

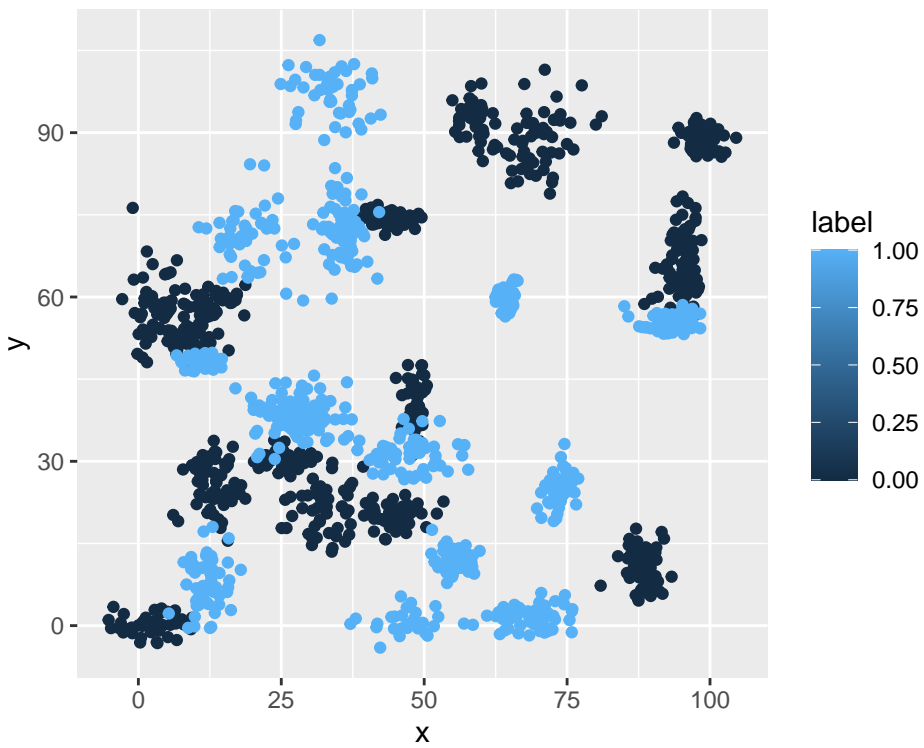
Intro To ML

Stephen Smitshoek

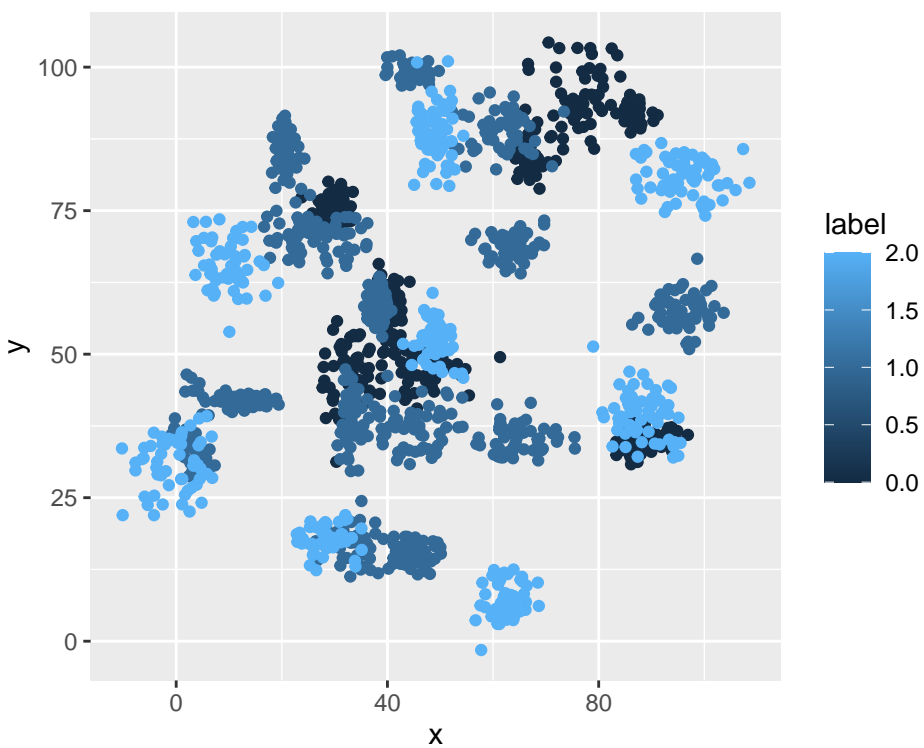
