Mid-term Multiple Linear Regression

Arvind G E

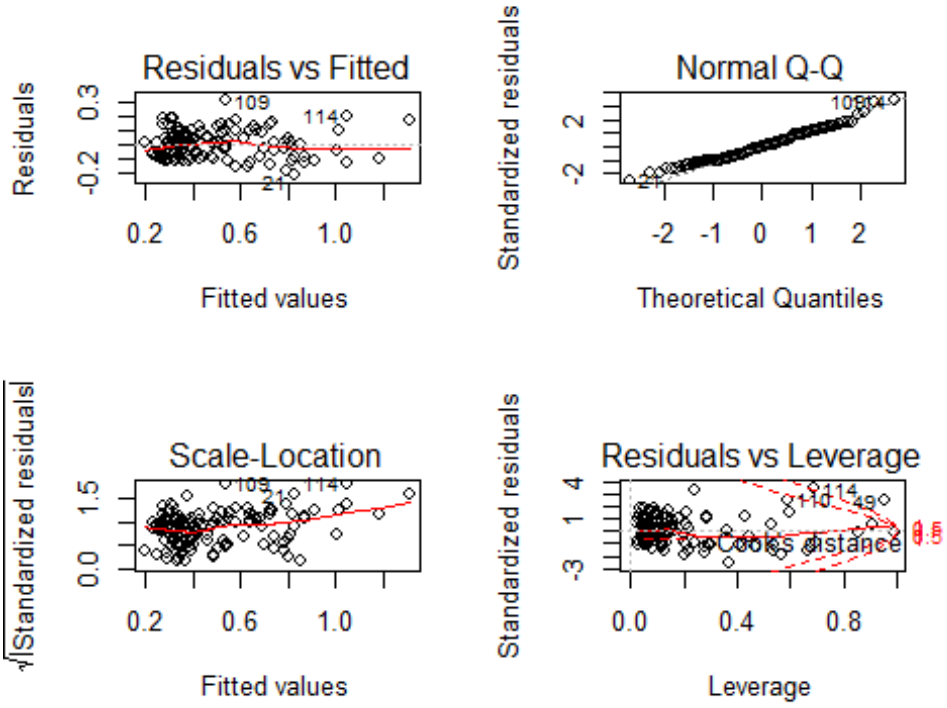
October 10, 2017

```
# Loading Given Data from Excel
library(readxl)

## Warning: package 'readxl' was built under R version 3.4.2

library(boot)
Training_set <- read_excel("E:/MS Study materials/613 Data Analysis/613 mid-
term/Training set.xlsx")
names(Training_set)

##  [1] "Ra"              "feed_rate"       "wheel_speed"
##  [4] "work_speed"      "peak_power"      "mean_power"
##  [7] "std_power"       "skewness_power"  "kurtosis_power"
## [10] "p2p_power"       "peak_mag_an"     "mean_an"
## [13] "skewness_an"     "kurtosis_an"     "total_energy_an"
## [16] "energy_band1_an" "energy_band2_an" "peak_mag_at"
## [19] "mean_at"         "skewness_at"     "kurtosis_at"
## [22] "total_energy_at" "energy_band1_at" "energy_band2_at"

# Model 1
# Fitting the response variable using all predictors to find their
significance and outlier points:
lm.fit = lm(Ra~.,data = Training_set)
summary(lm.fit)

##
## Call:
## lm(formula = Ra ~ ., data = Training_set)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.213340 -0.082835 -0.001831  0.080904  0.310042
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.144e-01  1.171e-01   1.831 0.069758 .
## feed_rate        1.750e-02  6.356e-02   0.275 0.783620
## wheel_speed      3.106e-03  1.741e-03   1.784 0.077137 .
## work_speed      -1.435e-04  1.209e-04  -1.187 0.237922
## peak_power      -5.642e-01  1.565e-01  -3.606 0.000467 ***
## mean_power       3.137e-01  1.451e-01   2.163 0.032704 *
## std_power        5.142e-01  2.887e-01   1.781 0.077643 .
## skewness_power   3.445e-01  1.411e-01   2.442 0.016167 *
```

