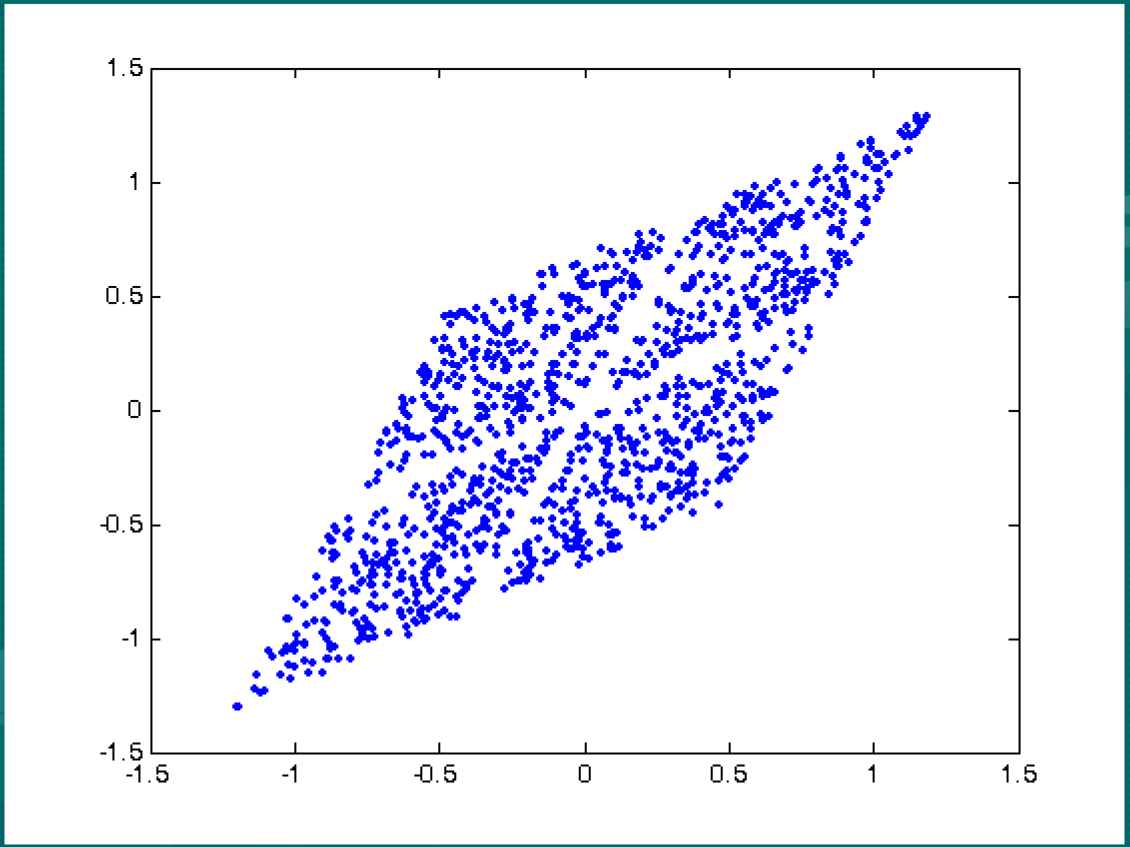
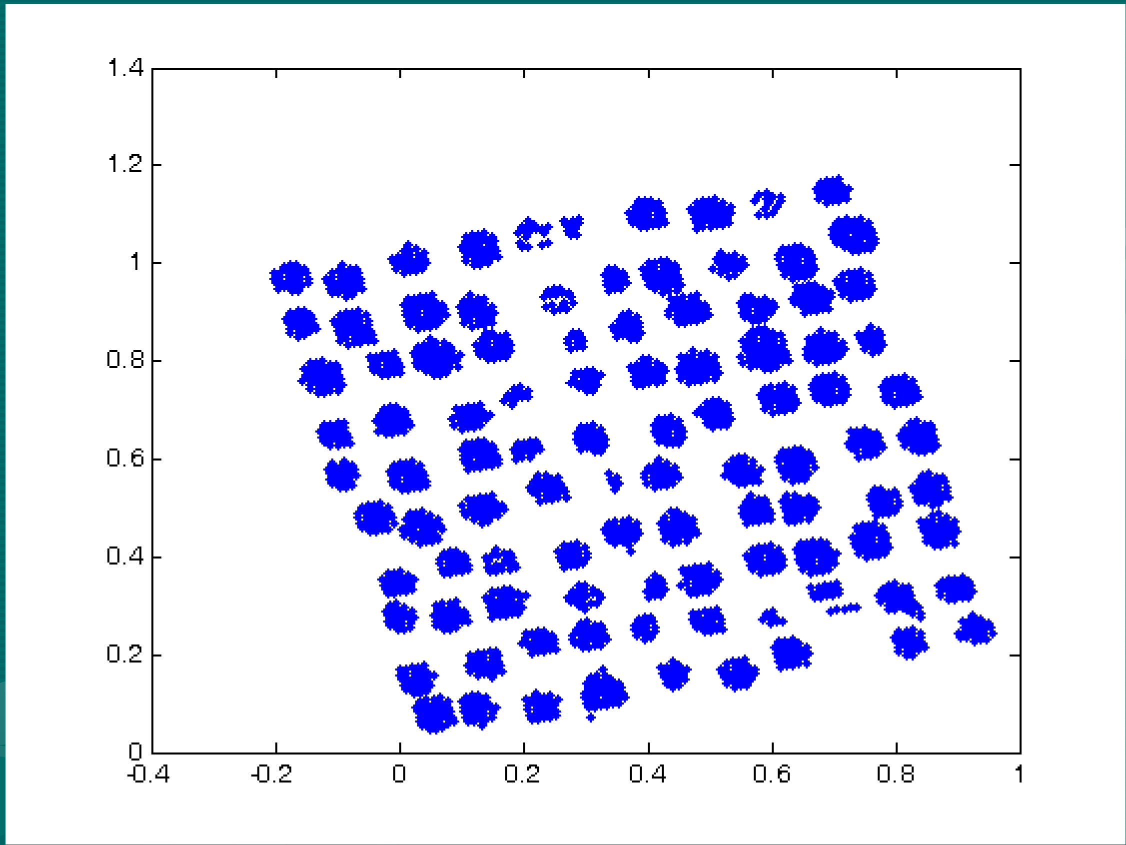
Unsupervised MFA clustering

