Tugas 2

3 April 2019

Nama : Patricia Joanne
NPM : 140810160065

Metode Elbow

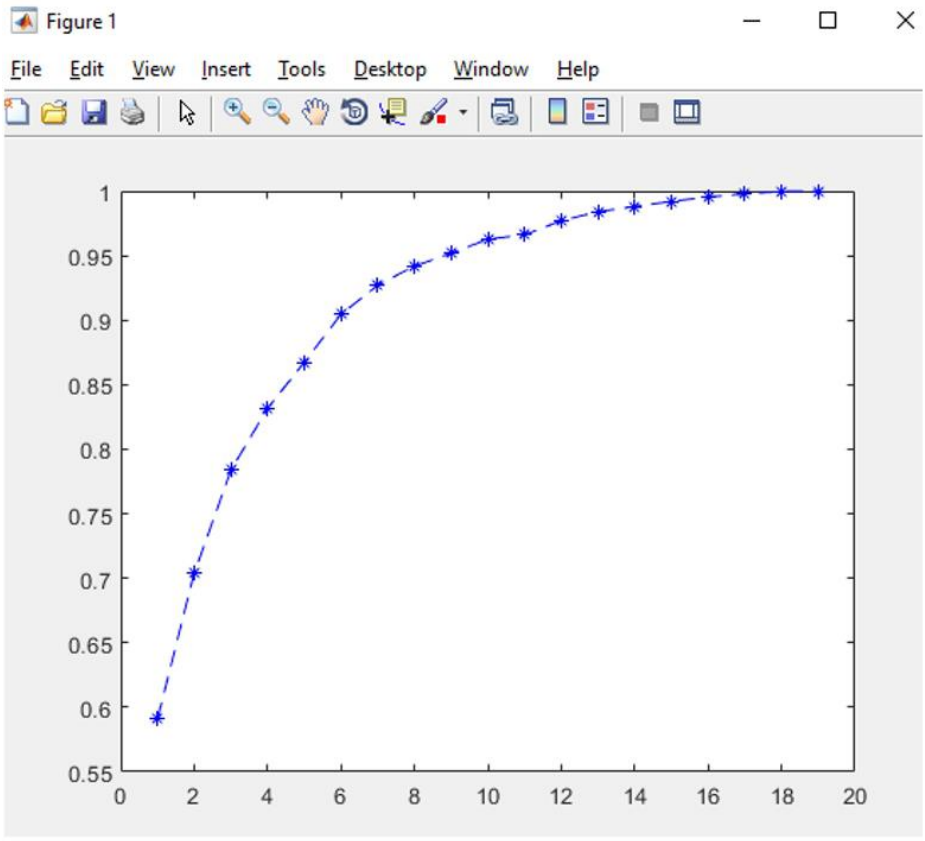
Program:

```
function [IDX,C,SUMD,K] = best_kmeans(X)
% [IDX,C,SUMD,K] = best_kmeans(X) partitions the points in the N-by-P data matrix X
% into K clusters. Rows of X correspond to points, columns correspond to variables.
% IDX containing the cluster indices of each point.
% C is the K cluster centroids locations in the K-by-P matrix C.
% SUMD are sums of point-to-centroid distances in the 1-by-K vector.
% K is the number of cluster centriods determined using ELBOW method.
% After find the best K clusters, IDX, C, SUMD are determined using kmeans function
in matlab.
dim=size(X);

% default number of test to get minimun under differnent random centriods
test_num=10;
distortion=zeros(dim(1),1);
for k_temp=1:dim(1)
    [~,~,sumd]=kmeans(X,k_temp,'emptyaction','drop');
    destortion_temp=sum(sumd);
    % try differnet tests to find minimun disortion under k_temp clusters
    for test_count=2:test_num
        [~,~,sumd]=kmeans(X,k_temp,'emptyaction','drop');
        destortion_temp=min(destortion_temp,sum(sumd));
    end
    distortion(k_temp,1)=destortion_temp;
end
variance=distortion(1:end-1)-distortion(2:end);
distortion_percent=cumsum(variance)/(distortion(1)-distortion(end));
plot(distortion_percent,'b*--');
[r,~]=find(distortion_percent>0.9);
```

```
K = r(1,1)+1;
[IDX,C,SUMD] = kmeans(X,K);
```

Screenshot:




```
ans =

     4
     3
     1
     5
     3
     4
     4
     3
     1
     1
     6
     3
     5
     5
     2
     2
     2
     7
     2
     2
```