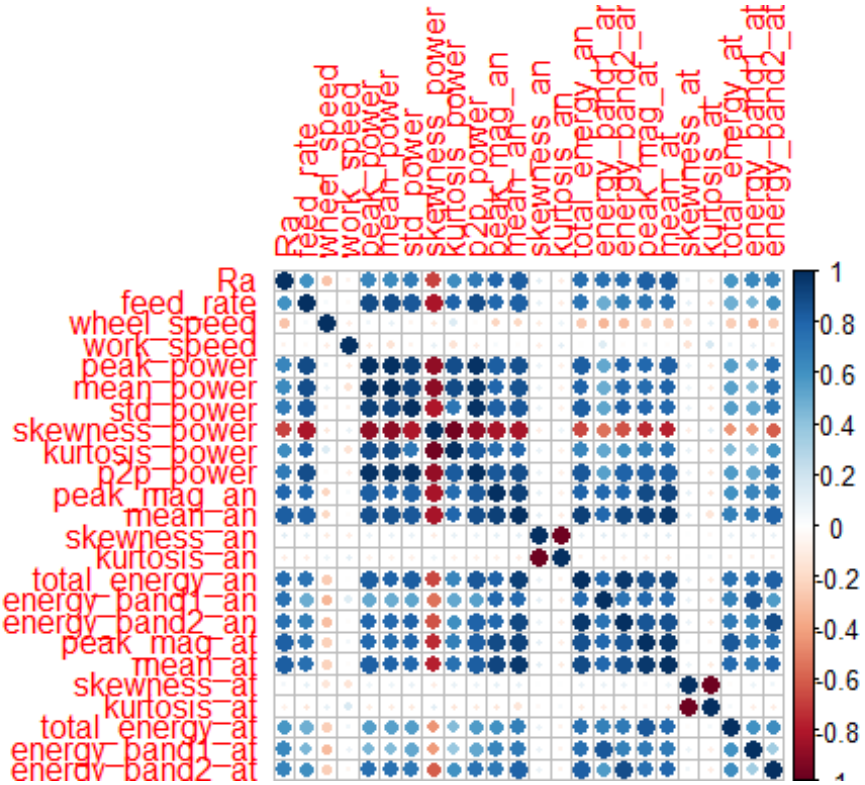
```
## kurtosis_power   1.402e-01  5.790e-02   2.422 0.017047 *
## p2p_power        1.898e-01  9.065e-02   2.093 0.038608 *
## peak_mag_an     -4.181e-02  1.714e-02  -2.440 0.016289 *
## mean_an          3.147e-01  1.558e-01   2.019 0.045847 *
## skewness_an     -2.272e-02  4.530e-02  -0.501 0.617016
## kurtosis_an     -6.761e-04  1.752e-03  -0.386 0.700307
## total_energy_an -7.308e-12  3.040e-12  -2.404 0.017865 *
## energy_band1_an  2.496e-11  1.653e-11   1.510 0.133772
## energy_band2_an  5.724e-11  4.310e-11   1.328 0.186877
## peak_mag_at      3.140e-02  3.030e-02   1.036 0.302248
## mean_at          5.083e-01  2.417e-01   2.103 0.037706 *
## skewness_at      8.396e-02  6.123e-02   1.371 0.173058
## kurtosis_at      9.153e-03  5.491e-03   1.667 0.098356 .
## total_energy_at -8.702e-12  4.093e-12  -2.126 0.035701 *
## energy_band1_at  3.888e-12  3.440e-11   0.113 0.910224
## energy_band2_at  1.545e-11  3.637e-11   0.425 0.671773
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1074 on 111 degrees of freedom
## Multiple R-squared:  0.8368, Adjusted R-squared:  0.8029
## F-statistic: 24.74 on 23 and 111 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2))
plot(lm.fit)
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```




```
# Checking which predictors have a correlation with the response and which
predictors have a high correlation amongst themselves
library(corrplot)

## Warning: package 'corrplot' was built under R version 3.4.2

corrplot(cor(data.frame(Training_set)))
```



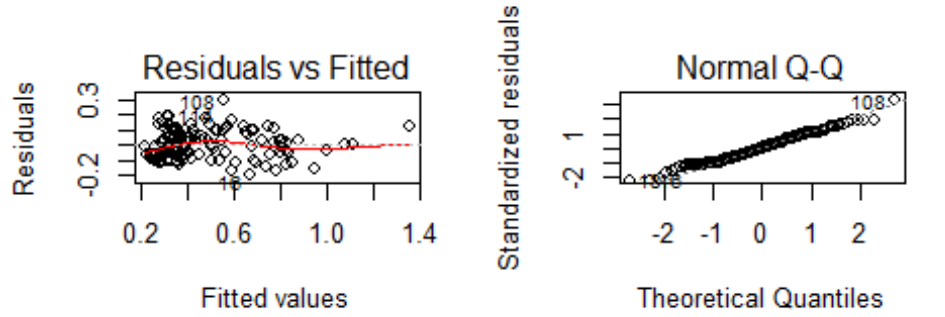

