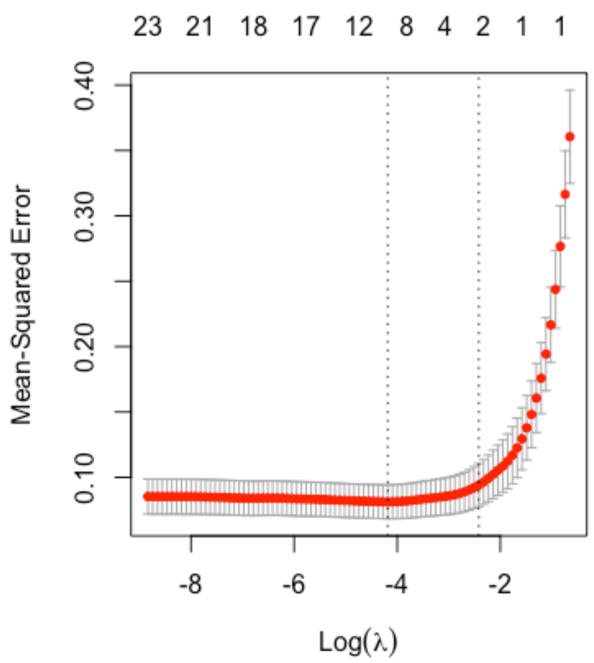
Part c:

Model	RMSE	Bias
Best of Part a	9.775889	0.9723511
Best of Part b	8.172926	0.9456576

The model from Part b(lwage) is more accurate.

Part d:





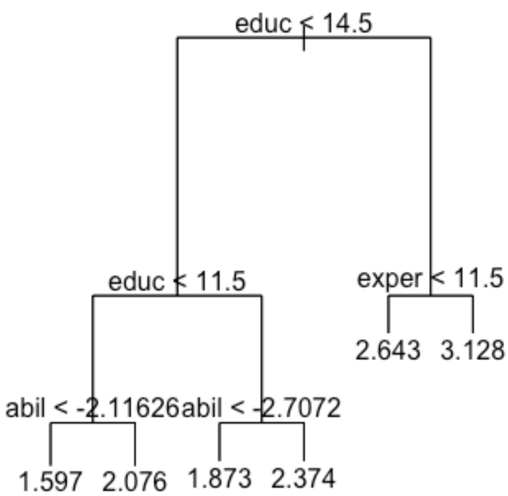
Part e:

Best LASSO:

$$lwage = 2.38 + \ 0.07abil \ + \ 0.11educ - 0.0\ brkhme14 \ + 0.03urban + \ 0.01ne18$$

RMSE: 0.5388025

Part f: Baseline: lwage ~ .-wage (include all variables without wage)



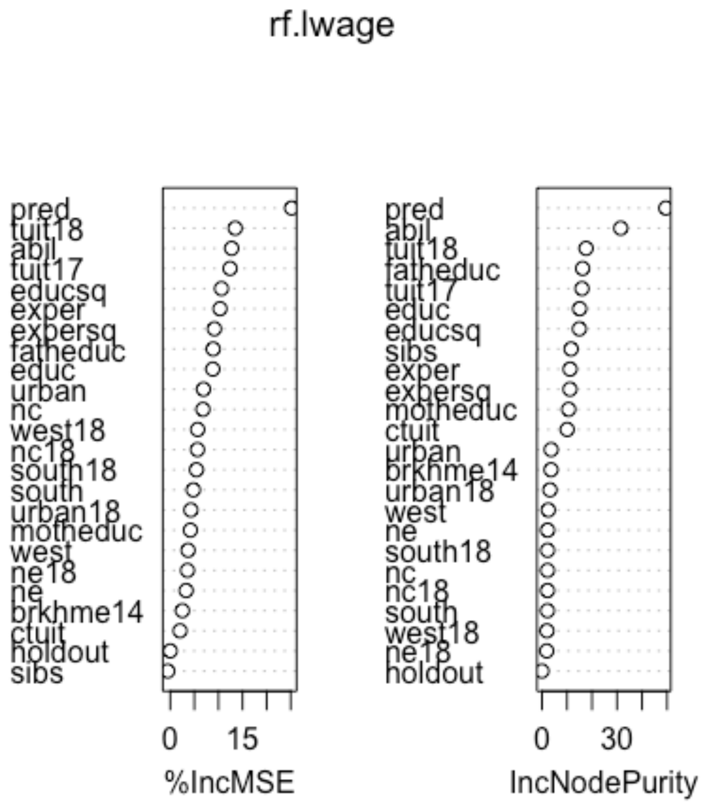
Interpretation:
When a man's education is below 11.5 years and ability is below -2.11626, his log of wage is likely between 1.597 and 2.076. If same above, except the ability is below -2.7072 but above -2.11626, his log of wage is likely between 1.873 and 2.374.
When a man's education is below 14.5 and experience is below 11.5, his log of wage is likely between 2.643 and 3.128.

Part g:

Baseline: lwage ~ .-wage

Forest	RMSE
1	0.512113
2	0.4563601
3	0.4645163
4	0.4492797
5	0.4587447
6	0.4376896

Forest 6 is the best.



Part h:

Baseline: lwage ~ .-wage

Tree	RMSE
1	0.5524145
2	0.509748
3	0.6749561
4	0.7073127
5	0.6039382

Tree 2 is the best.

Part i:

Model	RMSE
b	0.6779442
e	0.6740205
f	0.6822352
g	0.6747422
h	0.7188055

Model e (LASSO) is the best.