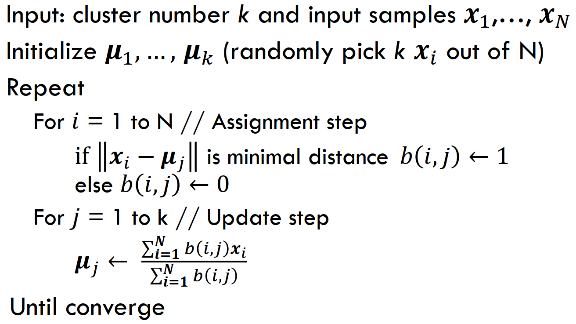
HW6  
Student ID: 109598001, Student Name: Hao-Ying Cheng

1. a dataset 𝑆 = {[0, 0] , [2,2] , [0,1] , [1,1]}, Use k = 2 and initial conditions 𝝁1 = [−1,−1] and 𝝁2 = [3,3] in the computation in one iteration.

先計算與mean之間的距離，依據最小距離而歸類於某一類別。

取自 Clustering Algorithm PPT p.7.

D1 (distance between 𝝁1) = [1.41, 4.24, 2.24, 2.83]

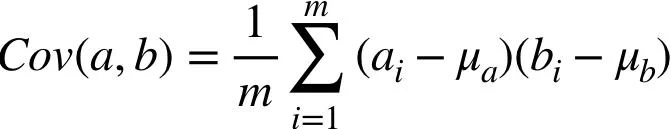
D2 (distance between 𝝁2) = [4.24, 1.41, 3.61, 2.83]

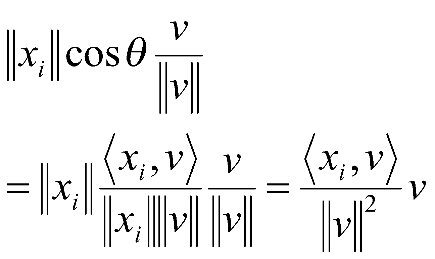
C1 (Class 1) = [[0,0], [0,1], [1,1]], C2 (Class 2) = [[2,2]]

New\_ 𝝁1 = [0.33, 0.67], New\_ 𝝁2 = [2, 2]

可參考jupyter file— HW6\_Problem1\_Kmeans.ipynb

2. 計算Principle Component Analysis (PCA) 資料集投影的數據點:

Covariance: 

投影向量: 

求Covariance的Eigenvalue & Eigenvector: 

Assume these data points for matrix A and the float limits 2 decimals (e.g., 3.067 🡺 3.06):

A = [[1,1], [4,4], [5,5]],

Mean(A) = M = [3.33, 3.33], C = AM [[-2.33, -2.33], [0.67, 0.67], [1.67, 1.67]]

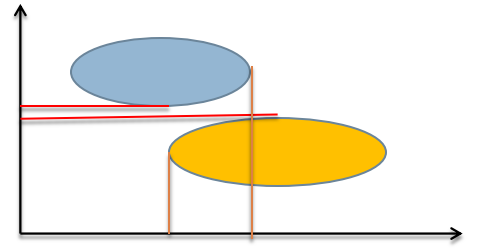
V = Cov(C) = [[4.33, 4.33], [4.33, 4.33]],

Eigenvector (V) = [[0.71, -0.71], [0.71, 0.71]], Eigenvalue(V) = [8.67, 0]

Project data points = Eigenvector \* Eigenvalue matrix = [[-3.3, 0], [0.94, 0], [2.36, 0]]

可參考jupyter file— HW6\_Problem2\_PCAManual.ipynb

3. In the breast cancer dataset, 總共分為2類classes. 若使用LDA方法做降維 (dimensionality reduction), 則最大維度數量為1個features. 假設總共k類classes需要分類, maximum dimension為k-1. 根據Dimensionality Reduction Techniques’s PPT p.57, LDA採用投影 (projection) 的方式,選擇適合的投影向量去做分類. 這表示投影維度空間勢必實際資料維度空間. 如下圖所示:

取自Dimensionality Reduction Techniques’ PPT p.52.

4. 請參閱jupyter file—HW6\_Problem4\_PCA.ipynb.

a. n\_components計算結果為 Pov ([5, 6, 7, 8, 9]) 皆大於0.9

5. 請參閱jupyter file—HW6\_Problem5\_ICA2PCA.ipynb

顯示為以PCA方法實作，結果org\_2.wav無法辨別。