

Idea: PCA找整個group變異最大的那條線 , LDA找最能分開class的那條線(已知class)

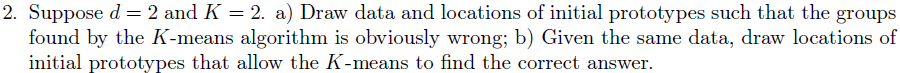
PCA

LDAA

1. (b)

LDAA

PCA



Idea: K-means假設class散布是等size的圓形(d=2)

3

2

1

K-means後(錯)

挑點1-2當起點

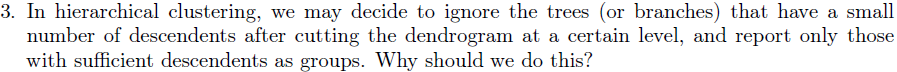
的

K-means後(對)

挑點1-3當起點

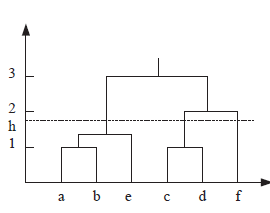
的

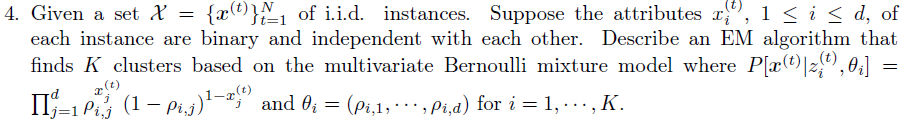
正確結果



Idea: cluster 是 一群有相同性質的object. ,

1. 如果分得太細(K太大), 會將原本具有相同性質但略有不同的object分成不同群,而這必非我們所要的,所以,我們必須依照使用上的需要來決定h為多少.
2. 分得太細的data,其代表與其他資料不太相同,且其數量稀少(在空間中稀疏),因此我們會將其當作noise看待.





E : formula ->

M : 1-- subject to

Use Lagrangian L= 

對->

for i=1,2…,K

2--\*

\*

對微分->

for i=1…K , j=1…d

至此得到新的,

Repeat 上述EM兩步驟 直到收斂為止