Continuous variables	Discrete variables	Discrete & Continuous variables
Newton method	MFA	EM
Newton-Gauss	MFA(LM) ?	MFA supervised learning
LM method	MFA +Newton's	MFA unsupervised learning

Unsupervised data



Unsupervised data



Gaussian mixtures

Assumption: unsupervised data are sampled from Gaussian mixtures

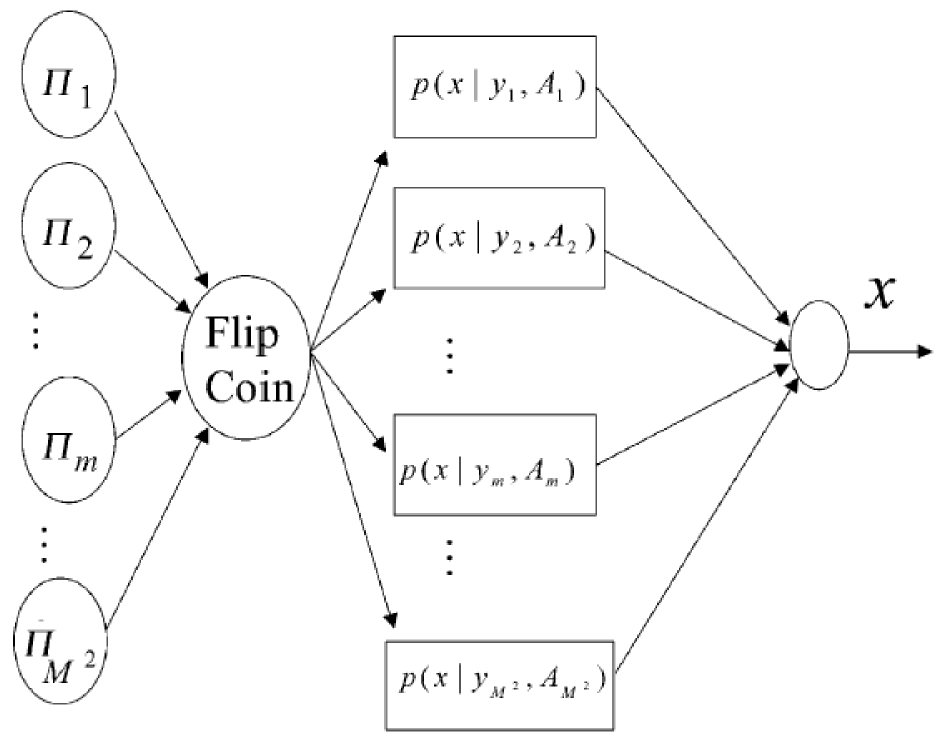


Fig. 1. The generative model.

