Yelizaveta Semikina
HW 5
STAT 385

Question 8
A

```
 X <- rnorm(100)
> noise_vect <- rnorm(100)
> x
  [1] -0.626453811  0.183643324 -0.835628612  1.595280802  0.329507772 -0.820468384
0.487429052
  [8]  0.738324705  0.575781352 -0.305388387  1.511781168  0.389843236 -0.621240581
-2.214699887
 [15]  1.124930918 -0.044933609 -0.016190263  0.943836211  0.821221195  0.593901321
0.918977372
 [22]  0.782136301  0.074564983 -1.989351696  0.619825748 -0.056128740 -0.155795507
-1.470752384
 [29] -0.478150055  0.417941560  1.358679552 -0.102787727  0.387671612 -0.053805041
-1.377059557
 [36] -0.414994563 -0.394289954 -0.059313397  1.100025372  0.763175748 -0.164523596
-0.253361680
 [43]  0.696963375  0.556663199 -0.688755695 -0.707495157  0.364581962  0.768532925
-0.112346212
 [50]  0.881107726  0.398105880 -0.612026393  0.341119691 -1.129363096  1.433023702
1.980399899
 [57] -0.367221476 -1.044134626  0.569719627 -0.135054604  2.401617761 -0.039240003
0.689739362
 [64]  0.028002159 -0.743273209  0.188792300 -1.804958629  1.465554862  0.153253338
2.172611670
 [71]  0.475509529 -0.709946431  0.610726353 -0.934097632 -1.253633400  0.291446236
-0.443291873
 [78]  0.001105352  0.074341324 -0.589520946 -0.568668733 -0.135178615  1.178086997
-1.523566800
 [85]  0.593946188  0.332950371  1.063099837 -0.304183924  0.370018810  0.267098791
-0.542520031
 [92]  1.207867806  1.160402616  0.700213650  1.586833455  0.558486426 -1.276592208
-0.573265414
 [99] -1.224612615 -0.473400636
> noise_vect
  [1] -0.62036668  0.04211587 -0.91092165  0.15802877 -0.65458464  1.76728727
0.71670748
  [8]  0.91017423  0.38418536  1.68217608 -0.63573645 -0.46164473  1.43228224
-0.65069635
 [15] -0.20738074 -0.39280793 -0.31999287 -0.27911330  0.49418833 -0.17733048
-0.50595746
 [22]  1.34303883 -0.21457941 -0.17955653 -0.10019074  0.71266631 -0.07356440
-0.03763417
```

