MA 543/DS 502 – HW 3 responses

Problem 1)

a)

Of the 3 selection models, it would be Best Subset selection with k predictors that would give the smallest *training* RSS; Best Subset selection looks at every possible selection model (all 2^p models) and chooses the very best one (ie smallest RSS). Both Forward and Backward stepwise selection models depend on the predictors chosen first, basing the chosen selection off of a heuristic path of chosen predictors, leading to potentially worse RSS. As the name would indicate, Best Subset selection gives the best (ie, smallest) training RSS.

b)

It's difficult to determine which selection model would have the best *test* RSS. There is a possibility of overfitting for each model, however, there are probably cases where either the Forward or Backward stepwise selection model would give a lower test RSS than a best subset selection model.

c)

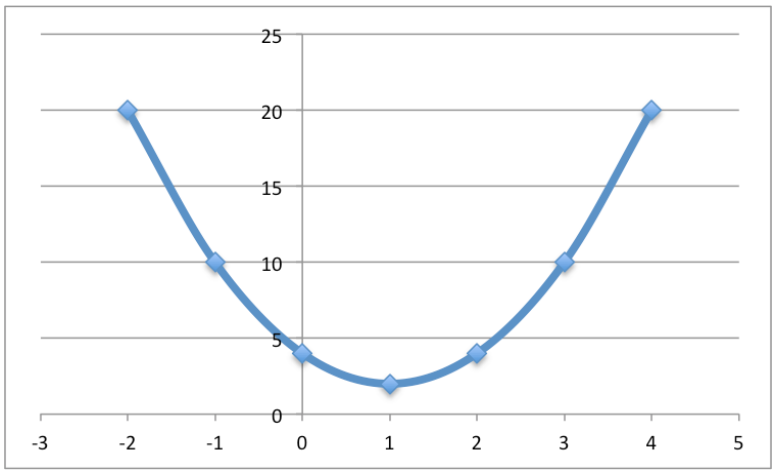
i) TRUE
ii) TRUE
iii) FALSE
iv) FALSE
v) FALSE

Problem 2)

a)

$y_1 = 2$
$\lambda = 1$

According to (6.14), this plot should be minimized at the β value of 2/(1+1) = 1. As shown below, this is in fact the case.



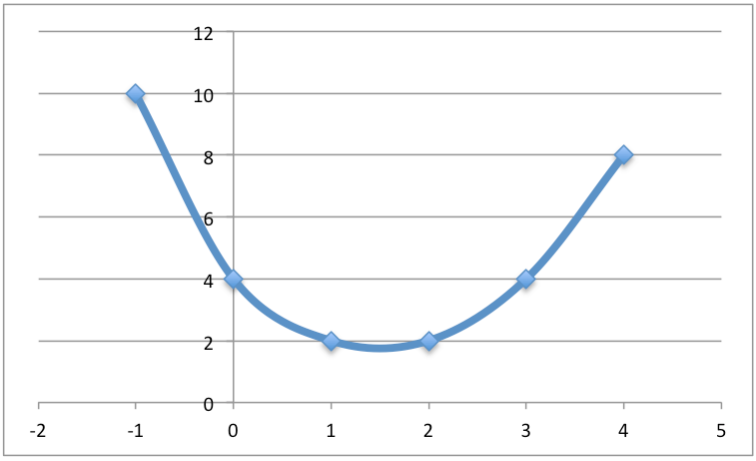
b)

Case 1:

$y_1 = 2$

$\lambda = 1$

Since 2 > .5 in this example, according to (6.15), this plot should be minimized at the β value of 2 - .5 = 1.5. As shown below, this is in fact the case.