30/05/2022

Data Graphs

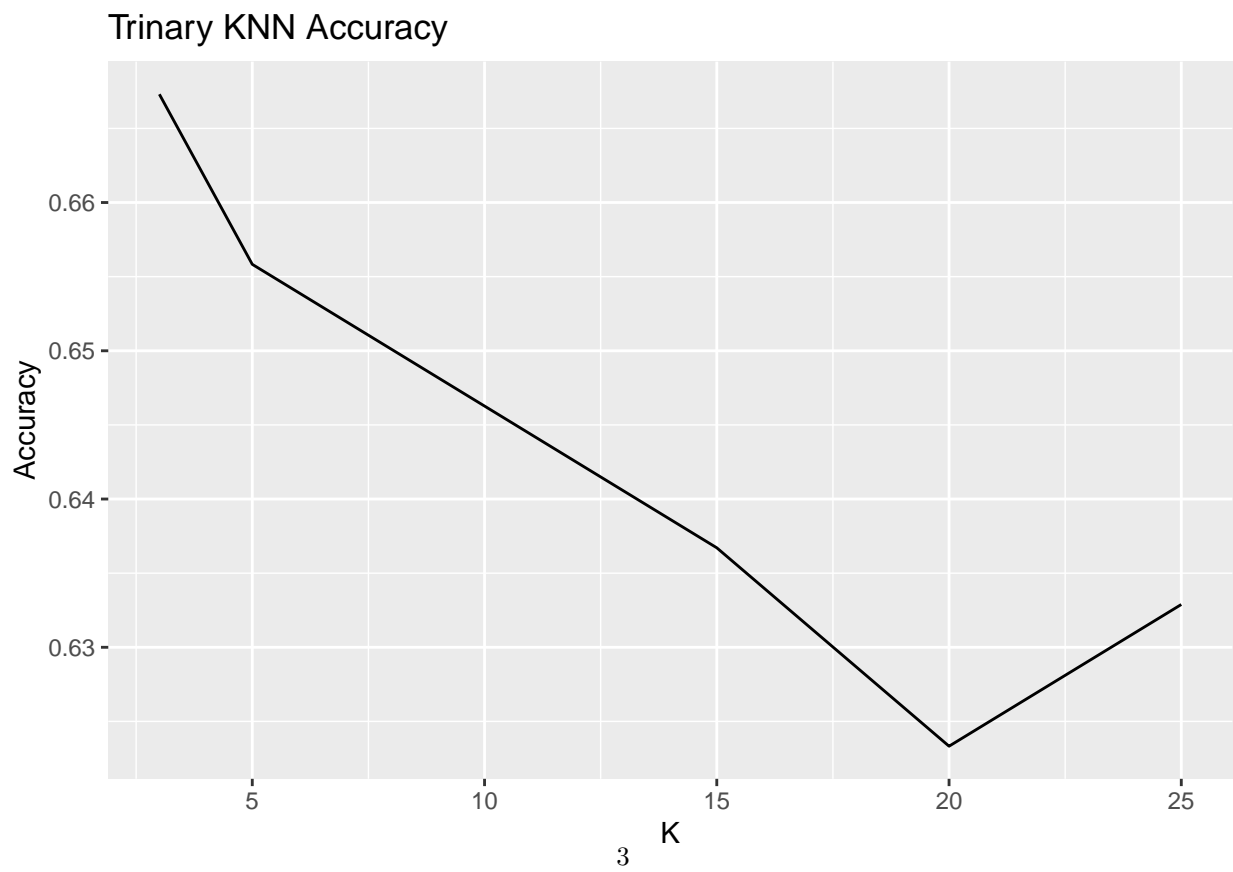
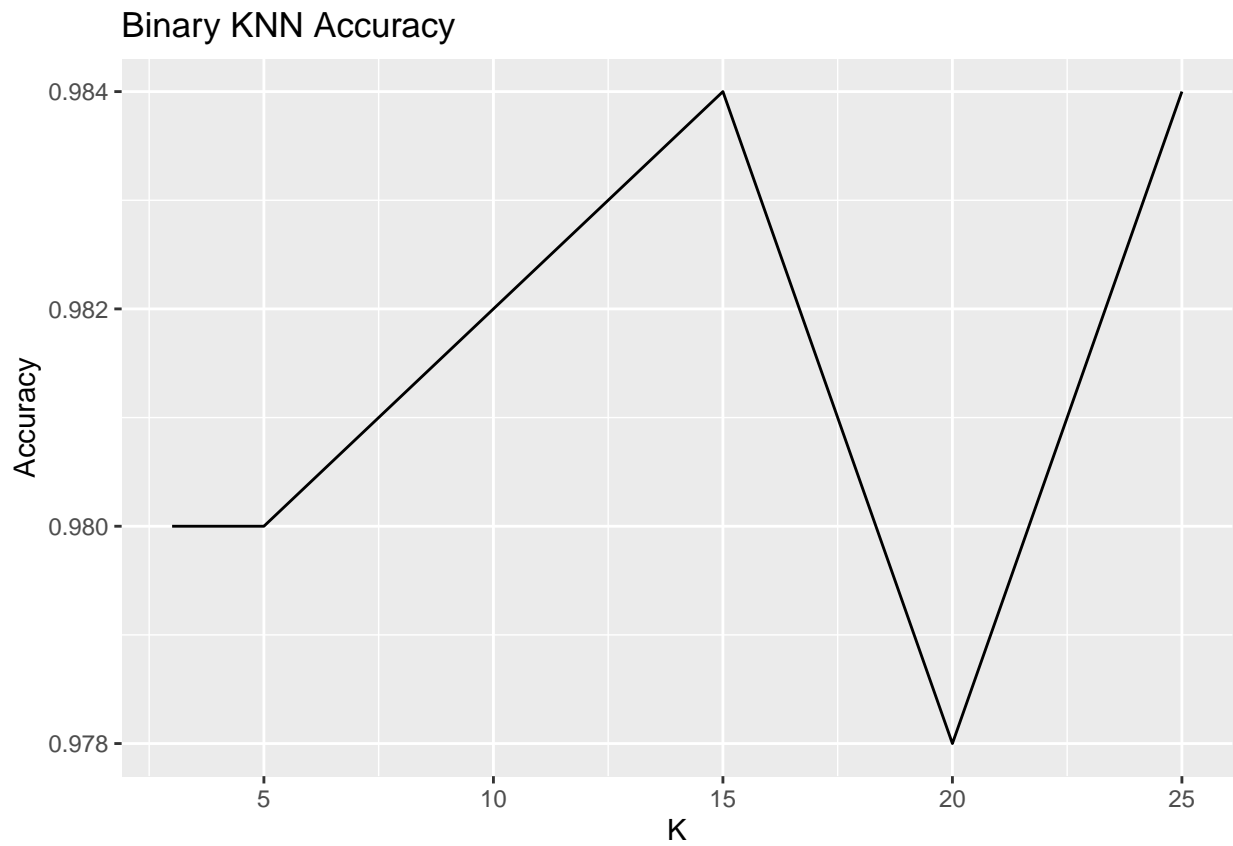
Binary Classifier



