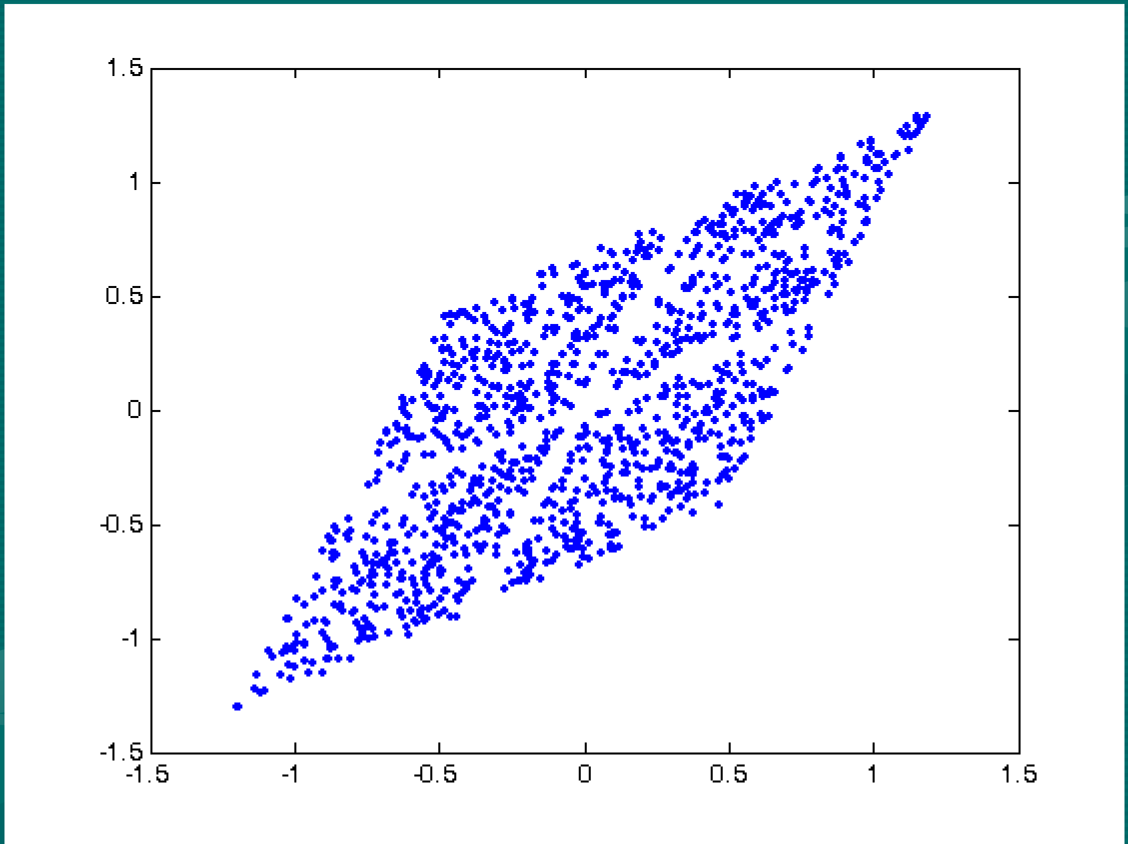
