1 GMM and EM algorithm

1-d data set: $\{-67, -48, 6, 8, 14, 16, 23, 24\}$

$K = 2$ (2 components)

1.1 There are 5 independent parameters. They are

i) $P(Y=1) \rightarrow$ Prior

ii) Mean of component 1 (μ_1)

iii) Mean of component 2 (μ_2)

iv) Variance of component 1 (σ_1^2)

v) Variance of component 2 (σ_2^2)

1.2 Initialization

Component 1: Prior : $P(Y=1) = 0.25$ Component 2: $P(Y=2) = 0.75$

$\mu_1 = -10$, $\mu_2 = 10$, $\sigma_1^2 = 20$, $\sigma_2^2 = 20$

E-Step :-

$$P(Y=i / X_j, \mu_1, \ldots, \mu_K) \propto \exp\left(\frac{-1}{2\sigma_i^2} \|X_j - \mu_i\|^2\right) P(Y=i)$$

$$P(Y=1 / X_1, \mu_1, \mu_2) \propto \exp\left(\frac{-1}{2(20)^2} \|-67+10\|^2\right)(0.25)$$

~~(after normalization)~~ $\simeq 1 \quad \therefore P(Y=2 / X_1, \mu_1, \mu_2) \simeq 0$

$$P(Y=1 / X_2, \mu_1, \mu_2) \propto \exp\left(\frac{-1}{2(20)^2} \|-48+10\|^2\right)(0.25)$$

$\simeq 1 \quad \therefore P(Y=2 / X_2, \mu_1, \mu_2) \simeq 0$

$$P(Y=1 / X_3, \mu_1, \mu_2) \propto \exp\left(\frac{-1}{2(20)^2} \|6+10\|^2\right)(0.25)$$

$= 8.26 \times 10^{-4}$

$\therefore P(Y=2 / X_3, \mu_1, \mu_2) = 9.99 \times 10^{-1} \simeq 1$

$P(Y=1 / X_4, \mu_1, \mu_2) \propto \exp\left(\frac{-1}{2(20)^2} \|8+10\|^2\right)(0.25)$

$\simeq 1.12 \times 10^{-4}$

$P(Y=2 / X_4, \mu_1, \mu_2) = 9.99 \times 10^{-1} \simeq 1$

$P(Y=1 / X_5, \mu_1, \mu_2) \propto \exp\left(\frac{-1}{2(20)^2} \|14+10\|^2\right)(0.25)$

$= 2.77 \times 10^{-7} \simeq 0$

$P(Y=2 / X_5, \mu_1, \mu_2) = 9.99 \times 10^{-1} \simeq 1$

$P(Y=1 / X_6, \mu_1, \mu_2) \propto \exp\left(\frac{-1}{2(20)^2} \|16+10\|^2\right)(0.25)$

$= 3.75 \times 10^{-8} \simeq 0$

$P(Y=2 / X_6, \mu_1, \mu_2) = 9.99 \times 10^{-1} \simeq 1$

$P(Y=1 / X_7, \mu_1, \mu_2) \propto \exp\left(\frac{-1}{2(20)^2} \|23+10\|^2\right)(0.25)$

$= 3.42 \times 10^{-11} \simeq 0$

$P(Y=2 / X_7, \mu_1, \mu_2) \propto \exp\left(\frac{-1}{2(20)^2} \|23-10\|^2\right)(0.75)$

$= 1$

$P(Y=1 / X_8, \mu_1, \mu_2) \propto \exp\left(\frac{-1}{2(20)^2} \|24+10\|^2\right)(0.25)$

$\simeq 1.26 \times 10^{-11} \simeq 0$

$P(Y=2 / X_8, \mu_1, \mu_2) = 1$

Scanned by CamScanner

M - Step

Updating all the parameters based on the results of E-Step.