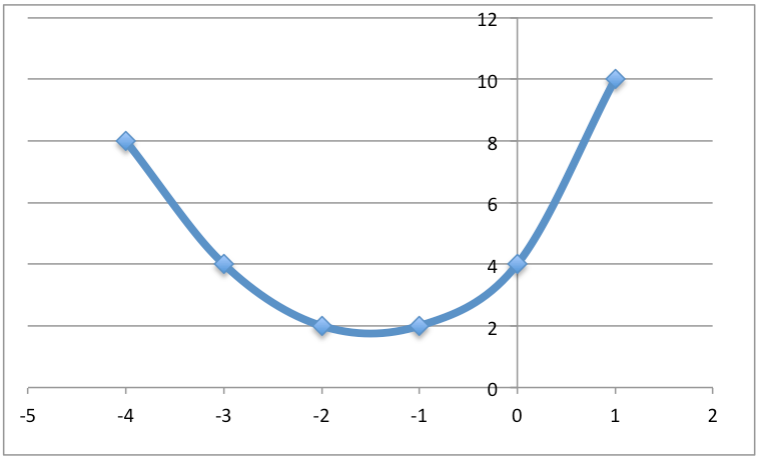


Case 2:

$y_1 = -2$

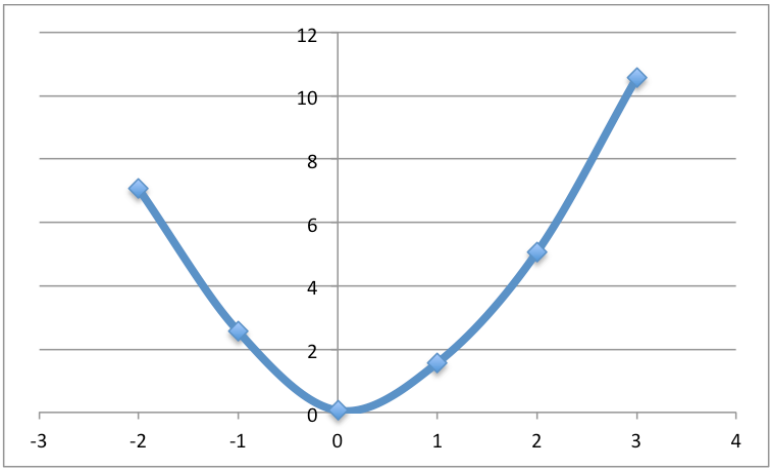
$\lambda = 1$

Since -2 < -.5 in this example, according to (6.15), this plot should be minimized at the β value of -2 + .5 = -1.5. As shown below, this is in fact the case.



Case 3:
$y_1 = .25$
$\lambda = 1$

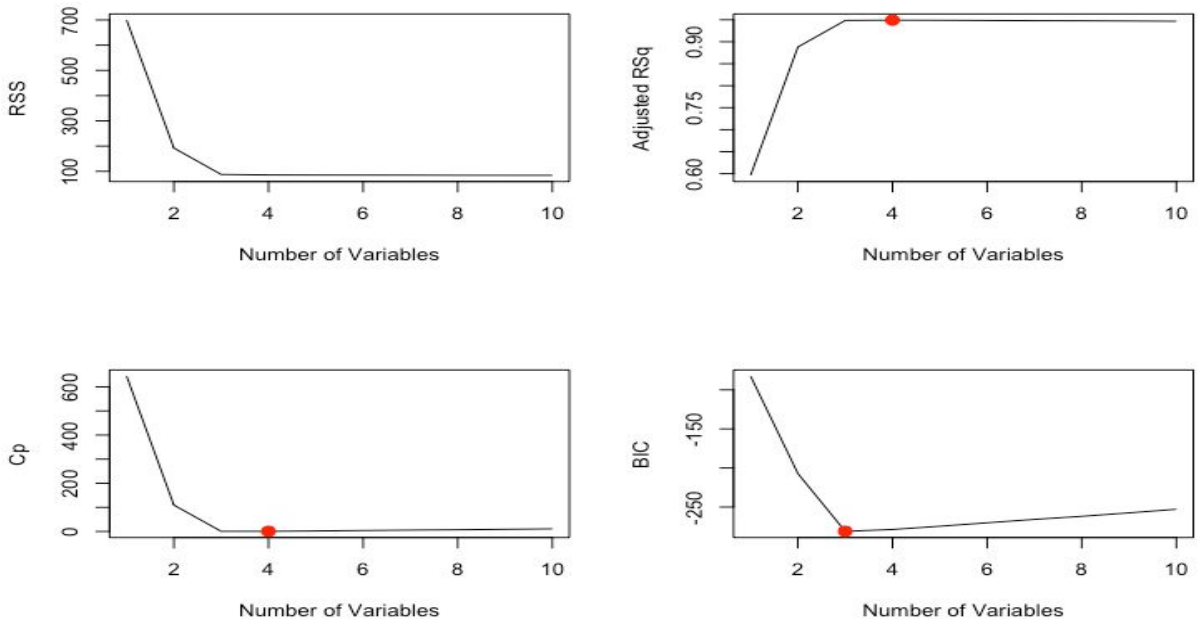
Since .25 < .5 in this example, according to (6.15), this plot should be minimized at the β value of 0. As shown below, this is in fact the case.



Problem 3)

b)
Our chosen constants are:
$\beta_0 = 3$
$\beta_1 = 2$
$\beta_2 = -2$
$\beta_3 = 0.8$

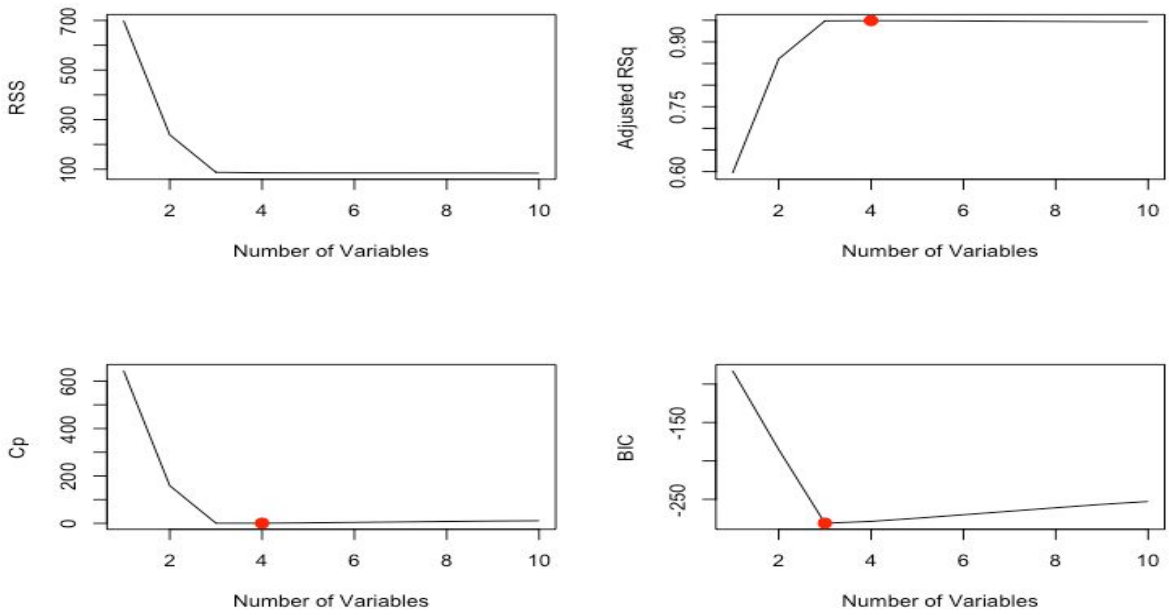
c) Best Subset:



From using best subset selection method on 10 predictor variables, and based on the results of AdjR2, Cp, and BIC:
AdjR2 and Cp both say that the best model is a 4-variable model, whereas BIC gets the correct best model with 3-variables:

