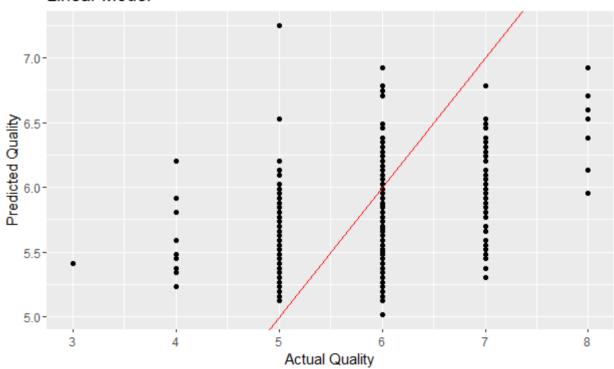
Part 1: Regression

Complete the following objectives utilizing the data set you selected as ideal for:

Linear Regression

a)

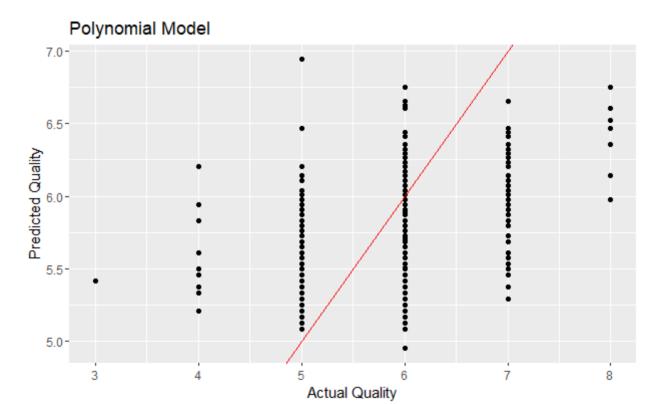
Linear Model



"Linear Model MSE: 0.459945869270605"

The linear regression model uses only the alcohol content to predict the quality of the wine. The model has an MSE of about 0.460 on the test data.

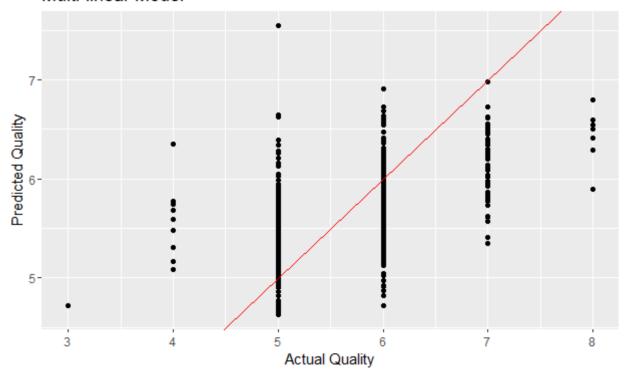
b)



Polynomial Model MSE: 0.458923441001458
This model uses a quadratic term (degree 2) for the alcohol content to predict wine quality. It has a slightly lower MSE of about 0.459 on the test data.

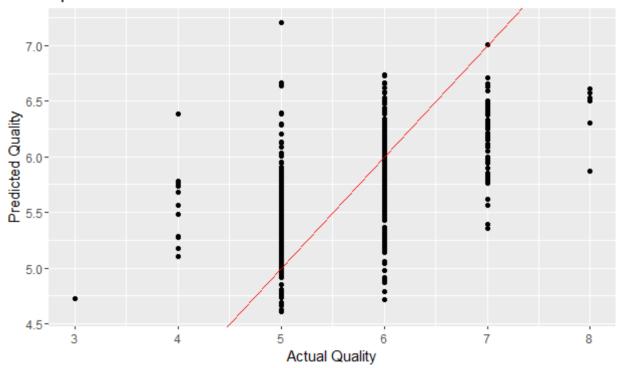
c)

Multi-linear Model



Multi-linear Model MSE: 0.426564653304327
This model uses all available features in the dataset to predict wine quality. It has a noticeably lower MSE of about 0.427 on the test data.

