Programming Assignment Unit 4

September 26, 2025

1 K-Nearest Neighbors – Unit 4 Assignment

Data Mining and Machine Learning
Author: Sana

1.1 Introduction

We classify points in 2D space into two classes, A and B, using the k-Nearest Neighbors (kNN) algorithm.

Class A: (0,0), (1,1), (2,2)

Class B: (6,6), (5.5,7), (6.5,5)

[ ]:
```
# Load the class package
library(class)

# Training data
train <- rbind(
  c(0,0),
  c(1,1),
  c(2,2),
  c(6,6),
  c(5.5,7),
  c(6.5,5)
)

# Classification labels: 3 A's then 3 B's
cl <- factor(c(rep("A",3), rep("B",3))) # This assigns each of the six training points to its class.

train
cl
```

A matrix: 6 × 2 of type dbl

0.0	0
1.0	1
2.0	2
6.0	6
5.5	7
6.5	5

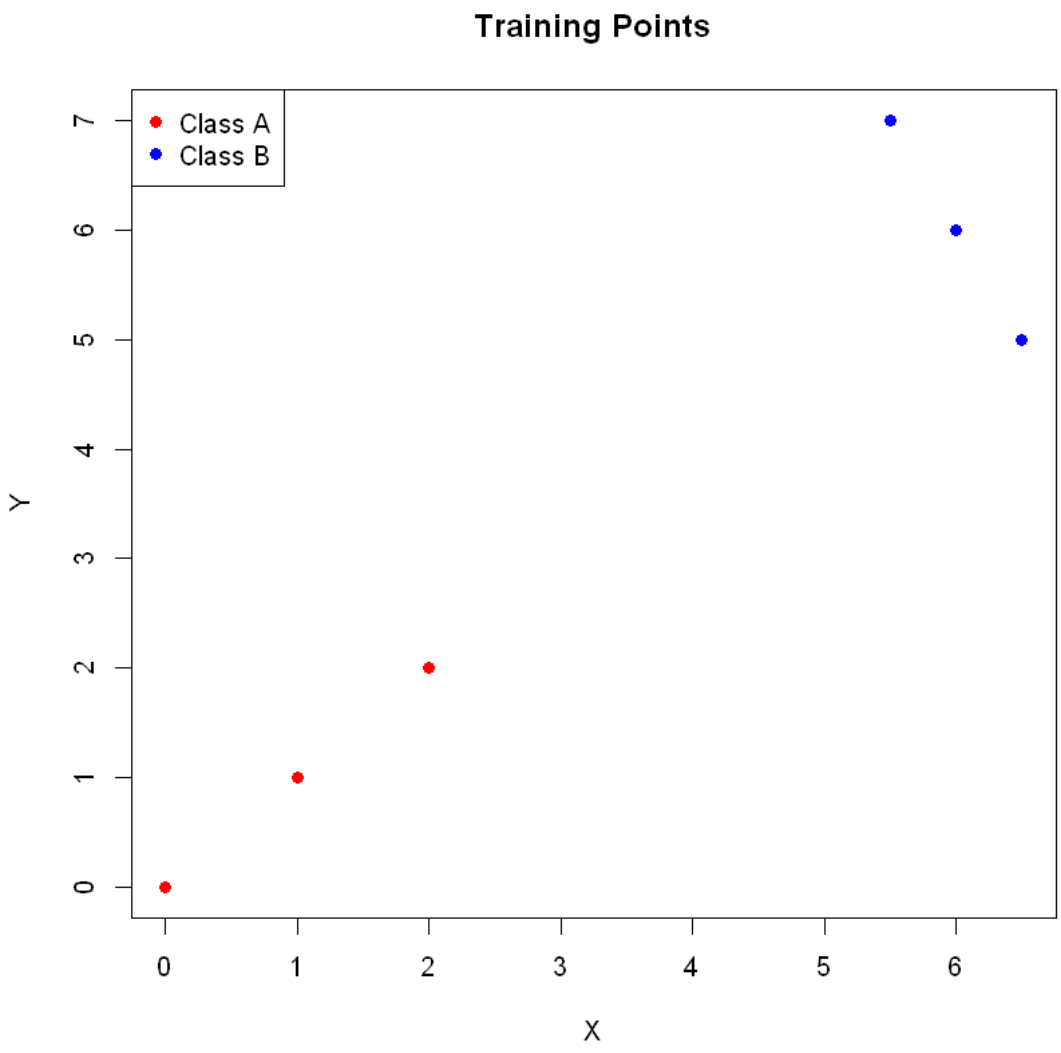
1. A 2. A 3. A 4. B 5. B 6. B

Levels: 1. 'A' 2. 'B'

1.2 Plot of Training Objects

The chart below shows the six training points, red for Class A and blue for Class B.

[3]:
```
plot(train, col=c(rep("red",3), rep("blue",3)),
     pch=19, xlab="X", ylab="Y", main="Training Points")
legend("topleft", legend=c("Class A","Class B"),
       col=c("red","blue"), pch=19)
```



Plotted on an X-Y chart, Class A forms a small cluster in the lower-left corner while Class B forms a cluster in the upper-right corner.

1.2.1 1. Test Point (4,4), k = 1

Classify the point (4,4) using k = 1 and summarize the result.

[4]:
```
test1 <- matrix(c(4,4), ncol=2)
pred1 <- knn(train, test1, cl, k=1)
summary(pred1)
```

A 0 **B** 1

The summary shows the predicted class.
The test point (4,4) is classified as **Class B**, because its single nearest neighbor is from Class B.

1.2.2 2. Test Point (3.5,3.5), k = 1

[5]:
```
test2 <- matrix(c(3.5,3.5), ncol=2)
pred2 <- knn(train, test2, cl, k=1)
summary(pred2)
```

A 1 **B** 0

The point (3.5,3.5) is visually nearer Class A points, so it is predicted as **Class A**.

1.2.3 3. Test Point (3.5,3.5), k = 3

[6]:
```
pred3 <- knn(train, test2, cl, k=3)
summary(pred3)
```

A 0 **B** 1

With k = 3, the majority vote of the three closest neighbors predicts **Class B**.

1.2.4 4. Classifying Four Test Points

Test points: (4,4), (3,3), (5,6), (7,7) using k = 3.

[8]:
```
testBatch <- matrix(c(4,4, 3,3, 5,6, 7,7), ncol=2, byrow=TRUE)
predBatch <- knn(train, testBatch, cl, k=3)
summary(predBatch)
table(predBatch)  # count per class
```

A 1 **B** 3

```
predBatch
A B
1 3
```

The table shows how many of the four test points are classified into each class. - 1 point is classified as Class A

- 3 points are classified as Class B

1.3 Conclusion

Using the `class::knn()` function in R:

- The two class clusters are well separated on the chart.
- Classification depends on both the distance of each test point and the number of neighbors **k**.
- Increasing **k** can change the result, as seen for point (3.5,3.5), which switches from Class A at k=1 to Class B at k=3.