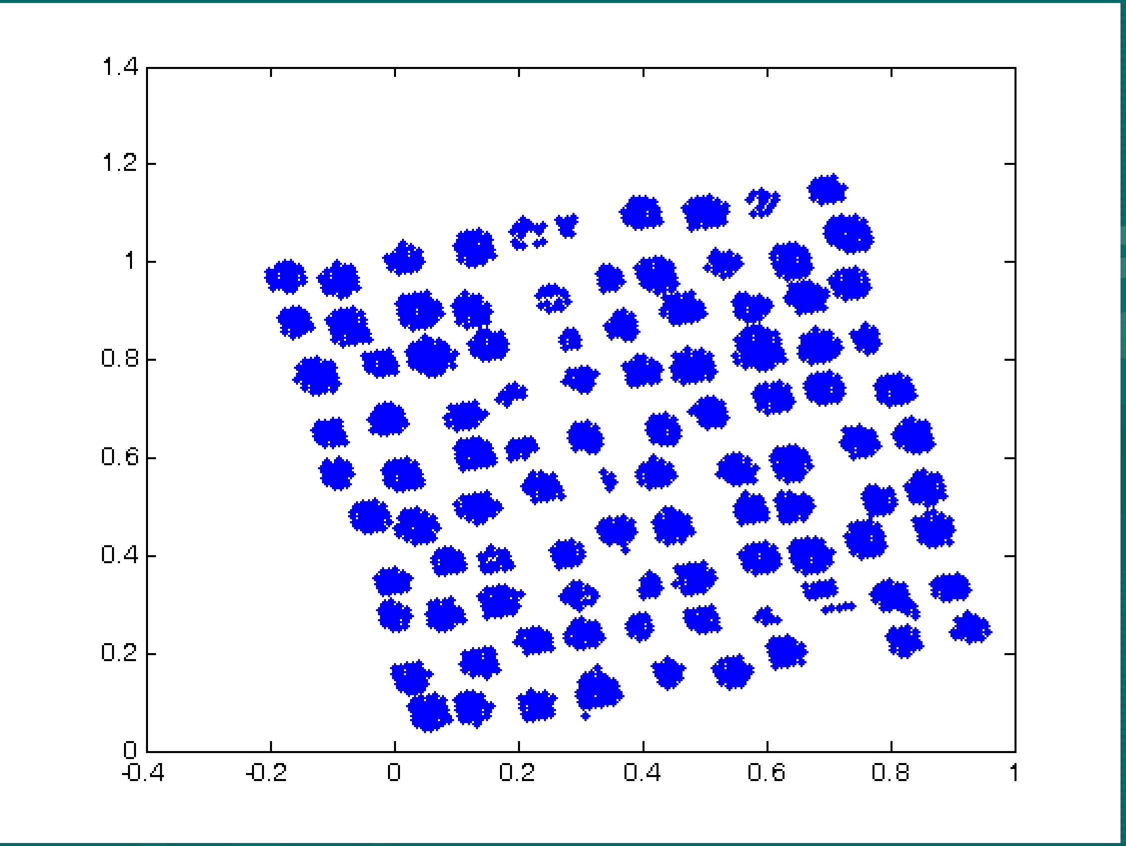
# EM clustering

