**HMW 2**

**SHAFAGH YAZDANI**

**3.1**

**Part a. & b**

We shuffle the data set first and then split the data set to 2 part , data training , data test, and thenrun the leave one out cross validation for KKnn.

**model<-train.kknn(V11~., data\_training,kmax=20, scale=TRUE)**

We shuffle the data set first and then split the data set to 3 part , data training . data validation, data test,

Applying 10 fold cross validation method and then test our best model with the test data we have found that

Using the kknn model the best accuracy with splitting data was

K= 22 with model accuracy of 0.9191919k

Here the output for different K value accuracy:

k= 1 modelaccuracy= 0.8484848

k= 2 modelaccuracy= 0.8484848

k= 3 modelaccuracy= 0.8484848

k= 4 modelaccuracy= 0.8484848

k= 5 modelaccuracy= 0.8989899

k= 6 modelaccuracy= 0.8989899

k= 7 modelaccuracy= 0.8989899

k= 8 modelaccuracy= 0.8989899

k= 9 modelaccuracy= 0.8989899

k= 10 modelaccuracy= 0.9090909

k= 11 modelaccuracy= 0.9090909

k= 12 modelaccuracy= 0.9090909

k= 13 modelaccuracy= 0.9090909

k= 14 modelaccuracy= 0.9090909

k= 15 modelaccuracy= 0.9090909

k= 16 modelaccuracy= 0.9090909

k= 17 modelaccuracy= 0.9090909

k= 18 modelaccuracy= 0.9090909

k= 19 modelaccuracy= 0.9090909

k= 20 modelaccuracy= 0.9090909

k= 21 modelaccuracy= 0.9090909

k= 22 modelaccuracy= 0.9191919

k= 23 modelaccuracy= 0.9191919

k= 24 modelaccuracy= 0.9191919

k= 25 modelaccuracy= 0.9191919

we did the same 10 fold cross validation model for ksvm with different c value

Setting default kernel parameters

c= 1e-06 modelaccuracy= 0.8181818 Setting default kernel parameters

c= 1e-05 modelaccuracy= 0.8181818 Setting default kernel parameters

c= 1e-04 modelaccuracy= 0.8181818 Setting default kernel parameters

c= 0.001 modelaccuracy= 0.9292929 Setting default kernel parameters

c= 0.01 modelaccuracy= 0.9393939 Setting default kernel parameters

c= 0.1 modelaccuracy= 0.9393939 Setting default kernel parameters

c= 1 modelaccuracy= 0.9393939 Setting default kernel parameters

c= 2 modelaccuracy= 0.9393939 Setting default kernel parameters

c= 4 modelaccuracy= 0.9393939 Setting default kernel parameters

c= 6 modelaccuracy= 0.9393939 Setting default kernel parameters

c= 8 modelaccuracy= 0.9393939 Setting default kernel parameters

c= 10 modelaccuracy= 0.9393939 Setting default kernel parameters

c= 50 modelaccuracy= 0.9393939 Setting default kernel parameters

c= 100 modelaccuracy= 0.9393939 Setting default kernel parameters

Question 4.1

One of the usages of clustering models is in healthcare study.

Specially in medicine, when know that response of individual to prescriptions varies a lot. knowing that how each person response to specific prescription can help the physician

and health system alot through optimization of the prescription and diminishing the drugs side effects.

So that each patient’s response can be observed and then the clustering model can be applied to

classify the different patients group based on their response to drug(specific prescriptions).

