CSE – 575

Homework #3

1 Gaussian Mixture Model & EM Algorithm

dataset : [−67, −48, 6, 8, 14, 16, 23, 24]

$k = 2$

1.1) Total No of parameters in this GMM ⇒ 4 + 1 ⇒ 5

And they are : $\left\{ \begin{matrix} \mu_1, \sigma_1^2 \\ \mu_2, \sigma_2^2 \end{matrix} \right\}$ as $k=2$; π_k ↓ mixing coeff. / Prior

1.2) Parameter Initialization ↓

As $k=2$, there are two components i.e. a & b

$\mu_a = 16$	$\mu_b = -56$	$P(a) = 0.5$
$\sigma_a^2 = 10$	$\sigma_b^2 = 12$	$P(b) = 0.5$

E-Step ↓

$$P(x_i|a) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left\{-\frac{(x_i-\mu_a)^2}{2\sigma_a^2}\right\}$$

$$P(x_i|b) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left\{-\frac{(x_i-\mu_b)^2}{2\sigma_b^2}\right\}$$

$$a_i = P(a|x_i) = \frac{P(x_i|a)\,P(a)}{P(x_i|a)P(a) + P(x_i|b)P(b)}$$

$b_i = 1 - a_i$

Now let's apply the E-step to each of the datapoint given in dataset: [−67, −48, 6, 8, 14, 16, 23, 24]

$x_1 = -67$

$$P(x_1|a) = \frac{1}{\sqrt{2\pi \times 10^2}} \exp\left\{-\frac{(-67-16)^2}{2\times 100}\right\} = \frac{1}{25}\exp\left\{-\frac{(83)^2}{200}\right\}$$

$$= \frac{1}{25}\times 0 = 0$$

$$P(x_1|b) = \frac{1}{\sqrt{2\pi\times 144}} \exp\left\{-\frac{(-67+56)^2}{2\times 144}\right\} = \frac{1}{30}\times 0.65 = 0.021$$

$$a_1 = P(a|x_1) = \frac{0}{0+0.5\times 0.021} = 0$$

$$b_1 = P(b|x_1) = 1 - a_1 = \underline{\underline{1}}$$

$x_2 = -48$

$$P(x_2|a) = \frac{1}{25}\exp\left\{-\frac{(-48-16)^2}{200}\right\} = \frac{1}{25}\times 0 = 0$$

$$P(x_2|b) = \frac{1}{30}\exp\left\{-\frac{(-48+56)^2}{288}\right\} = \frac{1}{25}\times 0.80 = 0.026$$

$$a_2 = P(a|x_2) = \frac{0}{0+0.5\times 0.026} = 0$$

$$b_2 = 1 - a_2 = \underline{\underline{1}}$$

Scanned with CamScanner

$x_4 = 8$

$$P(x_4|a) = \frac{1}{25} \exp\left\{-\frac{(8-16)^2}{200}\right\} = 0.029$$

$$P(x_4|b) = \frac{1}{30} \exp\left\{-\frac{(8+56)^2}{288}\right\} = 0$$

$$a_4 = P(a|x_4) = \frac{0.029 \times 0.5}{0.029 \times 0.5 + 0} = 1$$

$$b_4 = 1 - a_4 = 0$$

$x_3 = 6$

$$P(x_3|a) = \frac{1}{25} \exp\left\{-\frac{(6-16)^2}{200}\right\} = \frac{1}{25} \times 0.6 = 0.024$$

$$P(x_3|b) = \frac{1}{30} \exp\left\{-\frac{(6+56)^2}{288}\right\} = \frac{1}{30} \times 0 = 0$$

$$a_3 = P(a|x_3) = \frac{0.024 \times 0.5}{0.024 \times 0.5 + 0} = 1$$

$$b_3 = 1 - a_3 = 0$$

$x_5 = 14$

$$P(x_5|a) = \frac{1}{25} \exp\left\{-\frac{(14-16)^2}{200}\right\} = \frac{1}{25} \times 0.98 = 0.039$$

$$P(x_5|b) = \frac{1}{30} \exp\left\{-\frac{(14+56)^2}{288}\right\} = \frac{1}{30} \times 0 = 0$$

$$a_5 = P(a|x_5) = 1$$

$$b_5 = 1 - 1 = 0$$

$x6 = 16$

$$P(x6|a) = \frac{1}{25} \exp\left\{-\frac{(16-16)^2}{200}\right\} = \frac{1}{25} \times 1 = 0.04$$