$$Y = 3.0731622 + (2.7436826)*X^1 + (-2.1697963)*X^2 + (0.1394003)*X^5$$

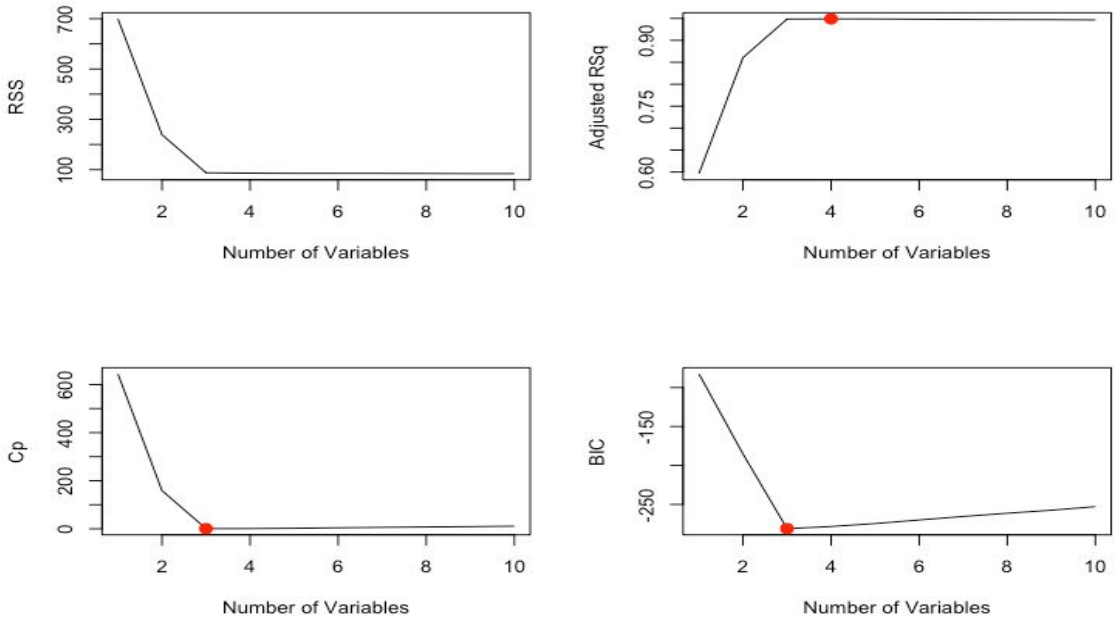
d) Forward Stepwise selection:



Best model is 3-variable with the following coefficients:

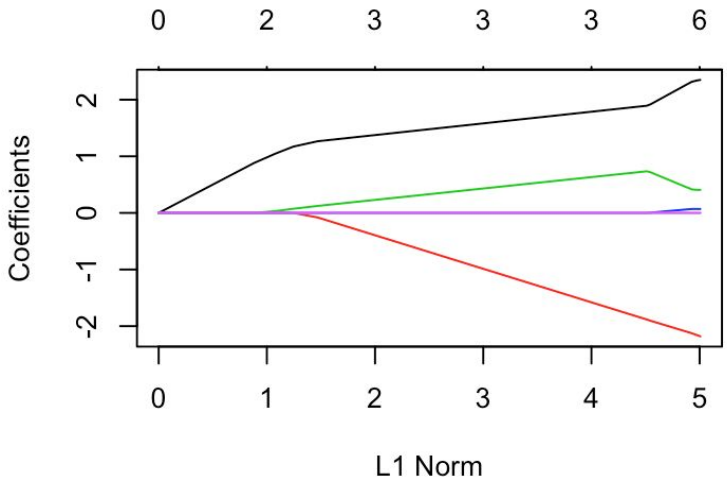
$Y = 3.0731622 + (2.7436826)*X^1 + (-2.1697963)*X^2 + (0.1394003)*X^5$

Backward Stepwise selection:



Best model is still 3-variable, with same coefficients as before.
The most accurate is the Best Subset selection method, as the other stepwise selection can potentially include the insignificant predictor variables.