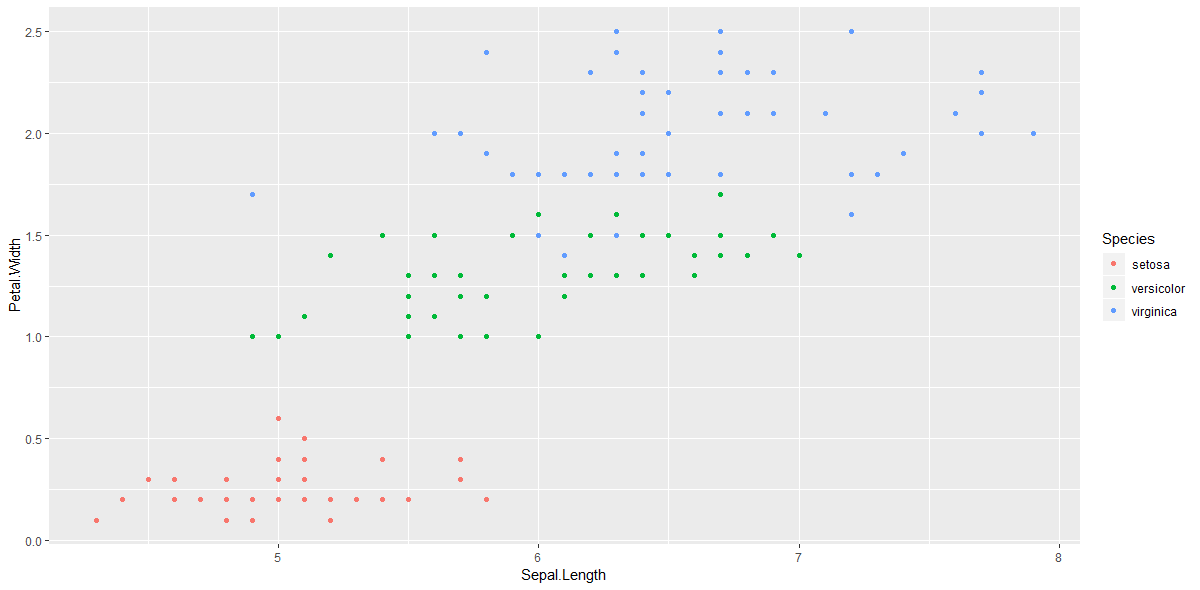
Many predictors can be used in this model. For example, for a drug like Metformin which is used for diabetic patients

predictors like Blood pressures, weight, heart rate, glucose level, mental health, etc can be used for

clustering model of patient’s response to drug.

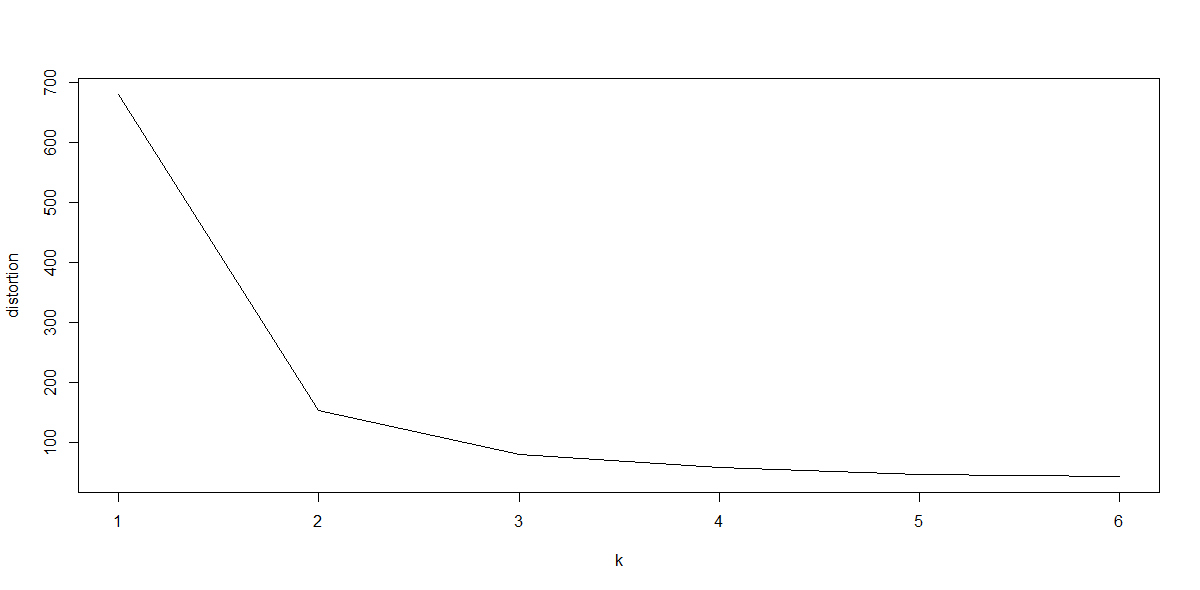
Question 4.2

This is plot shows our data set of iris flowers, we see that 2 species of flower have some overlap.

