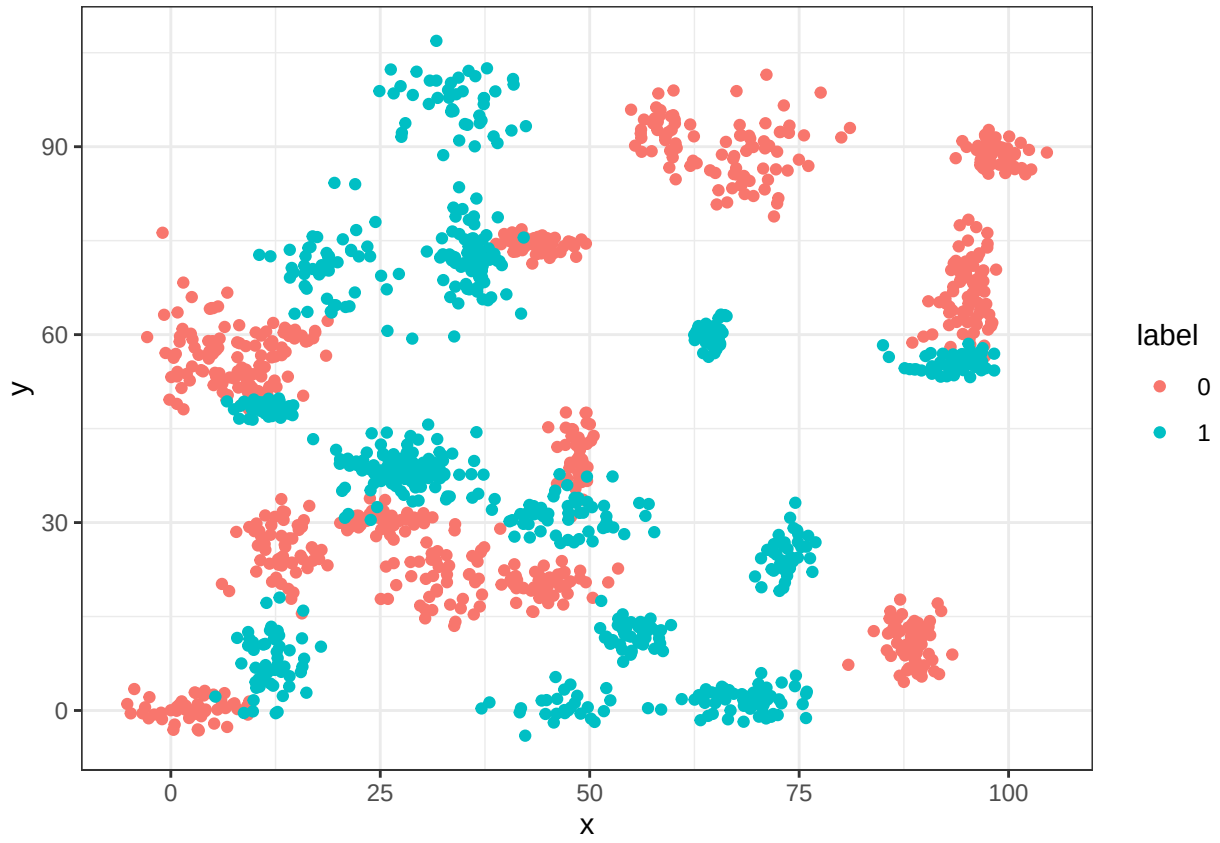
Week11_Classifier

Akila Selvaraj

2/27/2022

Binary Classifier Data

Scatter plot of binary classifier data



Splitting train and test data for binary classifier data

```
#Executing step by step
#k=3
binary_knn <- knn(bin_data.train[,c(2,3)], bin_data.test[,c(2,3)],
                  bin_data.train[,1], k=3, prob=TRUE)
tab <- table(binary_knn,bin_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
```

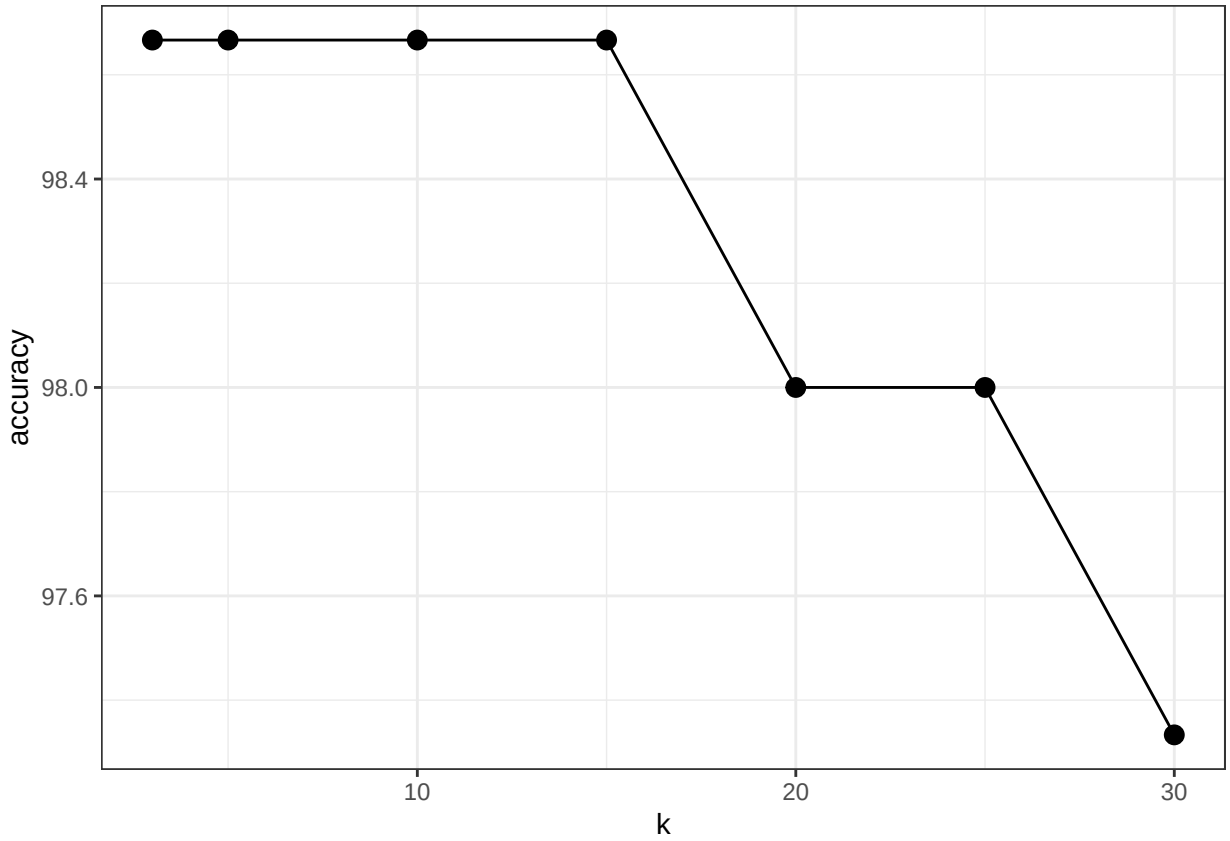
```
accuracy_df <- data.frame (k  = 3, accuracy = accuracy(tab))
#k=5
binary_knn_5 <- knn(bin_data.train[,c(2,3)], bin_data.test[,c(2,3)],
                    bin_data.train[,1], k=5, prob=TRUE)
tab <- table(binary_knn_5,bin_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=5, accuracy = accuracy(tab)))
#k=10
binary_knn_10 <- knn(bin_data.train[,c(2,3)], bin_data.test[,c(2,3)],
                    bin_data.train[,1], k=10, prob=TRUE)
tab <- table(binary_knn_10,bin_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=10, accuracy=accuracy(tab)))
#k=15
binary_knn_15 <- knn(bin_data.train[,c(2,3)], bin_data.test[,c(2,3)],
                    bin_data.train[,1], k=15, prob=TRUE)
tab_15 <- table(binary_knn_15,bin_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=15, accuracy=accuracy(tab)))
#k=20
binary_knn <- knn(bin_data.train[,c(2,3)], bin_data.test[,c(2,3)],
                    bin_data.train[,1], k=20, prob=TRUE)
tab <- table(binary_knn,bin_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=20, accuracy=accuracy(tab)))
#k=25
binary_knn <- knn(bin_data.train[,c(2,3)], bin_data.test[,c(2,3)],
                    bin_data.train[,1], k=25, prob=TRUE)
tab <- table(binary_knn,bin_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=25, accuracy=accuracy(tab)))
#k=30
binary_knn_30 <- knn(bin_data.train[,c(2,3)], bin_data.test[,c(2,3)],
                    bin_data.train[,1], k=30, prob=TRUE)
tab <- table(binary_knn_30,bin_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=30, accuracy=accuracy(tab)))
```

K vs Accuracy data

```
##    k accuracy
## 1  3 98.66667
## 2  5 98.66667
## 3 10 98.66667
## 4 15 98.66667
## 5 20 98.00000
## 6 25 98.00000
## 7 30 97.33333
```