e) Lasso



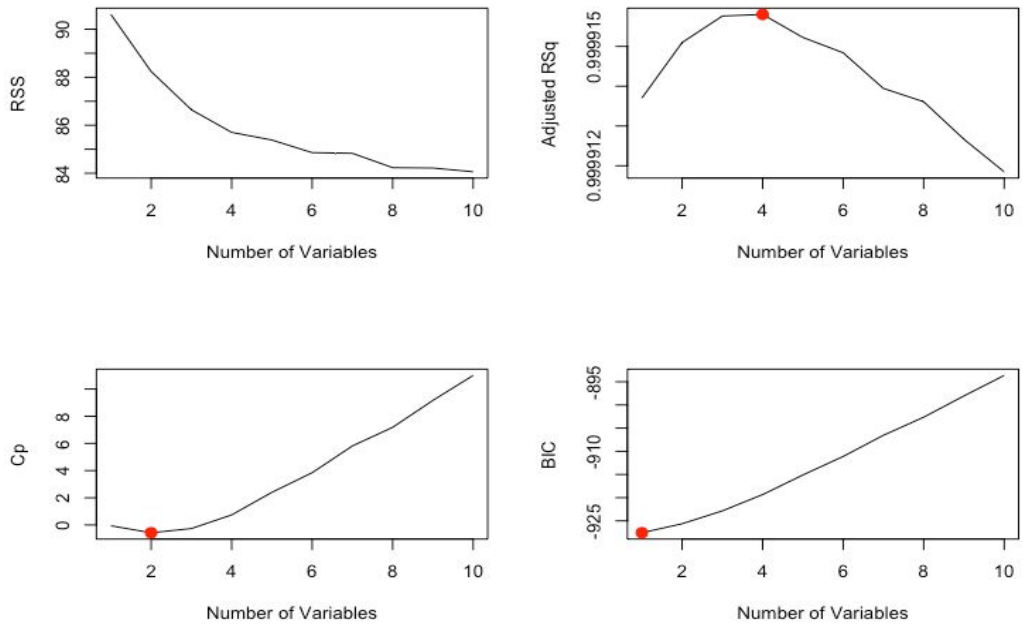
Based on the cross-validation technique, the best λ is: 0.008408922
And thus, Lasso regression says that the following model is the best, with respective coefficients:

$Y = 3.082721 + (2.349054)*X^1 + (-2.180709)*X^2 + (0.4071389)*X^3 + (6.923711e\text{-}02)*X^5 + (1.212715e\text{-}04)*X^8 + (7.137022e\text{-}05)*X^{10}$

While although Lasso correctly includes the necessary predictor variables (X^1,X^2,X^3), it also includes the unnecessary predictors (X^5, X^8, X^{10}).

f)
New model: $Y = \beta_0 + \beta_7 X^7 + \varepsilon$
Same β_0.
We chose $\beta_7 = -1.7$

Best Subset selection:



Comparing between AdjR2, Cp, and BIC, it is only BIC that shows the correct best model, with the following coefficient:

$$Y = 2.95894 + (-1.69923)*X^7$$

Lasso:
Best $\lambda = 3.002169$
Thus Lasso regression says the following model is the best:

$$Y = 2.749891 + (-1.649696)*X^7$$

Fortunately, both best subset selection and lasso methods correctly say that a 1-variable model is the best model for our case.
One can assume best subset selection has the best accuracy thanks to -1.69923 being approximately equal to $\beta_7 = -1.7$.

Problem 4)

b) Linear regression fit:

```
Call:
lm(formula = Apps ~ ., data = College.train)

Residuals:
    Min      1Q  Median      3Q     Max
-2485.0  -392.4   -57.9   261.6  6637.3

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -756.88732  535.44217  -1.414 0.158301
PrivateYes  -515.36459  190.44202  -2.706 0.007112 **
Accept         1.20916    0.07046  17.161  < 2e-16 ***
Enroll        -0.29026    0.23380  -1.241 0.215188
Top10perc     52.53768    7.69358   6.829 3.38e-11 ***
Top25perc    -16.38973    6.09706  -2.688 0.007500 **
F.Undergrad    0.09625    0.04103   2.346 0.019495 *
P.Undergrad    0.02831    0.05417   0.523 0.601540
Outstate      -0.02161    0.02599  -0.831 0.406262
Room.Board     0.13204    0.06968   1.895 0.058847 .
Books          0.16642    0.34092   0.488 0.625726
Personal       0.16409    0.09270   1.770 0.077519 .
PhD           -8.51810    6.73921  -1.264 0.207015
Terminal      -0.35379    7.43056  -0.048 0.962049
S.F.Ratio      3.79436   16.20785   0.234 0.815028
perc.alumni   -2.86286    5.56666  -0.514 0.607349
Expend         0.05855    0.01528   3.832 0.000148 ***
Grad.Rate      7.14929    3.91090   1.828 0.068323 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1001 on 382 degrees of freedom
Multiple R-squared:  0.9314,    Adjusted R-squared:  0.9283
F-statistic: 304.9 on 17 and 382 DF,  p-value: < 2.2e-16
```