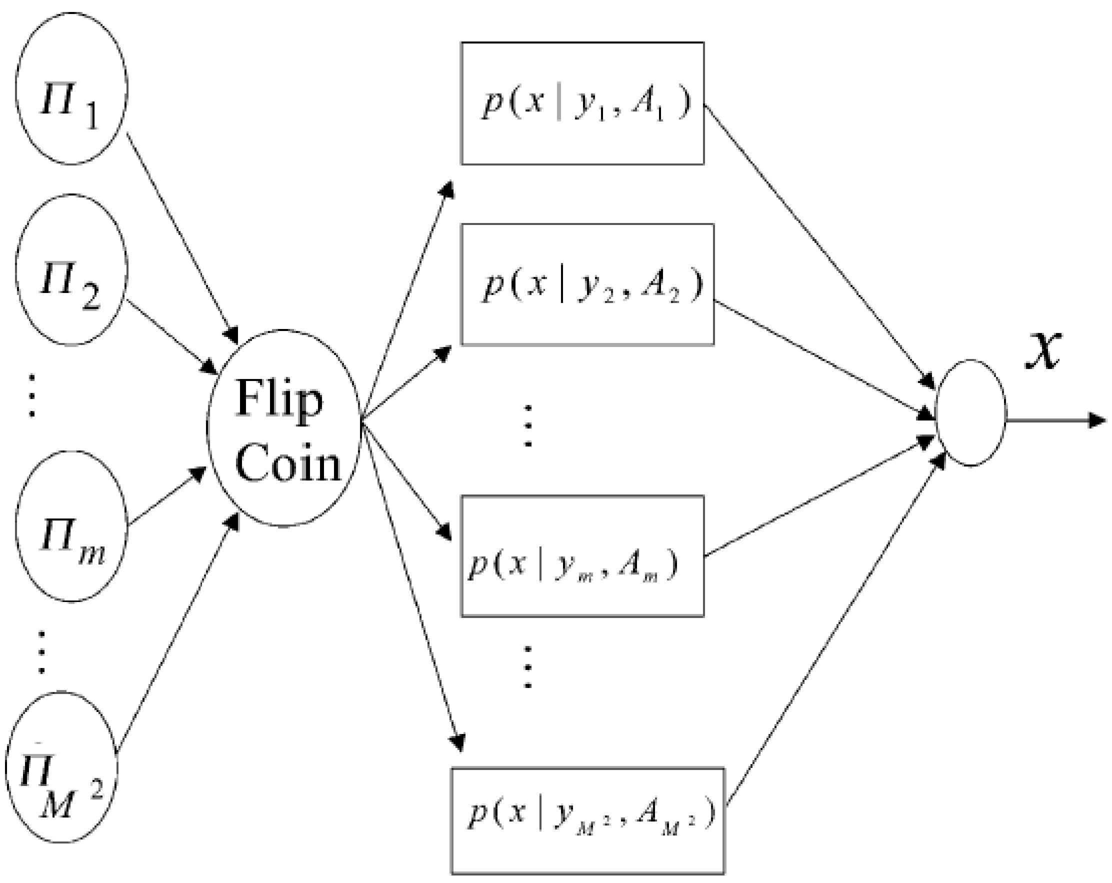


Fig. 1. The generative model.