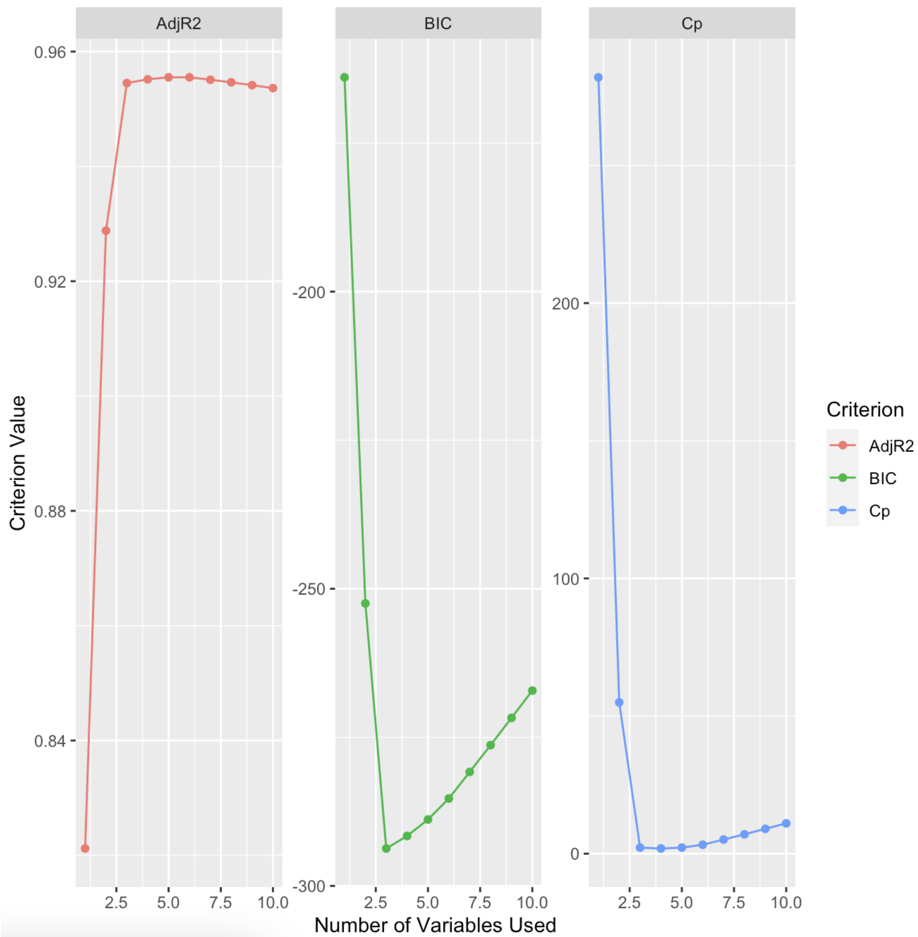
```
 [29] -0.68166048 -0.32427027  0.06016044 -0.58889449  0.53149619 -1.51839408
0.30655786
 [36] -1.53644982 -0.30097613 -0.52827990 -0.65209478 -0.05689678 -1.91435943
1.17658331
 [43] -1.66497244 -0.46353040 -1.11592011 -0.75081900  2.08716655  0.01739562
-1.28630053
 [50] -1.64060553  0.45018710 -0.01855983 -0.31806837 -0.92936215 -1.48746031
-1.07519230
 [57]  1.00002880 -0.62126669 -1.38442685  1.86929062  0.42510038 -0.23864710
1.05848305
 [64]  0.88642265 -0.61924305  2.20610246 -0.25502703 -1.42449465 -0.14439960
0.20753834
 [71]  2.30797840  0.10580237  0.45699881 -0.07715294 -0.33400084 -0.03472603
0.78763961
 [78]  2.07524501  1.02739244  1.20790840 -1.23132342  0.98389557  0.21992480
-1.46725003
 [85]  0.52102274 -0.15875460  1.46458731 -0.76608200 -0.43021175 -0.92610950
-0.17710396
 [92]  0.40201178 -0.73174817  0.83037317 -1.20808279 -1.04798441  1.44115771
-1.01584747
 [99]  0.41197471 -0.38107605
```

B

```
Y <- 3 + 1*X + 4*X^2 - 1*X^3 + noise_vect
> y
  [1] -1.4683186 -0.8976494 -1.6455447 -0.1628524 -0.9988923 -0.9684124 -0.5771086
-0.4032941
  [9] -0.6160630 -0.7321502 -0.4030435 -0.9204896 -0.9525497 -2.2700240 -0.4893797
-1.1206688
 [17] -1.0880933 -0.5978602 -0.4658423 -0.7473820 -0.6670007 -0.2731721 -1.0163624
-2.0395650
 [25] -0.7151348 -0.8498978 -1.0962889 -1.7447847 -1.4094901 -0.8720968 -0.3056201
-1.1986175
 [33] -0.6732901 -1.4065010 -1.6118903 -1.5916097 -1.2723890 -1.1617267 -0.6130110
-0.6326363
 [41] -1.5608517 -0.8325350 -1.0677614 -0.8375510 -1.6233579 -1.5414523 -0.2959174
-0.6113846
 [49] -1.3777482 -0.9695975 -0.6884003 -1.3106532 -0.9089572 -1.7970221 -0.6553532
-0.2785981
 [57] -0.9336035 -1.6773840 -1.0612469 -0.6002046  0.3070840 -1.0792818 -0.3905096
-0.7643933
 [65] -1.5264474 -0.3540782 -1.9662361 -0.6233462 -0.9594732  0.1381904 -0.1852506
-1.3285226
 [73] -0.5803871 -1.4863370 -1.7103169 -0.8629584 -1.0247360 -0.4806361 -0.7059812
-0.9927834
 [81] -1.5921652 -0.8216154 -0.3559753 -2.1285959 -0.5727712 -0.8732135 -0.1023033
-1.3436125
```

[89] -0.9225435 -1.0979780 -1.3155360 -0.2955632 -0.6027357 -0.4422999 -0.5086040 -0.9827529
[97] -1.2780067 -1.5405946 -1.5093126 -1.3319693

C



Adjusted R-squared (AdjR2): The plot shows that as the number of variables increases from 1 to 2 or 3, the adjusted R-squared increases, it means a better fit.

Bayesian Information Criterion (BIC): The BIC plot has a decrease as variables are added from 1 to 2 or 3, after which it begins to increase again. Lower BIC values generally indicate a better model.

Mallows's Cp (Cp): Similar to the BIC plot, the Cp plot decreases with the addition of the first few variables, indicating improved model fit. As more variables are added beyond this point, the Cp value begins to increase, which means the model is overfitted.

All three criteria seem to agree that a model with around 2 to 3 variables is optimal.

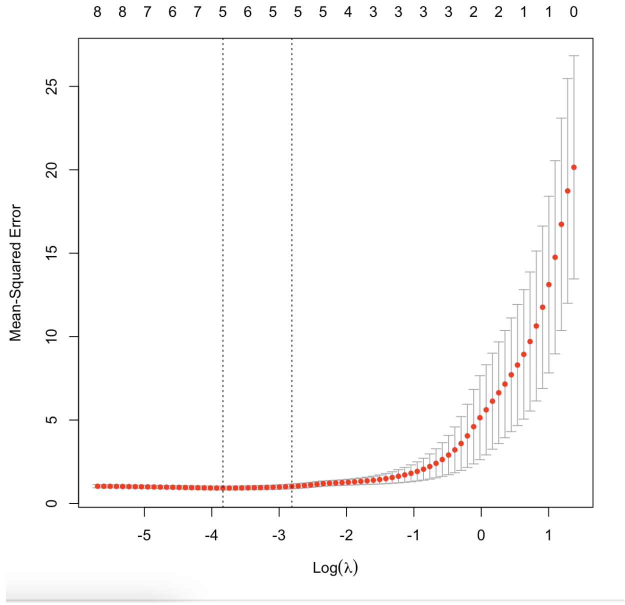
Best Model According to Cp:
Coefficients:
Intercept: 6.102647
X: -7.192951
X^2: 40.744047
X^3 : -14.709083
X^5: 1.480188

Best Model According to BIC:
Coefficients:
Intercept: 6.102647
X: -7.192951
X^2: 40.744047
X^3 : -14.709083

Best Model According to Adjusted R^2:
Coefficients:
Intercept: 6.102647
X: -7.192951
X^2: 40.744047
X^3 : -14.709083
X^4: 1.257095
X^5: 1.480188
Regsubsets has 3 parameters

D
The conclusion is that both backward and forward stepwise selection methods agree with the best subset selection method in terms of the Cp criterion.