Plot K vs accuracy

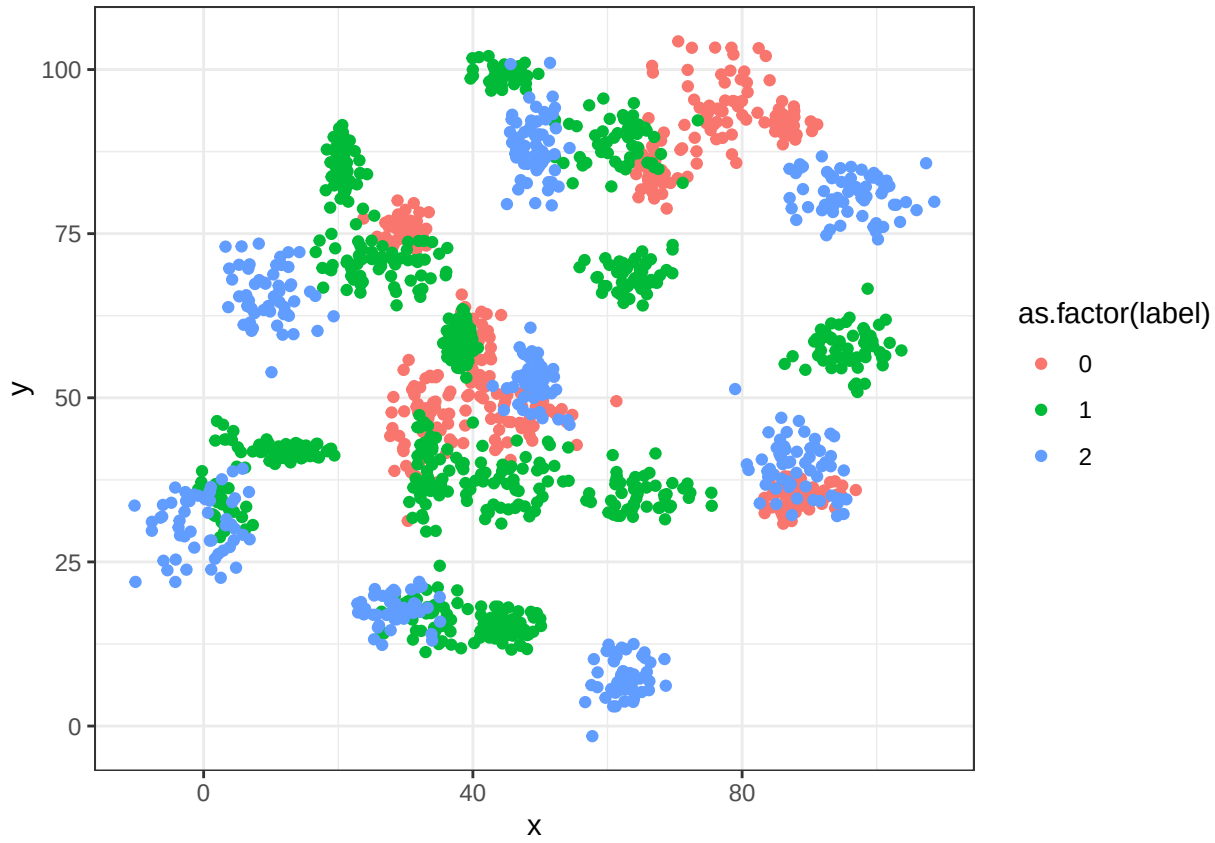
```
ggplot(data = accuracy_df, aes(x = k, y = accuracy)) +  geom_point(size=3) +   theme_bw() + geom_line()
```



Trinary classifier data

Scatter plot of trinary classifier data

```
## [1] "C:/Users/akila/Documents/Data Science/R_projects"
```