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

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

Fitting Gaussian mixtures

$$L = \sum_k L_k$$

$$L_k = \log \prod_{\boldsymbol{\delta}[t]=e_k} p_k(\mathbf{x}[t])$$

$$= \sum_{\boldsymbol{\delta}[t]=e_k} \ln p_k(\mathbf{x}[t])$$

$$= \sum_t \delta_k[t] \ln p_k(\mathbf{x}[t])$$

Objective function

$$L(\boldsymbol{\delta}, \mathbf{y}, \mathbf{A}) = \sum_k L_k$$

$$= \frac{1}{2} \sum_t \sum_k \delta_k [t] (\mathbf{x}[t] - \mathbf{y}_k)^T \mathbf{A} (\mathbf{x}[t] - \mathbf{y}_k) - \frac{N}{2} \log |\mathbf{A}|$$

$$L(\boldsymbol{\delta}, \mathbf{y}, \mathbf{A}) = \sum_k L_k$$

$$= \frac{1}{2}\sum_t \sum_k \delta_k[t](\mathbf{x}[t] - \mathbf{y}_k)^T \mathbf{A}(\mathbf{x}[t] - \mathbf{y}_k) - \frac{N}{2}\log|\mathbf{A}|$$

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

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

$$\Pr(\boldsymbol{\delta}[t] = \mathbf{e}_k^K) \propto \exp(-\frac{\beta}{2}(\mathbf{x}[t] - \mathbf{y}_k)^T \mathbf{A}(\mathbf{x}[t] - \mathbf{y}_k))$$

E step

$$u_k[t] = -\frac{\partial L}{\partial \langle \delta_k[t] \rangle} = \frac{-1}{2}\left(\mathbf{x}[t]-\mathbf{y}_k\right)\mathbf{A}(\mathbf{x}[t]-\mathbf{y}_k) \quad \text{(E1)}$$

$$v_k[t] \equiv \langle \delta_k[t] \rangle = \frac{\exp(\beta u_k[t])}{\sum_l \exp(\beta u_l[t])} \quad \text{(E2)}$$

M step

Solve

$$\frac{dL(\mathbf{y} \mid \mathbf{v}, \mathbf{A})}{d\mathbf{y}_k} = 0, \ \mathrm{k} = 1, \ldots, \mathrm{K} \quad \text{(M1)}$$

where

$$\mathbf{v} = \langle \boldsymbol{\delta} \rangle$$

M step

Solve

$$\frac{dL(\mathbf{A} \mid \mathbf{v}, \mathbf{y})}{dA_{ab}} = 0, \ \ \mathrm{a, b} = 1, \ldots, \mathrm{d} \quad \text{(M2)}$$

where

$$\mathbf{v} = \langle \boldsymbol{\delta} \rangle$$

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 (E1-2)
4. M step: use eq (M1) to update all y_k
   use eq(M2) to update matrix A
5. If the halting condition holds, halt
6. Go to step 3

Exercise

Derive updating rules for the M step

$$\frac{\partial L}{\partial y_k} = \frac{1}{2}\sum_t v_k[t](\mathbf{A} + \mathbf{A}^T)(\mathbf{x}[t] - \mathbf{y}_k) = \mathbf{0}$$

$$\mathbf{y}_k = \frac{\sum_t v_k[t]\mathbf{x}[t]}{\sum_t v_k[t]} \qquad \text{(M1)}$$

$$\frac{\partial L}{\partial A_{ab}} = \frac{-1}{2}\sum_t \sum_k v_k[t](x_a[t] - y_{ka})(x_b[t] - y_{kb}) + \frac{N}{2}[(\mathbf{A}^T)^{-1}]_{ab} = 0$$

$$\Rightarrow \quad \mathbf{A} = (\mathbf{W}^{-1})^T \qquad \text{(M2)}$$

$$W_{ab} = \frac{1}{N}\sum_t \sum_k v_k[t](x_a[t] - y_{ka})(x_b[t] - y_{kb})$$

Annealed EM

$$1.\ \text{Set } \beta \text{ to a sufficiently low value}$$

$$\mathrm{A} = 0.01 \times \mathrm{I}$$

$$\mathbf{y}_k \approx \frac{1}{N}\sum_t \mathbf{x}[t],\ v_k[t] \approx \frac{1}{K}$$

$$2.\ \text{E step: update v using (E1)}$$

$$3.\ \text{M step: update y using (M1)}$$

$$\text{update A using (M2)}$$

$$4.\ \text{If } \frac{1}{\mathrm{N}}\sum_{\mathrm{t}}\sum_k v_k[t]^2 \geq 0.98, \text{ halt}$$

$$\text{else } \beta \leftarrow 0.98\beta, \text{ goto step 2}$$

$$L(\boldsymbol{\delta}, \mathbf{y}, \mathbf{A}) = \sum_k L_k$$

$$= \frac{1}{2} \sum_t \sum_k \delta_k[t] (\mathbf{x}[t] - \mathbf{y}_k)^T \mathbf{A} (\mathbf{x}[t] - \mathbf{y}_k) - \frac{N}{2} \log|\mathbf{A}|$$

