Supervised and Unsupervised MFA learning

- Self-organization
- Classification
- Independent component analysis

Outline

- MFA learning
- Mixed integer and linear programming (MILP)
  - Classification
  - Independent component analysis
  - Self-organization
- MATLAB demo
  - Face recognition
  - Fetal ECG extraction
  - Automatic microarray image analysis
- Conclusions

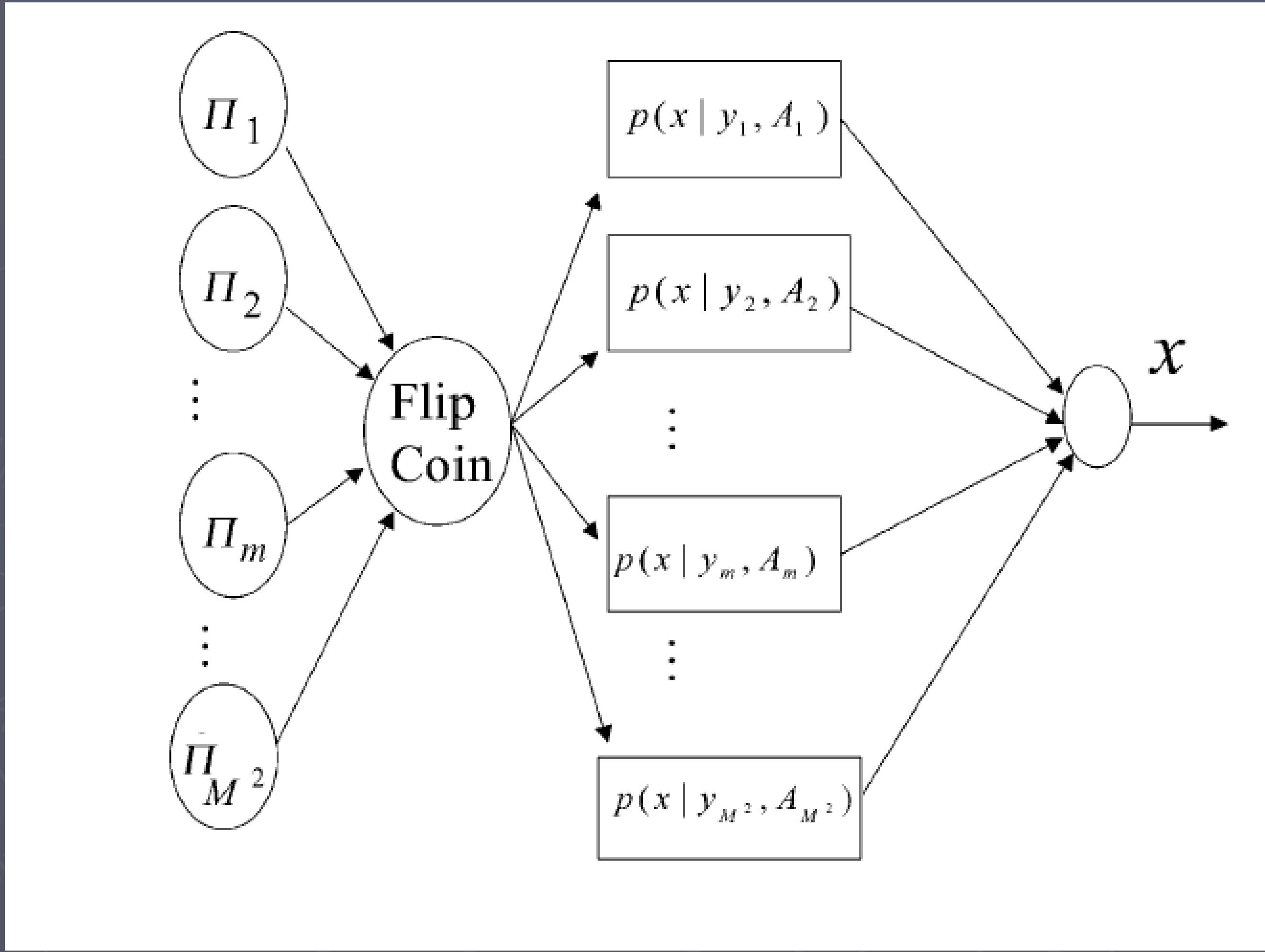
MILP

- Variables
  - Binary discrete variables
  - Real continuous variables
- The objective function is not differentiable with respect to discrete variables
- MILP is solved by MFA learning