Spline Model



Spline Model MSE: 0.423194625548574

This model uses a natural spline transformation of the alcohol content, along with all other features in the dataset, to predict wine quality. The model encountered a warning about potential issues with the rank-deficient fit, which suggests that some of the predictors may be highly correlated. Despite this, it has the lowest MSE of about 0.423 on the test data.

Part2: Feature Selection / Model Optimization Methods

- 1) Feature Selection
 - a. Perform a Forward Stepwise Selection

1 subsets of each size up to 20

Selection Al		•					
			+-1:+4	infant doaths	۸٦coh	ol percentage_exp	oondituro
		Muurt_Mor	carrey	" "	H II	" "	pena i cai e
		11 11			0.0	11 11	
		11%11		11 11		11 11	
3 (1)		11.8.11		11 11		11 11	
4 (1)		11411		11 11		11 11	
3 (1)		11411				 !! & !!	
0 (1)							
7 (1) "		n * n		" "	" "	11 ½ II	
8 (1) "*		11.5.11		" "	" "	11.7.11	
9 (1) "*		11.5.11		" "	11 % 11	11 ½ II	
10 (1) "*		11.4.11		" "	11.5.11	11.7.11	
11 (1)"				" "	0.80	11.7.11	
12 (1)"		11.8.11		11 % II	11.54.11	11.44.11	
13 (1) "*		11 % 11		11 % 11	11.54.11	11.5.11	
14 (1) "*		11.5.11		11 ½ II	11 % 11	11 × 11	
15 (1) "*		11.8.11		n * n	$u \otimes u$	11.8.11	
16 (1) "*		11.8.11		n * n	$u \otimes u$	11.8.11	
17 (1) "*		11 % 11		11 ½ II	11 % 11	11 2 11	
18 (1) "*		пуп		пжп	0.80	пжп	
19 (1) "*	. 11 11% 11	11.5.11		пжп	пжп	11.811	
20 (1) "*	. 11 11% 11	11.5.11		пжп	пжп	11.811	
	natitic B	Moseles BMI	· `undo	r-five_deaths`	Dolio -	Total_expenditure	Dinh+honia
1 (1) "'		Measles BMI		i - i i ve_ueatiis		i oca i _expendicui e	" "
2 (1) "'						1 11	
3 (1) "'						1 11	
4 (1) "'	1					1 11	шш
5 (1) "'		п п п п п п т п т т т т т т т т т т т т				1 11	11 11
6 (1) "'		п п п п м				1 11	11 11
7 (1) "'	ı	п п п п м				1 11	11%11
8 (1) "'	•	п п п п т				1 11	11.5.11
9 (1) "'	•	п п п п д г			0 0 0	1 11	пуп
10 (1)"'	1	п п п п д г			0 0 0	1 11	пуп
11 (1)"'	1	п п п*1	111411			1 11	пУп
12 (1)"'	•	п п п ч	1 11%11			1 11	11.74.11
13 (1)"'	•	п п п м	111411			1 11	11 % 11
14 (1)"'	1	п п п п д г	111411		11 11 1	1411	пУп
15 (1)"'	•	п п п ч	1 11%11		11.811	1411	11.4.11
16 (1) "*'	•	11 11 11%1	1 11%11		11%11 1	1511	11 % 11
17 (1) "*'	1	11 71 11 71	1 11%11		11%11 1	1411	пУп
18 (1) "*'	•	11.5.11	11111		11 % II 1	1411	11.7.11
19 (1) "*'	•	11.511 11.51	1 11%11		11 % II	1411	11.7.11
20 (1) "*'	•	11411 1141	1 11%11		11½ II - 1	14.11	11411

```
`HIV/AIDS` GDP Population `thinness__1-19_years` `thinness_5-9_years`
                                            11 11 11
      (1)
1
                                                                           ......
                                                                                                                            .....
                     0.38\,0
                                            2
         1)
      (
                     \Pi \not\cong \Pi
                                                                            .....
3
      (1
              )
                                            0 0 0 0
                                                                           11 11
                                                                                                                            11 11
                     \Pi \not\cong \Pi
4
      ( 1
              )
                     0.8 \pm 0
                                            0 0 0 0
                                                                           11 11
                                                                                                                            5
      (1
              )
                     \Pi \not\approx \Pi
6
      (1)
                                            0 0 0 0
                                                                           11 11
                                                                                                                            11 11
                     \Pi \not\cong \Pi
7
      (1)
                     0.8 \pm 0
                                                                           11 11
8
         1
      (
              )
                     0.8 \pm 0
9
      (1)
                                            11 11
                                                                           11 11
10
       (1)
                     \Pi \not\cong \Pi
                                                                                                                            11 11
                     \Pi \not\cong \Pi
                                                                            11 11
                                                                                                                            11 11
        (1)
11
                                                                           11 11
                                                                                                                            .....
        (1)
                     \Pi \not\approx \Pi
12
            1)
                     \Pi \not\cong \Pi
                                            0 0 0 0
                                                                           11 11
                                                                                                                            \Pi \not\approx \Pi
13
        (
                     пуп