# Unsupervised data



# Unsupervised data



# Gaussian mixtures

Assumption: unsupervised data are sampled from Gaussian mixtures

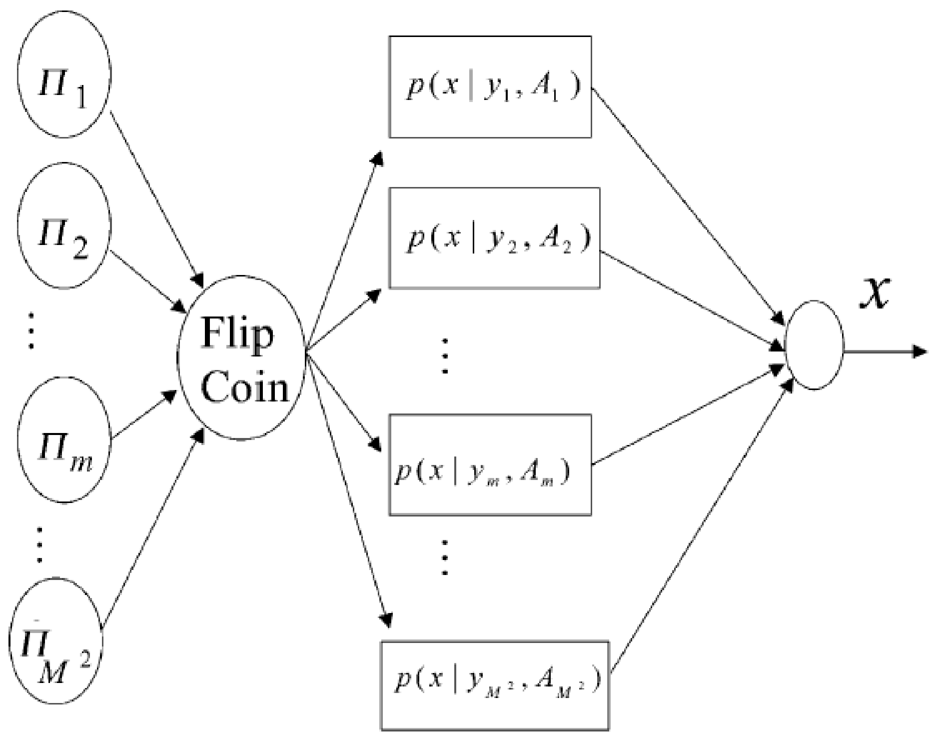



Fig. 1. The generative model.

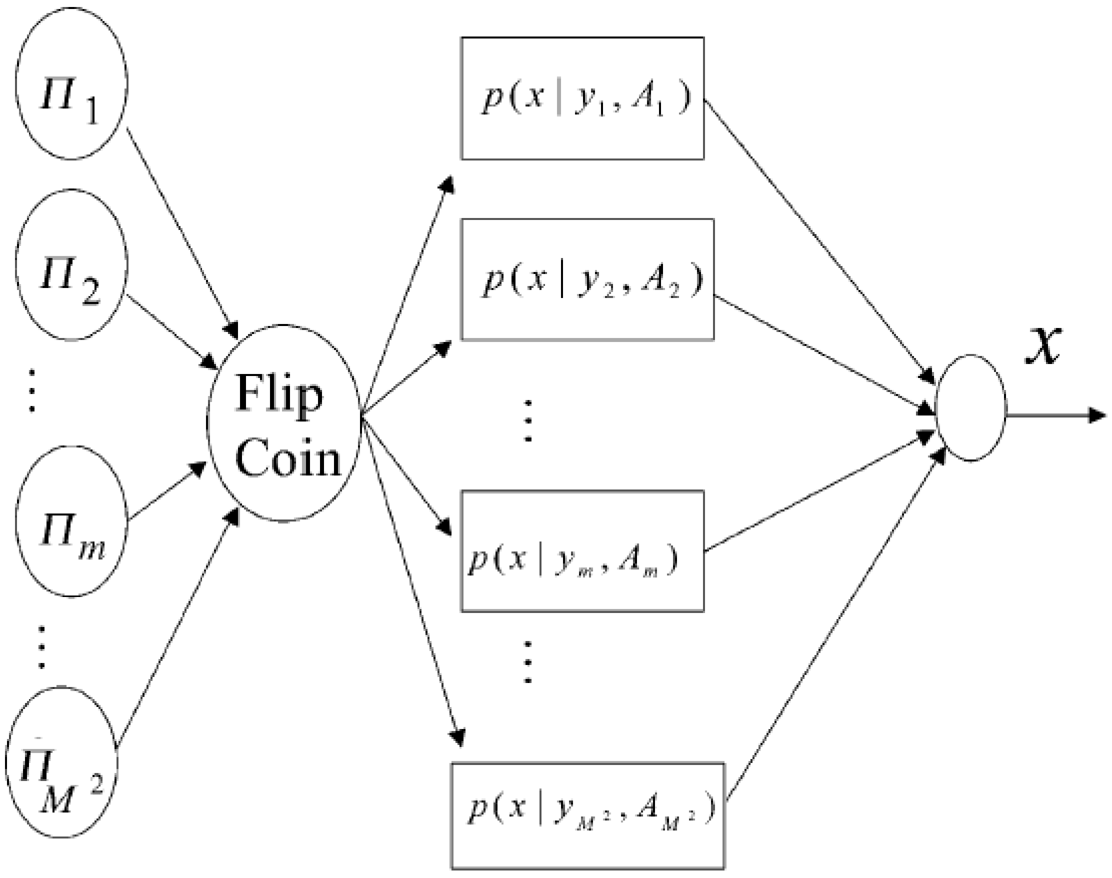



Fig. 1. The generative model.

