K-Nearest Neighbors Classification

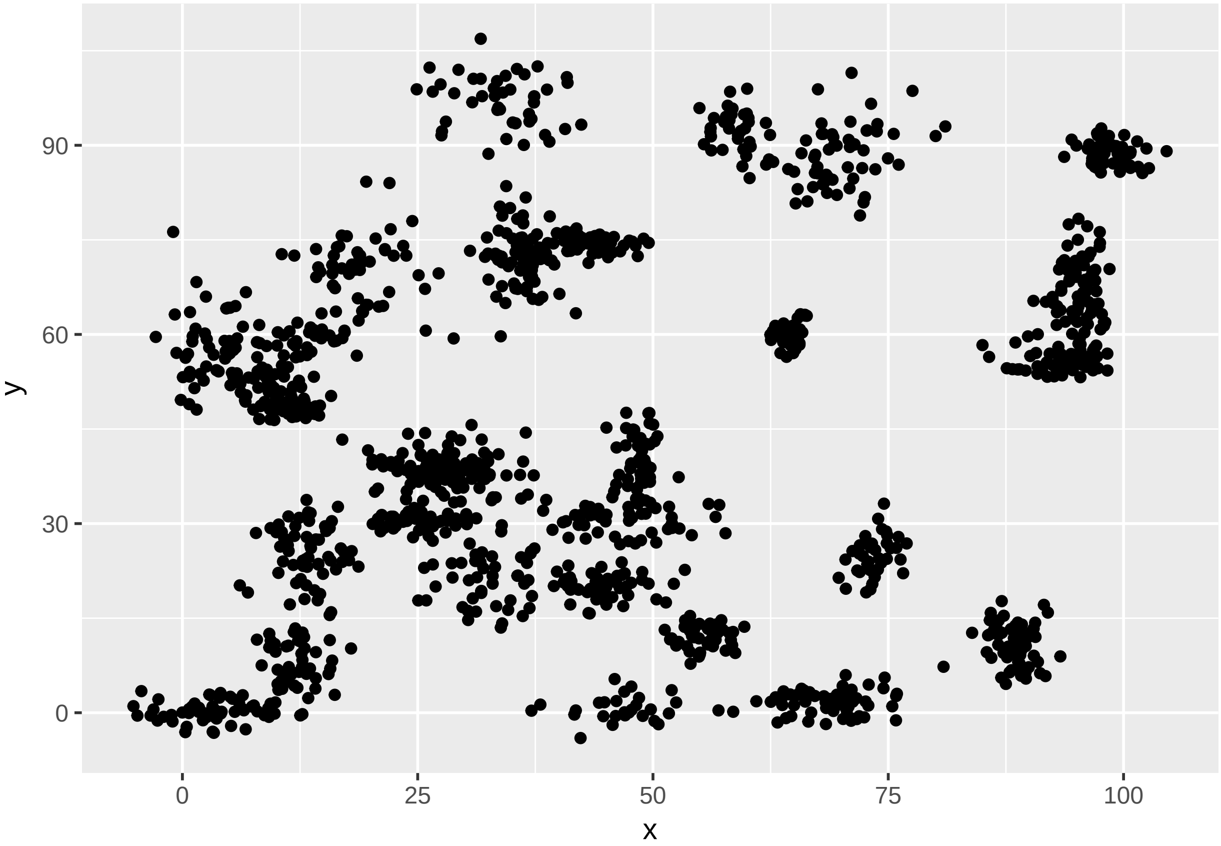
Demond Love

Regression algorithms are used to predict numeric quantity while classification algorithms predict categorical outcomes. The k nearest neighbors algorithm categorizes an input value by looking at the labels for the k nearest points and assigning a category based on the most common label.

```
library(ggplot2)
library(class)
setwd('/Users/Love/Documents/Projects')
binary = read.csv('./binary-classifier-data.csv')
trinary = read.csv('./binary-classifier-data.csv')
```

Plotting the data from each dataset using a scatter plot.

```
ggplot(binary, aes(x, y)) + geom_point()
```



```
# ggplot(trinary, aes(x, y)) + geom_point()
```

Preparing data to fit a model.

```
binary$label = as.factor(binary$label)

binary_train = binary[1:1198,2:3]
binary_test = binary[1199:1498,2:3]
binary_train_z = as.data.frame(lapply(binary_train, scale))
binary_test_z = as.data.frame(lapply(binary_test, scale))
binary_train_labels = binary[1:1198,1]
binary_test_labels = binary[1199:1498,1]
```

```
trinary$label = as.factor(trinary$label)

trinary_train = trinary[1:1198,2:3]
trinary_test = trinary[1199:1498,2:3]
trinary_train_z = as.data.frame(lapply(trinary_train, scale))
trinary_test_z = as.data.frame(lapply(trinary_test, scale))
trinary_train_labels = trinary[1:1198,1]
trinary_test_labels = trinary[1199:1498,1]
```