$P(Y=1) = \frac{2}{8} = 0.25$ ∴ $P(Y=2) = \frac{6}{8} = 0.75$

$$\mu_i = \frac{\sum_{j=1}^{m} P(Y=i/X_j)\, X_j}{\sum_{j=1}^{m} P(Y=i/X_j)}$$

$$\therefore \mu_1 = \{1 \times (-67) + 1(-48) + (8.26 \times 10^{-4})6 + (1.12 \times 10^{-4})8 + (2.77 \times 10^{-7})14 + (3.75 \times 10^{-8})16 + (3.42 \times 10^{-11})23 + (1.26 \times 10^{-11})24\} / \{1 + 1 + 8.26 \times 10^{-4} + 1.12 \times 10^{-4} + \cancel{1.26 \times 10^{-11}} + 2.77 \times 10^{-7} + 3.75 \times 10^{-8} + 3.42 \times 10^{-11} + 1.26 \times 10^{-11}\}$$

$$= -57.47$$

$$\mu_2 = \{2.39 \times 10^{-29}(-67) + (4.28 \times 10^{-21})(-48) + (9.99 \times 10^{-1})6 + (9.99 \times 10^{-1})(8) + (9.99 \times 10^{-1})(14) + (9.99 \times 10^{-1})16 + 1(23) + 1(24)\} / \{2.39 \times 10^{-29} + 4.28 \times 10^{-21} + 9.99 \times 10^{-1} + 9.99 \times 10^{-1} + 9.99 \times 10^{-1} + 9.99 \times 10^{-1} + 1 + 1\}$$

$$= 15.168$$

$$\sigma_i^2 = \frac{\sum_{j=1}^{m} P(Y=i/X_j)\,(X_j-\mu_i)^2}{\sum_{j=1}^{m} P(Y=i/X_j)}$$

$$\sigma_1^2 = \{1 \times (-67+57.47)^2 + 1\,(-48+57.47)^2 + (8.26\times10^{-4})(6+57.47)^2 + (1.12\times10^{-4})(8+57.47)^2 + (2.77\times10^{-7})(14+57.47)^2 + (3.75\times10^{-8})(16+57.47)^2 + (3.42\times10^{-11})(23+57.47)^2 + (1.26\times10^{-11})(24+57.47)^2\}$$

$$/\{1+1+8.26\times10^{-4} + 1.12\times10^{-4} + 2.77\times10^{-7} + 3.75\times10^{-8} + 3.42\times10^{-11} + 1.26\times10^{-11}\}$$

$$= 92.11$$

$$\sigma_2^2 = \{(2.39\times10^{-29})(-67-15.17)^2 + (4.28\times10^{-21})(-48-15.17)^2 + (9.99\times10^{-1})(6-15.17)^2 + (9.99\times10^{-1})(8-15.17)^2 + (9.99\times10^{-1})(14-15.17)^2 + (9.99\times10^{-1})(16-15.17)^2 + 1\,(23-15.17)^2 + 1\,(24-15.17)^2\}$$