E



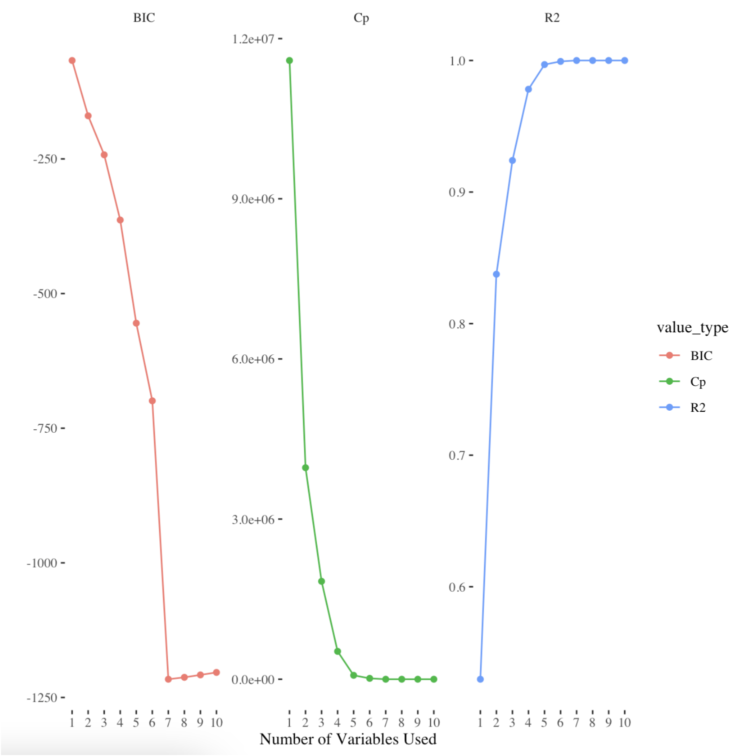
we can see that as λ increases (moving from left to right on the x-axis), the MSE initially decreases, and then starts to increase. The plot shows the behavior where a small λ can improve the model's prediction error by preventing overfitting. However, too much regularization (large λ) increases the bias, leading to underfitting and, consequently, a larger MSE. The optimal λ is the value that minimizes the cross-validation MSE.

coefficients:
Intercept: 3.1537591247

X: 0.8716827172
X^2: 3.6974703820
X^3 : -0.9540089306
X^4: 0.0067057108
X^5: implicitly zero due to lasso shrinkage
X^6: 0.0047086898
X^7: implicitly zero due to lasso shrinkage
X^8: 0.0003825871
X^9: implicitly zero due to lasso shrinkage
X^10: implicitly zero due to lasso shrinkage

Lasso regression has effectively identified key predictors for the response variable Y within the dataset, demonstrating its capability for feature selection by shrinking less important coefficients to zero. The relationships between the predictors and the response suggest a complex underlying structure that is not linear.

F



Best subset selection and lasso regression have identified a small subset of polynomial terms that are useful for predicting Y. The results from the lasso are particularly strong, showing that the model fits the data with high precision.

Question 9
A

B
RMSE: 1371.887
MAE: 645.0298
RMSE measures the deviation between the predicted values and the actual values. Lower RMSE means better performance. MAE measures the difference between the predicted and actual values. A lower MAE means better predictive accuracy.

C
RMSE: 796.2182
MAE: 341.7651
The ridge regression model with the selected lambda seems to have a lower RMSE and MAE compared to the linear model, which means that it may provide a better fit to the data.

D
RMSE: 140.5866
MAE: 79.6226
The Lasso model has only 2 non-zero coefficient estimates. This means that the Lasso model has performed feature selection and retained only two predictor variables in the final model, while setting the coefficients of the other variables to zero.

E
RMSE: 0.09816806
MAE: 0.07761307
RMSE and MAE values are close to zero, meaning that the PCR model is performing very well on the test data.
The value of M selected by cross-validation for the PCR model is 10. This means that the model achieved the best performance on the test data when using 10 principal components.

F
RMSE: 0.09492598
MAE: 0.07509352
RMSE and MAE values are close to zero, meaning that the PLS model is performing very well on the test data.
The value of M selected by cross-validation for the PLS model is 10. This means that the model achieved the best performance on the test data when using 10 components.

G
The results show that PCR and PLS outperform Linear Regression, Ridge Regression, and Lasso Regression in predicting the number of college applications received. PCR and PLS yield the lowest RMSE and MAE, indicating their high predictive accuracy. There is a significant difference between PCR/PLS and the other three methods, with PCR and PLS delivering the most accurate predictions.

Question 11
C
print(results)

```
  Model     RMSE
1 Lasso 2.719704
2 Ridge 2.727369
3   PCR 3.410396
```

Based on the RMSE values, the Lasso regression model appears to be the most suitable for predicting the per capita crime rate in the Boston dataset, as it provides the lowest prediction error. This choice may also involve feature selection, as Lasso tends to shrink some coefficients to zero, effectively reducing the feature space to the most relevant variables.