

Machine Intelligence and Expert Systems

Computer Assignment-2

Clustering

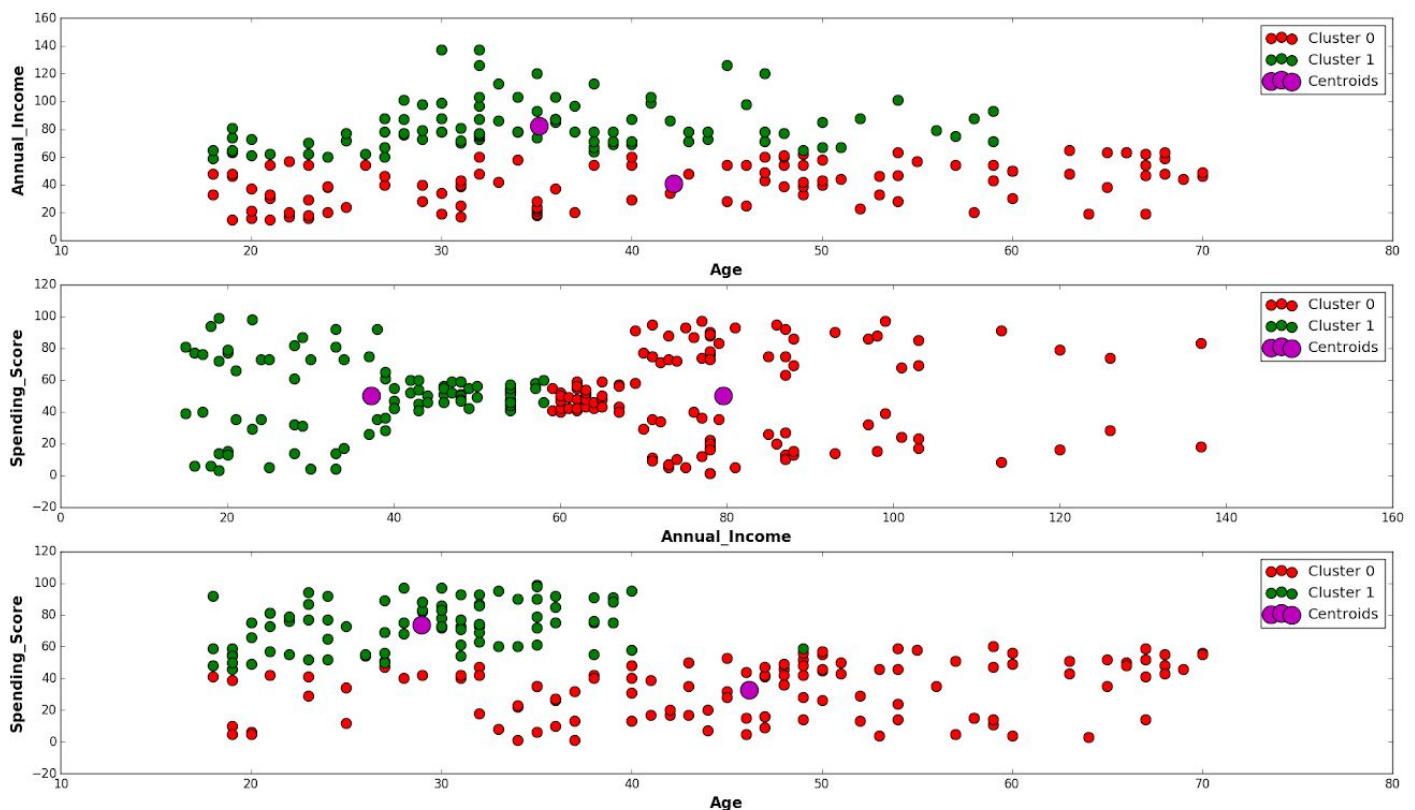
Name: Anand Jhunjunwala

