MSDS 6372: Unit 2 HW 2

One of the major advantages of LASSO regression is that the estimate of the regression coefficients (betas) are allowed to be biased whereas the OLS estimates are forced to be unbiased. This highlights the commonly referred to "variance / bias trade-off". Since MSE = Variance + Bias^2, it is easy to see that for the OLS unbiased estimates that the MSE(betas) = Variance(betas) since the Bias is zero. However, with the biased LASSO estimates, one is often able to reduce the variance of the estimate of the betasat the cost of introducing a little bias. Often the reduction in the variance is greater than the increase in the squared bias and we see a reduction in the MSE of the betas.

In short, this means that our estimate of the regression equation with the smaller MSE has greater probability of being close to the equation with the real betas: the real trend. This means that if we cross validate our model on a test set (that maintains the same trend as the training set but with different noise), the model with the smaller MSE has greater probability of capturing more of the true trend. Statistically, this will be reflected in statistics such as the ASE (Test) (Average Squared Error for the model trained on the training set and used to fit the test set) and R squared of the test set (basically any goodness of fit statistic that is **with respect to the model and the test set**.)

We will use the ASE (Test) statistic to provide evidence of preferred models for this data set: LASSO or OLS.

Attached HW 2 SAS Code has the data set from the paper you read as well as some code to divide the data set into a training and test set.

The assignment for this week is simple:

1. Run the code and make an argument / discussion as to which model / estimates (LASSO or OLS) will provide better predictability.
   Make sure and copy and paste all relevant output to support your decision. Don't overthink this. The answer can reference a single statistic.

LASSO

Root MSE	36.61967
Dependent Mean	945.10526
R-Square	0.7792
Adj R-Sq	0.7525
AIC	318.28350
AICC	320.99318
SBC	286.47143
ASE (Train)	1164.55253
ASE (Test)	651.75350
CV PRESS	28898

OSL

Root MSE	24.36490
Dependent Mean	945.10526
R-Square	0.9023
Adj R-Sq	0.8904
AIC	287.31794
AICC	290.02762
SBC	255.50587
ASE (Train)	515.53690
ASE (Test)	251.28571
CV PRESS	28898

Going by the R-square and Adj R-sq the OLS will be better for predictability.

2. Provide confidence intervals for all estimates from both the LASSO and OLS models. Also report the margin of error for each interval and comment on which margins of error seem to be smaller. This should be fun! It uses a cutting age function and method that is related to machine

learning topics (bootstrapping) called model averaging. Read the SAS documentation about it … it is a very strait forward yet powerful idea.

```
ods output FitStatistics = t0;
proc glmselect data=train testdata = test
          seed=1 plots(stepAxis=number)=(criterionPanel ASEPlot CRITERIONPANEL);
model SAT = lTakers     Income     Years     Public     Expend     Rank / selection=LASSO(
choose=CV stop=CV) CVdetails  ;
run;
quit;
ods graphics off;



ods output FitStatistics = t0;
proc glmselect data=train testdata=test
          seed=1 plots(stepAxis=number)=(criterionPanel ASEPlot CRITERIONPANEL);
model SAT = lTakers Years  Expend   Rank   / selection=stepwise( choose=CV stop=CV include =
4) CVdetails  ;
run;
quit;
ods graphics off;

*store the estimated r-square;
data _null_;
 set t0;
 if Label1 = "R-Square" then
 call symput('r2bar', cValue1);
run;


%let rep = 500;
proc surveyselect data= train out=bootsample
    seed = 1347 method = urs
          samprate = 1 outhits rep = &rep;
run;
ods listing close;


ods output  FitStatistics = t (where = (Label2 =  "R-Square"));
proc reg data = bootsample;
by replicate;
model SAT = lTakers     Income     Years     Public     Expend     Rank;
run;
```

```
quit;

  * converting character type to numeric type;
data t1;
set t;
r2 = cValue2 + 0;
run;


* creating confidence interval, normal distribution theory method;
* using the t-distribution;
%let alphalev = .05;
ods listing;
proc sql;
 select  &r2bar as r2,
      mean(r2) - &r2bar as bias,
                  std(r2) as std_err,
      &r2bar - tinv(1-&alphalev/2, &rep-1)*std(r2) as lb,
      &r2bar + tinv(1-&alphalev/2, &rep-1)*std(r2) as hb
 from t1;
quit;
```

LASSO

r2	bias	std_err	lb	hb
0.7792	0.141223	0.024737	0.730598	0.827802

OSL

r2	bias	std_err	lb	hb
0.9023	0.018123	0.024737	0.853698	0.950902

3. **Chapter 12.10**

Model Variables	n	p Regression Coefficients	RSS	DF	σ2 (RSS/DF	R2	Adj R2	Cp	BIC
None	28	1	8100	27	300	0	0	1	157.298
A		2	6240	26	240	0.22963	0.1992	-7	153.5499
B		2	5980	26	230	0.261728	0.23248	-8.3056	152.40083
C		2	6760	26	260	0.165432	0.1316	-4.222	155.71109
AB		3	5500	25	220	0.320988	0.26442	-8.093	153.4375
AC		3	5250	25	210	0.351852	0.298	-9.24	152.18147
BC		3	5750	25	230	0.290123	0.2308	-6.893	154.6377
ABC		4	5160	24	215	0.362963	0.2799	-7.18694	155.01043

The smallest estimate of pure error variance is Model AC with 210

The largest adjusted R is Model ABC with .2799

The Model with smallest CP is B with -8.3056 and the smallest BIC is AC with 152.18147

12.11

Best 1-predictor is Model B

F = ((8100 - 6240)/1)/ (6240/26) = 7.75

which has p-value = 0.009879 (using 1-pf(7.75,1,26) in R).

The best 2-predictor model with B in it, which is AB:

F = ((6240 -5500)/1) / (5500/25) = 3.364

which has p-value = 0.0786 (using 1-pf(3.364,1,25) in R).

We then stop and say the AB model is not better than just using B, so the model with B is our best predictive model through forward model selection (using the ESS F-statistic's p-value as our criterion).