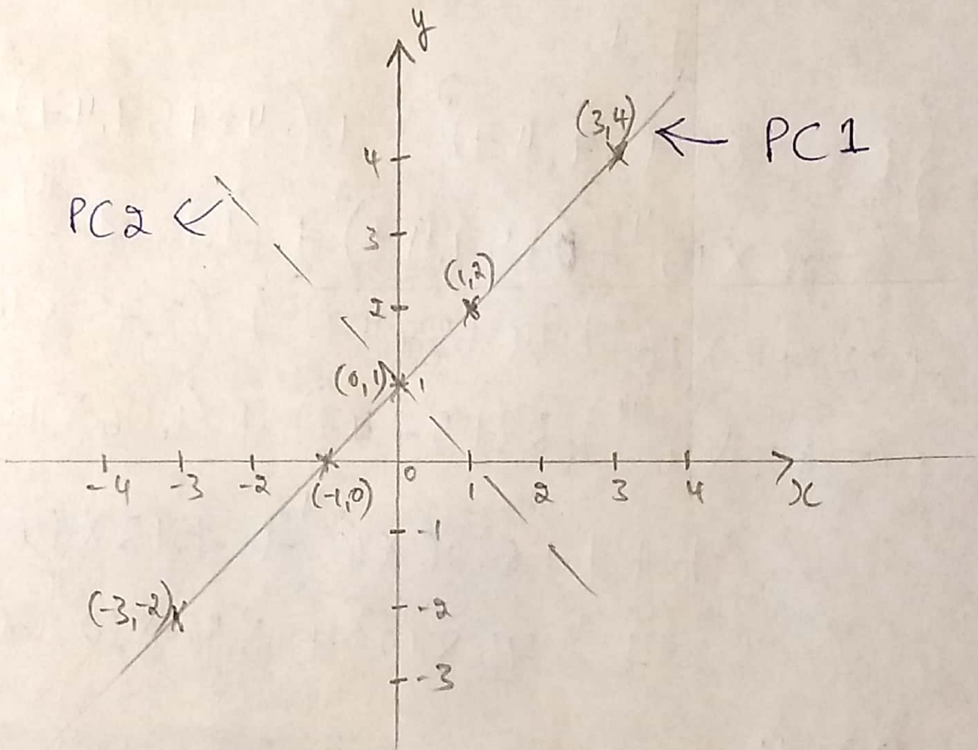
$$/\{2.39\times10^{-29} + 4.28\times10^{-21} + 9.99\times10^{-1} + 9.99\times10^{-1} + 9.99\times10^{-1} + 9.99\times10^{-1} + 1 + 1\}$$

$$= 46.13$$

2 Principal Component Analysis

2.1 Principal Components

2-d data set : $\{(0,1), (-1,0), (-3,-2), (1,2), (3,4)\}$



First Principal Component : This will be the line passing through all the data points. Entire variance will be explained by the first principal component itself. So the second principal component will be insignificant.

$$\frac{y-y_1}{x-x_1} = \frac{y_2-y_1}{x_2-x_1} \quad \left[\because \text{Equation of a line passing through } (x_1,y_1) \text{ \& } (x_2,y_2)\right]$$

$$\frac{y-1}{x-0} = \frac{0-1}{-1-0} \quad \left[(0,1) \text{ \& } (-1,0)\right]$$

$$\therefore y = x+1 \quad , \quad \text{Slope} = m_1 = 1 \quad \left[\because y = mx+c\right]$$

Angle from x-axis $= \tan^{-1}(m_1) = 45^\circ$

~~Second principal~~

Vector form of PC1 : $(\sin 45^\circ, \cos 45^\circ) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

Second principal Component :-

This will be orthogonal to first principal component and hence slope will be -1. i.e $m_2 = -1$

$$\frac{y - y_1}{x - x_1} = m_2$$

[Eq of a line with slope m & passing through (x_1, y_1)]

$$\frac{y - 1}{x - 0} = -1$$

$\therefore y = -x + 1$ is the equation of second PC.

$\angle$le between x-axis & 2^{nd} PC will be $\tan^{-1}(m_2)$

$$\tan^{-1}(m_2) = \tan^{-1}(-1) = 135^\circ$$

Vector form of PC2 : $(\sin 135^\circ, \cos 135^\circ) = \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)$

