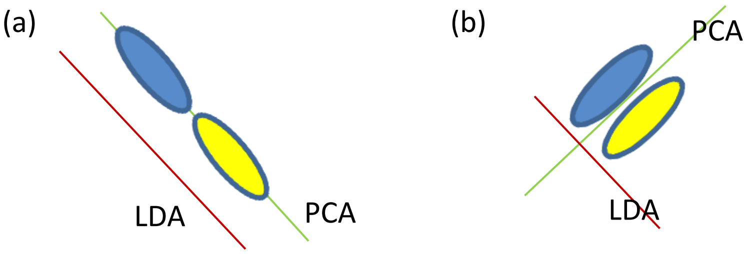
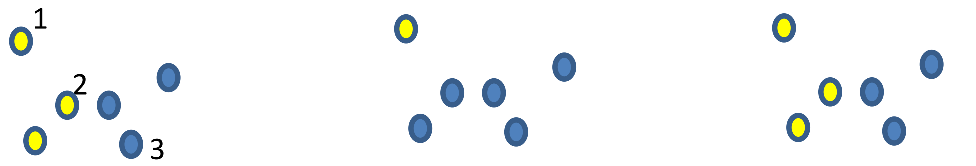
1. Suppose $d = 2$. Draw data of two Gaussian classes such that a) PCA and LDA find the same direction; b) PCA and LDA find orthogonal directions.

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)



2. Suppose $d = 2$ and $K = 2$. a) Draw data and locations of initial prototypes such that the groups found by the K-means algorithm is obviously wrong; b) Given the same data, draw locations of initial prototypes that allow the K-means to find the correct answer.

Idea: K-means 假設 class 散布是等 size 的圓形(d=2)

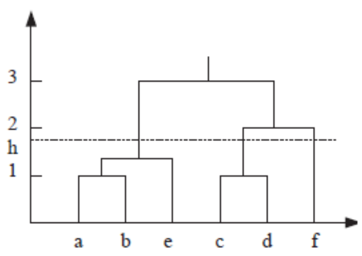


正確結果　　K-means 後(錯) 挑點 1-2 當起點　　K-means 後(對) 挑點 1-3 當起點

3. In hierarchical clustering, we may decide to ignore the trees (or branches) that have a small number of descendents after cutting the dendrogram at a certain level, and report only those with sufficient descendents as groups. Why should we do this?

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.



4. Given a set $\mathcal{X} = \{x^{(t)}\}_{t=1}^{N}$ of i.i.d. instances. Suppose the attributes $x_i^{(t)}$, $1 \le i \le d$, of each instance are binary and independent with each other. Describe an EM algorithm that finds K clusters based on the multivariate Bernoulli mixture model where $P[x^{(t)}|z_i^{(t)}, \theta_i] = \prod_{j=1}^{d} \rho_{i,j}^{x_j^{(t)}} (1-\rho_{i,j})^{1-x_j^{(t)}}$ and $\theta_i = (\rho_{i,1}, \cdots, \rho_{i,d})$ for $i = 1, \cdots, K$.

E : formula ->$P\left(z_i^{(t)} \middle| x^{(t)}, \theta^{old}\right) = \frac{P[x^t|z^t,\theta^{old}]*\pi_i^{old}}{\sum_{j=1}^{K} P[x^t|z_j^t,\theta^{old}]*\pi_j^{old}} \rightarrow h_i^t$(自定符號)

M : 1--$arg_{\pi_1,\ldots,\pi_k} MAX \sum_{t=1}^{N}\sum_{i=1}^{K} \ln(\pi_i) * h_i^t$ subject to $\sum_{i=1}^{K}\pi_i = 1$

Use Lagrangian L= $\sum_{t=1}^{N}\sum_{i=1}^{K} \ln(\pi_i)\, h_i^{(t)} - \alpha\left(\sum_{i=1}^{K}\pi_i - 1\right)$

對π_i微分->$\sum_{i=1}^{N}\frac{h_i^t}{\pi_i} - \alpha = 0 \rightarrow \alpha = \sum_{t=1}^{N}\sum_{i=1}^{K} P(z_i^t|x^t,\theta^{old}) = N$

$\pi_i = \frac{1}{N}\sum_{i=1}^{N}\frac{h_i^t}{\pi_i}$ for i=1,2…,K

2--$arg_{\theta_1,\ldots,\theta_k} MAX \sum_{t=1}^{N}\sum_{i=1}^{K} \ln(P[x^t|z_i^t,\theta_i])*h_i^t$

$\sum_{t=1}^{N}\sum_{i=1}^{K}\sum_{j=1}^{d} [x_j^t * ln(\rho_{i,j}) + (1 - x_j^t) * (ln(\rho_{i,j}))]*h_i^t$

對$\rho_{i,j}$微分-> $\sum_{t=1}^{N}\left[\frac{x_j^t}{\rho_{i,j}} - \frac{(1-x_j^t)}{ln(\rho_{i,j})}\right] * h_i^t = 0$ $\because$ 對$\rho_{i,j}$微分,只有$\rho_{i,j}$會留下.其他當常數微掉

$\sum_{t=1}^{N}(x_j^t - \rho_{i,j}) = 0$

$\rho_{i,j} = \frac{1}{N}\sum_{t=1}^{N} x_j^t$ for i=1…K , j=1…d

至此得到新的θ , π

Repeat 上述 EM 兩步驟 直到收斂為止