X={x[t]}t

# Gaussian pdf

$$P_k(x) = P(x|y_k, A_k)$$

$$= \frac{1}{(2\pi)^{d/2}\sqrt{|A_k^{-1}|}} \exp\left( \frac{(x - y_k)^t A_k (x - y_k)}{2} \right)$$

# Weight sum of Gaussian pdfs

$$p(\mathbf{x}) = \sum_k \pi_k \, p_k(\mathbf{x} \mid \mathbf{A}, \mathbf{y}_k)$$

# Uncorrelated data

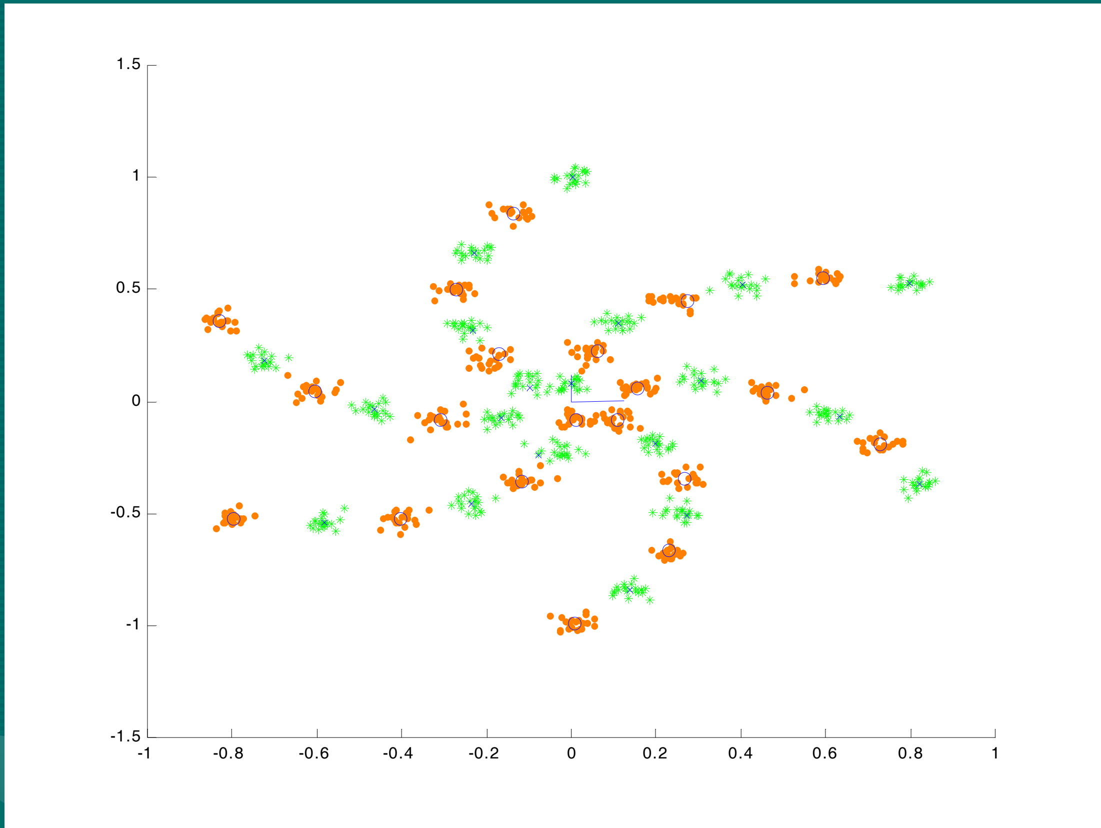
$$p_k(\mathbf{x} \mid \sigma_k, \mathbf{y}_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(\frac{-(\mathbf{x}-\mathbf{y}_k)^T(\mathbf{x}-\mathbf{y}_k)}{2\sigma_k^2}\right)$$