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

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

$$\mathrm{C} = \frac{1}{\sum_{k=1}^{K} \exp(\beta u_k[t])}$$

$$\Pr(\boldsymbol{\delta}[t] = \mathbf{e}_k^K) = \frac{\exp(\beta u_k[t])}{\sum_{j=1}^{K} \exp(\beta u_j[t])}$$

$$\text{Expectation of } \boldsymbol{\delta}[t] = \sum_{k=1}^{K} \mathbf{e}_k^K \Pr(\boldsymbol{\delta}[t] = \mathbf{e}_k^K)$$

$$\text{Entropy of } \boldsymbol{\delta}[t] = -\sum_{k=1}^{K} \Pr(\boldsymbol{\delta}[t] = \mathbf{e}_k^K) \ln \Pr(\boldsymbol{\delta}[t] = \mathbf{e}_k^K)$$

$$\Pr(\boldsymbol{\delta}[t]=\mathbf{e}_k^K)=\frac{\exp(\beta u_k[t])}{\sum_{j=1}^{K}\exp(\beta u_j[t])}\equiv v_k[t]=\langle\delta_k[t]\rangle$$

$$\mathrm{H_t}\equiv \text{Entropy of }\boldsymbol{\delta}[t]=-\sum_{k=1}^{K}\Pr(\boldsymbol{\delta}[t]=\mathbf{e}_k^K)\ln\Pr(\boldsymbol{\delta}[t]=\mathbf{e}_k^K)$$

$$=-\sum_{k=1}^{K}v_k[t](\beta u_k[t]-\ln\sum_{j=1}^{K}\exp(\beta u_j[t]))$$

$$=-\beta\sum_{k=1}^{K}v_k[t]u_k[t]+\sum_{k=1}^{K}v_k[t]\ln\sum_{j=1}^{K}\exp(\beta u_j[t])$$

Free energy

- Combination of mean energy and negative entropy

Randomization

- Boltzmann assumption
  - $\boldsymbol{\delta}$ is regarded as a random vector

$$\Pr(\boldsymbol{\delta}) \propto \exp(-\beta \mathrm{E}(\boldsymbol{\delta}))$$

- Free energy

$$\mathrm{F} = \left\langle \mathrm{E}(\boldsymbol{\delta}) \right\rangle - \frac{1}{\beta} H(\boldsymbol{\delta})$$

$$F = \langle E(\boldsymbol{\delta}) \rangle - \frac{1}{\beta} H(\boldsymbol{\delta})$$

$$\approx E(\{\langle \boldsymbol{\delta}[t] \rangle\}) - \frac{1}{\beta} \sum_t H(\boldsymbol{\delta}[t])$$

Derived based on

Kullback - Leiberg(KL) divergence

$$F \approx E(\{\langle \boldsymbol{\delta}[t] \rangle\}) - \frac{1}{\beta}\sum_t H_t$$

$$= E(\mathbf{v}) + \sum_t \sum_k v_k[t] u_k[t]$$

$$- \frac{1}{\beta}\sum_t \ln \sum_j \exp(\beta u_j[t])$$

Mean field equation

$$\frac{\partial \mathrm{F}}{\partial v_k[t]} = 0, \frac{\partial \mathrm{F}}{\partial u_k[t]} = 0, \forall k, t$$

$$u_k[t] = -\frac{\partial \mathrm{E}(\mathbf{v})}{\partial v_k[t]},$$

$$v_k[t] = \frac{\exp(\beta u_k[t])}{\sum_j \exp(\beta u_j[t])}$$

Free energy

$$L(\boldsymbol{\delta}, \mathbf{y}, \mathbf{A}) = \sum_k L_k$$

$$= \frac{1}{2} \sum_t \sum_k \delta_k[t](\mathbf{x}[t] - \mathbf{y}_k)^T \mathbf{A}(\mathbf{x}[t] - \mathbf{y}_k) - \frac{N}{2} \log|\mathbf{A}|$$