$$P(x6|b) = \frac{1}{25} \exp\left\{-\frac{(16+56)^2}{288}\right\} = 0$$

$a6 = 1$

$b6 = 0$

$x_7 = 23$

$$P(x_7|a) = \frac{1}{25} \exp\left\{-\frac{(23-16)^2}{200}\right\} = \frac{1}{25} \times 0.78 = 0.031$$

$$P(x_7|b) = \frac{1}{30} \exp\left\{-\frac{(23+56)^2}{288}\right\} = \frac{1}{30} \times 0 = 0$$

$a_7 = 1$

$b_7 = 0$

$x8 = 24$

$$P(x8|a) = \frac{1}{25} \exp\left\{-\frac{(24-16)^2}{200}\right\} = \frac{1}{25} \times 0.72 = 0.029$$

$$P(x8|b) = \frac{1}{25} \exp\left\{-\frac{(24+56)^2}{288}\right\} = \frac{1}{30} \times 0 = 0$$

$a8 = 1$

$b8 = 0$

M-Step

After E-Step, we have

$a_1 = a_2 = 0$ | $b_1 = b_2 = 1$

$a_3 = a_4 = a_5 = a_6 = a_7 = a_8 = 1$ | $b_3 = b_4 = b_5 = b_6 = b_7 = b_8 = 0$

$$\mu_a = \frac{a_1 x_1 + a_2 x_2 - - - a_8 x_8}{a_1 + a_2 - - - a_8}$$

$$= \frac{0+0+6+8+14+16+23+24}{6} = \underline{15.166}$$

$$\sigma_a^2 = \frac{a_1(\mu_a - x_1)^2 + a_2(\mu_a - x_2)^2 + - + a_8(\mu_a - x_8)^2}{a_1 + a_2 \;-\; - \; + a_8}$$

$$= \frac{0+0+84.02+51.35+1.35+0.69+61.37+78.03}{6} = \underline{46.136}$$

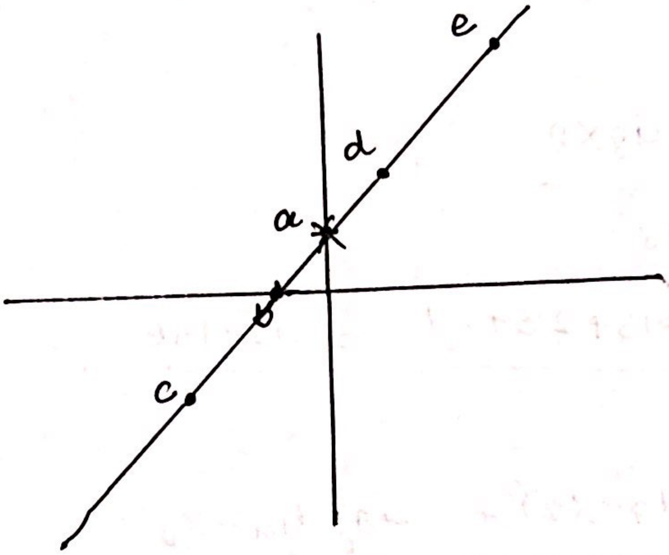
$$\mu_b = \frac{b_1 x_1 + b_2 x_2 + - - - b_8 x_8}{b_1 + b_2 + \cdot - - b_8}$$

$$= \frac{-67-48}{2} = \underline{\underline{-57.5}}$$

$$\sigma_b^2 = \frac{90.25+90.25}{2} = \underline{\underline{90.25}}$$

2) Principle Component Analysis

2.1) dataset : $\{\overset{a}{(0,1)}, \overset{b}{(-1,0)}, \overset{c}{(-3,-2)}, \overset{d}{(1,2)}, \overset{e}{(3,4)}\}$



$$\text{Mean} = \left(\frac{0+(-1)-3+1+3}{5}, \frac{1+0-2+2+4}{5}\right)$$

$$= (0, 1) \equiv \underline{a}$$

Recenter →

$$X_c \leftarrow X - \bar{X}$$

$a' = (0,0)$, $b' = (-1,-1)$, $c' = (-3,-3)$, $d' = (1,1)$, $e = (3,3)$

Principle component 1 →

$$= \frac{1}{\sqrt{1^2+1^2}}(1,1) \Rightarrow \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