2.2 Reconstruction error

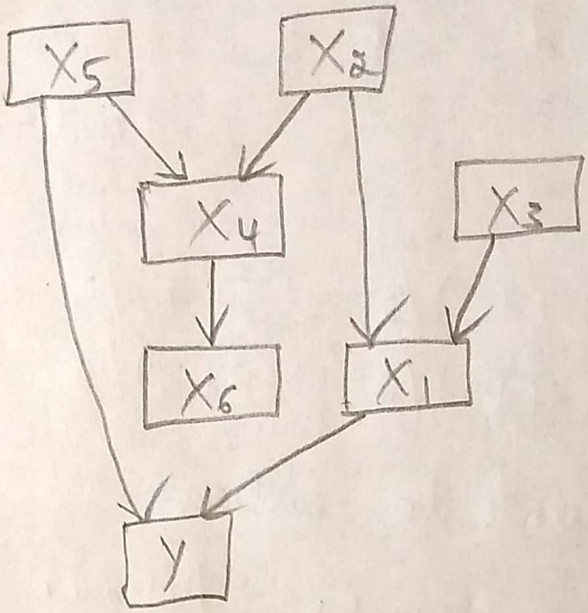
'n' principle components can be computed.

If we take top n [i.e all] components for ~~reconstru~~ reconstruction, then reconstruction error will be 0.

Reconstruction error, $error_K = m \sum_{j=K+1}^{n} u_j^T \Sigma u_j$ if K components out of 'n' are used for reconstruction.

Since $K = n$, in this case, $error_n = 0$.

3 Graphical Models



Sol^n According to representation theorem,

$$P(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{N} P(X_i \mid \text{Parent}_{X_i})$$

$$\therefore P(Y, X_1, X_2, X_3, X_4, X_5, X_6) =$$

$$= P(Y \mid X_5, X_1) . P(X_1 \mid X_2, X_3) P(X_2) . P(X_3) . P(X_4 \mid X_5, X_2) . P(X_5) .$$

$$. P(X_6 \mid X_4)$$

4. K-Means

4.1

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

$k = 2$

Cluster 1: a, c
Cluster 2: b, d

Iteration 1:- Centroid 1 : $(9+3)/2 \,,\, (3+7)/2 = 6,5$
Centroid 2 : $(7+5)/2 \,,\, (9+3)/2 = 6,6$

Re-assigning points to clusters

	Distance from C_1	Distance from C_2	Cluster
a	$\sqrt{(3-6)^2+(3-5)^2} = \sqrt{13}$	$\sqrt{(3-6)^2+(3-6)^2} = \sqrt{18}$	C_1
b	$\sqrt{(7-6)^2+(9-5)^2} = \sqrt{17}$	$\sqrt{(7-6)^2+(9-6)^2} = \sqrt{10}$	C_2
c	$\sqrt{(9-6)^2+(7-5)^2} = \sqrt{13}$	$\sqrt{(9-6)^2+(7-6)^2} = \sqrt{10}$	C_2
d	$\sqrt{(5-6)^2+(3-5)^2} = \sqrt{5}$	$\sqrt{(5-6)^2+(3-6)^2} = \sqrt{10}$	C_1

Cluster 1 : a, d
Cluster 2 : b, c

Iteration 2:- Centroid 1 : $(3+5)/2, (3+3)/2 = 4,3$
Centroid 2 : $(7+9)/2, (9+7)/2 = 8,8$

Re-assigning points to clusters

	Distance from C_1	Distance from C_2	Cluster
a	$\sqrt{(3-4)^2+(3-3)^2} = \sqrt{1}$	$\sqrt{(3-8)^2+(3-8)^2} = \sqrt{50}$	C_1
b	$\sqrt{(7-4)^2+(9-3)^2} = \sqrt{45}$	$\sqrt{(7-8)^2+(9-8)^2} = \sqrt{2}$	C_2
c	$\sqrt{(9-4)^2+(7-3)^2} = \sqrt{41}$	$\sqrt{(9-8)^2+(7-8)^2} = \sqrt{2}$	C_2
d	$\sqrt{(5-4)^2+(3-3)^2} = \sqrt{1}$	$\sqrt{(5-8)^2+(3-8)^2} = \sqrt{34}$	C_1

Cluster 1 : a, d
Centroid 1 : $(3+5)/2, (3+3)/2 = 4,3$
Cluster 2 : b, c
Centroid 2 : $(7+9)/2, (9+7)/2 = 8,8$

Since the clusters & centroids have converged, stopping the iteration.

