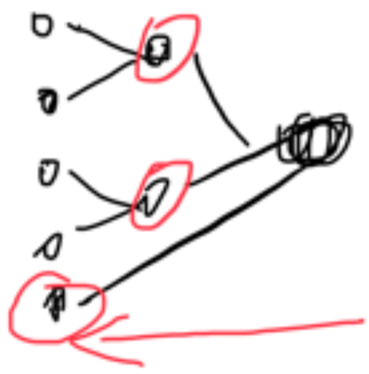
6/17 Wed

① Hierarchical clustering

- Bot-up Top_down
- Distance function between cluster

Bot-up



Top_Down

Single link   average   Complete link

shortest   Maximal

Spread out   Compact

② Density cluster

DBSCAN

- MinPts Eps
- Core, Border
  outlier
- DBSCAN

Traverse points

Expand cluster

MinPts Eps

DBSCAN { detect arbitrary shape
can not handle varying density.

GMM

$$P(X) = \sum_{k=0}^{k} \pi_k N(u_k, \delta_k)$$

Z_k

$$P(X) = \sum_{k=0}^{k} P(Z_k) P(X|Z_k)$$

$$P(Z_{nk}|X) = ?$$

$$\begin{cases} P(Z_{nk}) = \pi_k \rightarrow \text{weights} \\ P(X|Z_{nk}) = N(X|u_k, \delta_k) \end{cases}$$

$$P(Z_{nk}|X) = \frac{P(Z_{nk}, X)}{P(X)} = \frac{P(X|Z_{nk})P(Z_{nk})}{P(X)}$$

$$= \frac{\pi_k N(X|u_k, \delta_k)}{\sum P(X, Z_{kn})}$$

$$= \frac{\pi_k N(X|u_k, \delta_k)}{\sum_{j=0}^{k} \pi_j N(X|u_j, \delta_j)}$$

$$\boxed{\tau(Z_{nk})} = P(Z_{nk}|X) = \frac{\pi_k N(X|u_k, \delta_k)}{\sum_{j=0}^{k} \pi N(\quad)}$$

Learn parameter $\theta(u, \delta, \pi)$

$$\arg\max[P(X|\theta)] \overset{\log}{=} \prod_{i}^{n} P(X_i|\theta)$$

$$\overset{\log}{=} \prod^{N} \sum^{k} \pi_k N(X|u_k, \delta_k)$$

$$= \sum^{N} \log \sum^{k} \pi_k N(X|u_k, \delta_k)$$

$$\frac{\partial\left(\sum^{N} \log \sum^{k} \pi\right)}{\partial u_k} = \sum^{N} \frac{\quad}{\sum^{k} \pi_k N(X|u_k, \delta_k)}$$

$$N(x|\mu,\delta) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{(x-\mu)^2}{2\delta^2}\right)$$

$$\frac{\partial N}{\partial \mu} = N \cdot \frac{x-\mu}{\delta^2}$$

$$= \sum \frac{N(x|\mu_k,\delta_k)\,\frac{x-\mu}{\delta^2}}{\sum^{k} \pi_k N(x|\mu_k,\delta_k)} = 0$$

$$\tau(z_{nk}) = \frac{\pi_k N(x|\mu_k,\delta_k)}{\sum \pi_j N(x_j|\mu_{jk},\delta_j)}$$

$$= \sum^{N} \tau(z_{nk})\,\frac{x_n-\mu}{\delta^2} = 0$$

$$\Rightarrow \sum \tau(z_{nk})\,x_n = \sum \tau(z_{nk})\,\mu$$

$$\Rightarrow \boxed{\mu = \frac{\sum \tau(z_{nk})\,x_n}{\sum \tau(z_{nk})}}$$