```
# Checking Variation Inflation Factors:
library(car)

##
## Attaching package: 'car'

## The following object is masked from 'package:boot':
##
##     logit

vif(lm(Ra~.,data = Training_set))

##       feed_rate     wheel_speed      work_speed      peak_power
##        6.871219        3.127190        1.403378     1842.728282
##      mean_power       std_power  skewness_power  kurtosis_power
##     1197.719034      246.476243      102.257820      136.967761
##       p2p_power     peak_mag_an         mean_an     skewness_an
##      507.961888       37.693134      173.249407       21.656602
##     kurtosis_an total_energy_an energy_band1_an energy_band2_an
##       21.669448       99.360849       20.102023       43.232926
##     peak_mag_at         mean_at     skewness_at     kurtosis_at
##       49.335228       89.727256       38.051326       38.576425
## total_energy_at energy_band1_at energy_band2_at
##       13.788874       19.940620       23.676798

# Removing data points (49,114) because they lie outside Cook's distance in
'Residuals Vs Leverage' plot
Training_set = Training_set[-c(49,114),]
```

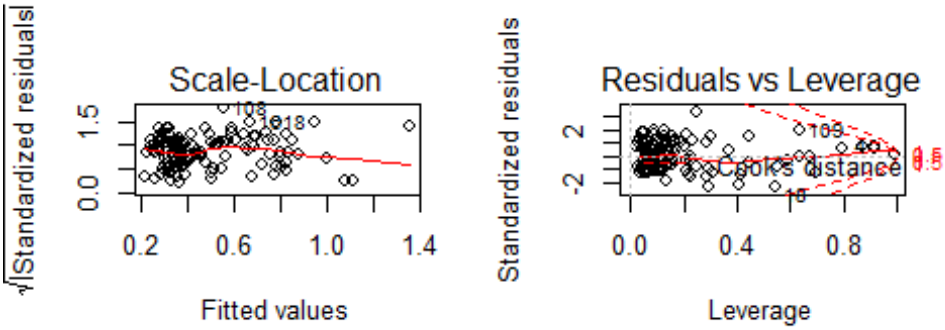
```r
lm.fit = lm(Ra~.,data = Training_set)
summary(lm.fit)
```

```
##
## Call:
## lm(formula = Ra ~ ., data = Training_set)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.194950 -0.073721 -0.003603  0.073001  0.289506
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.469e-01  1.135e-01   2.176 0.031731 *
## feed_rate        2.835e-02  6.015e-02   0.471 0.638360
## wheel_speed      3.351e-03  1.651e-03   2.029 0.044872 *
## work_speed      -1.496e-04  1.157e-04  -1.293 0.198605
## peak_power      -5.939e-01  1.482e-01  -4.008 0.000112 ***
## mean_power       3.435e-01  1.373e-01   2.503 0.013807 *
## std_power        4.444e-01  2.818e-01   1.577 0.117682
## skewness_power   3.103e-01  1.340e-01   2.315 0.022510 *
## kurtosis_power   1.291e-01  5.512e-02   2.343 0.020957 *
## p2p_power        1.983e-01  8.784e-02   2.258 0.025968 *
## peak_mag_an     -2.750e-02  1.664e-02  -1.652 0.101399
## mean_an          5.062e-02  2.288e-01   0.221 0.825309
## skewness_an     -1.632e-02  4.284e-02  -0.381 0.704076
## kurtosis_an     -4.880e-04  1.657e-03  -0.295 0.768872
## total_energy_an -1.834e-12  3.931e-12  -0.467 0.641726
## energy_band1_an  1.155e-11  1.686e-11   0.685 0.494729
## energy_band2_an -5.865e-11  5.497e-11  -1.067 0.288338
## peak_mag_at      2.950e-02  2.879e-02   1.025 0.307760
## mean_at          9.290e-01  4.338e-01   2.141 0.034472 *
## skewness_at      5.328e-02  5.841e-02   0.912 0.363662
## kurtosis_at      6.748e-03  5.224e-03   1.292 0.199134
## total_energy_at -2.253e-11  1.054e-11  -2.138 0.034777 *
## energy_band1_at  3.596e-11  3.761e-11   0.956 0.341064
## energy_band2_at  1.111e-10  5.025e-11   2.212 0.029081 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1014 on 109 degrees of freedom
## Multiple R-squared:  0.8453, Adjusted R-squared:  0.8126
## F-statistic: 25.89 on 23 and 109 DF,  p-value: < 2.2e-16
```

