Jia Li Dong

SVM Fit into Project

Introduction:

The contents and scope in which I shall test my project is finding out how the different percentages of each color (white, black, red, blue, yellow, green, purple, and cyan) in one picture affect how many likes one can garner on a single picture in WorldCosplay.net, which is a site consisting of artists and photographers that do costume play. In this SVM fit, because there's a huge percentage of either the color white or the color black that influences the number of likes a picture gains, I'm trying to see whether a certain percentage of certain colors influences the number of likes in a picture.

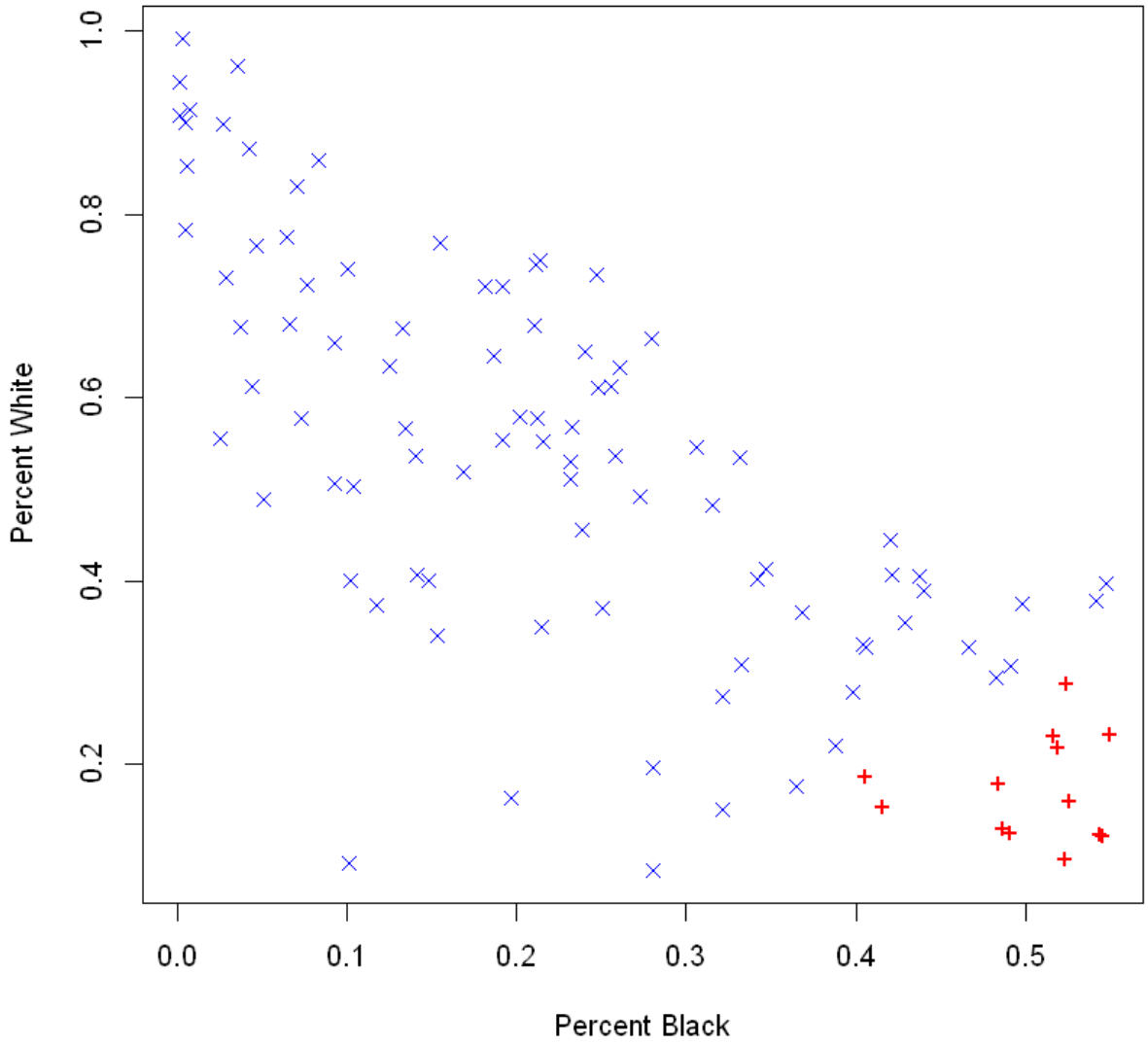
Code:

```
library(e1071)
library(data.table)
library(ggplot2)
library(dplyr)
library(e1071)
library(rpart)
library(caret)
library(tidyverse)
library(ggplot2)
library(stringr)
colorstats <- read.csv('StatsA.csv')
colorstats %>%
  select(Percentage.Black, Percentage.White, Number.of.Likes) ->stat


stat %>% drop_na() ->cs


cs$Number.of.Likes = as.factor(cs$Number.of.Likes)
svm.fit = svm(Number.of.Likes ~ Percentage.Black + Percentage.White, cs, kernel = "linear", cost = 0.01)
preds = predict(svm.fit, cs)
plot(cs[preds == 16, ]$Percentage.Black, cs[preds ==16, ]$Percentage.White, col = "blue",
```

