CSE 472

Machine Learning Sessional

Dimensionality Reduction using Principal Component Analysis and Clustering using Expectation-maximization Algorithm
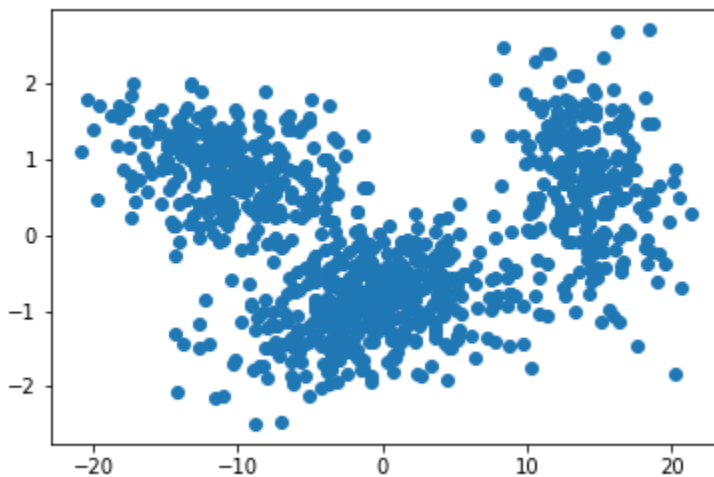
Zahin Wahab

*St ID:***1505031**

Principal Component Analysis (PCA):

Two highest eigenvalues: 98.20975315996711,1.0677144032146537

PCA plot:

Expectation-maximization(EM) algorithm:

From the above PCA plot, the number of Gaussian distributions, K= 3
Dimension of vector, D= 2

Initial means: $\begin{bmatrix} -4.26209577 & 0.02871596 \\ 7.80382482 & 2.04719386 \\ -6.31811969 & 0.50086761 \end{bmatrix}$

Initial Co-variances: $\begin{bmatrix} \begin{bmatrix} 1.16393397e+02 & -1.22512675e-01 \\ -1.22512675e-01 & 1.06853983e+00 \end{bmatrix} \\ \begin{bmatrix} 1.59170396e+02 & 1.59919342e+01 \\ 1.59919342e+01 & 5.26291230e+00 \end{bmatrix} \\ \begin{bmatrix} 1.38168348e+02 & -3.16770918e+00 \\ -3.16770918e+00 & 1.31883388e+00 \end{bmatrix} \end{bmatrix}$

Initial mixing coefficients: $\begin{bmatrix} 0.33 & 0.33 & 0.34 \end{bmatrix}$

Final means: $\begin{bmatrix} -0.65490696 & -0.91103357 \\ 14.23271415 & 0.67616318 \\ -10.41170435 & 0.8563192 \end{bmatrix}$

Final Co-variances: $\begin{bmatrix} \begin{bmatrix} 22.04644472 & 0.83026857 \\ 0.83026857 & 0.26001577 \end{bmatrix} \\ \begin{bmatrix} 6.92272453 & -0.47558014 \\ -0.47558014 & 0.66884288 \end{bmatrix} \\ \begin{bmatrix} 16.4855099 & -0.61339994 \\ -0.61339994 & 0.24986888 \end{bmatrix} \end{bmatrix}$

Final mixing coefficients: $\begin{bmatrix} 0.46002022 & 0.24035382 & 0.29962595 \end{bmatrix}$

Gaussian Distribution of PCA analysed data (K=3) using EM algorithm: