

تمرین

پاسخ سوال اول)

تعیین تعداد بهینه خوشه ها در پروژه های داده کاوی همواره یکی از چالش های اصلی پیاده سازی الگوریتم های خوشه بندی مبتنی بر تقسیم یا به عبارتی Partitioning Clustering مانند k-means است و همچنین جواب بهینه برای تعیین تعداد خوشه ها در الگوریتم های خوشه بندی کار ساده ای نیست و تابع مقادیر پارامتر ها و مقادیر داده ها و میزان فاصله تا مراکز خوشه های اولیه می باشد یک روش استفاده از الگوریتم Agglomerative می باشد.

پاسخ سوال دوم)

مقادیر اولیه ی خوشه ها (کلاسها) همانطور که در صورت مساله داده شده بود به صورت رندوم می باشد و ممکن است فاصله ی این مراکز تا بقیه ی داده ها به گونه ای قرار گیرند که در برخی حالات ، بعضی از خوشه ها خالی از داده باشند و این حالت در مواقعی که مقادیر رندوم اولیه مراکز دسته ها در مکان مناسبی قرار نگیرند حالت بهینه ای از این الگوریتم بوجود نمی آید.

پاسخ سوال سوم)

تمرین متلب :

```
function y=kMeansCluster(m,k,isRand)
    if isRand
        p = randperm(size(m,1));
        for i=1:k
            c(i,:)=m(p(i),:);
        end
    else
        for i=1:k
            c(i,:)=m(i,:);
        end
    end
end
```

```

temp=zeros(maxRow,1);
while 1
    d=DistMatrix(m,c);
    [~,g]=min(d,[1,2]);
    if g==temp
        break;
    else
        temp=g;
    end
    for i=1:k
        f=find(g==i);
        if f
            c(i,:)=mean(m(g==i,:),1);
        end
    end
end
y=[m,g];
end

```