Principle component 2 → (orthogonal to PC 1)

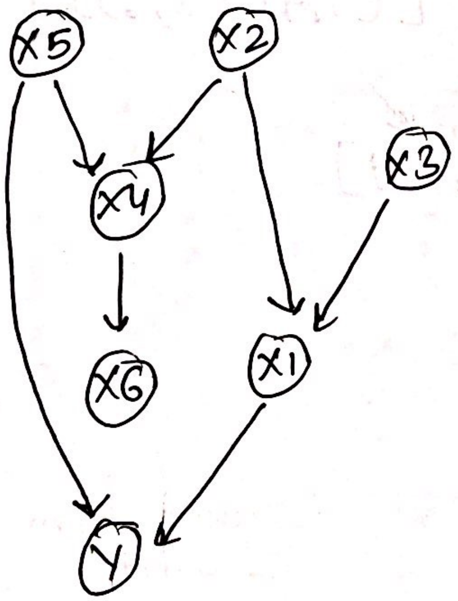
$$= \frac{1}{\sqrt{1^2+1^2}}(-1,1) \Rightarrow \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \text{ or } \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)$$

2.2) Reconstruction Error :-

$[\equiv]_{m \times n}$

for m example of a n-dimensional dataset, we can compute n principle components and if we are going to use all n for reconstruction of the data set then, the reconstruction error will be zero as we all selecting all possible n-dimension for the reconstruction.

3) Graphical Model :-



$$P(Y, X1, X2, X3, X4, X5, X6)$$
$$= P(Y \mid X5, X1) \cdot P(X1 \mid X2, X3) \cdot P(X2) \cdot P(X3) \cdot P(X4 \mid X2, X5) \cdot P(X5) \cdot P(X6 \mid X4)$$

4 K-Means

4.1

a	b	c	d
3	7	9	5
3	9	7	3

Given

cluster-1 ⇒ a, c

cluster-2 ⇒ b, d

Iteration-1

clusters = [[(3,3), (9,7)], [(7,9), (5,3)]]

Recenter-step

centroid = [[(6,5)], [(6,6)]]

Iteration-2

classify Step →

	distance from centroid 1	distance from centroid 2
a (3,3)	min($\sqrt{13}$ ✓	$\sqrt{18}$)
b (9,7)	min($\sqrt{15}$	$\sqrt{10}$ ✓)
c (7,9)	min($\sqrt{17}$	$\sqrt{10}$ ✓)
d (5,3)	min($\sqrt{5}$ ✓	$\sqrt{10}$)

So, we have new clusters ⇒

clusters = [[(3,3), (5,3)], [(9,7), (7,9)]]

Recenter step

centroids = [[(4, 3)], [(8, 8)]]

Iteration 3]

Classify step ⇒

		distance from centroid 1	distance from centroid 2
a	(3,3)	min($\sqrt{1}$ ✓	$5\sqrt{2}$)
b	(9,7)	min($\sqrt{41}$	$\sqrt{2}$ ✓)
c	(7,9)	min($\sqrt{45}$	$\sqrt{2}$ ✓)
d	(5, 3)	min($\sqrt{1}$ ✓	$\sqrt{34}$)

new clusters]

clusters = [[(3,3), (5,3)], [(9,7), (7,9)]]

Recenter step

centroids = [[(4, 3)], [(8, 8)]]

Since new clusters and new centroids in iteration 3 are equal to Iteration 2, we can say that it has converged and above clusters will be the final clusters.

4.2) Potential functn :-

$$F(\mu, C) = \sum_{j=1}^{m} \| \mu_{c(j)} - x_j \|^2$$

$$= [(1^2+0) + (1^2+0) + (1^2+1^2) + (1^2+1^2)]$$

$$= [1+1+2+2]$$

$$= 6$$

4.3) K- Means Implementation :-

Pseudo code :-

Step-1 ⇒ Process the dataset, ignore the 1st & last column as it is not needed for clustering.

Step-2 ⇒ Randomly pick the K- centroids

Step-3 ⇒ Now perform the classify & recenter step until it is converged. i.e.

classify step → Assign each of the example / data point to its nearest centroid based on ecludian distance. i.e.

$$c^{(t)}(j) \leftarrow \arg\min_i \|\mu_i - x_j\|^2$$

Recenter step → Now once each data point is assigned to some cluster, find the new centroid for each cluster which is nothing but the average of all points inside that cluster. i.e.

$$\mu_i^{(t+1)} \leftarrow \arg\min \sum_{j:c(j)=i} \|\mu - x_j\|^2$$

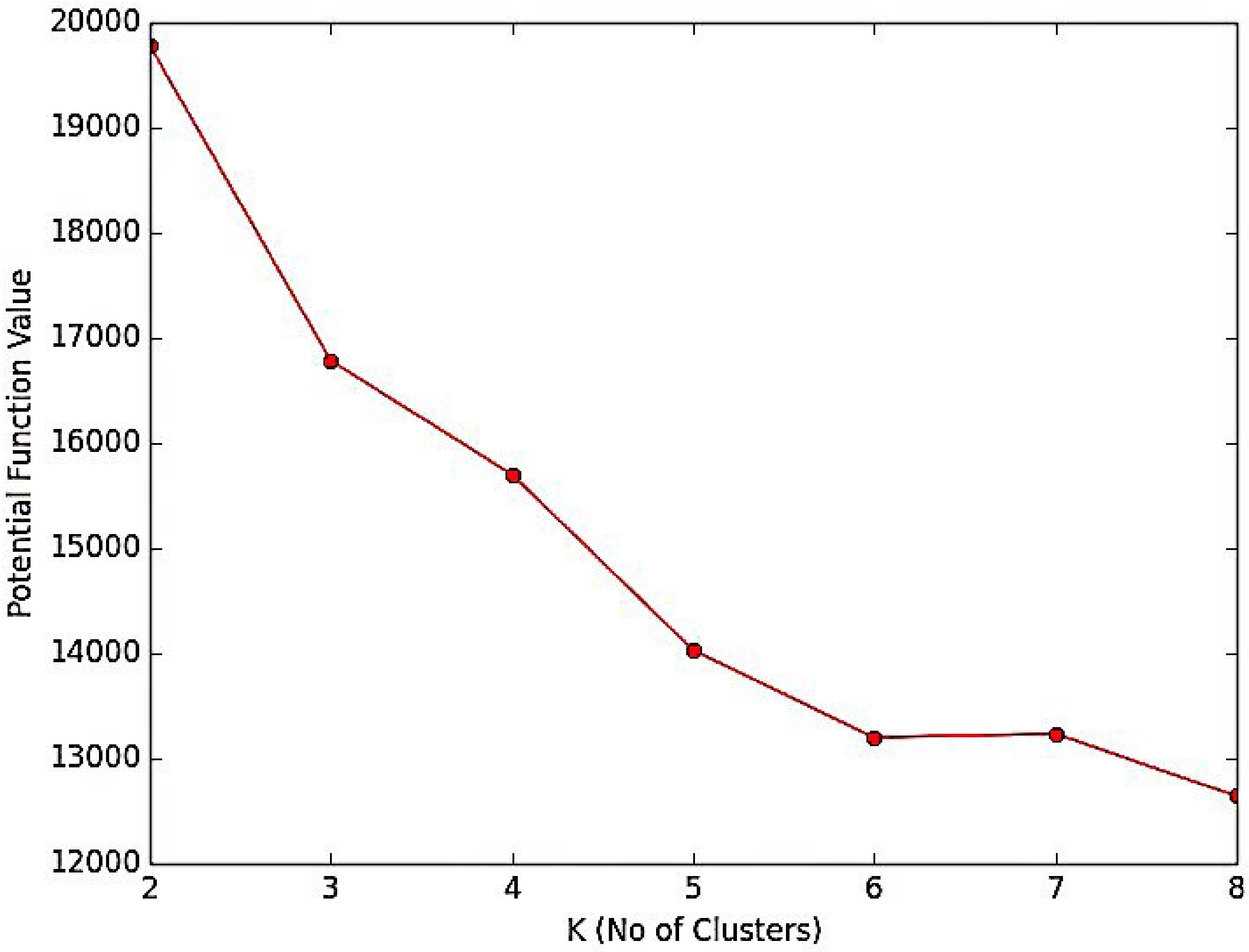
Keep performing the classify & recenter step until it is converged.

