

K-Means Clustering

Importing the dataset

```
dataset = read.csv('Mall_Customers.csv')
```

```
dataset = dataset[4:5]
```

Splitting the dataset into the Training set and Test set

```
install.packages('caTools')
```

```
library(caTools)
```

```
set.seed(123)
```

```
split = sample.split(dataset$DependentVariable, SplitRatio = 0.8)
```

```
training_set = subset(dataset, split == TRUE)
```

```
test_set = subset(dataset, split == FALSE)
```

Feature Scaling

```
training_set = scale(training_set)
```

```
test_set = scale(test_set)
```

Using the elbow method to find the optimal number of clusters

```
set.seed(6)
```

```
wcss = vector()
```

```
for (i in 1:10) wcss[i] = sum(kmeans(dataset, i)$withinss)
```

```
plot(1:10, wcss, type = 'b', main = paste('The Elbow Method'), xlab = 'Number of clusters', ylab = 'WCSS')
```

Fitting K-Means to the dataset

```
set.seed(29)
```

```
kmeans = kmeans(x = dataset, centers = 5)
```

```
y_kmeans = kmeans$cluster
```

Visualising the clusters

```
library(cluster)
```

```
clusplot(dataset, y_kmeans, lines = 0, shade = TRUE, color = TRUE, labels = 2, plotchar = FALSE,
```

```
span = TRUE, main = paste('Clusters of customers'), xlab = 'Annual Income', ylab = 'Spending Score')
```