$$\mathrm{F}(\mathbf{v}, \mathbf{u}, \mathbf{y}, \mathbf{A})$$

$$= \mathrm{E}(\mathbf{v}, \mathbf{y}, \mathbf{A}) + \sum_t \sum_k v_k[t] u_k[t]$$

$$- \frac{1}{\beta} \sum_t \ln \sum_j \exp(\beta u_j[t])$$

Saddle point

$$\frac{\partial \mathrm{F}}{\partial v_k[t]} = 0, \frac{\partial \mathrm{F}}{\partial u_k[t]} = 0, \forall k, t$$

$$\frac{\partial \mathrm{F}}{\partial \mathbf{y}_k} = \frac{dL(\mathbf{y} \mid \mathbf{v}, \mathbf{A})}{d\mathbf{y}_k} = 0, \quad \text{(M1)}$$

$$\frac{\partial \mathrm{F}}{\partial A_{ab}} = \frac{dL(\mathbf{A} \mid \mathbf{v}, \mathbf{y})}{dA_{ab}} = 0, \quad \text{(M2)}$$

Updating rules

$$u_k[t] = -\frac{\partial L}{\partial \langle \delta_k[t] \rangle} = \frac{-1}{2}(\mathbf{x}[t]-\mathbf{y}_k)\mathbf{A}(\mathbf{x}[t]-\mathbf{y}_k) \quad \text{(E1)}$$

$$v_k[t] \equiv \langle \delta_k[t] \rangle = \frac{\exp(\beta u_k[t])}{\sum_l \exp(\beta u_l[t])} \quad \text{(E2)}$$

$$\mathbf{y}_k = \frac{\sum_t v_k[t]\mathbf{x}[t]}{\sum_t v_k[t]} \quad \text{(M1)}$$

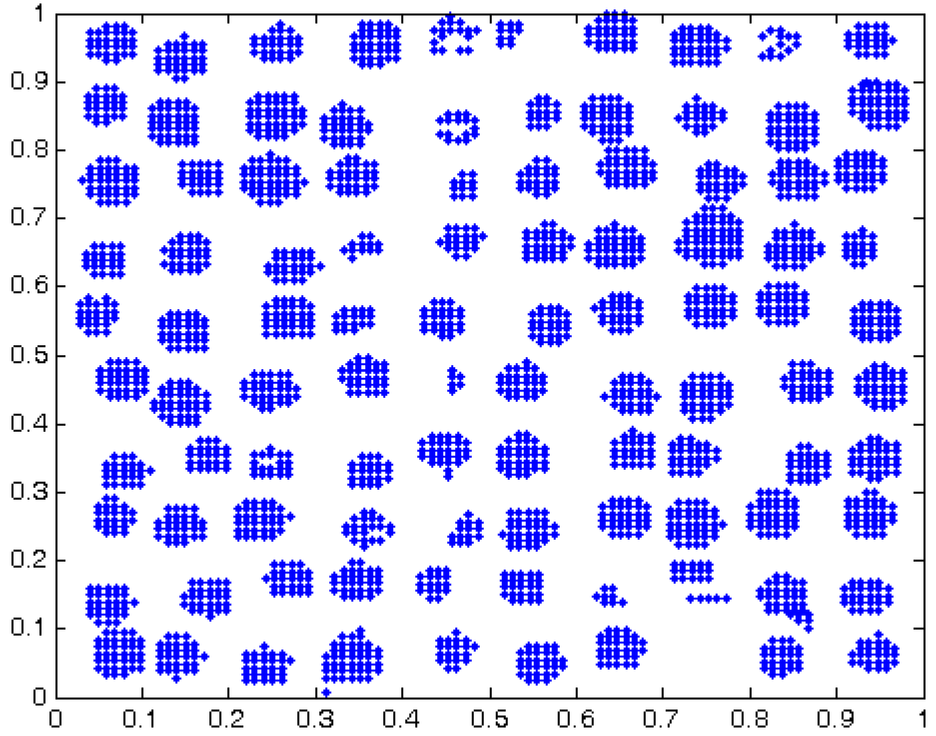
$$\mathbf{A} = (\mathbf{W}^{-1})^T \quad \text{(M2)}$$

$$W_{ab} = \frac{1}{N}\sum_t \sum_k v_k[t](x_a[t]-y_{ka})(x_b[t]-y_{kb})$$

Exercise

- Implement unsupervised MFA learning for data clustering

Data set



Data set