Linear fit gives:
Test MSE = 1520331
Test R2 = 0.9051744

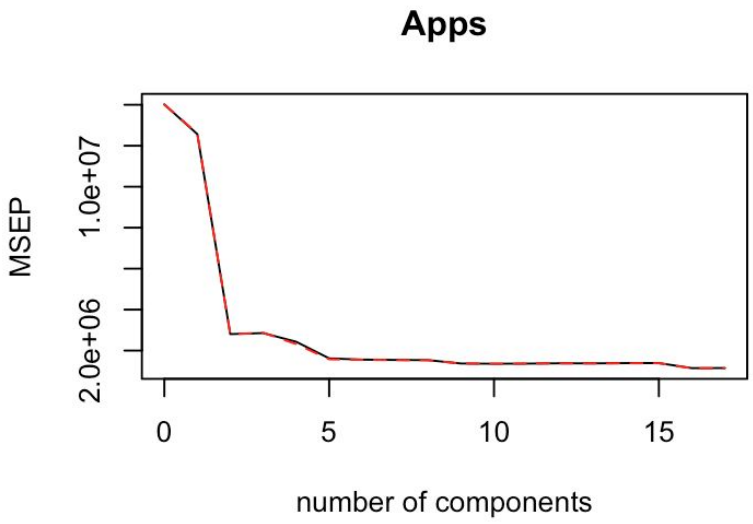
c) Ridge Regression fit:

Ridge regression gives:
Test MSE = 1520405
Test R2 = 0.9051698

d) Lasso:

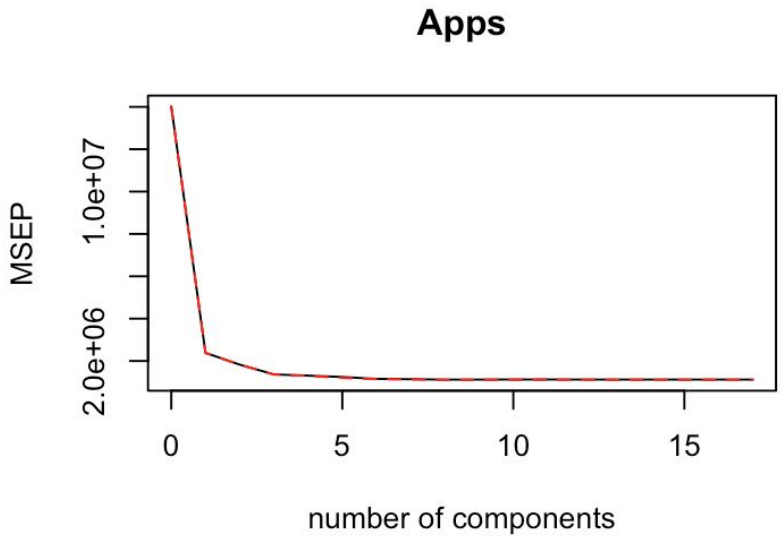
Lasso gives:
Test MSE = 1520433
Test R2 = 0.9051681
And number of non-zero coefficients = 19

e) PCR:



PCR gives:
Chosen M from cross-validation = 17
Test MSE = 1520331
Test R2 = 0.9051744

f) PLS:



PLS gives:
Chosen M from CV = 10
Test MSE = 1538778
Test R2 = 0.9040239

g)
From performing all of the above selection and fit methods, and inspecting the respective performance and test errors, roughly all models are appropriate and accurate for the College Apps dataset.
More specifically, all models give a test R^2 error of about 0.90. Thus, all models have a high accuracy of predicting the College.test dataset.