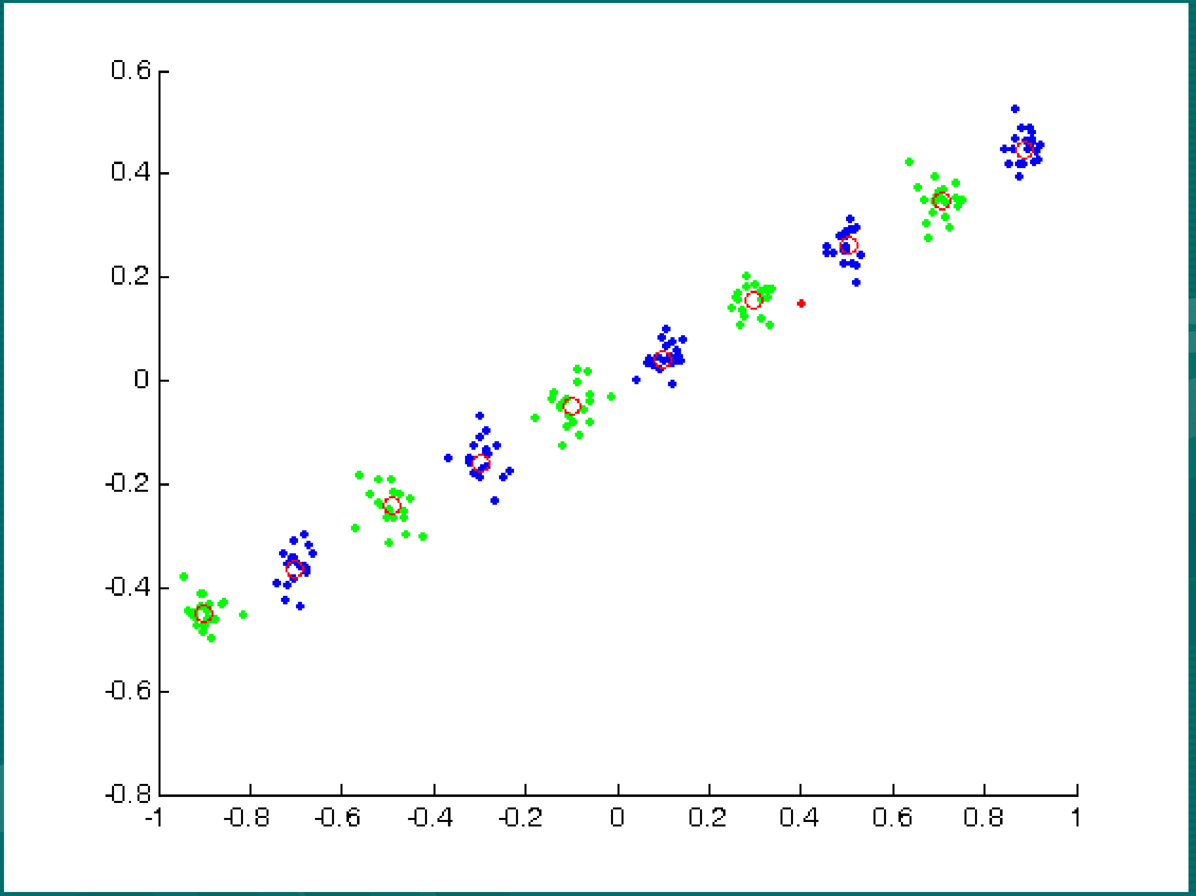
# Data creation

- Write a matlab function to create uncorrelated data, $X=\{x[t]\}_t$

- X is a sample from Gaussian mixtures
- $\mathbf{x} \in R^d$ d=2,

```
range=2.5;
y=[1 1;1 -1;-1 1;-1 -1]*range;
K=4;N=100;
X=[];
for k=1:K
    X=[X; randn(N,2)-repmat(y(k,:),N,1)];
end
plot(X(:,1),X(:,2),'.'); hold on;
plot(X(200,1),X(200,2),'ro');
```


1.5
1
0.5
0
-0.5
-1
-1.5
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1


0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

# Membership

$$e_k^K = [0\,, 0, \cdots, 0, 1, 0, \cdots, 0, 0]^T$$

$$\text{pos} \quad 1 \;\; 2 \cdots \text{k} - 1, \text{k}, \text{k} + 1, \cdots, \text{K}$$

