Quadratic Regression Analysis in RStudio

1 Data Set :

Weight	Lenght
1.0	5.29
1.5	6.31
2.0	7.28
2.5	8.33
3.0	9.30
3.5	10.32

So this is the data set which we have analyse using quadratic linear model.So, in RStudio we will read this dataset from the file " Spring.csv " using the command " *data=read.csv("Spring.csv",header=TRUE)* ". Now once system have read the data, we can do our analysis. We will use *dim()* function to find out the dimension of the matrix which is essentially the data-set stored in the form of matrix.

In this figure below we can see how to upload and get dimension of the dataset.

```
> data=read.csv(file.choose(),header=TRUE)
> data
  Weight Length
1    1.0   5.29
2    1.5   6.31
3    2.0   7.28
4    2.5   8.33
5    3.0   9.30
6    3.5  10.32
> dim(data)
[1] 6 2
```

Figure 1

Now we fit the *lm()* function on the dataset, particularly the quadratic regression.
Let's check the below figure here we have used *I()* along the higher powers of explanatory variables.

```
> model=lm(data$Length ~ data$Weight + I((data$Weight)^2))
> model

Call:
lm(formula = data$Length ~ data$Weight + I((data$Weight)^2))

Coefficients:
       (Intercept)         data$Weight  I((data$Weight)^2)
         3.283e+00           2.010e+00           4.679e-16

> model.matrix(model)
  (Intercept) data$Weight I((data$Weight)^2)
1           1         1.0               1.00
2           1         1.5               2.25
3           1         2.0               4.00
4           1         2.5               6.25
5           1         3.0               9.00
6           1         3.5              12.25
attr(,"assign")
[1] 0 1 2
```

Figure 2

So we have used *I()* function within *lm()* function because here we are inside the formulae object i.e., some expression which contain tilde sign '∼' then after this sign all operations doesn't mean there usual meaning and can considered as considered as boolean expression. This means ($weight + weight^2$) will give us the result as ($weight$) something like a boolean.

Let's see this the *lm()* function without the use of *I()* function. What we will get?

```
> modelwrong=lm(data$Length ~ data$Weight + (data$Weight)^2)
> modelwrong

Call:
lm(formula = data$Length ~ data$Weight + (data$Weight)^2)

Coefficients:
(Intercept)  data$Weight
      3.283        2.010

> modellinear=lm(data$Length ~ data$Weight)
> modellinear

Call:
lm(formula = data$Length ~ data$Weight)

Coefficients:
(Intercept)  data$Weight
      3.283        2.010
```

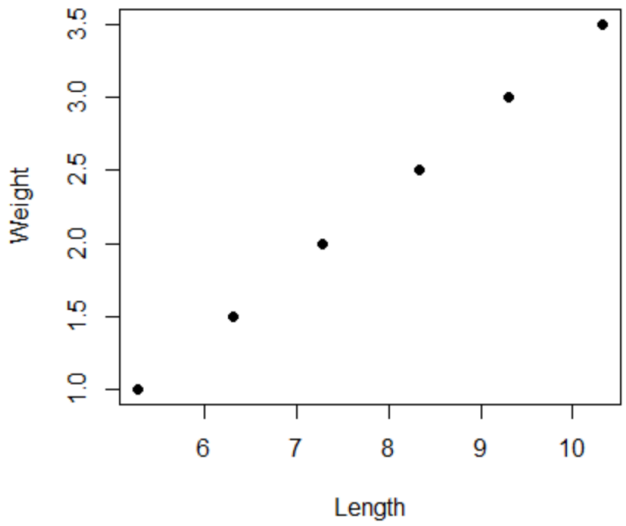
Figure 3

```
> model.matrix(modelwrong)
  (Intercept) data$Weight
1           1         1.0
2           1         1.5
3           1         2.0
4           1         2.5
5           1         3.0
6           1         3.5
attr(,"assign")
[1] 0 1
> model.matrix(modellinear)
  (Intercept) data$Weight
1           1         1.0
2           1         1.5
3           1         2.0
4           1         2.5
5           1         3.0
6           1         3.5
attr(,"assign")
[1] 0 1
```

Figure 4

So here in figure 3, we have denoted lm() function without use of I() function as 'modelwrong'. Here in 'modelwrong' we can see that there is no term of (weight^2) in the coefficients, additionally if we compare the model matrix used in calculations within lm() function of 'model' in figure 2 and other two models of figure 3, we found that there is no (weight^2) column in for 'modelwrong' and 'modellinear', this clarifies what we were talking about boolean kinda expression in earlier paragraph.
Now let's see the coefficients of 'model' function. So, for $weight^2$ we have a coefficient of 4.67*e$^-$16, this means ($weight^2$) is not much significant here. Also from figure 3, we can check 'modellinear' function which shows a linear regression between two variables.
After comparing coeffients of 'model' and 'modellinear' we can say that there is no need to fit a quadratic regression as square term is of no significance in this situation.

Let's see the plot between two variables :



Here we can clearly see that only linear variable can easily explain the relation between weight and length so here in this particular example we don't need quadratic regression but in this way we should conduct our analysis where we should compare the results from different degree of regressions and choose the appropriate model while keeping in mind about over-fitting and under-fitting if we are using Regression analysis.