Termination project report

Name: Esraa Aldreabi

B#: B00575499

In this project I used dataset for mall customers. The dataset contains information about customers who subscribe to the mall card. The dataset shape is 200 rows and 5 column. The columns are: customer id, gender, age, annual income and spending score. The spending score column values is between(1-100) computed for each customer based on spending, income and number of visit.

The problem statement : Segment customer based on annual income and spending score. Since we don't know how many segments are there this is a cluster problem.

I used K-means clustering algorithm. K-means is a unsupervised machine learning algorithm that groups data into a specified number (k) of clusters.

The algorithm steps are:

1. Choose the number of cluster
2. Select random k points (centroids)
3. Assign each data point to the closest centroid (that’s forms k clusters)
4. Compute and place the new centroid of each cluster
5. Reassign each data point to the new closest centroid
6. If any reassignment took place go to step 4 otherwise finish

I used the elbow method to find the optimal number of clusters. As we can see in figure 1 the elbow (the point of inflection on the curve) shows the optimal number of clusters is 5.

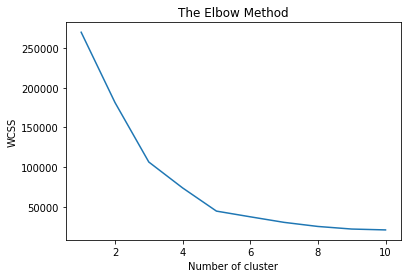


Figure 1: The Elbow Method

The next step we apply the k-means algorithm with 5 clusters and the results is shown in figure 2.

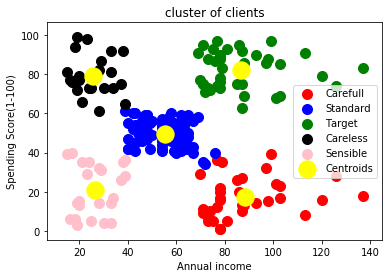


Figure 2: the results

Figure 2 shows the five clusters. We can name the customers groups as follow:

1. Careful: This group contains the customers with high income and low spending score
2. Standard: this group contains customers with medium income and medium spending score
3. Target: This group contains customers with high income and high spending score
4. Careless: This group contains customers with low income and high spending score
5. Sensible: This group contains customers with low income and low spending score

This results can help the mall manager to know which group of customers he can target.

Possible assumptions:

1. Target group: consider loyal to the mall maybe send them more marketing brochures, issue high limit credit cards, gift cards.
2. Carful group: the manager may consider send them discounts coupons
3. Careless group: The manager may be carful about issuing a credit card to them or issue one with low limit.