Unsupervised Learning :

$$D = \{x_i\}_{i=1}^{n}$$

* Clustering
* Dimensionality reduction.
* Outlier Analysis.

Clustering :

$D = \{x_i\}_{i=1}^{n}$ $\quad x_i \in \mathbb{R}^d$

$D \rightarrow \{c_1, c_2, \ldots c_k\}$

cluster representative
centroid

$c_k = \{ x_i \mid \text{Distance}(x_i, \mu_k) = \text{minimum} \}$ (*)

2 techniques

Partitioning

ex. K-means
K-mediod

Hierarchical

Top down

ex. Diana

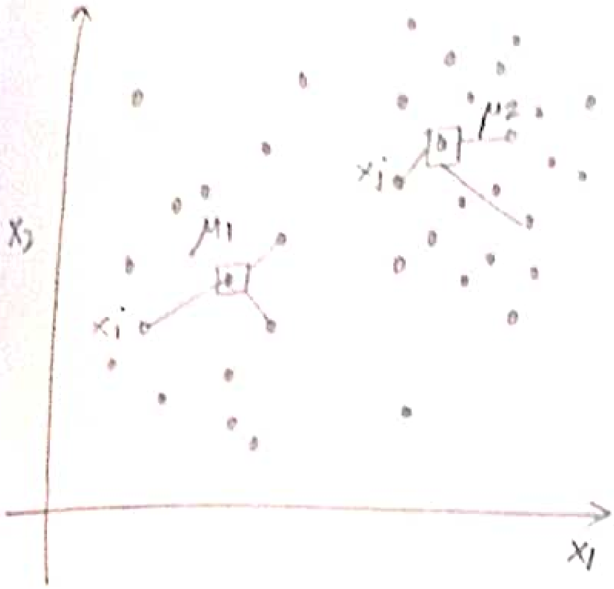
Issue! Divisibility Analysis

Bottom up.

ex. AgNes

Agglomerative nesting

K-Means :



assume $k = 2$

iteration 1

if no domain knowledge choose μ_1, μ_2 randomly from dataset

assign x_i to μ_k based on distance. (*)

iteration 2

$\mu_1 = \text{mean}(x_i \in C_1)$
$\mu_2 = \text{mean}(x_i \in C_2)$

Thus from iteration 2.
μ_1, μ_2 centroids need not be actual datapoints

Apply K-Means clustering alg. for foll. data and formulate 2 clusters. Use Euclidean distance as distance metrix. Use random centroids.

	x_1	x_2	x_3	x_4	x_5
Distance	71.24	52.53	64.54	55.69	54.58
Speed	28	25	27	22	25

①

$\mu_1 = (64.54, 27) = x_3$
$\mu_2 = (55.69, 22) = x_4$

$d(x_1, \mu_1) = \cancel{\sqrt{5.89}}\ \sqrt{45.89}$
$d(x_1, \mu_2) = \sqrt{114.3}$

$d(x_2, \mu_1) = \cancel{\sqrt{145.2}}\ \sqrt{153.2}$
$d(x_2, \mu_2) = \sqrt{18.9}$

$d(x_5, \mu_1) = \sqrt{103.2}$
$d(x_5, \mu_2) = \sqrt{10.23}$

$C_1 = \{x_1, x_3\}$

$C_2 = \{x_2, x_4, x_5\}$

$\mu_1 = (67.89, 27.5)$
$\mu_2 = (54.26, 24)$

Measuring the performance of KMeans

Distortion/ inertia. $= \Sigma\,(d(x_i, \mu_i))^2$

$net_j = \sum_{i=0}^{d} x_i w_{ij}$

<u>one hot encoding</u> for output classes in multiclass classification

eg: Iris dataset.
(3 classes).

∴ output

class 1	class 2	class 3
1	0	0
0	1	0
0	0	1.

} one hot encoding.

K-means clustering :

Assumption :
i. closeness assumption

Form $c_1, \dots, c_k$

min distortion / intracluster distance.

$$D = \sum_{x_i \in \mu_j} \| x_i - \mu_j \|^2$$

Algorithm :

1. initialize the centroids $\mu_1, \dots, \mu_k$ randomly

2. while ! convergence → no change in centroids in successive iterations

   a) for each $x_i \in D$ // cluster formation

   x_i is assigned $\mu_i = \underset{\mu_j}{\text{argmin}} \| x_i - \mu_j \|^2 \quad 1 \le j \le k.$

   b) for each cluster c_j. // centroid updation

$$\mu_j = \frac{\sum_{x_i \in \mu_j} x_i}{n_j}$$

1 iteration : $O(nkd)$

Limitations :

& each cluster ~ $N(\mu, \varepsilon)$

i. works only for spherical clusters.

ii. Convergence depends on selection of initial centroids.