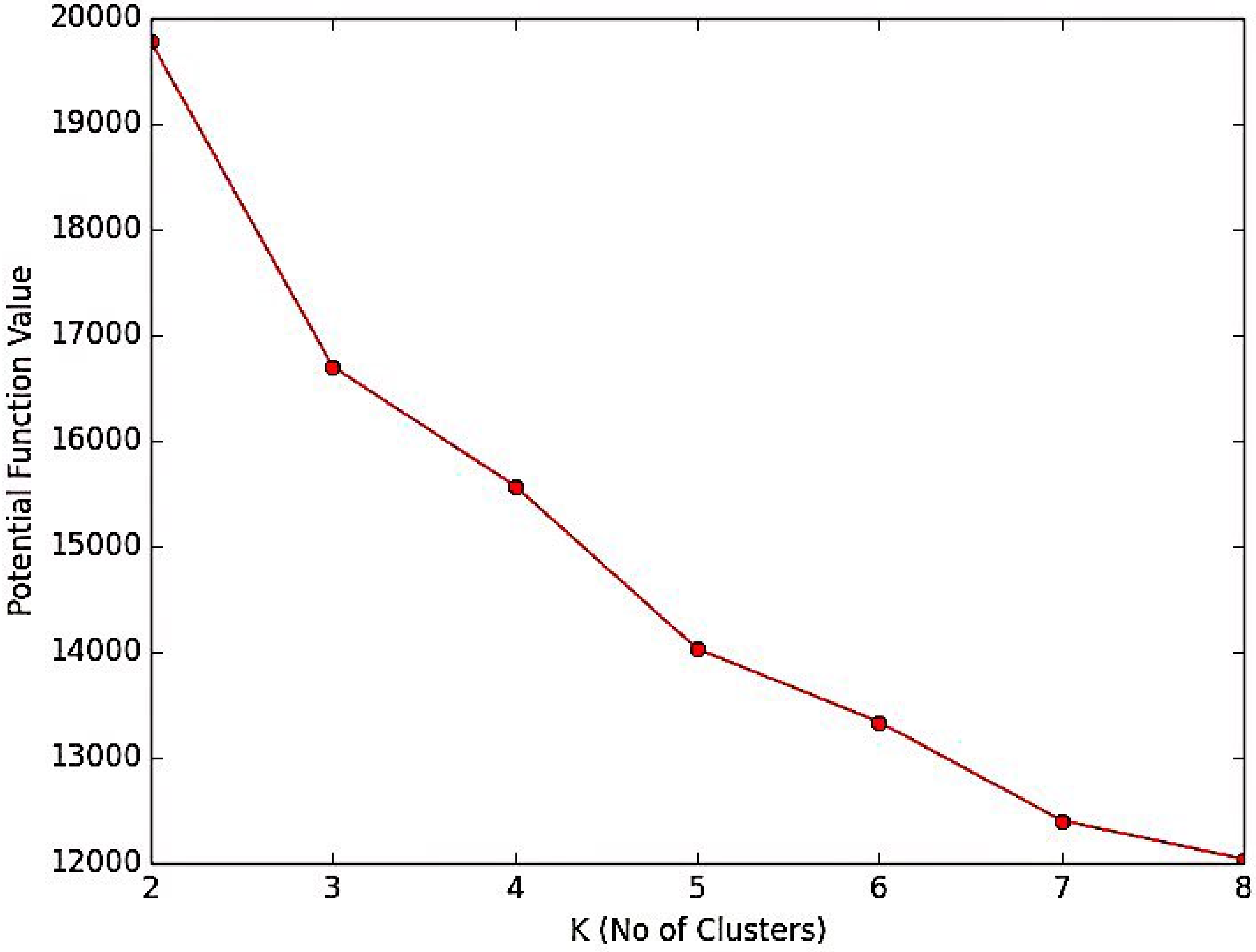
Step-4 find the potential funcn i.e.

