KNN with R

Umakanth Prakash

```
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
library(ggplot2)
library(plotly)
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

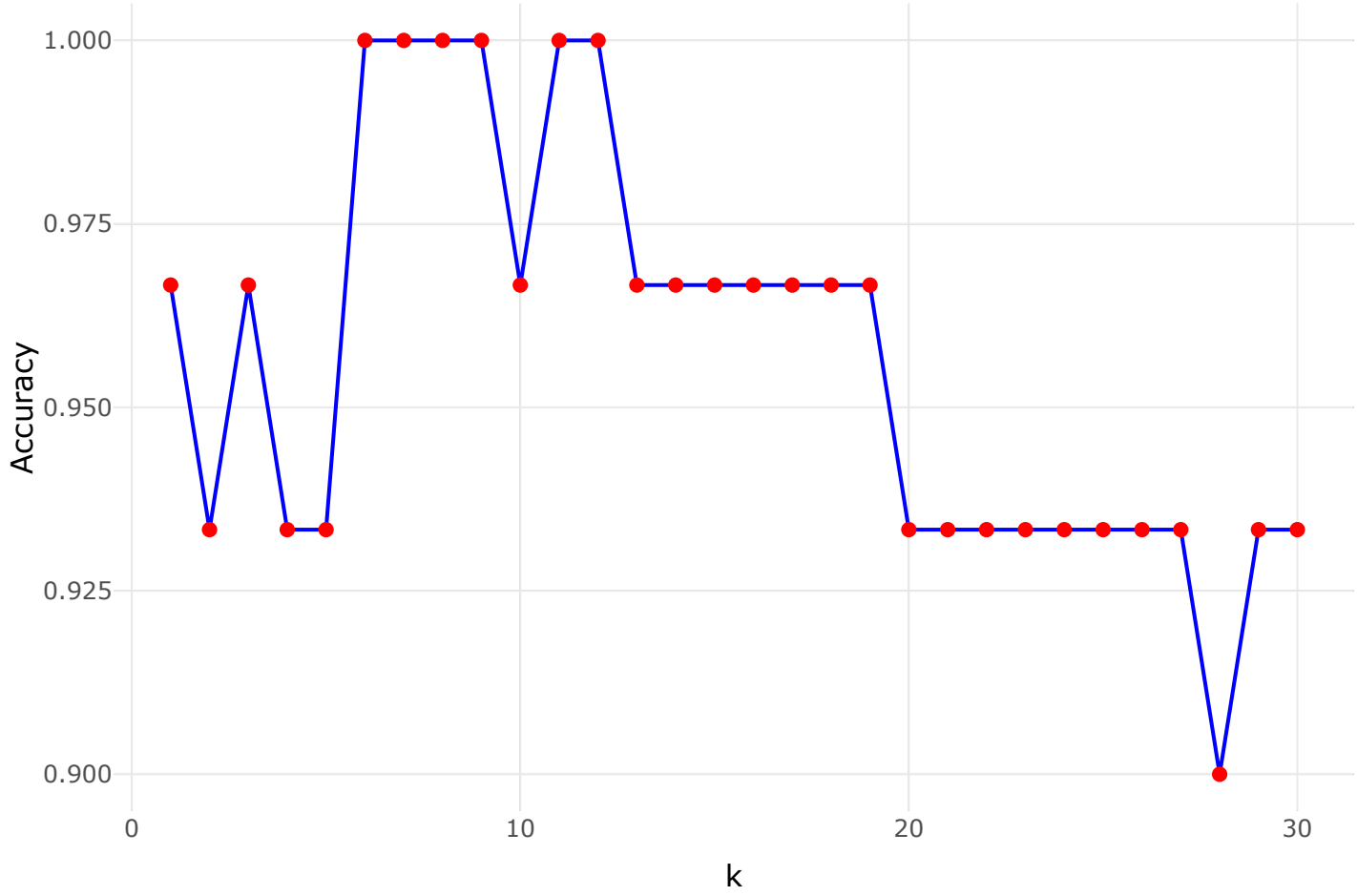
```
iris <- read.csv('iris.csv')

print(head(iris))
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
print(unique(iris$Species))
```

```
## [1] "setosa"     "versicolor" "virginica"
```

```
iris_filtered <- iris[iris$Species %in% c("virginica", "versicolor"), ]

X <- iris_filtered[, 1:4]
y <- iris_filtered$Species

# Lets split the data into training and testing sets
set.seed(1)
train_index <- createDataPartition(y, p = 0.7, list = FALSE)
X_train <- X[train_index, ]
X_test <- X[-train_index, ]
y_train <- y[train_index]
y_test <- y[-train_index]

k_values <- 1:30

accuracies <- c()

# doing cross validation for each k
for (k in k_values) {
  # Here we are training the model
  knn_model <- knn3Train(X_train, X_test, y_train, k = k)

  # and we will calculate the accuracy
  accuracy <- mean(knn_model == y_test)
  accuracies <- c(accuracies, accuracy)
}

# Now we will plot the accuracy vs k values
plot_data <- data.frame(
  k = k_values,
  Accuracy = accuracies
)

plot <- ggplot(plot_data, aes(x = k, y = Accuracy)) +
  geom_line(color = "blue") +
  geom_point(color = "red") +
  labs(title = "Accuracy vs. k Value (Euclidean Distance)", x = "k", y = "Accurac
y") +
  theme_minimal()

# using plotly for interactive plot
interactive_plot <- ggplotly(plot)

interactive_plot
```




```
# We will use the k value 5 to train the model
best_k <- 5
final_knn_model <- knn3Train(X_train, X_test, y_train, k = best_k)

# lets make prediction on our test set
y_pred <- final_knn_model

y_pred <- factor(y_pred, levels = levels(y_test))

# generating the confusion matrix
conf_matrix <- table(Predicted = y_pred, Actual = y_test)
print("Confusion Matrix:")
```

```
## [1] "Confusion Matrix:"
```

```
print(conf_matrix)
```

```
## < table of extent 0 x 2 >
```

```
# calculating the accuracy once again
accuracy <- mean(y_pred == y_test)
print(paste("Accuracy with k =", best_k, ":", accuracy))
```

```
## [1] "Accuracy with k = 5 : NA"
```