KNN also called as mixture of gaussians.

as each cluster (as assumed) follows gaussian distribution.

Spectral Clustering

i. Based on spectral properties of graph.

ii. Given dataset D is represented as a graph G.

iii. Nodes : Datapoints.
Edges : Connection between the nodes.

iv. G may be directed / undirected.
* adjacency matrix.
* Similarity matrix.
* K-neerest neighbour.

Algorithm :

Input : Dataset.
Output : clusters.

1. Formulate graph A.

2. Find degree matrix D as a diagonal matrix in which diagonals are degree.

3. Find Laplacian $L = D - A$
in which Diagonals → degree
off diagonals → negative edge waits.

4. Find eigen values of L.
\# zero eigen values = \# connected components / clusters.

First non-zero eigen value = Fiedler value.
smallest
gives the graph cut*

* no. of edges to be removed to get connected components

5. Find the eigen vector $\underline{v}$ for the fiedler value.
fiedler vector.

6. For each value x_m of eigen vector V.

a) if $x_m > 0$

$x_m \rightarrow C_1$

if $x_m < 0$

$x_m \rightarrow C_2$.

For the data set used in k means clustering

$$
\begin{array}{cc}
 & \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\
A = & \begin{bmatrix}
0 & \sqrt{359} & \sqrt{45.89} & \sqrt{277} & \sqrt{286} \\
\sqrt{359} & 0 & \sqrt{148} & \sqrt{18.9} & \sqrt{4.2} \\
\sqrt{45.89} & \sqrt{148} & 0 & \sqrt{103} & \sqrt{103.2} \\
\sqrt{277} & \sqrt{18.9} & \sqrt{103} & 0 & \sqrt{10.2} \\
\sqrt{286} & \sqrt{4.2} & \sqrt{103.2} & \sqrt{10.2} & 0
\end{bmatrix}
\end{array}
$$

adjacency matrix

$$
A = \begin{bmatrix}
0 & 18.95 & 6.77 & 16.67 & 16.93 \\
18.95 & 0 & 12.17 & 4.36 & 2.05 \\
6.77 & 12.17 & 0 & 10.16 & 11.14 \\
16.67 & 4.36 & 10.16 & 0 & 3.19 \\
16.93 & 2.05 & 11.14 & 3.19 & 0
\end{bmatrix}
$$

2 nearest neighbours. wrong/sus.

$$
A = \begin{bmatrix}
0 & 0 & 6.77 & 16.67 & 0 \\
0 & 0 & 0 & 4.36 & 2.05 \\
6.77 & 0 & 0 & 10.16 & 0 \\
16.67 & 4.36 & 10.16 & 0 & 3.19 \\
0 & 2.05 & 0 & 3.19 & 0
\end{bmatrix}
$$

$$D = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

D - A

$$L = \begin{bmatrix} 4 & -18.95 & -6.77 & -16.67 & -16.93 \\ -18.95 & 4 & -12.17 & -4.36 & -2.05 \\ -6.77 & -12.17 & 4 & -10.16 & -11.14 \\ -16.67 & -4.36 & -10.16 & 4 & -3.19 \\ -16.93 & -2.05 & -11.14 & -3.19 & 4 \end{bmatrix}$$

Properties of symmetric vectors :

i. all eigen values are real & positive.

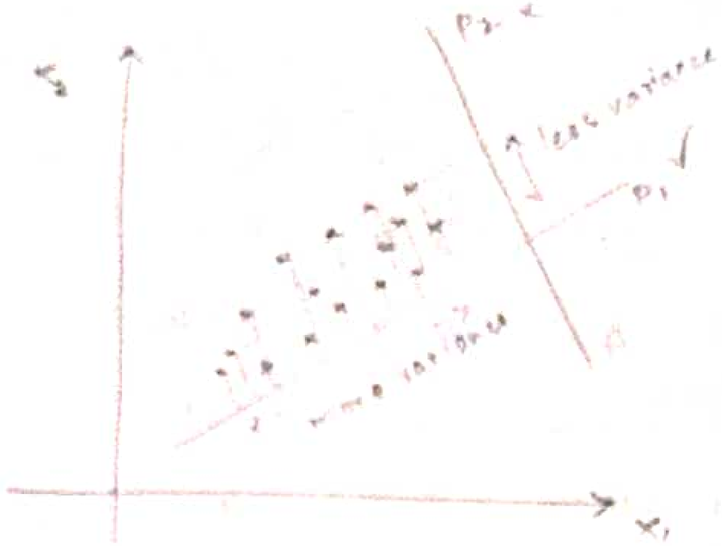
ii. all eigen vectors of eigen values are real & orthogonal.