Splitting train and test data for binary classifier data

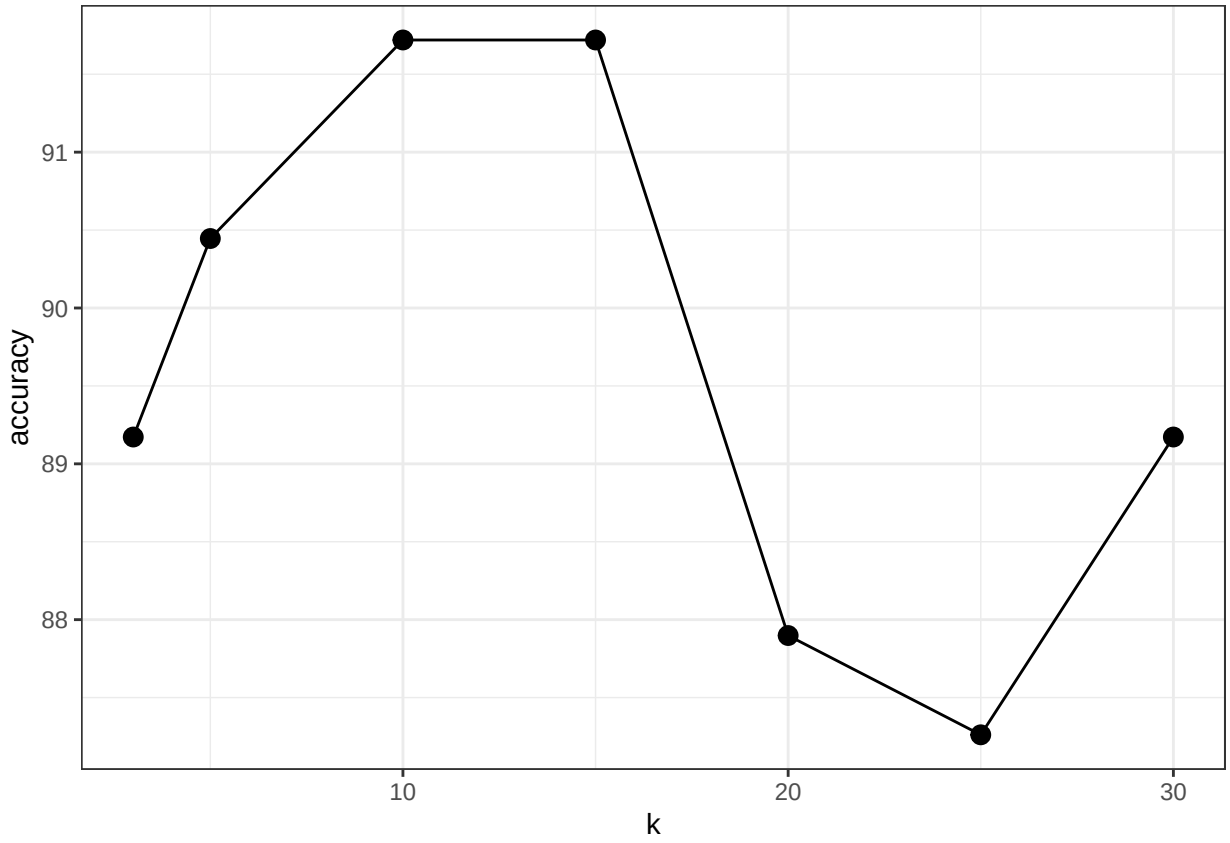
#Executing step by step by changing k values

```
#k=3
trinary_knn <- knn(tri_data.train[,c(2,3)], tri_data.test[,c(2,3)],
                tri_data.train[,1], k=3, prob=TRUE)
tab <- table(trinary_knn,tri_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- data.frame (k  = 3, accuracy = accuracy(tab))
#k=5
trinary_knn <- knn(tri_data.train[,c(2,3)], tri_data.test[,c(2,3)],
                  tri_data.train[,1], k=5, prob=TRUE)
tab <- table(trinary_knn,tri_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=5, accuracy = accuracy(tab)))
#k=10
trinary_knn <- knn(tri_data.train[,c(2,3)], tri_data.test[,c(2,3)],
                  tri_data.train[,1], k=10, prob=TRUE)
tab <- table(trinary_knn,tri_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=10, accuracy=accuracy(tab)))
#k=15
trinary_knn <- knn(tri_data.train[,c(2,3)], tri_data.test[,c(2,3)],
                  tri_data.train[,1], k=15, prob=TRUE)
tab_15 <- table(trinary_knn,tri_data.test[,1])
```

```
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=15, accuracy=accuracy(tab)))
#k=20
trinary_knn <- knn(tri_data.train[,c(2,3)], tri_data.test[,c(2,3)],
                   tri_data.train[,1], k=20, prob=TRUE)
tab <- table(trinary_knn,tri_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=20, accuracy=accuracy(tab)))
#k=25
trinary_knn <- knn(tri_data.train[,c(2,3)], tri_data.test[,c(2,3)],
                   tri_data.train[,1], k=25, prob=TRUE)
tab <- table(trinary_knn,tri_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=25, accuracy=accuracy(tab)))
#k=30
trinary_knn <- knn(tri_data.train[,c(2,3)], tri_data.test[,c(2,3)],
                   tri_data.train[,1], k=30, prob=TRUE)
tab <- table(trinary_knn,tri_data.test[,1])
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
accuracy_df <- rbind(accuracy_df, data.frame(k=30, accuracy=accuracy(tab)))
```

Plot K vs accuracy

```
ggplot(data = accuracy_df, aes(x = k, y = accuracy)) +  geom_point(size=3) +   theme_bw() + geom_line()
```



Linear classifier accuracy on datasets

Accuracy looks good in linear classifier and the plot confirms the same.

Accuracy difference between regression and linear classifier

Accuracy of logistic regression classifier from last week was just 0.5. But in linear classifier, accuracy has been increased above 80 which is good. Regression algorithm perform better when we try to predict continuous quantity numerical value. As we have to predict discrete classical label, Linear classification works better and the accuracy has been increased too.