Fundamentals of Data Science

Semester B 20-21

Tutorial 8

1a.(i) The first cluster is {6,12,18,24,30}, and its centroid is given by $\frac{6+12+18+24+30}{5}=18$

The second cluster is {42,48}, and its centroid is given by $\frac{42+48}{2}=45$

The sum squared error is

$(6-18)^2+(12-18)^2+(18-18)^2+(24-18)^2$

$+(30-18)^2+(42-45)^2+(48-45)^2=378$

(ii) The first cluster is {6,12,18,24}, and its centroid is given by $\frac{6+12+18+24}{4}=15$

The second cluster is {30,42,48}, and its centroid is given by $\frac{30+42+48}{3}=40$

The sum squared error is

$(6-15)^2+(12-15)^2+(18-15)^2+(24-15)^2$

$+(30-40)^2+(42-40)^2+(48-40)^2=348$

b. There will be no change to the clusters generated in both cases (i) and (ii).

2 The first cluster is {A}, and its centroid is (2,10).

The second cluster is {C,D,E,F,H}, and its centroid is $\left(\frac{8+5+7+6+4}{5},\frac{4+8+5+4+9}{5}\right)=(6,6)$.

The third cluster is {B,G}, and its centroid is $\left(\frac{2+1}{2},\frac{5+2}{2}\right)=(1.5,3.5)$.