Feature Selection Techniques in Regression Model with Example in R

https://ashutoshtripathi.com/

Feature selection is a way to reduce the number of features and hence reduce the computational complexity of the model. Many times feature selection becomes very useful to overcome with overfitting problem. It helps us in determining the smallest set of features that are needed to predict the response variable with high accuracy. If we ask the model, does adding new features, necessarily increase the model performance significantly? If not then why to add those new features which are only going to increase model complexity.

So now let's understand how we can select the important set of features out of total available features in the given data set.

It is always better to understand with an example. So let's look at the mtcars data set below in R:

```
> head(mtcars)
                  X  mpg cyl disp  hp drat    wt  qsec       vs am gear carb
1         Mazda RX4 21.0   6  160 110 3.90 2.620 16.46 V-shaped  1    4    4
2     Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02 V-shaped  1    4    4
3        Datsun 710 22.8   4  108  93 3.85 2.320 18.61 straight  1    4    1
4    Hornet 4 Drive 21.4   6  258 110 3.08 3.215 19.44 straight  0    3    1
5 Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02 V-shaped  0    3    2
6           Valiant 18.1   6  225 105 2.76 3.460 20.22 straight  0    3    1
> 
```

We will remove column x as it contains only car models and it will not add much value in prediction.

```
> mtcars = read.csv(file = "[redacted]/mtcars.csv",header=TRUE,sep=",")
> mtcars$X=NULL
> str(mtcars)
'data.frame':   32 obs. of  11 variables:
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : int  6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp  : int  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs  : Factor w/ 2 levels "straight","V-shaped": 2 2 1 1 2 1 2 1 1 1 ...
 $ am  : int  1 1 1 0 0 0 0 0 0 0 ...
 $ gear: int  4 4 4 3 3 3 3 4 4 4 ...
 $ carb: int  4 4 1 1 2 1 4 2 2 4 ...
> ## ---------------------------------------------------------------------------
> head(mtcars)
   mpg cyl disp  hp drat    wt  qsec       vs am gear carb
1 21.0   6  160 110 3.90 2.620 16.46 V-shaped  1    4    4
2 21.0   6  160 110 3.90 2.875 17.02 V-shaped  1    4    4
3 22.8   4  108  93 3.85 2.320 18.61 straight  1    4    1
4 21.4   6  258 110 3.08 3.215 19.44 straight  0    3    1
5 18.7   8  360 175 3.15 3.440 17.02 V-shaped  0    3    2
6 18.1   6  225 105 2.76 3.460 20.22 straight  0    3    1
> tail(mtcars)
    mpg cyl  disp  hp drat    wt qsec       vs am gear carb
27 26.0   4 120.3  91   NA 2.140 16.7 V-shaped  1    5    2
28 30.4   4  95.1 113 3.77 1.513 16.9 straight  1    5    2
29 15.8   8    NA 264 4.22 3.170 14.5 V-shaped  1    5    4
30 19.7   6 145.0 175 3.62 2.770 15.5 V-shaped  1    5    6
31 15.0   8 301.0 335 3.54 3.570 14.6 V-shaped  1    5    8
32 21.4   4 121.0 109 4.11 2.780 18.6 straight  1    4    2
> 
```

In the above data there are 12 features (x, mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb) and we want to predict the mpg (miles per gallon) hence it becomes our target/response variable.

Let's randomly select any of the predictor variable and try to fit the model for predicting mpg. Let's start with attribute "wt" then:

```
> LinearReg = lm(mpg ~ wt, data = mtcars)
> summary(LinearReg)

Call:
lm(formula = mpg ~ wt, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5783 -2.4766 -0.0902  1.4931  6.8458 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  37.2932     1.9056  19.571  < 2e-16 ***
wt           -5.3359     0.5679  -9.396 2.66e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.091 on 29 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.7528,	Adjusted R-squared:  0.7442 
F-statistic: 88.29 on 1 and 29 DF,  p-value: 2.658e-10

> 
```

Three stars (or asterisks) represent a highly significant p-value. Consequently, a small p-value for the intercept and the slope indicates that we can reject the null hypothesis which allows us to conclude that there is a strong relationship between mpg and weight. Typically, a p-value of 5% (.05) or less is a good cut-off point. In our model example, the p-values are very close to zero. Also R-squared value .74 tells us that around 74% of variance in target variable is explained by the model hence model is also significant.

Now let's fit the model with two variables "wt" and "hp" (horse power) as below: (note we can go with any two randomly picked predictors as we are just trying to understand what happens if we go with hit and trial method)

```
> MultiLinearReg = lm(mpg ~ wt+hp, data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ wt + hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.0225 -1.6664 -0.0960  0.9602  5.7892 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 37.241799   1.606254  23.185  < 2e-16 ***
wt          -3.817199   0.639579  -5.968 1.99e-06 ***
hp          -0.032715   0.009138  -3.580  0.00128 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.605 on 28 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.8304,	Adjusted R-squared:  0.8183 
F-statistic: 68.55 on 2 and 28 DF,  p-value: 1.63e-11

> 
```

Now R-squared value has increased to .81 from .74. Which means model has become more significant. Also looking into the no of stars against wt and "hp" we can say both are strongly related to target variable and hence both are important.

There might be the case that by adding new variable impact of already added variables decreases, in that case if p value crosses the upper threshold of .05 for any old variables then it means that variable now has become insignificant then we remove that variable.

now add one more variable "qsec" and analyse the model summary as below:

```
> MultiLinearReg = lm(mpg ~ wt+hp+qsec, data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ wt + hp + qsec, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.9847 -1.5012 -0.4675  1.1674  5.6921 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 26.52810    8.46623   3.133  0.00425 ** 
wt          -4.41916    0.76237  -5.797 4.17e-06 ***
hp          -0.01700    0.01507  -1.128  0.26943    
qsec         0.57466    0.44226   1.299  0.20522    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.579 on 26 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.8455,	Adjusted R-squared:  0.8277 
F-statistic: 47.43 on 3 and 26 DF,  p-value: 1.107e-10

> |
```

Logically by adding new variable it should not reduce the impact of already added variables but in this case as we can see in above image that variable “hp” and “qsec” both become insignificant (p-value > .05 also there is no star).

Now let’s add all the variable and see what happens:

```
> MultiLinearReg = lm(mpg ~ ., data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6074 -0.9126 -0.2565  0.8726  4.1079 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -16.48709   25.58316  -0.644    0.530
cyl           1.17550    1.50475   0.781    0.449
disp          0.01744    0.02903   0.601    0.558
hp           -0.01745    0.02987  -0.584    0.569
drat          3.62707    2.62587   1.381    0.190
wt           -3.22226    2.73952  -1.176    0.261
qsec          0.81826    0.82806   0.988    0.341
vsV-shaped   -0.90970    3.43725  -0.265    0.795
am            1.46786    2.55847   0.574    0.576
gear          4.03646    2.65074   1.523    0.152
carb         -1.39833    1.52270  -0.918    0.375

Residual standard error: 2.775 on 13 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.8889,	Adjusted R-squared:  0.8034 
F-statistic:  10.4 on 10 and 13 DF,  p-value: 0.0001112

> |
```

From above summary we see that none of the variable is significant as all p values are greater than the threshold limit .05, also summary has not produces any stars as significant code. This is kind of surprising. if no variable is significant then how to fit the model?

So if we do hit and try method with all combinations of the variable then there will be total 2^k – 1 linear models we have to try and see which are the significant features. Isn't this a time consuming job, of-course yes. So what to be done now? Here comes the feature selection techniques which helps us in finding the smallest set of features which produces the significant model fit. So in Regression very frequent used techniques for feature selection are as following:

- **Stepwise Regression**
- **Forward Selection**
- **Backward Elimination**

1. Stepwise Regression

In Stepwise regression technique we start fitting the model with each individual predictor and see which one has the lowest p-value. Then pick that variable and then fit the model using two variable one which we already selected in previous step and taking one by one all remaining ones. Again we select the one which has the lowest p-value. Also keep in mind that by adding the new variable, impact of already selected variable in previous step should still be significant. We keep this iteration until we get a combination whose p-value is less than the threshold of .05.