```r
par(mfrow=c(2,2))
plot(lm.fit)
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



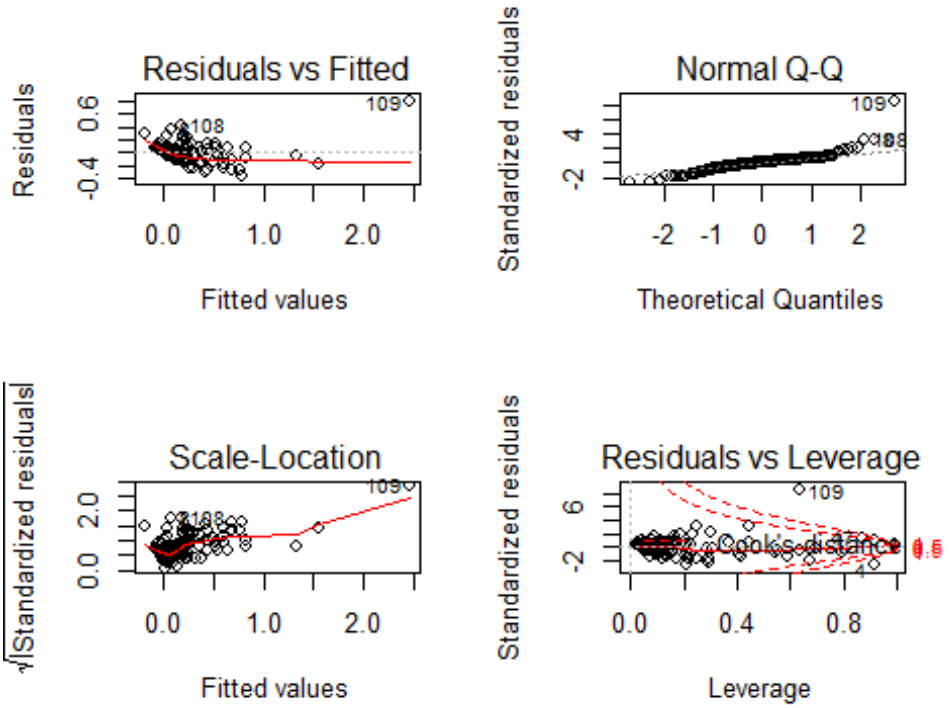



```r
# Model 2
# Response Ra seems to have a non-linear relationship with variables, so I am
lm.fit = lm(Ra^3~.,data = Training_set)
summary(lm.fit)

##
## Call:
## lm(formula = Ra^3 ~ ., data = Training_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35341 -0.06208  0.01316  0.05868  0.78060
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.283e-01  1.732e-01   1.895 0.060695 .
## feed_rate        3.179e-01  9.181e-02   3.463 0.000764 ***
## wheel_speed      7.583e-03  2.521e-03   3.008 0.003262 **
## work_speed       2.359e-04  1.766e-04   1.336 0.184360
## peak_power      -1.350e+00  2.262e-01  -5.970 3.01e-08 ***
## mean_power       9.547e-01  2.095e-01   4.557 1.36e-05 ***
## std_power       -2.522e-01  4.301e-01  -0.587 0.558751
## skewness_power  -8.729e-03  2.046e-01  -0.043 0.966049
## kurtosis_power  -5.299e-02  8.413e-02  -0.630 0.530108
## p2p_power        5.356e-01  1.341e-01   3.994 0.000118 ***
## peak_mag_an     -5.696e-02  2.540e-02  -2.242 0.026961 *
## mean_an          3.491e-01  3.493e-01   0.999 0.319772
```

```
## skewness_an     -1.031e-01  6.540e-02  -1.577 0.117733
## kurtosis_an     -3.860e-03  2.529e-03  -1.527 0.129769
## total_energy_an  4.063e-12  6.001e-12   0.677 0.499817
## energy_band1_an -5.245e-11  2.574e-11  -2.038 0.044004 *
## energy_band2_an -4.345e-10  8.390e-11  -5.179 1.03e-06 ***
## peak_mag_at      1.684e-01  4.395e-02   3.833 0.000212 ***
## mean_at         -1.114e-01  6.622e-01  -0.168 0.866701
## skewness_at      1.369e-02  8.916e-02   0.154 0.878271
## kurtosis_at      9.614e-04  7.973e-03   0.121 0.904251
## total_energy_at -2.808e-11  1.609e-11  -1.745 0.083766 .
## energy_band1_at  2.328e-10  5.740e-11   4.056 9.42e-05 ***
## energy_band2_at  5.604e-10  7.671e-11   7.305 4.82e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1548 on 109 degrees of freedom
## Multiple R-squared:  0.8442, Adjusted R-squared:  0.8113
## F-statistic: 25.67 on 23 and 109 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2))
plot(lm.fit)
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```