Fitting various models and measuring the accuracy of each. Fitting a k nearest neighbors model for each dataset for k=3, k=5, k=10, k=15, k=20, and k=25. Also plotting the results in a graph where the x-axis is the different values of k and the y-axis is the accuracy of the model.

```
binary_test_pred3 = knn(train = binary_train, test = binary_test, cl = binary_train_labels, k = 3)
binary_test_pred5 = knn(train = binary_train, test = binary_test, cl = binary_train_labels, k = 5)
binary_test_pred10 = knn(train = binary_train, test = binary_test, cl = binary_train_labels, k = 10)
binary_test_pred15 = knn(train = binary_train, test = binary_test, cl = binary_train_labels, k = 15)
binary_test_pred20 = knn(train = binary_train, test = binary_test, cl = binary_train_labels, k = 20)
binary_test_pred25 = knn(train = binary_train, test = binary_test, cl = binary_train_labels, k = 25)
```

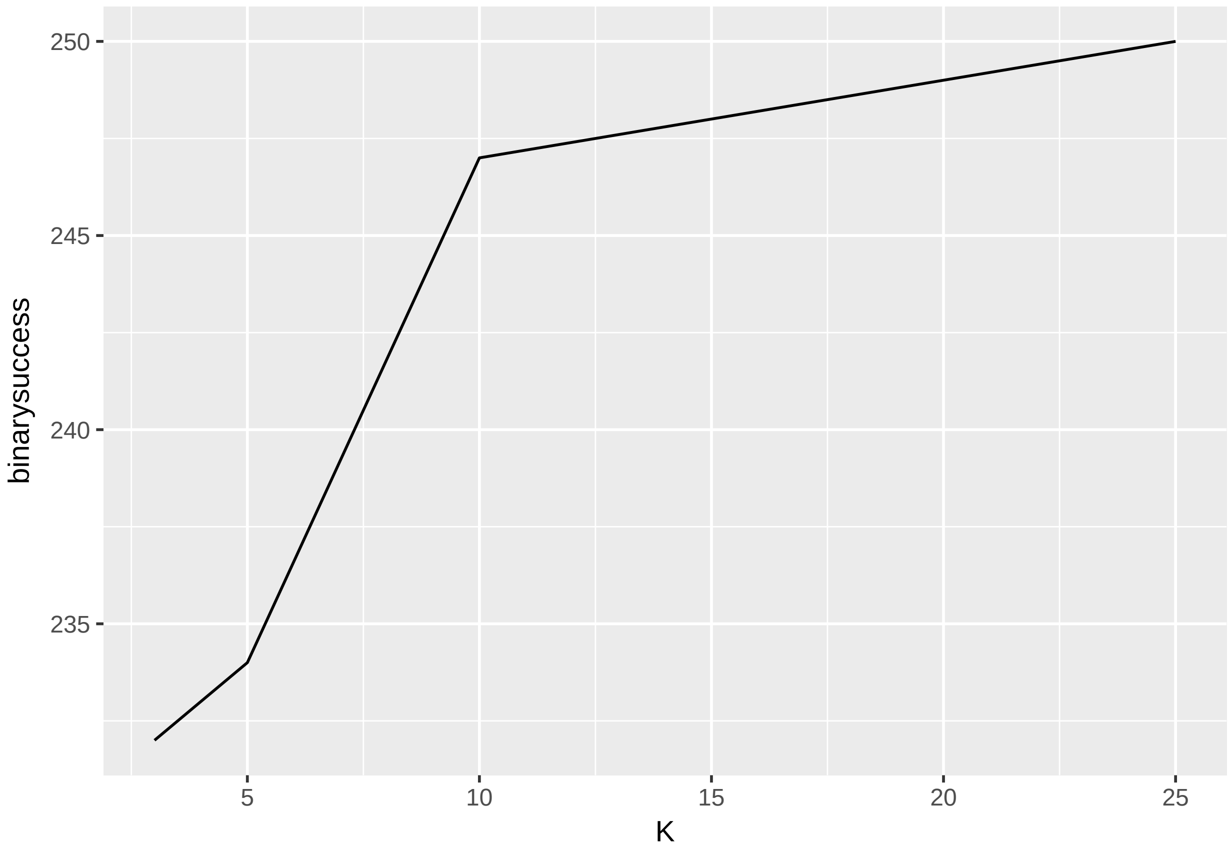
```
trinary_test_pred3 = knn(train = trinary_train_z, test = trinary_test, cl = trinary_train_labels, k = 3
trinary_test_pred5 = knn(train = trinary_train_z, test = trinary_test, cl = trinary_train_labels, k = 5
trinary_test_pred10 = knn(train = trinary_train_z, test = trinary_test, cl = trinary_train_labels, k =
trinary_test_pred15 = knn(train = trinary_train_z, test = trinary_test, cl = trinary_train_labels, k =
trinary_test_pred20 = knn(train = trinary_train_z, test = trinary_test, cl = trinary_train_labels, k =
trinary_test_pred25 = knn(train = trinary_train_z, test = trinary_test, cl = trinary_train_labels, k =
```

```
binarysuccess3 = sum(binary_test_pred3 != binary_test_labels)
binarysuccess5 = sum(binary_test_pred5 != binary_test_labels)
binarysuccess10 = sum(binary_test_pred10 != binary_test_labels)
binarysuccess15 = sum(binary_test_pred15 != binary_test_labels)
binarysuccess20 = sum(binary_test_pred20 != binary_test_labels)
binarysuccess25 = sum(binary_test_pred25 != binary_test_labels)
binarysuccess = c(binarysuccess3, binarysuccess5, binarysuccess10, binarysuccess15, binarysuccess20, bi
successes = as.data.frame(binarysuccess)
```