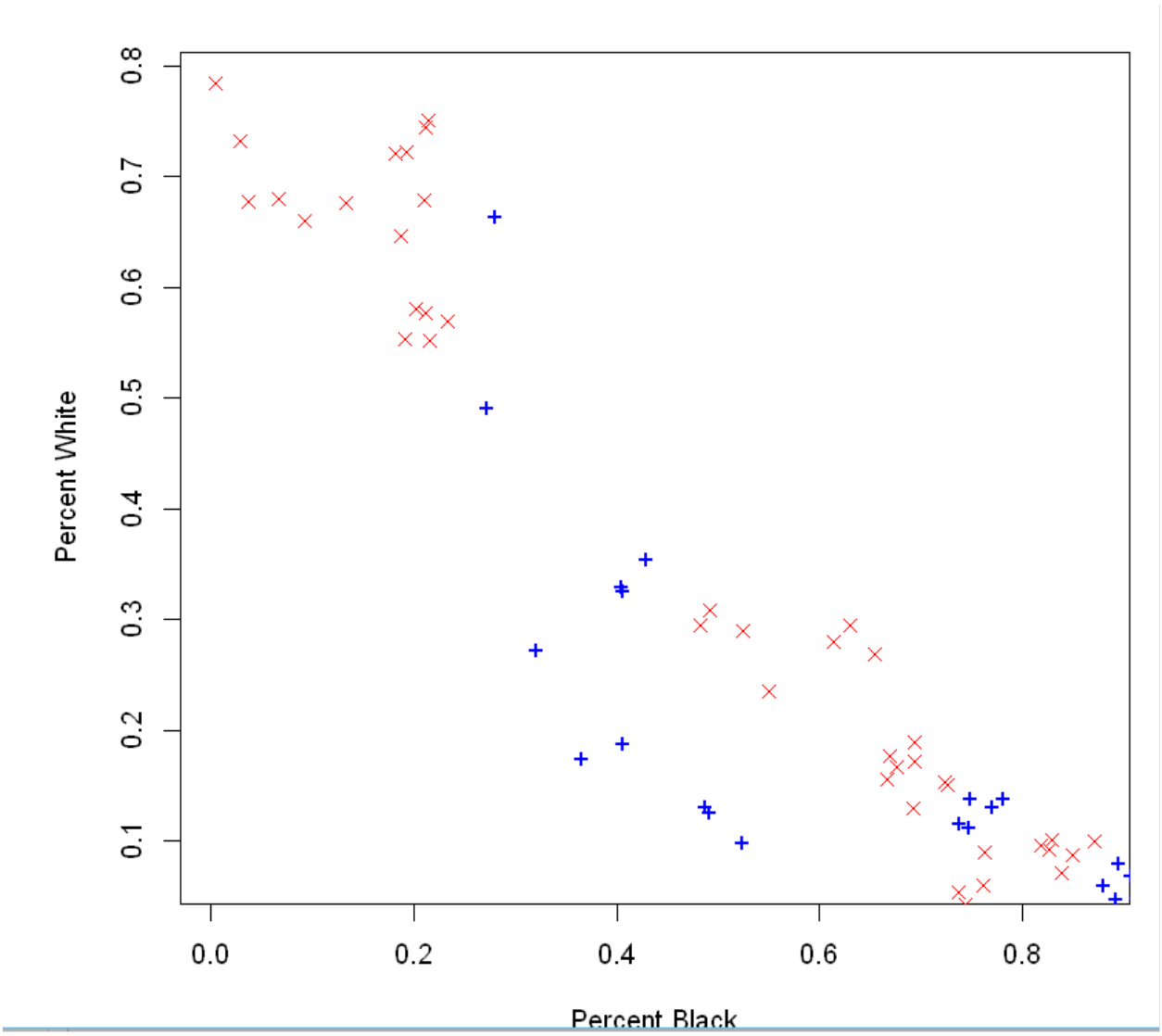
```
    pch = 4, xlab = "Percent Black", ylab = "Percent White")
points(cs[preds == 3, ]$Percentage.Black, cs[preds == 3, ]$Percentage.White, col = "red", pch = "+")
points(cs[preds == 5, ]$Percentage.Black, cs[preds == 5, ]$Percentage.White, col = "red", pch = "+")
points(cs[preds == 20, ]$Percentage.Black, cs[preds == 20, ]$Percentage.White,  col = "blue", pch = 4)
```



Explanation: This is the code for the linear kernel. As you can see, even at a very low cost, only a certain number (very few) of the points are classified to the red class. In terms of predicting whether it is accurate, it turns out not to be, as the last graph in this document shows the true picture in terms of how the percentage of either black or white plays into how many likes there are in the picture. For prediction purposes, I tested the most common values in terms of the number of likes, as it is rare to obtain a number of likes of above 21. Anything below 16 likes would be classified as red, and 16 and above likes would be classified as blue.

Radial Kernel below:

```
svmnl.fit = svm(Number.of.Likes ~ Percentage.Black + Percentage.White, cs, kernel = "radial", gamma = 10)
preds = predict(svmnl.fit, cs)
plot(cs[preds == 16, ]$Percentage.Black, cs[preds == 16, ]$Percentage.White, col = "red", pch = 4,
    xlab = "Percent Black", ylab = "Percent White")
points(cs[preds == 3, ]$Percentage.Black, cs[preds == 3, ]$Percentage.White, col = "blue", pch = "+")
points(cs[preds == 5, ]$Percentage.Black, cs[preds == 5, ]$Percentage.White, col = "blue", pch = "+")
points(cs[preds == 11, ]$Percentage.Black, cs[preds == 11, ]$Percentage.White, col = "blue", pch = "+")
points(cs[preds == 20, ]$Percentage.Black, cs[preds == 20, ]$Percentage.White,  col = "red", pch = 4)
points(cs[preds == 21, ]$Percentage.Black, cs[preds == 21, ]$Percentage.White,  col = "red", pch = 4)
```



Explanation: So in this graph, it shows a better depiction of the true graph (which is the very last graph below, which means that radial is a better representation of the true data.