Standard basis

$$\Xi = \{\mathbf{e}_1^K, \cdots, \mathbf{e}_k^K, \cdots, \mathbf{e}_K^K\}$$

# Membership vector

$$\boldsymbol{\delta}[t] \in \Xi = \{e_1^K, \cdots, e_k^K, \cdots, e_K^K\}$$

$$\boldsymbol{\delta}[t] = e_k^K \Leftrightarrow \text{x[t] is generated by the kth pdf}$$

$$\Pr(\xi_i = e_k^K) \propto \exp(-\beta \| \mathbf{x}_i - \mathbf{y}_k \|^2)$$

$$\sum_{k=1}^{K} \Pr(\xi_i = e_k^K) = 1$$

$$\Pr(\xi_i = e_k^K) = ?$$

$$\Pr(\xi_i = e_k^K) = C\exp(-\beta \| \mathbf{x}_i - \mathbf{y}_k \|^2)$$

$$C\sum_{k=1}^{K} \exp(-\beta \| \mathbf{x}_i - \mathbf{y}_k \|^2) = 1$$

$$C = \frac{1}{\sum_{k=1}^{K} \exp(-\beta \| \mathbf{x}_i - \mathbf{y}_k \|^2)}$$

$$\Pr(\xi_i = e_k^K) = \frac{\exp(-\beta \| \mathbf{x}_i - \mathbf{y}_k \|^2)}{\sum_{k=1}^{K} \exp(-\beta \| \mathbf{x}_i - \mathbf{y}_k \|^2)}$$

$$\left\langle \xi_i \right\rangle = \sum_{k=1}^{K} \Pr(\xi_i = \mathbf{e}_k^K)\mathbf{e}_k^K = \sum_{k=1}^{K} \frac{\exp(-\beta \| \mathbf{x}_\mathrm{i} - \mathbf{y}_\mathrm{k} \|^2)}{\sum_{h=1}^{\mathrm{K}} \exp(-\beta \| \mathbf{x}_\mathrm{i} - \mathbf{y}_\mathrm{h} \|^2)} \mathbf{e}_k^K$$

$$= \left( \frac{\exp(-\beta \| \mathbf{x}_\mathrm{i} - \mathbf{y}_1 \|^2)}{\sum_{\mathrm{k}=1}^{\mathrm{K}} \exp(-\beta \| \mathbf{x}_\mathrm{i} - \mathbf{y}_\mathrm{k} \|^2)}, \frac{\exp(-\beta \| \mathbf{x}_\mathrm{i} - \mathbf{y}_2 \|^2)}{\sum_{\mathrm{k}=1}^{\mathrm{K}} \exp(-\beta \| \mathbf{x}_\mathrm{i} - \mathbf{y}_\mathrm{k} \|^2)}, \cdots, \frac{\exp(-\beta \| \mathbf{x}_\mathrm{i} - \mathbf{y}_\mathrm{K} \|^2)}{\sum_{\mathrm{k}=1}^{\mathrm{K}} \exp(-\beta \| \mathbf{x}_\mathrm{i} - \mathbf{y}_\mathrm{k} \|^2)} \right)$$

# E1

$$v_{ik} \equiv \langle \xi_{ik} \rangle = \Pr(\boldsymbol{\xi}_i = \mathrm{e}_k^K) = \frac{\exp(-\beta \| \mathbf{x}_i - \mathbf{y}_k \|^2)}{\sum_{k=1}^{K} \exp(-\beta \| \mathbf{x}_i - \mathbf{y}_k \|^2)} \quad \text{(E1)}$$