                                                                           \mathbf{H} = \mathbf{H}
                                                                                                                            пуп
            1
14
        (
                )
                                                                           ......
        (1)
                     0.8 \pm 0
                                                                                                                            H \otimes H
15
                     0.8 \pm 0
                                            11 11
                                                                           11 11
                                                                                                                            \Pi \not \approx \Pi
        (1)
16
                    \Pi \not \cong \Pi
                                                                                                                            \Pi \not \cong \Pi
17
        (1)
                                            0 0 0 0
                     \Pi \not\cong \Pi
                                                                           \Pi \not\simeq \Pi
                                                                                                                           \Pi \not\simeq \Pi
18
       (1)
                     \Pi \not\cong \Pi
                                            \mathbf{u} + \mathbf{u} = \mathbf{u} - \mathbf{u}
                                                                           \Pi \not\simeq \Pi
                                                                                                                            \Pi \not\cong \Pi
19
        (
            1)
                                            пун пун
                     \Pi \not \cong \Pi
                                                                           \Pi \not\cong \Pi
                                                                                                                            \Pi \not\approx \Pi
20
        (1)
                       Income_composition_of_resources Schooling
       (1)
1
                                                                                                 . .
                       \Pi \not\cong \Pi
       (1)
2
                                                                                                 11 11
                       \Pi \not\cong \Pi
3
       (1)
                       \Pi \not \propto \Pi
                                                                                                 \Pi \not \cong \Pi
4
       (
           1)
                       пуп
                                                                                                 пуп
5
       (1)
                       \Pi \not \approx \Pi
                                                                                                 \Pi \not \cong \Pi
6
       (
           1)
                       \Pi \not \simeq \Pi
                                                                                                 пуп
7
       (1)
                       \Pi \not\cong \Pi
                                                                                                 пуп
8
       (1)
                       \Pi \not\approx \Pi
                                                                                                 \Pi \not \approx \Pi
9
       (1)
                       \Pi \not\approx \Pi
                                                                                                 \Pi \not \approx \Pi
         (1)
10
                       \Pi \not\cong \Pi
             1
                  )
                                                                                                 \Pi \not \approx \Pi
11
         (1)
                       \Pi \not\cong \Pi
                                                                                                 \Pi \not\cong \Pi
12
                       \Pi \not \simeq \Pi
                                                                                                 пуп
13
         (1)
                       \Pi \not \approx \Pi
                                                                                                 пуп
             1
14
         (
                  )
         (1)
                       пуп
                                                                                                 пуп
15
16
         (
             1)
                                                                                                 \Pi \not \approx \Pi
                       \Pi \not\simeq \Pi
                                                                                                 H \otimes H
         (1)
17
                       \Pi \not\cong \Pi
                                                                                                 пуп
         (1)
18
                       \Pi \not \cong \Pi
                                                                                                 \Pi \not \cong \Pi
         (1
19
                  )
         (1)
                       \Pi \not\approx \Pi
                                                                                                 \Pi \not \approx \Pi
20
```