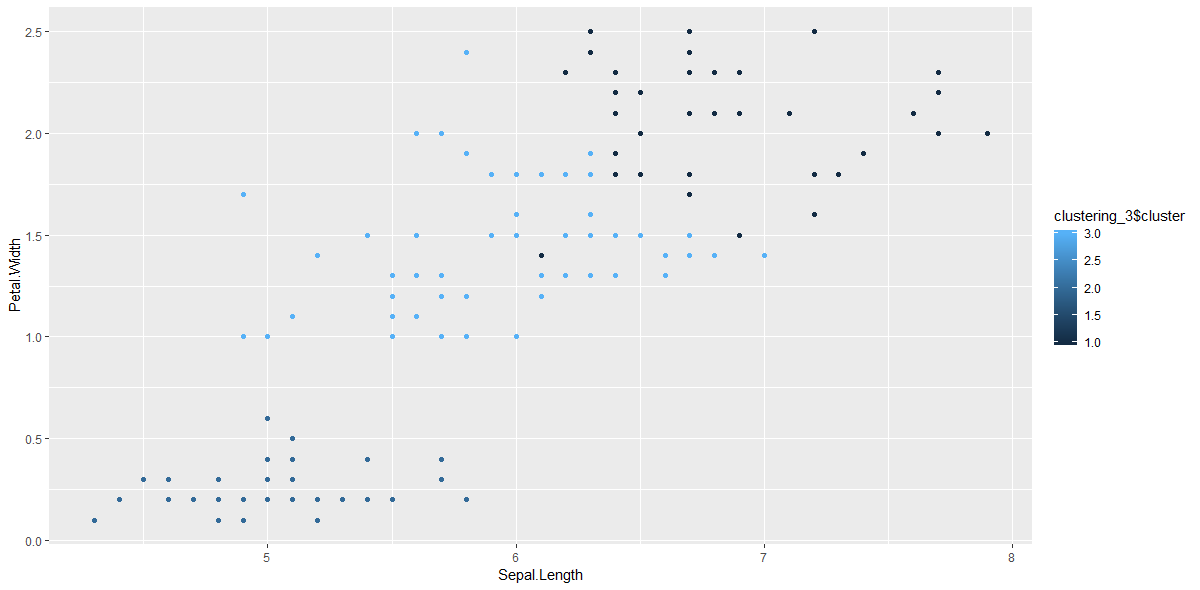


In the elbow diagram of number of k shows that both K=2 and K=3 are good candidate. However, the best k mean clustering model is the K=3. Since the whitin sum of square does not reduce significantly after K=3

We used whitin sum of square(distortion) as metric for choosing our clustering model.



As we see in this two ggplots with 3 clusters and 2 cluster. We notice that 3 clustering model have classified most of these species correctly.



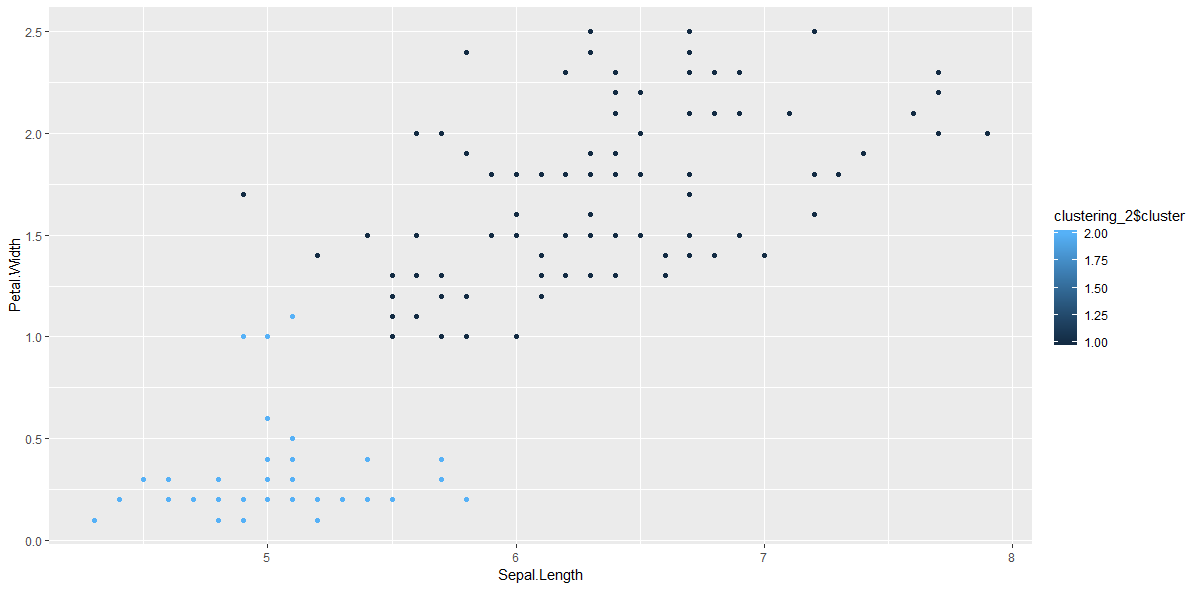


Table for species with 3 cluster shows that most of our species have been classified correctly.

setosa versicolor virginica

1 0 2 36

2 50 0 0

3 0 48 14