# M1

$$E(\xi, Y) = \sum_{i=1}^{N} \sum_{k=1}^{K} \xi_{ik} (\mathbf{x}_i - \mathbf{y}_k)^T (\mathbf{x}_i - \mathbf{y}_k)$$

$$E(\langle \xi \rangle, Y) = \sum_{i=1}^{N} \sum_{k=1}^{K} \langle \xi_{ik} \rangle (\mathbf{x}_i - \mathbf{y}_k)^T (\mathbf{x}_i - \mathbf{y}_k)$$

$$\frac{\partial E(\langle \xi \rangle, Y)}{\partial \mathbf{y}_k} = -2 \sum_{i=1}^{N} \langle \xi_{ik} \rangle (\mathbf{x}_i - \mathbf{y}_k) = 0$$

$$\sum_{i=1}^{N} \langle \xi_{ik} \rangle \mathbf{y}_k = \sum_{i=1}^{N} \langle \xi_{ik} \rangle \mathbf{x}_i \Rightarrow \mathbf{y}_k = \frac{\sum_{i=1}^{N} \langle \xi_{ik} \rangle \mathbf{x}_i}{\sum_{i=1}^{N} \langle \xi_{ik} \rangle}$$

$$v_{ik} \equiv \left\langle \xi_{\mathrm{ik}} \right\rangle = \Pr(\xi_{\mathrm{i}} = \mathrm{e}_{\mathrm{k}}^{\mathrm{K}}) = \frac{\exp(-\beta \| \mathbf{x}_{\mathrm{i}} - \mathbf{y}_{\mathrm{k}} \|^2)}{\sum_{\mathrm{k}=1}^{\mathrm{K}} \exp(-\beta \| \mathbf{x}_{\mathrm{i}} - \mathbf{y}_{\mathrm{k}} \|^2)} \quad \text{(E1)}$$

$$\mathbf{y}_k = \frac{\sum_{i=1}^{N} \left\langle \xi_{ik} \right\rangle \mathbf{x}_i}{\sum_{i=1}^{N} \left\langle \xi_{ik} \right\rangle}$$

```
while sat < 0.99
    % E step: update v
    [v] = update_v(N,temp,v,loop,X,Y);
    % M step: update Y
    sat = sum(sum(v.*v))/N;
    temp  = temp*tscale;
    ii=ii+1;
    if mod(ii,10)==0
    fprintf('Tmp:%7.5f sat:%7.5f \n',temp,sat);
    end
end
```

# Fitting Criterion

$$L(\boldsymbol{\delta}, \mathbf{y}) = \frac{1}{2} \sum_t \sum_k \delta_k[t](\mathbf{x}[t] - \mathbf{y}_k)^T (\mathbf{x}[t] - \mathbf{y}_k)$$

$$\Pr(\boldsymbol{\delta}[t] = \mathbf{e}_k^K) \propto \exp(\beta u_k[t])$$

$$u_k[t] = -\frac{dL(\langle \boldsymbol{\delta} \rangle \mid \mathbf{y})}{d\langle \delta_k[t] \rangle}$$

$$\Pr(\boldsymbol{\delta}[\mathrm{t}] = \mathrm{e}_\mathrm{k}^\mathrm{K}) \propto \exp(-\beta \| \mathbf{x}_\mathrm{i} - \mathbf{y}_\mathrm{k} \|^2)$$

$$v_{ik} = \frac{\exp(-\beta \| \mathbf{x}_i - \mathbf{y}_k \|^2)}{\sum_{j=1}^{K} \exp(-\beta \| \mathbf{x}_i - \mathbf{y}_j \|^2)} \quad \text{(E1)}$$

$$\Leftrightarrow$$

$$u_{ik} = -\| \mathbf{x}_i - \mathbf{y}_k \|^2$$

$$v_{ik} = \frac{\exp(\beta u_{ik})}{\sum_{j=1}^{K} \exp(\beta u_{ij})}$$

$$\mathbf{y}_k = \frac{\sum_{i=1}^{N} v_{ik} \mathbf{x}_i}{\sum_{i=1}^{N} v_{ik}} \quad \text{(M1)}$$

# EM method

1. Set each $v_{ik}$ near zero randomly,
   set $y_k$ near the mean of all $x_i$
2. Fix beta to one
3. E step: determine all $v_{ik}$ using eq (E1)
4. M step: use eq (M1) to update all $y_k$
5. If the halting condition holds, halt
6. Go to step 3