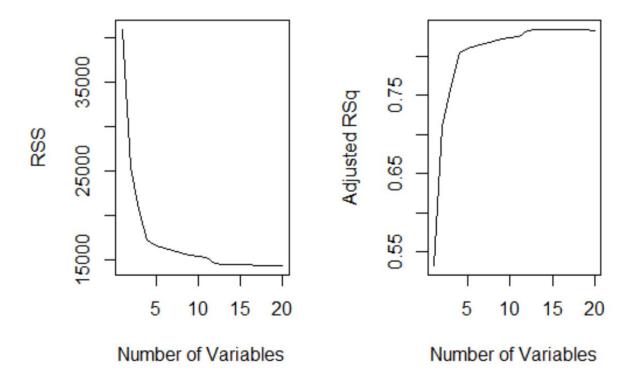
b. Perform a backward Stepwise Selection

1 subsets of each size up to 20
Selection Algorithm: backward
Year Status Adult_Mortality infant_deaths Alcohol percentage_expenditure

4 / 4 \	Year	Status	H441 C_M	orcarrey	inrant_deaths	, Alconor	" "	expenditure
1 (1)								
2 (1)						" "	" "	
3 (1)						" "	" "	
4 (1)			n % n					
5 (1)			11 % II			" "		
6 (1)			11 % 11				" "	
7 (1)		" "	11 % 11					
8 (1)	" "	" "			u.×.u	" "	11 % 11	
9 (1)		" "				" "	11 % 11	
10 (1)		" "	п У п		n <u>×</u> n		II % II	
11 (1)	11 % 11	11 % 11			п×п	11 % 11	11 % 11	
12 (1)		11 % 11			11.7.11	11.5.11	n % n	
13 (1)	11 % II	11 % 11			11.4.11	11.4.11	п % п	
14 (1)	11 % II				n.×n	11.74.11	п¾п	
15 (1)	11 % 11	11 % 11			11.7.11	11.4.11	п % п	
16 (1)	11 % 11	11 % 11			11.7.11	11.5.11	п У п	
17 (1)	11 % 11	11 % 11	11 % II		11.5.11	11.8.11	11 % 11	
18 (1)	11.54.11	11.5.11	11.4.11		11.4.11	11.8.11	11 % II	
19 (1)	11.54.11	11.54.11	11.54.11		11.4.11	11.8.11	n * n	
20 (1)	п У п	11 % 11	11.54.11		11.4.11	11 % 11	11 % 11	
	Hepat ⁻	itis B M	easles B	MI `under	-five_deaths` I	Polio Tota	al_expenditur	e Diphtheria
1 (1)		"			,		_ '	11 11
1 (1) 2 (1)	n ii n n	"	" "		,			" "
1 (1) 2 (1) 3 (1)	11 11 11 11	"	" "	" " " " " "	,		_ '	" " " "
1 (1) 2 (1) 3 (1) 4 (1)	" " " " " "	"	11 11 11 11 11 11	11 11 11 11 11 11	,	" " " " " " " " " " " " " " " " " " "	- '	" "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1)	11 11 11 11	"	" " " " " " " " " " " " " " " " " " "	" " " " " " " " " " " " " " " " " " "	,		- '	" " " "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1)	" " " " " "	"	" " " " " " " " " " " " " " " " " " "	" " " " " " " " " " " " " " " " " " "	,		- '	" " " " " "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1)	" " " " " " " " " " " " " " " " " " "	""	" " " " " " " " " " " " " " " " " " "	" " " " " " " " " " " " " " " " " " "	,		- '	" " " " " " " "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1)	n ii	"" "" "" "" "" ""	" " " " " " " " " " " " " " " " " " "	" " " " " " " " " " " " " " " " " " "			- '	" " " " " " " " " "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1) 