R Notebook

This is the “Iris” dataset. Originally published at UCI Machine Learning Repository: Iris Data Set, this small dataset from 1936 is often used for testing out machine learning algorithms and visualizations (for example, Scatter Plot). Each row of the table represents an iris flower, including its species and dimensions of its botanical parts, sepal and petal, in centimeters.

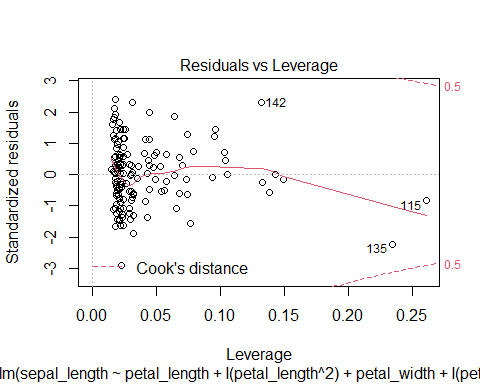
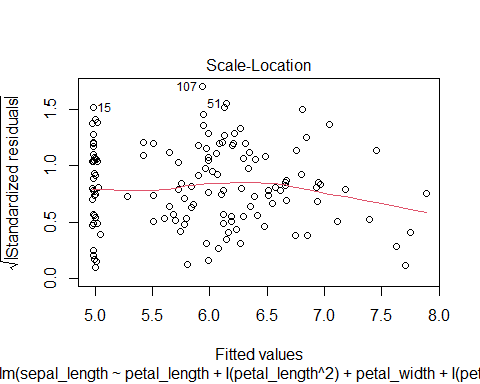
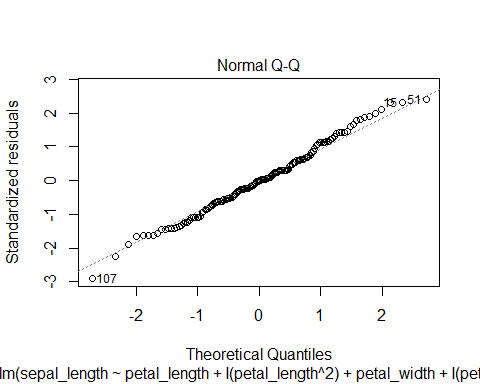
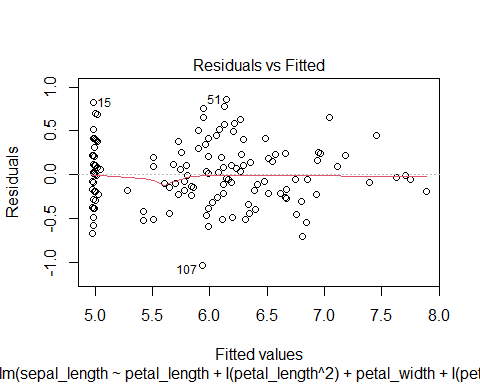
Iris = readr::read\_csv(  
 "https://gist.githubusercontent.com/curran/a08a1080b88344b0c8a7/raw/0e7a9b0a5d22642a06d3d5b9bcbad9890c8ee534/iris.csv", show\_col\_types = FALSE)

Q10: Fit a complete second order model for predicting a sepal\_length based on petal\_length and petal\_width and examine the residuals. Examine the conditions for your model. Discuss any issues that you see in the conditions.

Iris10 = lm(sepal\_length~petal\_length + I(petal\_length^2) + petal\_width + I(petal\_width^2)   
 + petal\_length\*petal\_width, data = Iris)  
summary(Iris10)

##   
## Call:  
## lm(formula = sepal\_length ~ petal\_length + I(petal\_length^2) +   
## petal\_width + I(petal\_width^2) + petal\_length \* petal\_width,   
## data = Iris)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.03541 -0.22444 -0.00044 0.21695 0.85811   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.06985 0.23422 21.646 <2e-16 \*\*\*  
## petal\_length -0.17961 0.30746 -0.584 0.560   
## I(petal\_length^2) 0.07061 0.09779 0.722 0.471   
## petal\_width 0.00520 0.68842 0.008 0.994   
## I(petal\_width^2) -0.27946 0.51092 -0.547 0.585   
## petal\_length:petal\_width 0.13631 0.44288 0.308 0.759   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3613 on 144 degrees of freedom  
## Multiple R-squared: 0.816, Adjusted R-squared: 0.8096   
## F-statistic: 127.7 on 5 and 144 DF, p-value: < 2.2e-16

plot(Iris10)



Overall, the residuals are constantly distributed with no obvious pattern. When fitted value is equal to 5.0, the residuals seem to be more compact.

Linear condition is satisfied since the fitted line is generally linear.

According to the qqplot, the points are generally lying on the fitted line, indicating satisfied normality condition.