iii. Covariance matrix Σ is symmetric. and sum of diagonals is sus!
sum of eigen values = sum of variance of 'd' features.

$$\lambda_1 + \lambda_2 + \cdots + \lambda_d = \sum_{j=1}^{d} \sigma_j^2 .$$

Dimensionality Reduction :

* PCA (Principal Component Analysis) (unsupervised)
* LDA (Linear Discriminant Analysis) (supervised)



find new dimensions

dimensions/axes should be orthogonal

a choice would be orthogonal eigen vectors (for real symmetric matrix)

The vectors chosen as new axes must be in direction of max variance.

Input $x = \{x_1, x_2, \ldots, x_n\}$ $[x_1 \ x_2 \ldots x_n]$

$x_i \in \mathbb{R}^d$

$k \times 1$ $\quad z = W^T X \longrightarrow (k \times d)(d \times 1)$

$z \in \mathbb{R}^k \quad , \ k < d.$

1. $\mu = \{\mu_1, \mu_2, \ldots, \mu_d\}$

2. mean subtracted data $\quad x' = x - \mu$ $\quad$ brings data around the origin

3. Calculate covariance matrix Σ $\quad$ because will be zero

$$\Sigma_{ij} = \sum_{k=1}^{n} \frac{(x_{ki} - \mu_i)(x_{kj} - \mu_j)}{n-1}$$

$$= \frac{\sum_{k=1}^{n} (x_{ki})(x_{kj})}{n-1}$$

4. Eigen values of Σ : $\lambda_1, \lambda_2, \ldots, \lambda_d$.

5. Arrange the Eigen values in descending order.

6. Choose the 'k' largest eigen values, $\lambda_1, \lambda_2, \ldots, \lambda_k$ and find corresponding eigen vectors. $W_1, W_2, \ldots, W_k$.

$$W = [W_1 \; W_2 \; \ldots \; W_k] \qquad |W_i| = d \times 1.$$

$$|W| = d \times k.$$

7. $$\underset{k \times n}{Z} = (\underset{k \times d}{W^T} \underset{d \times n}{X})^T$$

	x_1	x_2
P_1	2	6
P_2	1	7
$\mu =$	1.5	6.5

$$\frac{0.25 + 0.25}{2-1}$$

mean subtracted data:

	x_1	x_2
P_1	0.5	-0.5
P_2	-0.5	0.5

Covariance matrix. $\Sigma_{2 \times 2}$

$$\Sigma = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}$$

compute eigen values:

$$|\Sigma - \lambda I| = 0.$$

$$\begin{vmatrix} 0.5 - \lambda & -0.5 \\ -0.5 & 0.5 - \lambda \end{vmatrix} = 0$$

$$(0.5 - \lambda)^2 - 0.25 = 0.$$

$$(0.5 - \lambda)^2 = 0.25.$$

$$0.5 - \lambda = \pm\sqrt{0.25}$$
$$= \pm 0.5$$

here.
$d = 2$
$k = 1$

$\boxed{\lambda = 0.}$

$\boxed{\lambda = 1.}$ ✓ choose this λ_1

Calculate eigen vector for $\underline{\lambda = 1}$.

$$(\Sigma - \lambda I)X = 0.$$

$$\begin{bmatrix} -0.5 & -0.5 \\ -0.5 & -0.5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$-0.5x_1 \quad -0.5x_2 = 0.$$

$$-0.5x_1 \quad -0.5x_2 = 0.$$

$$-0.5x_1 = 0.5x_2$$

$$-x_1 = x_2$$

if $x_1 = k$, $x_2 = -k$.

$x_1 = 1$
$x_2 = -1$. $\Rightarrow$ unit vector. $\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

$$\therefore \quad W = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$$

$$z = W^T X.$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{array}{c} \begin{array}{cc} P_1 & P_2 \end{array} \\ \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix} \end{array}$$

$$= \frac{1}{\sqrt{2}} \quad -\frac{1}{\sqrt{2}}$$