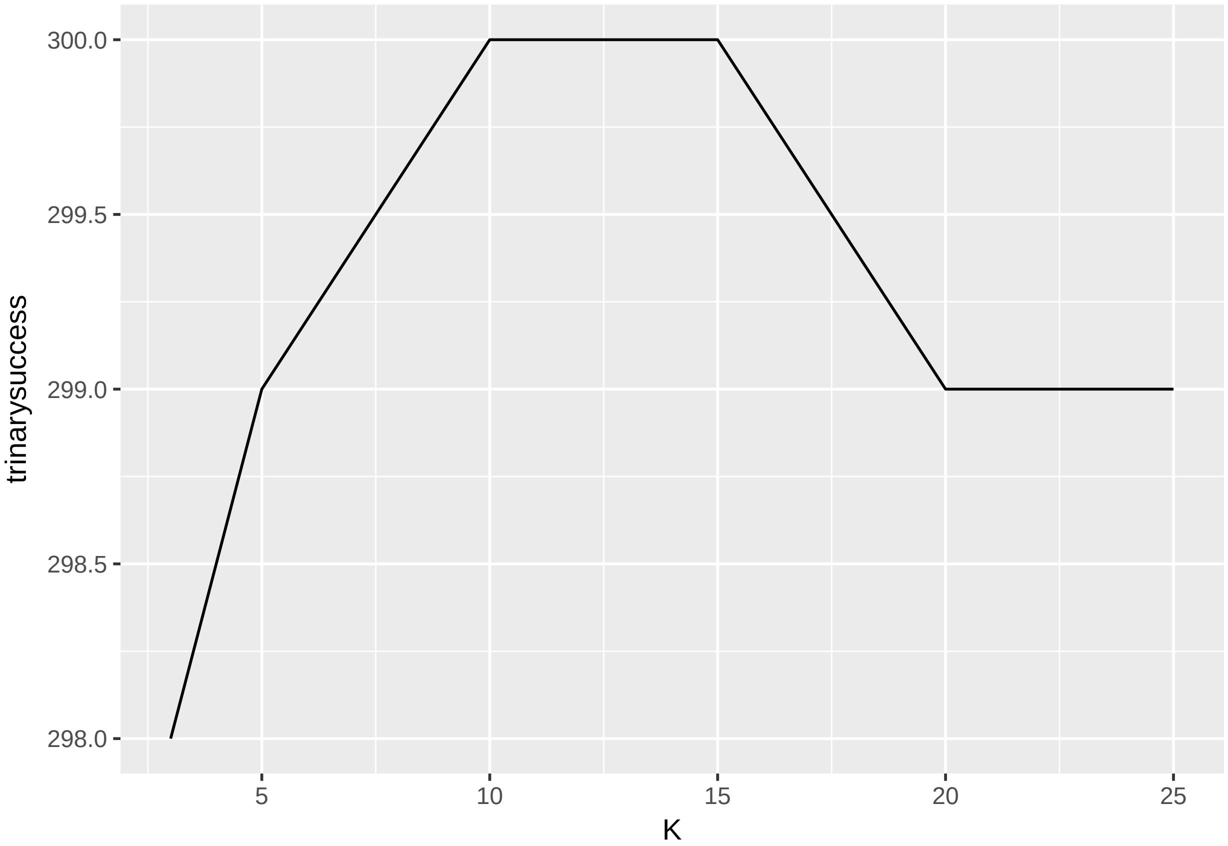
```
trinarysuccess3 = sum(trinary_test_pred3 != trinary_test_labels)
trinarysuccess5 = sum(trinary_test_pred5 != trinary_test_labels)
trinarysuccess10 = sum(trinary_test_pred10 != trinary_test_labels)
trinarysuccess15 = sum(trinary_test_pred15 != trinary_test_labels)
trinarysuccess20 = sum(trinary_test_pred20 != trinary_test_labels)
```

```r
trinarysuccess25 = sum(trinary_test_pred25 != trinary_test_labels)
trinarysuccess = c(trinarysuccess3, trinarysuccess5, trinarysuccess10, trinarysuccess15, trinarysuccess2
successes$trinarysuccess = trinarysuccess
successes$K = c(3, 5, 10, 15, 20, 25)
```

```r
ggplot(successes, aes(x = K, y = binarysuccess)) + geom_line()
```



```r
ggplot(successes, aes(x = K, y = trinarysuccess)) + geom_line()
```


300.0
299.5
299.0
298.5
298.0
trinarysuccess
5
10
15
20
25
K