```
## The residual plots don't look as good as they were for response 'Ra'
```

```r
# Model 3
## Using cube root values of predictors which had a good significance with Ra
when it was raised to power 3 and using other predictors in their linear
form.
lm.fit = lm(Ra~I(feed_rate^(1/3)) + wheel_speed + work_speed+
I(peak_power^(1/3))  + std_power + skewness_power + kurtosis_power +
p2p_power + log(peak_mag_an) + energy_band2_an + peak_mag_at  + mean_at +
skewness_at + kurtosis_at + total_energy_at + energy_band1_at  +
I(energy_band2_at^(1/3)), data = Training_set)
summary(lm.fit)

##
## Call:
## lm(formula = Ra ~ I(feed_rate^(1/3)) + wheel_speed + work_speed +
##     I(peak_power^(1/3)) + std_power + skewness_power + kurtosis_power +
##     p2p_power + log(peak_mag_an) + energy_band2_an + peak_mag_at +
##     mean_at + skewness_at + kurtosis_at + total_energy_at +
energy_band1_at +
##     I(energy_band2_at^(1/3)), data = Training_set)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.228571 -0.065777 -0.000098  0.067026  0.314268
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.844e+00  2.546e-01   7.242 5.41e-11 ***
## I(feed_rate^(1/3))         3.045e-01  7.892e-02   3.858 0.000189 ***
## wheel_speed                6.704e-03  1.897e-03   3.535 0.000589 ***
## work_speed                -1.004e-04  1.144e-04  -0.878 0.381992
## I(peak_power^(1/3))       -1.889e+00  2.695e-01  -7.008 1.76e-10 ***
## std_power                  4.211e-01  2.598e-01   1.621 0.107717
## skewness_power            -4.404e-02  1.095e-01  -0.402 0.688162
## kurtosis_power             5.307e-02  4.393e-02   1.208 0.229515
## p2p_power                  3.550e-02  7.079e-02   0.502 0.616972
## log(peak_mag_an)          -6.181e-02  2.813e-02  -2.197 0.030025 *
## energy_band2_an           -4.225e-12  2.610e-11  -0.162 0.871662
## peak_mag_at                3.166e-02  2.769e-02   1.143 0.255276
## mean_at                    1.028e+00  1.831e-01   5.616 1.38e-07 ***
## skewness_at                1.197e-02  5.523e-02   0.217 0.828821
## kurtosis_at                3.068e-03  4.913e-03   0.624 0.533601
## total_energy_at           -2.332e-11  5.790e-12  -4.027 0.000102 ***
## energy_band1_at            6.753e-12  1.559e-11   0.433 0.665715
## I(energy_band2_at^(1/3))   1.841e-04  7.908e-05   2.328 0.021648 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1026 on 115 degrees of freedom
## Multiple R-squared:  0.8329, Adjusted R-squared:  0.8082
## F-statistic: 33.72 on 17 and 115 DF,  p-value: < 2.2e-16
```

```r
# Model 4
## There are many predictors with very low significance. So removing few
variables with low significance.
lm.fit = lm(Ra~I(feed_rate^(1/3)) + wheel_speed + work_speed+
I(peak_power^(1/3))  + std_power +  kurtosis_power + log(peak_mag_an) +
peak_mag_at  + mean_at + skewness_at + kurtosis_at + total_energy_at +
energy_band1_at  + I(energy_band2_at^(1/3)), data = Training_set)
summary(lm.fit)

