# EM algorithm

useful only when there is missing data

I/P: Data set, $x$, $k$, $T_{min}$

O/P: $\hat{\theta}$

① Initialisation step: Use K-means algorithm

we will get $\mu_j^{(0)}$, $\sigma_j^{(0)}$ and $p_j^{(0)}$ and value of $k$

② Expectation: $P(j|x_i) = \dfrac{P(j) \cdot P(x_i|\theta_j)}{\sum_{j=1}^{K}\sum_{i=1}^{N} P(j) \cdot P(x_i|\theta_j)}$

③ Maximisation:

$$\mu_j^{(t)} = \frac{\sum_{i=1}^{N} P(j|x_i)\, x_i}{\sum_{j=1}^{N} P(j|x_i)}$$

$$\sigma_j^{(t)} = \frac{\sum_{i=1}^{N} P(j|x_i)(x_i - \mu_j)}{\sum_{i=1}^{N} P(j|x_i)}$$

$$p_j^{(t)} = \frac{1}{N}\sum_{i=1}^{N} P(j|x_i)$$

if $|\mu_j^{(t)} - \mu_j^{(t-1)}| < 1$

stop, else go to step ②

# Akaike Information Criterion (AIC)

**Model selection for mixture models**

$K_{min} = 1$

1. Apply ~~algo~~ EM algorithm for each possible values of obtained k upto $k_{max}$, get the optimised values of $\mu_j, \sigma_j, p_j$

20%

60% 2. Sub. the values in the AIC formula

20%

$$AIC = -\log likelihood + N_p/2$$

$N_p$ – no. of free parameters

if $N_p = \{\mu_j, \sigma_j, p_j\} = 3k \Rightarrow 3k-1 \because \sum_{i=1}^{k} p_j = 1$

then $AIC = -\log likelihood + \frac{3k-1}{2}$

where k is the no. of clusters

3. Choose the model and the k value for which AIC returns minimum value.

MDL – Minimum Description Length

$$MDL(K) = -\log likelihood + \frac{N_p}{2} \ln(N)$$

# Kullback Leibler Divergence

$$D(P \| Q) = \int_{-\infty}^{\infty} P(x) \log \frac{P(x)}{Q(x)} \, dx$$

$$\Rightarrow \sum_{i=1}^{N} P(i) \log \frac{P(i)}{Q(i)}$$

KL divergence is used to compute distance b/w two probability distribution.

# Distance for Binary Variables

| | Sphere | Shape | Sweet | Sour | Crunchy |
|---|---|---|---|---|---|
| Apple | yes | ~~yes~~ | yes | yes | yes |
| Banana ~~Orange~~ | NO | | yes | NO | NO |

Apple = (1, 1, 1, 1)

Banana ~~Orange~~ = (0, 1, 0, 0)

p → No of variables that positive for both objects

q → No of variables that positive for ith objects & negative for jth object

r → No of variables that negative for ith objects & positive for jth objects

s → No of variables that negative for both objects.

t = total No of variables = p+q+r+s

| | | object j | |
|---|---|---|---|
| | | Yes | No |
| obj i | Yes | p | q |
| | NO | r | s |

P = 1 , q = 3 , r = 0 , s = 0

$$\boxed{\text{Jaccard's distance} = d_{ij} = \frac{q+r}{p+q+r}}$$

$$\boxed{d_{ij} = \frac{3+0}{1+3+0} = \frac{3}{4}}$$

Euclidean Distance → Is appropriate when we have continuous numerical variables.

$$= \left[\sum_{k=1}^{len} abs\left(x_{1k} - x_{2k}\right)^{p}\right]^{1/p}$$

$p = 2.$

$$= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

## Mahalanobis Distance

- It accounts for the fact that the variances in each direction are different.
- It accounts for the co-variance between variables.
- It reduces to the familiar Euclidean distance for uncorrelated variables with unit variance.

$$\boxed{D_M(\vec{x}) = \sqrt{(\vec{x}-\vec{\mu})^T \vec{S} (\vec{x}-\vec{\mu})}}$$

$S \rightarrow$ Co-variance Matrix

$\mu \rightarrow$ mean

## Manhattan Distance

→ If you want to place less Emphasis on outliers this will try to reduce all Errors Equally since the gradient has constant magnitude.

$$\boxed{D = \left| (x_2 - x_1) + (y_2 - y_1) \right|}$$