Roll Number: 17EC35032

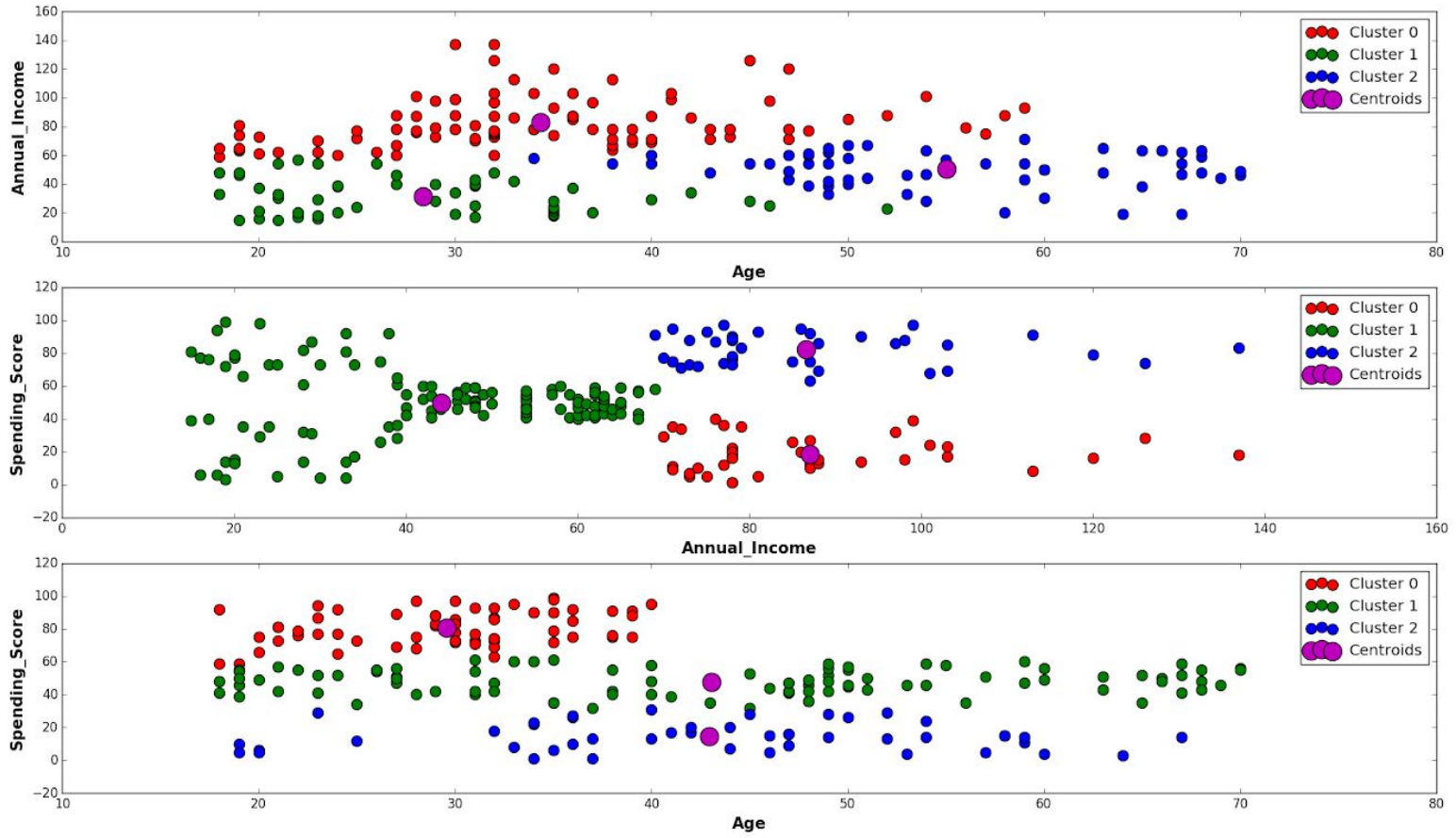
Aim: Using K-means clustering, cluster the Mall Customers dataset.