##
## Call:
## lm(formula = Ra ~ I(feed_rate^(1/3)) + wheel_speed + work_speed +
##     I(peak_power^(1/3)) + std_power + kurtosis_power + log(peak_mag_an) +
##     peak_mag_at + mean_at + skewness_at + kurtosis_at + total_energy_at +
##     energy_band1_at + I(energy_band2_at^(1/3)), data = Training_set)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.236971 -0.067678 -0.000601  0.063619  0.313263
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.788e+00  1.949e-01   9.175 1.79e-15 ***
## I(feed_rate^(1/3))          3.007e-01  7.540e-02   3.988 0.000116 ***
## wheel_speed                 6.690e-03  1.688e-03   3.962 0.000127 ***
## work_speed                 -1.010e-04  1.093e-04  -0.925 0.356958
## I(peak_power^(1/3))        -1.856e+00  2.325e-01  -7.981 1.06e-12 ***
## std_power                   5.494e-01  8.051e-02   6.823 4.07e-10 ***
## kurtosis_power              7.407e-02  1.288e-02   5.750 7.13e-08 ***
## log(peak_mag_an)           -5.952e-02  2.535e-02  -2.348 0.020514 *
## peak_mag_at                 3.290e-02  2.565e-02   1.282 0.202226
## mean_at                     1.023e+00  1.781e-01   5.743 7.37e-08 ***
## skewness_at                 7.281e-03  5.327e-02   0.137 0.891514
## kurtosis_at                 2.704e-03  4.754e-03   0.569 0.570564
## total_energy_at            -2.380e-11  4.793e-12  -4.965 2.34e-06 ***
## energy_band1_at             7.457e-12  1.396e-11   0.534 0.594281
## I(energy_band2_at^(1/3))    1.756e-04  5.422e-05   3.238 0.001561 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1014 on 118 degrees of freedom
## Multiple R-squared:  0.8325, Adjusted R-squared:  0.8126
## F-statistic: 41.88 on 14 and 118 DF,  p-value: < 2.2e-16

# Checking Variation Inflation Factor:
vif(lm(Ra~I(feed_rate^(1/3)) + wheel_speed + work_speed+ I(peak_power^(1/3))
+ std_power +  kurtosis_power + log(peak_mag_an) + peak_mag_at  + mean_at +
skewness_at + kurtosis_at + total_energy_at + energy_band1_at  +
I(energy_band2_at^(1/3)), data = Training_set))
```

```
##       I(feed_rate^(1/3))                wheel_speed                 work_speed
##                 9.474610                   3.240723                   1.261894
##      I(peak_power^(1/3))                  std_power             kurtosis_power
##                69.392437                  20.004840                   7.486699
##         log(peak_mag_an)                peak_mag_at                    mean_at
##                10.474366                  35.985843                  52.420113
##              skewness_at                kurtosis_at           total_energy_at
##                32.244848                  32.100877                  13.845259
##          energy_band1_at I(energy_band2_at^(1/3))
##                 2.964327                   7.366284
```

```
#Model 5:
## Further removal of insignificant predictors and variables with high
Variation Inflation Factor.
lm.fit = lm(Ra~I(feed_rate^(1/3)) + wheel_speed + work_speed+
I(peak_power^(1/3))  + std_power +  kurtosis_power + log(peak_mag_an) +
peak_mag_at  + mean_at + skewness_at + kurtosis_at + total_energy_at +
energy_band1_at  + I(energy_band2_at^(1/3)), data = Training_set)
summary(lm.fit)
```