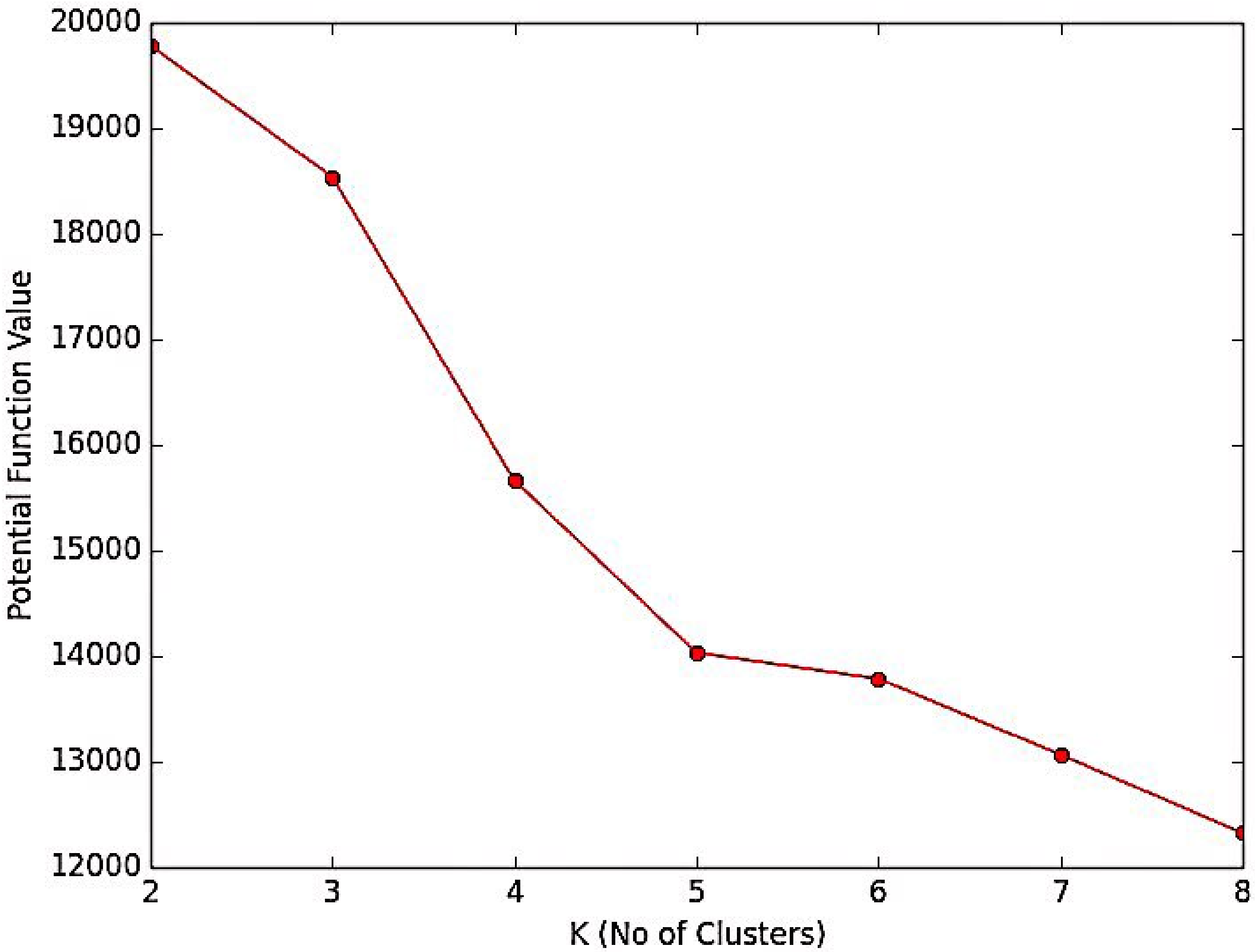
$$F(\mu, C) = \sum_{j=1}^{m} \|\mu_{c(j)} - x_j\|^2$$