Results:

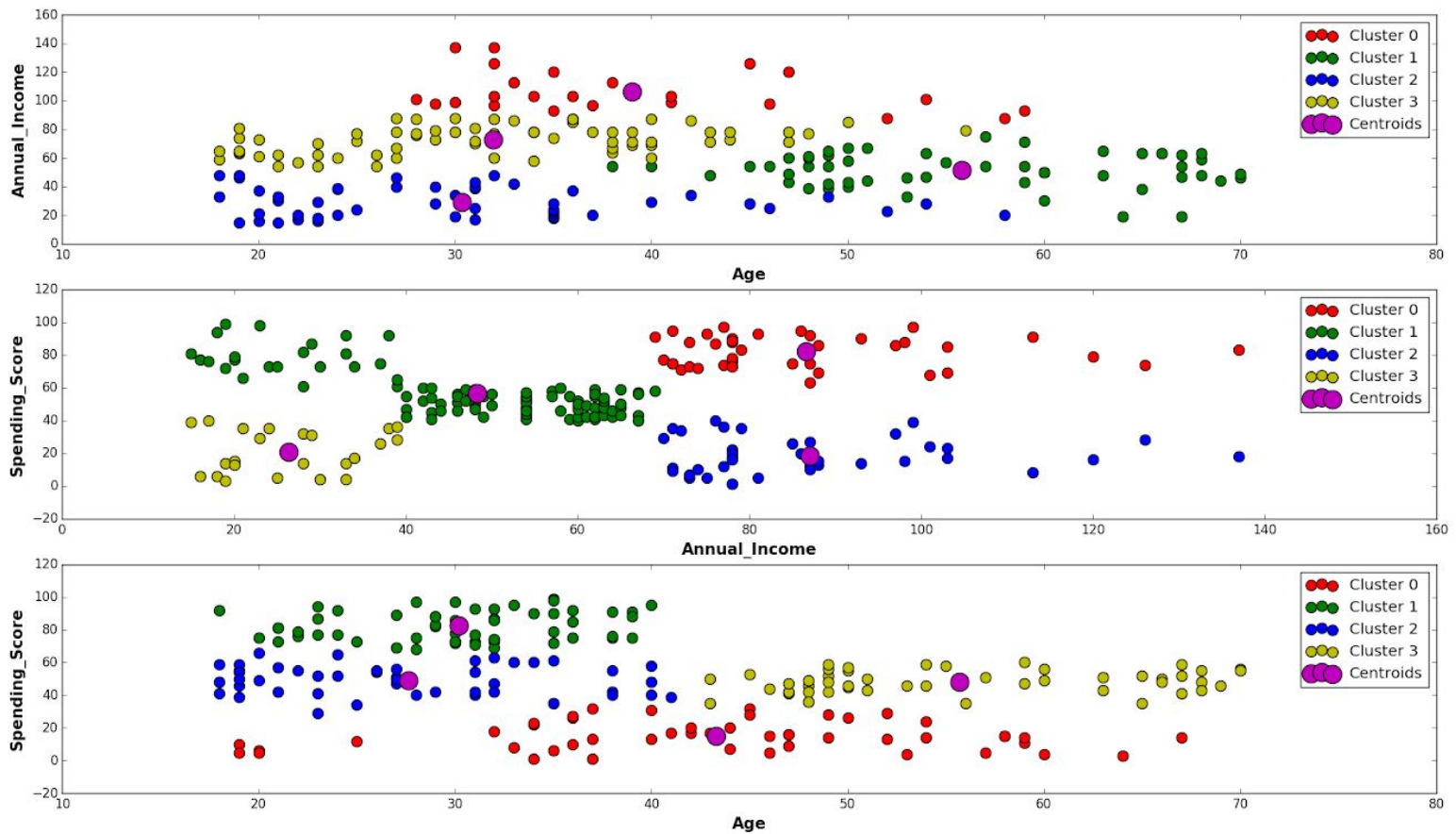
Cluster of Clients with K=2



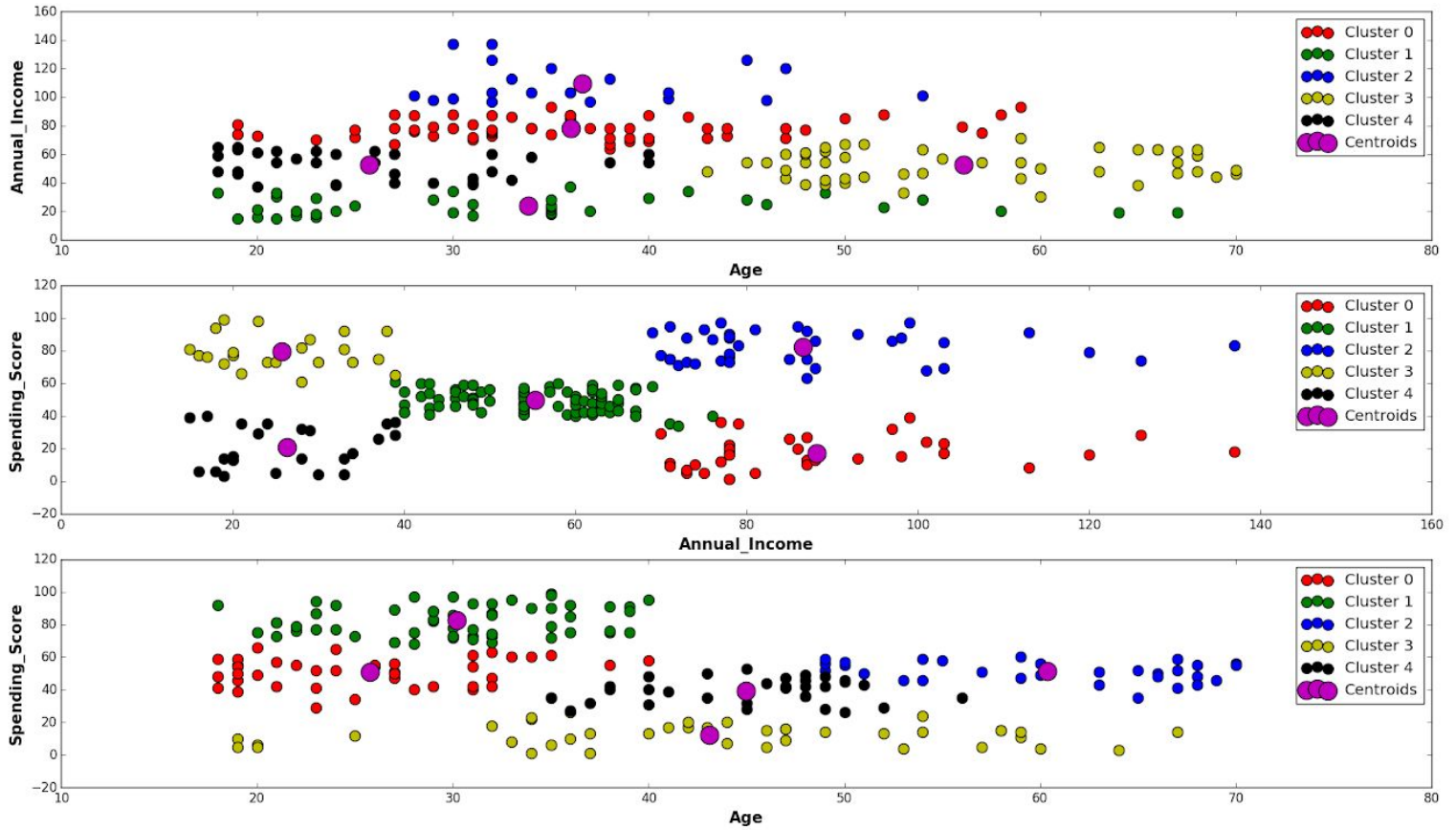
Cluster of Clients with K=3



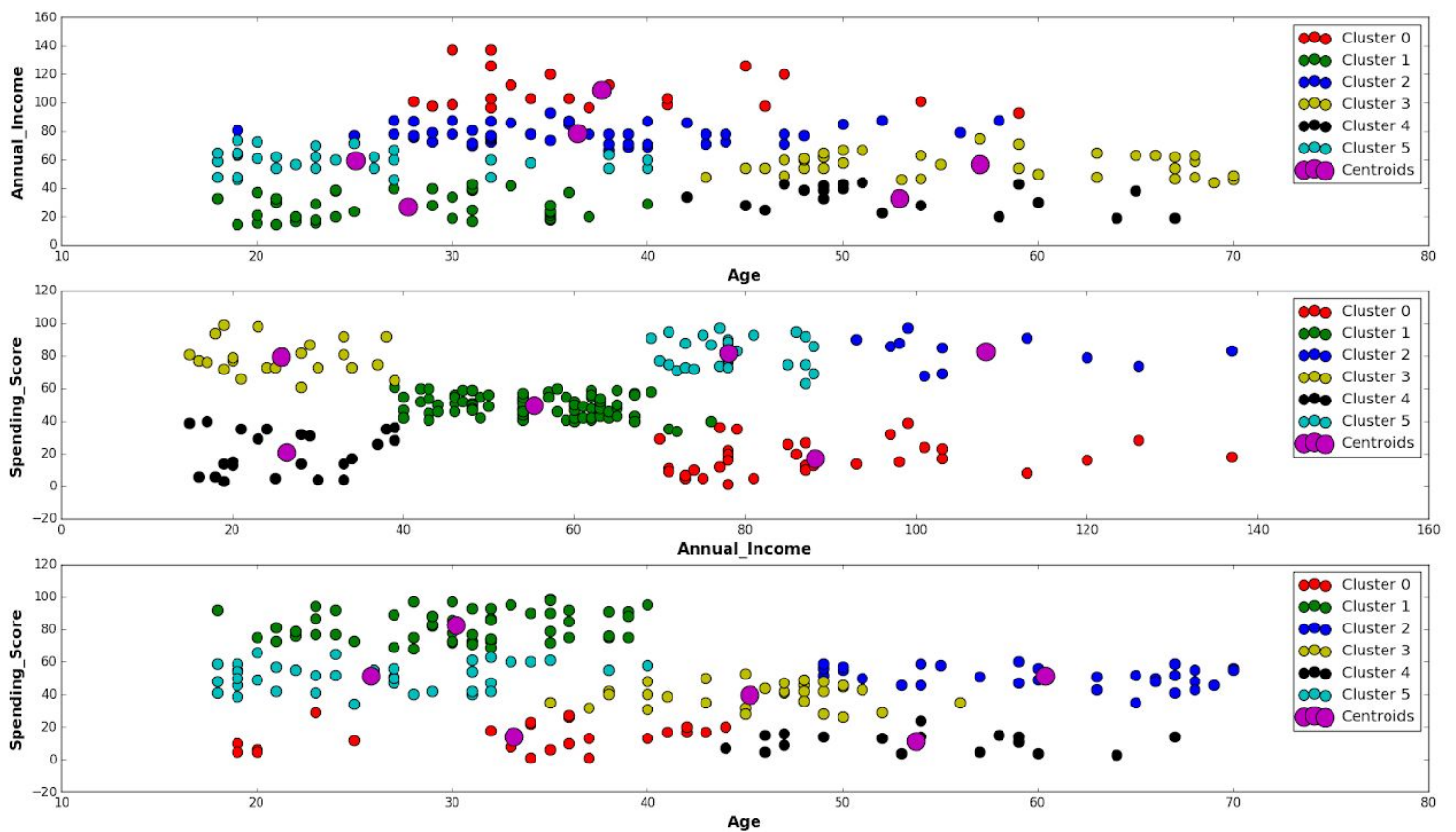
Cluster of Clients with K=4



Cluster of Clients with K=5



Cluster of Clients with K=6



Discussion:

- For clusters with features Annual_Income and Age:
 - Best Clustering was observed with K=3, with further increase in value of K very small clusters starts to form with very less intercluster distance indicating bad clustering
 - For clusters with features Spending_Score and Annual_Income:
 - From the shape of the plot it can be easily seen that it can be divided into 5 separate clusters and algorithm does the same for K=5 as can be seen from the plot
 - For all other values of K, the data is either over-clustered or under-clustered.
 - For clusters with features Spending_Score and Age:
 - However clear cluster boundaries are not seen in the plotted data at first gaze but with K=4 cluster boundaries are clearly visible.
 - For all other values of K, the data is either over-clustered or under-clustered.
 - With above observations I can conclude that clustering of data highly depends on the value of K and features taken to form clusters and the best cluster is observed with features Spending_Score and Annual_Income with K =5.
-