9 (1)	" " " " " " " " " " " " " " " " " " "	"" "" "" "" "" "" ""	" " " " " " " " " " " " " " " " " " "	" " " " " " " " " " " " " " " " " " "			- '	" " " " " " " " " " " "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1) 9 (1) 10 (1)		"" "" "" "" "" "" "" "" "" "" "" ""		" " " " " " " " " " " " " " " " " " "				" " " " " " " " " " " " " " " "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1) 9 (1) 10 (1) 11 (1)		"" "" "" "" "" "" "" "" "" "" "" ""		" " " " " " " " " " " " " " " " " " "				
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1) 9 (1) 10 (1) 11 (1) 12 (1)		"" "" "" "" "" "" "" "" "" "" "" "" ""		" " " " " " " " " " " " " " " " " " "				
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1) 9 (1) 10 (1) 11 (1) 12 (1) 13 (1)		"" "" "" "" "" "" "" "" "" "" "" "" ""		H H H H H H H H H H H H H H H H H H H				" " " " " " " " " " " " " " " " " "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1) 9 (1) 10 (1) 11 (1) 12 (1) 13 (1) 14 (1) 15 (1)		"" "" "" "" "" "" "" "" "" "" "" "" ""						" " " " " " " " " " " " " " " " " " "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1) 9 (1) 10 (1) 11 (1) 12 (1) 13 (1) 14 (1) 15 (1) 16 (1)		"" "" "" "" "" "" "" "" "" "" "" "" ""						
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1) 9 (1) 10 (1) 11 (1) 12 (1) 13 (1) 14 (1) 15 (1) 16 (1) 17 (1)		"" "" "" "" "" "" "" "" "" "" "" "" ""	" " " " " " " " " " " " " " " " " " "					" " " " " " " " " " " " " " " " " " "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1) 9 (1) 10 (1) 11 (1) 12 (1) 13 (1) 14 (1) 15 (1) 16 (1) 17 (1) 18 (1)		"" "" "" "" "" "" "" "" "" "" "" "" ""						" " " " " " " " " " " " " " " " " " "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1) 9 (1) 10 (1) 11 (1) 12 (1) 13 (1) 14 (1) 15 (1) 16 (1) 17 (1) 18 (1) 19 (1)		"" "" "" "" "" "" "" "" "" "" "" "" ""						" " " " " " " " " " " " " " " " " " "
1 (1) 2 (1) 3 (1) 4 (1) 5 (1) 6 (1) 7 (1) 8 (1) 9 (1) 10 (1) 11 (1) 12 (1) 13 (1) 14 (1) 15 (1) 16 (1) 17 (1) 18 (1) 19 (1)		"" "" "" "" "" "" "" "" "" "" "" "" ""						" " " " " " " " " " " " " " " " " " "