Let's understand this whole process using one example:

Step 1:

We fit the model with one predictor and target. We tried each predictor one by one and below each row represents the model fit with respective t-score, p-value and R-squared value. As we see mpg ~ wt fit has lowest p-value (also should be less than .05) so will select wt and go to step 2.

Target Variable y	Predictor variables x	t-score	p-value	Adjusted R-squared
mpg	cyl	-8.92	6.11E-10	0.7171
mpg	disp	-8.507	2.26E-09	0.7041
mpg	hp	-6.742	1.79E-07	0.5892
mpg	drat	4.837	3.99E-05	0.4274
mpg	wt	-9.396	2.66E-10	0.7442
mpg	qsec	2.478	0.0193	0.1463
mpg	vs	-4.864	3.42E-05	0.4223
mpg	am	3.971	0.000433	0.3299
mpg	gear	2.808	0.00898	0.1918
mpg	carb	-3.752	0.000782	0.3035

Step 2:

Now we will fit the model with two predictors. One we have already selected as wt in step 1 and for second predictor we will try one by one with all remaining predictors. And will again select those which have lowest p-value. in this case we got “wt” and “cyl”.

Target Variable y	Predictor variables x1	Predictor variables x2	t-score	p-value	Adjusted R-squared
mpg	wt	cyl	-3.655	0.00105	0.8207
mpg	wt	disp	-1.598	0.1216	0.7724
mpg	wt	hp	-3.58	0.00128	0.8183
mpg	wt	drat	1.069	0.29461	0.737
mpg	wt	qsec	3.648	0.00111	0.8259
mpg	wt	vs	-2.887	0.00742	0.7958
mpg	wt	am	-0.058	0.954	0.7317
mpg	wt	gear	-0.305	0.763	0.7076
mpg	wt	carb	-2.472	0.02	0.7865

Step 3:

Now will try to fit with 3 predictors two already selected in step 2 and third will try with remaining ones. But here we see none of the p-value is less than .05, hence none are significant.

All p-values are >.05 the upper limit

Target Var	Predictor variables x1	Predictor variables x2	Predictor variables x3	t-score	p-value	Adjusted R-squared
mpg	wt	cyl	disp	0.634	0.53149	0.8162
mpg	wt	cyl	hp	-1.585	0.1246	0.8299
mpg	wt	cyl	drat	0.19	0.85115	0.8084
mpg	wt	cyl	qsec	1.25	0.22238	0.8383
mpg	wt	cyl	vs	-0.574	0.57055	0.8163
mpg	wt	cyl	am	0.059	0.9532	0.814
mpg	wt	cyl	gear	-0.501	0.6206	0.7946
mpg	wt	cyl	carb	-1.458	0.15693	0.8248

As all p-values are greater than .05 hence none of the three combination features are going to be significant. Hence we stop here.

So using Stepwise regression we have got smallest set {wt, cyl} of features which have significance impact in final model fit. It does not mean other features do not have impact but they have very less impact which can be neglected if we are getting significant model fit with only two variables.

So here we have observed that our search space has reduced drastically as compared to hit and trial method where we have to compare the 2^10 – 1 = 1023 models.

2. Forward Selection

Forward selection is almost similar to Stepwise regression however only difference is that in forward selection we only keep adding the features. We do not delete the already added feature. in every iteration we add only those feature which increases the overall model fit.

3. Backward Elimination

In backward elimination in first step we include all predictors and in subsequent steps, keep on removing the one which has highest p-value (>.05 the threshold limit). After few iterations it will produce the final set of features which are enough significant to predict the outcome with desired accuracy.

We will take same example of mtcars data set and go step by step as following:

Step 1:

In step 1 we build the model with all the features available in the data set. Then observe few things:

```
> MultiLinearReg = lm(mpg ~ ., data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6074 -0.9126 -0.2565  0.8726  4.1079 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -16.48709   25.58316  -0.644    0.530
cyl           1.17550    1.50475   0.781    0.449
disp          0.01744    0.02903   0.601    0.558
hp           -0.01745    0.02987  -0.584    0.569
drat          3.62707    2.62587   1.381    0.190
wt           -3.22226    2.73952  -1.176    0.261
qsec          0.81826    0.82806   0.988    0.341
vsV-shaped   -0.90970    3.43725  -0.265    0.795
am            1.46786    2.55847   0.574    0.576
gear          4.03646    2.65074   1.523    0.152
carb         -1.39833    1.52270  -0.918    0.375

Residual standard error: 2.775 on 13 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.8889,	Adjusted R-squared:  0.8034 
F-statistic:  10.4 on 10 and 13 DF,  p-value: 0.0001112
```