Trinary Classifier



Accuracy



Linear Model vs Machine Learning

A linear model would not be a very good choice for this data. The data does not follow any type of linear trend, instead it seems that there are very clear groups of data determined by their x and y values. The accuracy is much higher for the k nearest neighbors method because the labels are determined by the x and y attributes and not any sort of linear trend.

Code

```

library(ggplot2)
library(class)
library(caTools)

setwd("C:\\Users\\sksmi\\PeytoAccess\\Personal\\Bellevue\\DSC520\\dsc520")

binary <- read.csv("data\\binary-classifier-data.csv")
trinary <- read.csv("data\\trinary-classifier-data.csv")

ggplot(binary, aes(x=x, y=y)) + geom_point(aes(color=label))
ggplot(trinary, aes(x=x, y=y, color=label)) + geom_point()

bin_split <- sample.split(binary, SplitRatio = 0.7)
tri_split <- sample.split(trinary, SplitRatio = 0.7)

bin_train <- subset(binary, bin_split==TRUE)
bin_test <- subset(binary, bin_split==FALSE)

tri_train <- subset(trinary, tri_split==TRUE)
tri_test <- subset(trinary, tri_split==FALSE)

bin_acc <- data.frame()
for (k in c(3, 5, 10, 15, 20, 25)) {
  bin_knn <- knn(train=bin_train, test=bin_test, cl=bin_train$label, k=k)
  bin_cm <- table(bin_test$label, bin_knn)
  acc <- (bin_cm[1,1] + bin_cm[2,2]) / sum(bin_cm)
  bin_acc <- rbind(bin_acc, c(k, acc))
}
names(bin_acc) <- c("K", "Accuracy")

tri_acc <- data.frame()
for (k in c(3, 5, 10, 15, 20, 25)) {
  tri_knn <- knn(train=tri_train, test=tri_test, cl=tri_train$label, k=k)
  tri_cm <- table(tri_test$label, tri_knn)
  acc <- (tri_cm[1,1] + tri_cm[2,2]) / sum(tri_cm)
  tri_acc <- rbind(tri_acc, c(k, acc))
}
names(tri_acc) <- c("K", "Accuracy")

ggplot(data=bin_acc, aes(x=K, y=Accuracy)) + geom_point()
ggplot(data=tri_acc, aes(x=K, y=Accuracy)) + geom_point()

```