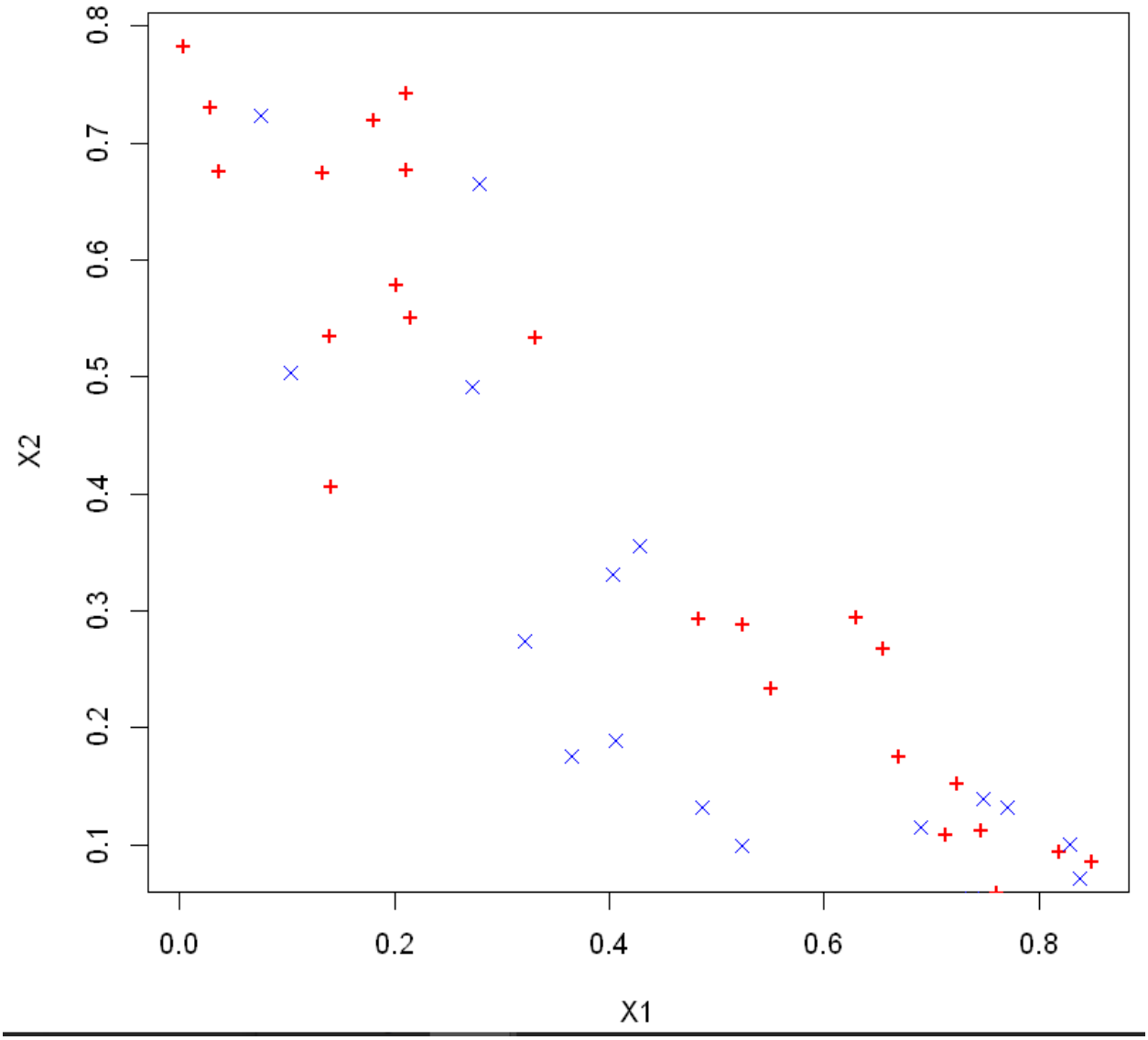
True Graphical Representation of the whole data:

```
plot(cs$Percentage.Black[cs$Number.of.Likes==16],
cs$Percentage.White[cs$Number.of.Likes==16],
     col = "red", xlab = "X1", ylab = "X2", pch = "+")
points(cs$Percentage.Black[cs$Number.of.Likes == 11],
cs$Percentage.White[cs$Number.of.Likes == 11], col = "blue", pch = 4)
points(cs$Percentage.Black[cs$Number.of.Likes == 5],
cs$Percentage.White[cs$Number.of.Likes == 5], col = "blue", pch = 4)
```

```
points(cs$Percentage.Black[cs$Number.of.Likes == 3],
cs$Percentage.White[cs$Number.of.Likes == 3], col = "blue", pch = 4)

points(cs$Percentage.Black[cs$Number.of.Likes == 20],
cs$Percentage.White[cs$Number.of.Likes == 20], col = "red", pch = "+")

points(cs$Percentage.Black[cs$Number.of.Likes == 21],
cs$Percentage.White[cs$Number.of.Likes == 21], col = "red", pch = "+")
```



Conclusion:

As you can see from the radial SVM and the true data, one may conclude that the number of likes that one gains in a single picture of work might not depend on the percentage of whites or blacks there are in the picture (in which, if the picture tends to be darker, or lighter, it might not have an effect on the number of likes). However, one can see that as one's picture has a higher

percentage of whites, there’s a tendency of more people liking the picture as opposed to pictures who are filtered to the darker side. The radial kernel with the proper gamma is able to make a better representation of the true data as opposed to the linear kernel.