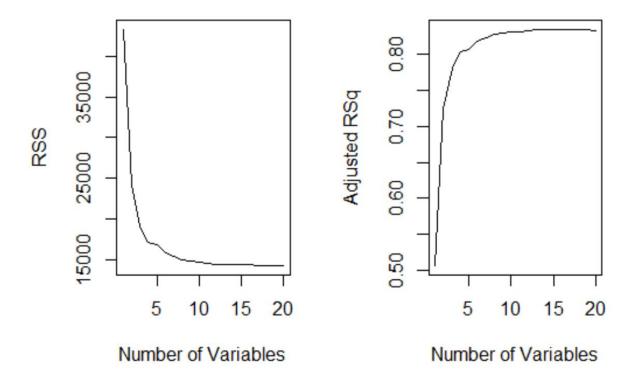
```
`HIV/AIDS` GDP Population `thinness__1-19_years`
                                                                                                                           `thinness_5-9_years
                     11 11
                                            . . . . . .
                                                                            11 11
      (1)
1
                     \Pi \otimes \Pi
                                            . . . . . .
                                                                            11 11
                                                                                                                             11 11
      (1)
2
                     \Pi \not\cong \Pi
                                            11 11
                                                    11 11
                                                                            ......
                                                                                                                            11 11
3
      (1)
                     \Pi \otimes \Pi
                                                                            11 11
                                                                                                                             4
      (1
              )
                     \Pi \not\cong \Pi
                                            11 11
                                                   11 11
                                                                            11 11
                                                                                                                             11 11
5
      (1)
                     \Pi \not\cong \Pi
                                                11
                                                                            11 11
                                                                                                                             11
6
      (1)
                                            11 11
                                                                            .....
                                                                                                                            11 11
7
                     \Pi \not\cong \Pi
      (1)
                     пжп
                                            11
                                                11
                                                                            11 11
                                                                                                                             11 11
8
      (1
              )
                                            11 11
                                                                            11 11
                                                                                                                             (1)
                     \Pi \not\cong \Pi
9
                     \Pi \not \cong \Pi
                                                                            11 11
                                                                                                                             11 11
10
        (1)
                                                11
                                                                            11 11
       (1)
                     \Pi \not\cong \Pi
                                                                                                                            11 11
11
                     \Pi \not \cong \Pi
                                            0 0 0 0
                                                                            11 11
                                                                                                                            11 11
12
            1)
                                                                            .....
                     \Pi \not\cong \Pi
                                                                                                                            \Pi \not \cong \Pi
13
        (1)
        (1)
                     \Pi \not\cong \Pi
                                            11
                                                11
                                                                            11 11
                                                                                                                            \Pi \not\cong \Pi
14
        (1)
                     0.8 \pm 0
                                                                            11 11
                                                                                                                            \Pi \not \approx \Pi
15
                     \Pi \not\approx \Pi
                                            0.00
                                                                            11 11
                                                                                                                            пжп
16
        (
            1)
                                                                                                                            H \otimes H
                     \Pi \not\cong \Pi
17
        (1)
                     \Pi \not\cong \Pi
                                            0.00
                                                                            \Pi \not \approx \Pi
                                                                                                                            \Pi \not\approx \Pi
       (1)
18
                    \Pi \otimes \Pi
                                            \mathbf{u} + \mathbf{u} = \mathbf{u} + \mathbf{u}
                                                                           \Pi \otimes \Pi
                                                                                                                            \Pi \not\cong \Pi
19
       (1)
                                            пун пун
                                                                           0.8 \pm 0
                                                                                                                            \Pi \not\approx \Pi
20
       (1)
                     Income_composition_of_resources Schooling
                                                                                          пуп
1
      (1)
                     11 11
                                                                                          пуп
2
      (1)
                     пуп
3
      (1)
                     пуп
                                                                                          \Pi \not \approx \Pi
4
      (1)
                     \Pi \not\cong \Pi
                                                                                          пуп
5
      (1)
                     \Pi \not \approx \Pi
                                                                                          пуп
6
      (1
              )
                                                                                          \Pi \not \simeq \Pi
7
      (1)
                     пуп
                     \Pi \not \approx \Pi
                                                                                          пуп
8
      (1)
      (1)
                     \Pi \not\approx \Pi
                                                                                          \Pi \not\approx \Pi
9
                                                                                          пуп
       (1)
                     \Pi \not \approx \Pi
10
                     \Pi \not\approx \Pi
                                                                                          \Pi \not \cong \Pi
11
       (1)
        (1)
                     \Pi \not\cong \Pi
                                                                                          пуп
12
                     \Pi \not \approx \Pi
                                                                                          \Pi \not \approx \Pi
           1)
13
        (
        (1)
                                                                                          \Pi \not \cong \Pi
                     \Pi \not\cong \Pi
14
                     \Pi \not \approx \Pi
                                                                                          \Pi \not \approx \Pi
        (1)
15
16
        (1)
                     пуп
                                                                                          \Pi \not\cong \Pi
                     \Pi \not\cong \Pi
                                                                                          \Pi \otimes \Pi
17
        (1)
                     \Pi \not\cong \Pi
                                                                                          \Pi \not \approx \Pi
       (1)
18
                                                                                          пуп
                    пуп
19
       (1)
       (1)"*"
                                                                                          \Pi \not \approx \Pi
20
```