```
##
## Call:
## lm(formula = Ra ~ I(feed_rate^(1/3)) + wheel_speed + work_speed +
##     I(peak_power^(1/3)) + std_power + kurtosis_power + log(peak_mag_an) +
##     peak_mag_at + mean_at + skewness_at + kurtosis_at + total_energy_at +
##     energy_band1_at + I(energy_band2_at^(1/3)), data = Training_set)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.236971 -0.067678 -0.000601  0.063619  0.313263
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.788e+00  1.949e-01   9.175 1.79e-15 ***
## I(feed_rate^(1/3))         3.007e-01  7.540e-02   3.988 0.000116 ***
## wheel_speed                6.690e-03  1.688e-03   3.962 0.000127 ***
## work_speed                -1.010e-04  1.093e-04  -0.925 0.356958
## I(peak_power^(1/3))       -1.856e+00  2.325e-01  -7.981 1.06e-12 ***
## std_power                  5.494e-01  8.051e-02   6.823 4.07e-10 ***
## kurtosis_power             7.407e-02  1.288e-02   5.750 7.13e-08 ***
## log(peak_mag_an)          -5.952e-02  2.535e-02  -2.348 0.020514 *
## peak_mag_at                3.290e-02  2.565e-02   1.282 0.202226
## mean_at                    1.023e+00  1.781e-01   5.743 7.37e-08 ***
## skewness_at                7.281e-03  5.327e-02   0.137 0.891514
## kurtosis_at                2.704e-03  4.754e-03   0.569 0.570564
## total_energy_at           -2.380e-11  4.793e-12  -4.965 2.34e-06 ***
## energy_band1_at            7.457e-12  1.396e-11   0.534 0.594281
## I(energy_band2_at^(1/3))   1.756e-04  5.422e-05   3.238 0.001561 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.1014 on 118 degrees of freedom
## Multiple R-squared:  0.8325, Adjusted R-squared:  0.8126
## F-statistic: 41.88 on 14 and 118 DF,  p-value: < 2.2e-16
```

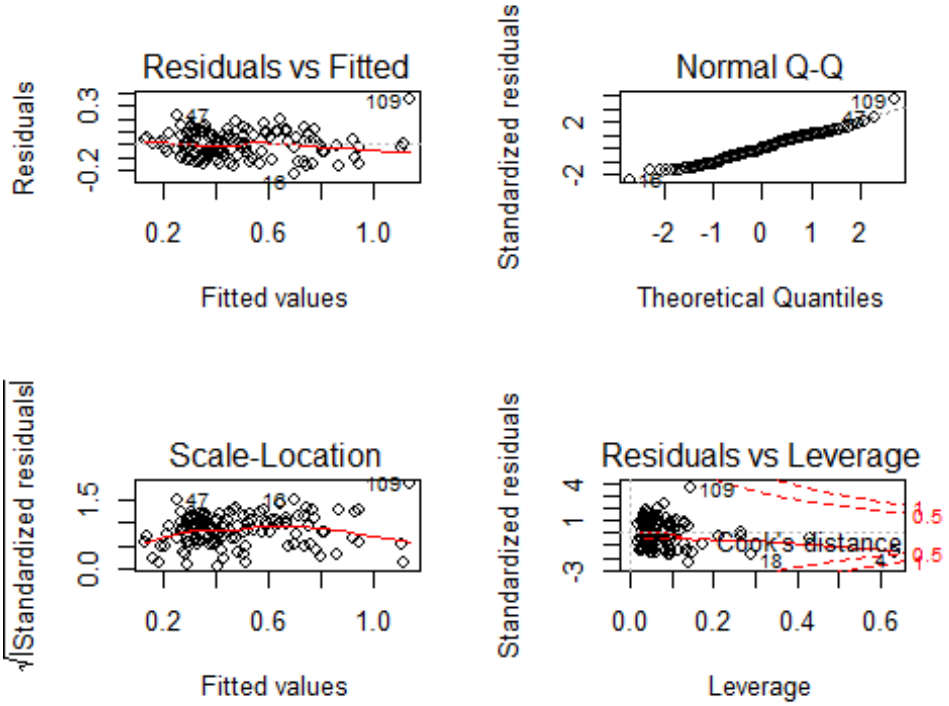
```
vif(lm(Ra~I(feed_rate^(1/3)) + wheel_speed + work_speed+ I(peak_power^(1/3))
+ std_power +  kurtosis_power + log(peak_mag_an) + peak_mag_at  + mean_at +
skewness_at + kurtosis_at + total_energy_at + energy_band1_at  +
I(energy_band2_at^(1/3)), data = Training_set))
```

```
##       I(feed_rate^(1/3))              wheel_speed               work_speed
##                 9.474610                 3.240723                 1.261894
##      I(peak_power^(1/3))                std_power           kurtosis_power
##                69.392437                20.004840                 7.486699
##         log(peak_mag_an)              peak_mag_at                  mean_at
##                10.474366                35.985843                52.420113
##              skewness_at              kurtosis_at          total_energy_at
##                32.244848                32.100877                13.845259
##          energy_band1_at I(energy_band2_at^(1/3))
##                 2.964327                 7.366284
```

```
# Model 6:
##Removing few more predictors, all the current predictors are significant.
lm.fit = lm(Ra~I(feed_rate^(1/3)) + wheel_speed + I(peak_power^(1/3))  +
std_power + kurtosis_power + log(peak_mag_an) + mean_at + total_energy_at  +
I(energy_band2_at^(1/3)), data = Training_set)
summary(lm.fit)
```