4.2 Potential Function

$$L(K) = \sum_{j=1}^{m} \|\mu_{c(j)} - x_j\|^2$$

= Sum of squared distances from points to their centroids

$= (\text{dist from } a \text{ to } c_1)^2 + (b \text{ to } c_2)^2 + (c \text{ to } c_2)^2 + (d \text{ to } c_1)^2$

$= (\sqrt{1})^2 + (\sqrt{2})^2 + (\sqrt{2})^2 + (\sqrt{1})^2$

$= 1 + 2 + 2 + 1$

$L(K) = \underline{\underline{6}}$

Scanned by CamScanner

4.3) **K-Means Clustering**

Pseudocode:

Repeat for different values of K

Centroids ← Randomly select K centroids for the dataset

Previous Centroids ← None

While Previous Centroids NOT EQUAL TO Centroids AND Number_of_iterations LESS THAN 1000

Previous Centroids ← Centroids

Assign all data points a cluster based on the centroid the points are closest to

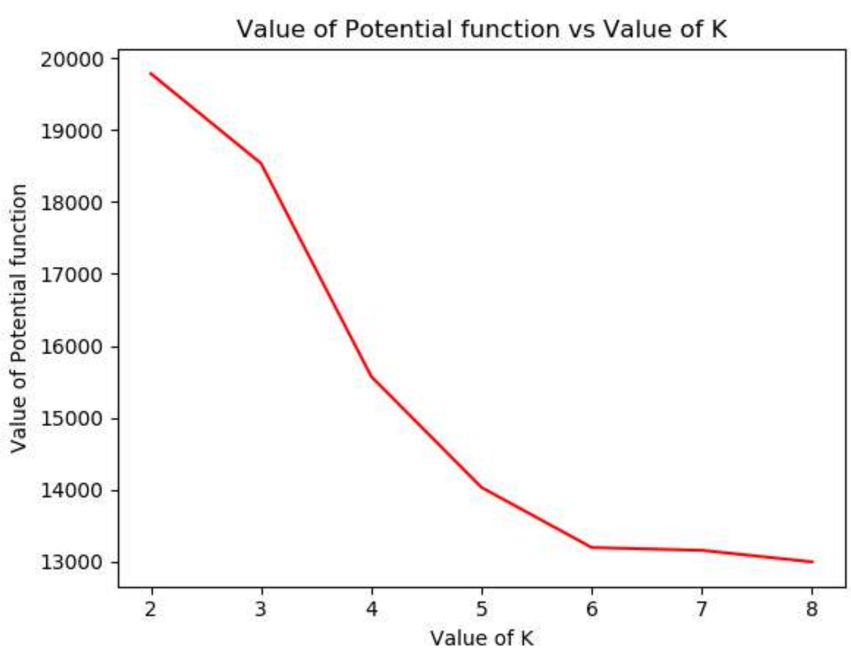
Centroids ← Re-Compute the centroids for all the clusters using average values

Number_of_iterations ← Number_of_iterations + 1

F(mu, C) ← Compute potential function as a sum of square of distance from the point to its centroid for all data points

Plot the graph of value of K vs value of Potential function

Graph:



If you were to pick the optimal value of K based on this curve, would you pick the one with the lowest value of the potential function? Why?

No, I won't pick the value of K at which potential function is lowest as the optimal value of K. Value of potential function will be lowest (0) when K is equal to number of samples. This doesn't mean optimal value of K is number of samples. So, elbow method helps to determine the optimal value of K. The above line graph resembles the shape of an arm. Elbow method says that, the value of K at which one can observe the clear elbow like shape (Elbow point) is the optimal value of K. Elbow point specifies the point after which the results are not good. Hence in the graph, I can observe an elbow point at K=6 and hence I will choose K=6 as the optimal value of K for this dataset.