- 2) Using the models generated for the feature selection, generate the plots of RSS and Adjusted R2
 - a. Forward Features



Forward selected features have a huge improvement in the RSS and adjusted R squared up to 4 features. There is continued improvement up 11 or 12 components. Additional features saw little to no improvement.

b. Backward Features



Backward selected features have a huge improvement in the RSS and adjusted R squared up to 4 features. There is continued improvement up 10 or 11 components. Additional features saw little to no improvement.

```
PCR
            X dimension: 1156 20
Y dimension: 1156 1
Data:
Fit method: svdpc
Number of components considered: 20
3 comps
7.947
7.944
                         1 comps
8.718
8.717
                                                           4 comps
7.938
                                                                      5 comps
7.483
                                                                                 6 comps
5.639
                                                                                             7 comps
5.486
5.483
               8.722
8.722
8.722
ps 9 comps
5.071
5.068
                                       8.035
7.982
                                                             7.935
adjcv
                                                                         7.497
                                                                                    5.637
        8 comps
5.322
                                           11 comps
4.980
                                                                     13 comps
                                                                                  14 comps
                                                                                              15 comps
4.243
4.238
                                                        12 comps
                               10 comps
                                   5.056
                                                            4.865
                                                                         4.446
                                                                                     4.424
adjcv
                                                                                     4.420
           5.341
                                                4.976
                                                                         4.440
                                   5.053
                                                            4.890
         16 comps
                         comps
3.779
                     17
                                  18 comps
                                              19 comps
                                                           20 comps
                                                  3.767
3.763
                                                               3.603
3.598
             4.243
                                      3.776
adjcv
             4.239
                         3.776
                                      3.773
TRAINING: % variance explained
                     1 comps
100.00000
                                              3 comps
100.00
                                                                     5 comps
100.00
                                   2 comps
                                                          4 comps
                                                                                6 comps
                                                                                           7 comps
                                                                                                        100.00
                                                           100.00
                                                                                            100.00
                                     100.0
                                                                                 100.00
Life_expectancy
                        0.09701
                                       15.8
                                                 17.83
                                                            18.14
                                                                        28.11
                                                                                   58.72
                                                                                              61.05
                                                                                                         63.38
                                                             comps
100.0
                     9 comps
                                10 comps
                                             11 comps
                                                                      13 comps
                                                                                   14 comps
                                                                                               15 comps
                       100.00
                                   100.00
                                                100.00
                                                                         100.00
                                                                                     100.00
                                                                                                  100.00
Life_expectancy
                        66.79
                                    67.16
                                                 68.23
                                                               70.9
                                                                          74.87
                                                                                      75.14
                                                                       20 comps
100 00
                                 17 comps
100.00
                     16 comps
                                              18 comps
100.00
                                                           19 comps
                                                             100.00
                        100.00
Life_expectancy
                         77.17
                                     81.77
                                                  81.83
                                                               81.99
                                                                           83.64
```

This is the output of a Principal Component Regression (PCR) analysis, which is a technique used to predict a dependent variable (in this case, life expectancy) using multiple independent variables. The analysis was performed with 20 components, and the root mean square error of prediction (RMSEP) decreased as more components were used, indicating an improvement in prediction accuracy. The variance explained by the model for life expectancy increased with each component, reaching a maximum of 83.64% with 20 components, which implies that 83.64% of the variability in life expectancy can be explained by the independent variables used in the model.

Part 3: Classification

1) Generate two Classification Models for based on your given data.

These models will include Logistic Regression, Linear Discriminant Analysis, as given in the chapter 4 lab exercise.

Logistic_predictions Confusion Matrix and Statistics

```
Reference
Prediction 0 1
0 94 4
1 4 69
```

Accuracy: 0.9532 95% CI: (0.9099, 0.9796)

No Information Rate : 0.5731 P-Value [Acc > NIR] : <2e-16

карра: 0.9044

Mcnemar's Test P-Value : 1

Sensitivity: 0.9592
Specificity: 0.9452
Pos Pred Value: 0.9592
Neg Pred Value: 0.9452
Prevalence: 0.5731
Detection Rate: 0.5497
Detection Prevalence: 0.5731
Balanced Accuracy: 0.9522

'Positive' Class: 0

lda_predictions

Confusion Matrix and Statistics

Reference Prediction 0 1 0 97 8 1 1 65

Accuracy: 0.9474

95% CI: (0.9024, 0.9757)

No Information Rate : 0.5731 P-Value [Acc > NIR] : <2e-16

Kappa : 0.8911

Mcnemar's Test P-Value: 0.0455

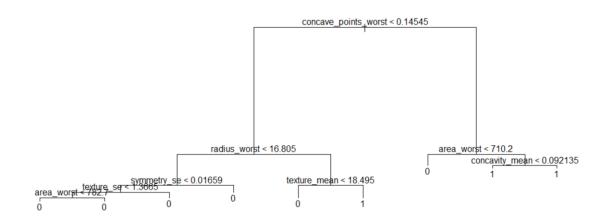
Sensitivity: 0.9898 Specificity: 0.8904 Pos Pred Value : 0.9238 Neg Pred Value : 0.9848 Prevalence: 0.5731 Detection Rate: 0.5673

Detection Prevalence: 0.6140 Balanced Accuracy: 0.9401

'Positive' Class: 0

These are the confusion matrices and associated statistics for two different predictive models: logistic regression and linear discriminant analysis (LDA). The logistic regression model achieved an accuracy of 95.32% with a kappa s tatistic of 0.9044, indicating high agreement between predicted and actual values. The LDA model's accuracy was slightly lower at 94.74%, with a kappa statistic of 0.8911, but it had a higher sensitivity, meaning it was more effect ive at correctly identifying positive instances of the class labeled '0'.