```
##
## Call:
## lm(formula = Ra ~ I(feed_rate^(1/3)) + wheel_speed + I(peak_power^(1/3)) +
##     std_power + kurtosis_power + log(peak_mag_an) + mean_at +
##     total_energy_at + I(energy_band2_at^(1/3)), data = Training_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.22472 -0.06934 -0.00145  0.07309  0.34601
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.826e+00  1.760e-01  10.376  < 2e-16 ***
## I(feed_rate^(1/3))        2.868e-01  7.499e-02   3.825 0.000207 ***
## wheel_speed               7.289e-03  1.672e-03   4.359 2.73e-05 ***
## I(peak_power^(1/3))      -1.902e+00  2.214e-01  -8.593 3.20e-14 ***
## std_power                 5.815e-01  7.247e-02   8.024 6.89e-13 ***
## kurtosis_power            7.440e-02  1.270e-02   5.859 3.98e-08 ***
## log(peak_mag_an)         -4.420e-02  2.197e-02  -2.011 0.046471 *
## mean_at                   1.112e+00  1.410e-01   7.884 1.45e-12 ***
## total_energy_at          -1.957e-11  4.468e-12  -4.380 2.52e-05 ***
## I(energy_band2_at^(1/3))  1.819e-04  4.963e-05   3.665 0.000366 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.103 on 123 degrees of freedom
## Multiple R-squared:  0.8201, Adjusted R-squared:  0.807
## F-statistic: 62.31 on 9 and 123 DF,  p-value: < 2.2e-16

cv.error.10 = cv.glm(Training_set,lm.fit,K=5)$delta[1]

par(mfrow=c(2,2))
plot(lm.fit)
```




```
## Plots look fairly good

vif(lm(Ra~I(feed_rate^(1/3)) + wheel_speed + I(peak_power^(1/3))  + std_power
+ kurtosis_power + log(peak_mag_an) + mean_at + total_energy_at  +
I(energy_band2_at^(1/3)), data = Training_set))

##       I(feed_rate^(1/3))              wheel_speed      I(peak_power^(1/3))
##                 9.096624                 3.086582                61.067302
##                std_power           kurtosis_power         log(peak_mag_an)
##                15.736620                 7.063098                 7.642553
##                  mean_at         total_energy_at I(energy_band2_at^(1/3))
##                31.903211                11.681314                 5.991241

# VIF value has come down. It's still high. But removal of more variables
than this reduces the R squared considerably.
```

```r
# Cross Validating:
lm.fit = glm(Ra~I(feed_rate^(1/3)) + wheel_speed + I(peak_power^(1/3))  +
std_power + kurtosis_power + log(peak_mag_an) + mean_at + total_energy_at  +
I(energy_band2_at^(1/3)), data = Training_set)
summary(lm.fit)
```

```
##
## Call:
## glm(formula = Ra ~ I(feed_rate^(1/3)) + wheel_speed + I(peak_power^(1/3))
+
##     std_power + kurtosis_power + log(peak_mag_an) + mean_at +
##     total_energy_at + I(energy_band2_at^(1/3)), data = Training_set)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.22472  -0.06934  -0.00145   0.07309   0.34601
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.826e+00  1.760e-01  10.376  < 2e-16 ***
## I(feed_rate^(1/3))        2.868e-01  7.499e-02   3.825 0.000207 ***
## wheel_speed               7.289e-03  1.672e-03   4.359 2.73e-05 ***
## I(peak_power^(1/3))      -1.902e+00  2.214e-01  -8.593 3.20e-14 ***
## std_power                 5.815e-01  7.247e-02   8.024 6.89e-13 ***
## kurtosis_power            7.440e-02  1.270e-02   5.859 3.98e-08 ***
## log(peak_mag_an)         -4.420e-02  2.197e-02  -2.011 0.046471 *
## mean_at                   1.112e+00  1.410e-01   7.884 1.45e-12 ***
## total_energy_at          -1.957e-11  4.468e-12  -4.380 2.52e-05 ***
## I(energy_band2_at^(1/3))  1.819e-04  4.963e-05   3.665 0.000366 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.01060093)
##
##     Null deviance: 7.2489  on 132  degrees of freedom
## Residual deviance: 1.3039  on 123  degrees of freedom
## AIC: -215.68
##
## Number of Fisher Scoring iterations: 2
```

```r
set.seed(2)
cv.error.10 = cv.glm(Training_set,lm.fit,K=5)$delta[1]
cv.error.10
```

```
## [1] 0.01217624
```

```r
#Model 7
## Comparing another model where I remove one predictor, 'peak_mag_an' and
comparing with cross-validation error
lm.fit = glm(Ra~I(feed_rate^(1/3)) + wheel_speed + I(peak_power^(1/3))  +
std_power + kurtosis_power + mean_at + total_energy_at  +
```

```
I(energy_band2_at^(1/3)), data = Training_set)
set.seed(2)
cv.error.10 = cv.glm(Training_set,lm.fit,K=5)$delta[1]
cv.error.10

## [1] 0.01175294
```