Eliminate “vsv” due to highest p value

Here we see that none of the p-value is <.05 hence it seems no variable is significant. But lets follow backward elimination by removing highest p-value variable and see what happens

Step 2

```
> mtcars$vs=NULL
> MultiLinearReg = lm(mpg ~ ., data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6558 -0.8959 -0.2943  1.0279  4.1240 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) -16.66123   24.71069  -0.674   0.5111  
cyl           1.01596    1.33216   0.763   0.4583  
disp          0.01489    0.02646   0.563   0.5826  
hp           -0.01374    0.02549  -0.539   0.5983  
drat          3.53910    2.51674   1.406   0.1815  
wt           -3.01234    2.53360  -1.189   0.2542  
qsec          0.82554    0.79964   1.032   0.3194  
am            1.20830    2.28319   0.529   0.6049  
gear          4.27737    2.40545   1.778   0.0971 .
carb         -1.59482    1.28452  -1.242   0.2348  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.682 on 14 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.8883,	Adjusted R-squared:  0.8164 
F-statistic: 12.37 on 9 and 14 DF,  p-value: 3.083e-05
```

Eliminate “am” due to highest p value

Still none of the variable is significant

Step 3

```
> mtcars$am=NULL
> MultiLinearReg = lm(mpg ~ ., data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.9462 -1.0742 -0.0731  1.1979  4.2844 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept) -15.636110  23.578329  -0.663   0.5167  
cyl           1.267286   1.198088   1.058   0.3059  
disp          0.004696   0.020479   0.229   0.8216  
hp           -0.009594   0.023209  -0.413   0.6848  
drat          3.696523   2.322859   1.591   0.1311  
wt           -2.165285   2.115192  -1.024   0.3212  
qsec          0.555553   0.681414   0.815   0.4269  
gear          4.956920   2.089229   2.373   0.0305 *
carb         -2.016243   1.123927  -1.794   0.0917 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.57 on 16 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:  0.8846,	Adjusted R-squared:  0.8269 
F-statistic: 15.33 on 8 and 16 DF,  p-value: 3.758e-06
```

Eliminate disp due to highest p value

gear has become significant as p-value < .05 (see star mark)

Step 4

```
> mtcars$disp=NULL
> MultiLinearReg = lm(mpg ~ ., data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0212 -1.8901 -0.4226  1.2236  5.3323 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 15.98889   20.55440   0.778   0.4467  
cyl         -0.05221    1.06356  -0.049   0.9614  
hp          -0.01803    0.02011  -0.897   0.3818  
drat         0.86212    2.11249   0.408   0.6880  
wt          -2.77002    1.36497  -2.029   0.0575 .
qsec         0.44899    0.70073   0.641   0.5298  
gear         1.75291    1.63189   1.074   0.2969  
carb        -0.54200    0.73414  -0.738   0.4699  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.783 on 18 degrees of freedom
  (6 observations deleted due to missingness)
Multiple R-squared:  0.8506,	Adjusted R-squared:  0.7926 
F-statistic: 14.65 on 7 and 18 DF,  p-value: 2.927e-06
```

Eliminate cyl due to highest p value

gear has again become insignificant, wt has one dot against its p values means it is near to significant

Step 5

```
> mtcars$cyl=NULL
> MultiLinearReg = lm(mpg ~ ., data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0194 -1.9234 -0.4116  1.2018  5.3065 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 15.20899   12.69615   1.198   0.2457  
hp          -0.01841    0.01804  -1.021   0.3202  
drat         0.89477    1.95173   0.458   0.6518  
wt          -2.79486    1.23404  -2.265   0.0354 *
qsec         0.46980    0.54319   0.865   0.3979  
gear         1.78156    1.48343   1.201   0.2445  
carb        -0.54112    0.71439  -0.757   0.4581  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.709 on 19 degrees of freedom
  (6 observations deleted due to missingness)
Multiple R-squared:  0.8506,	Adjusted R-squared:  0.8034 
F-statistic: 18.03 on 6 and 19 DF,  p-value: 6.465e-07
```