2) Generate a tree classifier for the classes and predictors used above.



tree_predictions

Confusion Matrix and Statistics

Reference Prediction 0 1 0 97 8 1 1 65

Accuracy: 0.9474 95% CI: (0.9024, 0.9757)

No Information Rate: 0.5731 P-Value [Acc > NIR] : <2e-16

Kappa: 0.8911

Mcnemar's Test P-Value: 0.0455

Sensitivity: 0.9898 Specificity: 0.8904 Pos Pred Value: 0.9238 Neg Pred Value: 0.9848 Prevalence: 0.5731 Detection Rate: 0.5673

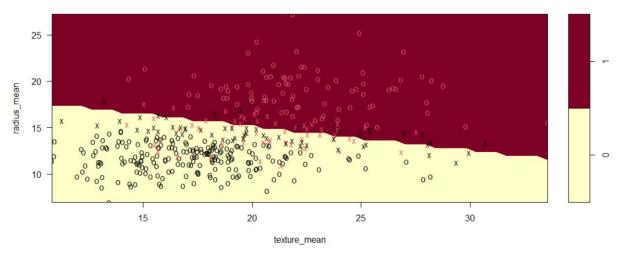
Detection Rate: 0.56/3
Detection Prevalence: 0.6140
Balanced Accuracy: 0.9401

'Positive' Class: 0

This is the confusion matrix and associated statistics for a predictive model, likely a decision tree given the name "tree_predictions". The model has an accuracy of 94.74%, and a kappa statistic of 0.8911, indicating a high level of agreement between the model predictions and the actual values. The model demonstrates high sensitivity (98.98%), meaning it correctly identified most positive instances of the class labeled '0'.

3) Construct a Support Vector Classifier for some set of classes.

SVM classification plot



This output represents the process and results of training a Support Vector Machine (SVM) model to predict the 'diagnosis' based on 'radius_mean' and 'texture_mean'. In the hyperparameter tuning stage, various cost values were tested for the SVM's regularization parameter. The best performance, an error rate of approximately 0.1106, was achieved with a cost value of 5. The final model, trained with the optimal cost, used a linear kernel and resulted in 113 support vectors split almost evenly between the two classes, 0 and 1.

Part 4: Models

a) For the first scenario, you might consider using a Multiple Linear Regression model. This model is used to understand the relationship between two or more variables and an outcome. The model could help your friend determine how each factor (like total inventory, number of employees, annual

operation budget) contributes to the company's success (total profits). The coefficients of the regression model will tell you the degree of impact each predictor variable has on the outcome variable.

- b) In this scenario I would try to use a Classification model, such as logistic regression, decision trees, or even ensemble methods like random forest or gradient boosting. These methods can be used to predict a binary outcome (like whether a household will use the coupon or not), and they can handle a mix of different types of variables. This approach would allow you to predict the probability of each postal code's households using the coupon, and you could then target the ones with the highest probabilities.
- c) A Dimensionality Reduction technique like Principal Component Analysis (PCA) could be a good starting point. This technique can reduce the thousands of searches and behavior features into a smaller set of new variables (called principal components) that still capture most of the information in the original features. This can make subsequent analyses more manageable and less resource intensive.
- d) In the last scenario, a Support Vector Machine (SVM) or a Random Forest classifier might be suitable. These are powerful machine learning models that can handle high-dimensional data and are capable of modeling complex decision boundaries, which seems necessary given the significant overlap between classes. SVM is particularly good when classes are overlapping and works well in high-dimensional space, while Random Forest can model non-linear decision boundaries and provide feature importance.