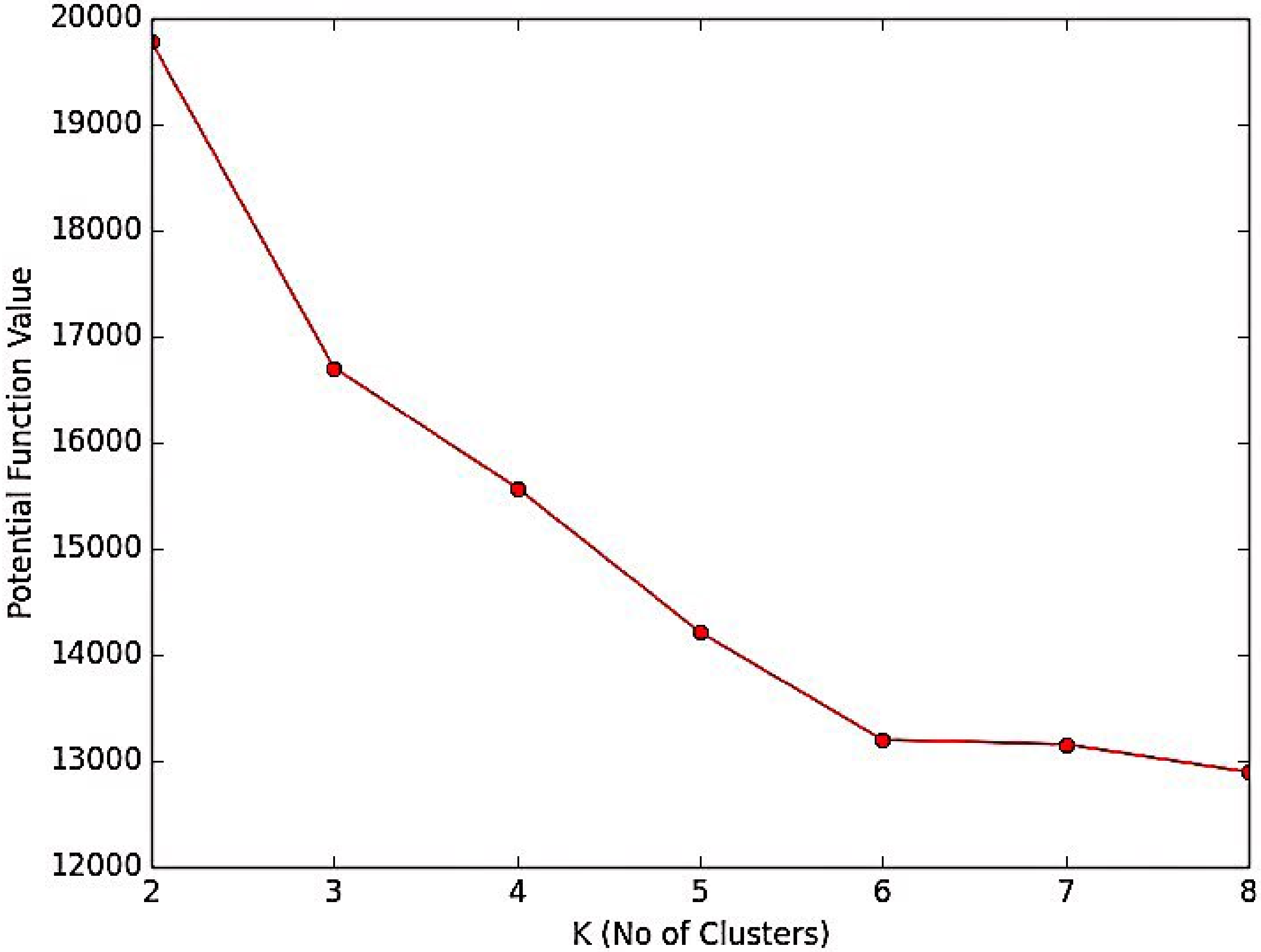
find the $\mathcal{L}(K)$ v K plot on next page.

We will use the "elbow" method for picking ~~the~~ optimal value of K. And based on the graph plot on next page, we will be choosing K=6 as our optimal value as per potential funcn value decreases abruptly at K=6 & stabilizes after that. We won't be choosing the ~~minimum~~ K value with min. potential functn. This K can change but we will always follow the elbow rule to determine it.


20000
19000
18000
17000
16000
15000
14000
13000
12000
Potential Function Value
2
3
4
5
6
7
8
K (No of Clusters)


20000
19000
18000
17000
16000
15000
14000
13000
12000
Potential Function Value
2
3
4
5
6
7
8
K (No of Clusters)


20000
19000
18000
17000
16000
15000
14000
13000
12000
Potential Function Value
2
3
4
5
6
7
8
K (No of Clusters)


20000
19000
18000
17000
16000
15000
14000
13000
12000
Potential Function Value
2
3
4
5
6
7
8
K (No of Clusters)

```
[Ashishs-MacBook-Pro:HW3 ashish$ python KMeans.py                                  ]
[19782.729804245933, 16713.777153738938, 15575.500891706823, 14032.417674323131,
 13354.78176780874, 13050.624206971424, 13031.282286224101]
```