Eliminate drat due to highest p value

Now thing to observe again, suddenly wt feature becomes significant (see star mark)

Step 6

```
> mtcars$drat=NULL
> MultiLinearReg = lm(mpg ~ ., data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.9829 -1.5504 -0.1811  0.9476  5.5610 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 17.27415   10.85202   1.592   0.1264  
hp          -0.01784    0.01683  -1.060   0.3011  
wt          -3.04070    1.08588  -2.800   0.0107 *
qsec         0.53597    0.48384   1.108   0.2805  
gear         1.90387    1.17976   1.614   0.1215  
carb        -0.47693    0.64761  -0.736   0.4696  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.597 on 21 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.8538,	Adjusted R-squared:  0.819 
F-statistic: 24.53 on 5 and 21 DF,  p-value: 4.166e-08
```

Now wt has become more significant as p-value reduces further from .0354 to .0107 (see star mark)

Eliminate carb due to highest p value

Step 7

```
> mtcars$carb=NULL
> MultiLinearReg = lm(mpg ~ ., data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.6584 -1.6494 -0.0659  0.6059  5.5750 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) 15.79469   10.13664   1.558   0.1328   
hp          -0.02041    0.01530  -1.334   0.1953   
wt          -3.42940    0.93666  -3.661   0.0013 **
qsec         0.71294    0.44205   1.613   0.1204   
gear         1.51020    0.88816   1.700   0.1025   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.542 on 23 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.8498,	Adjusted R-squared:  0.8237 
F-statistic: 32.53 on 4 and 23 DF,  p-value: 3.663e-09
```

Eliminate hp due to highest p value

Now wt has become even more significant as p-value reduces further from .0107 to .0013 (see two stars)

Step 8

```
> mtcars$hp=NULL
> MultiLinearReg = lm(mpg ~ ., data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
   Min     1Q Median     3Q    Max 
-3.791 -1.526 -0.258  1.056  5.683 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   8.5652     8.7039   0.984 0.334900    
wt           -4.3162     0.6704  -6.438 1.17e-06 ***
qsec          1.1586     0.2941   3.940 0.000614 ***
gear          1.2723     0.8841   1.439 0.163035    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.583 on 24 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.8382,	Adjusted R-squared:  0.818 
F-statistic: 41.44 on 3 and 24 DF,  p-value: 1.197e-09
```

Now “wt” and “qsec” both have become significant as p-value is very less from .05 (see three stars against both)

Eliminate “gear” due to highest p value

Step 9

```
> mtcars$gear=NULL
> MultiLinearReg = lm(mpg ~ ., data = mtcars)
> summary(MultiLinearReg)

Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.4892 -1.9124 -0.1948  1.4334  5.8729 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  19.0350     5.2788   3.606  0.00124 ** 
wt           -5.0834     0.4870 -10.438 5.61e-11 ***
qsec          0.9737     0.2669   3.648  0.00111 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.592 on 27 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.8379,	Adjusted R-squared:  0.8259 
F-statistic:  69.8 on 2 and 27 DF,  p-value: 2.143e-11
```

Now intercept has also become significant along with “wt” and “qsec”. So wt and qsec are the final features we got after backward elimination

At the end we got {wt, qsec} as smallest set of features. So now lets see the interesting thing, here whether backward elimination produces the same set of features which we got using Stepwise regression. Using Stepwise regression we have got {wt, cyl} as the best possible smallest set of features.

One more thing we can conclude that it is not always true that we will get same set of features with all the feature selection techniques. We have to select different techniques smartly based on the business problem and our understanding.

So that's all about these three feature selection techniques. There are other techniques which are also equally important to understand, those I will be writing in my upcoming posts.

If you like my articles, I would love to connect with you guys.

Follow me on my Blog: https://ashutoshtripathi.com

Join me on LinkedIn: https://www.linkedin.com/in/ashutoshtripathi1/

For Infographics on ML & DS topics please follow me on Instagram:

Instagram Username: ashutosh_ai

Thank You