Learning self-organizing map

- MILP
- Solve a MILP by the AEM method

A generative model for Gaussian mixtures



Multivariate Gaussian

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

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

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

$$\Xi = \{ e_1^K, \cdots, e_k^K, \cdots, e_K^K \}$$

Membership vector

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

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

Fitting Gaussian mixtures

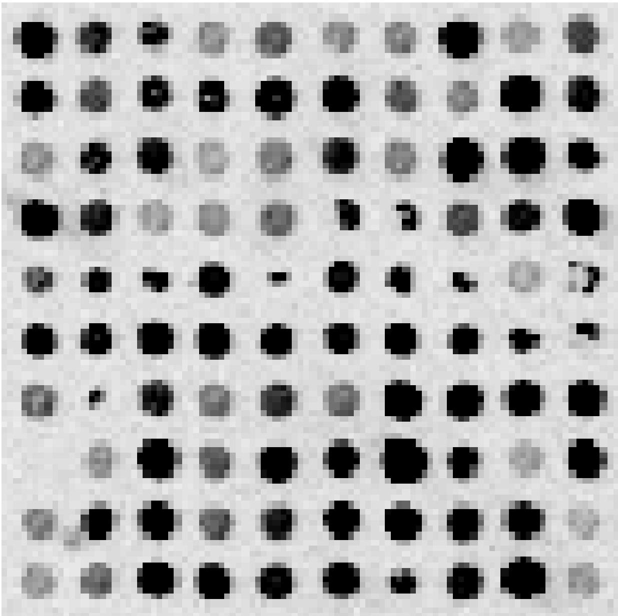
$$L = \sum_k L_k$$

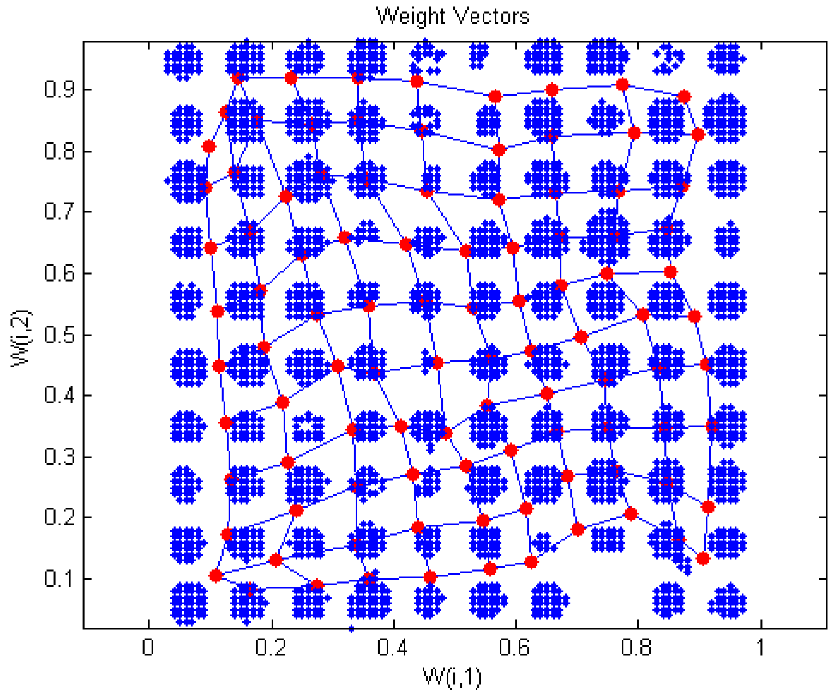
$$L_k = \log \prod_{i:\boldsymbol{\delta}_i = e_k} p_k(\mathbf{x}_i)$$

$$= \sum_{i:\boldsymbol{\delta}_i = e_k} \ln p_k(\mathbf{x}_i)$$

$$= \sum_i \delta_{ik} \ln p_k(\mathbf{x}_i)$$

Griding





Kohonen SOM algorithm

SOM

Minimize

$$E = E_1 + CE_2 = \frac{1}{2}\sum_i \sum_k \delta_{ik}\|x_i - y_k\|_A^2 - \frac{N}{2}\log|A|$$

$$+ \frac{C}{2}\sum_k \sum_{j\in NB(k)} \|y_k - y_j\|_A^2$$

subject to

$$\sum_k \delta_{ik} = 1, \qquad \forall i$$

where C is a weighting constant.

Energy function

$$E = E_1 + CE_2 = \frac{1}{2}\sum_i \sum_k \delta_{ik}\|x_i - y_k\|_A^2 - \frac{N}{2}\log|A|$$

$$+ \frac{C}{2}\sum_k \sum_{j \in NB(k)} \|y_k - y_j\|_A^2$$

MILP

- E is not differentiable with respect to discrete variables

$$\Omega = \{\delta_{ik}\}_{ik}$$

- Boltzmann assumption

$$\Pr(\Omega) \propto \exp(-\beta E(\Omega; \mathbf{A}, \mathbf{Y}))$$

Free energy

- Mean energy combined with negative entropy
- Entropy:

$$\textit{entropy} \text{ of } \Omega$$
$$\equiv H(\Omega)$$
$$= -\sum_{\{\Omega\}} \Pr(\Omega) \ln \Pr(\Omega)$$

Approximation to mean energy

$$
\begin{aligned}
&\text{mean energy} \\
&= \sum_{\{\Omega\}} \mathrm{E}(\Omega)\,\Pr(\Omega) \\
&\approx E(\langle \Omega \rangle)
\end{aligned}
$$

Free energy

Peterson and Söderberg (1989)

$$\psi(A, Y, \langle\delta\rangle, u) = E(A, Y, \langle\delta\rangle) + \sum_i \sum_m \langle\delta_{im}\rangle u_{im} - \frac{1}{\beta} \sum_i \ln\left(\sum_m \exp(\beta u_{im})\right)$$

Unsupervised MFA

- E step

$$\frac{\partial \psi}{\partial \langle \delta_{im} \rangle} = 0 \text{ for all } i, m$$

$$\frac{\partial \psi}{\partial u_{\mathrm{im}}} = 0 \text{ for all } i, m$$

- M step

$$\Delta y_m \propto -\frac{\partial E}{\partial y_m}$$

$$\Delta A_{ab} \propto -\frac{\partial E}{\partial A_{ab}}$$

⇒

$$\Delta y_m = 0$$

$$\Delta A_{ab} = 0$$

E step

$$u_{ik} = -\frac{\partial L}{\partial \langle \delta_{ik} \rangle} = \frac{-1}{2}\left(\mathbf{x}_i - \mathbf{y}_k\right)\mathbf{A}(\mathbf{x}_i - \mathbf{y}_k)$$

$$\langle \delta_{ik} \rangle = \frac{\exp(\beta u_{ik})}{\sum_l \exp(\beta u_{il})} \qquad \text{(E1)}$$

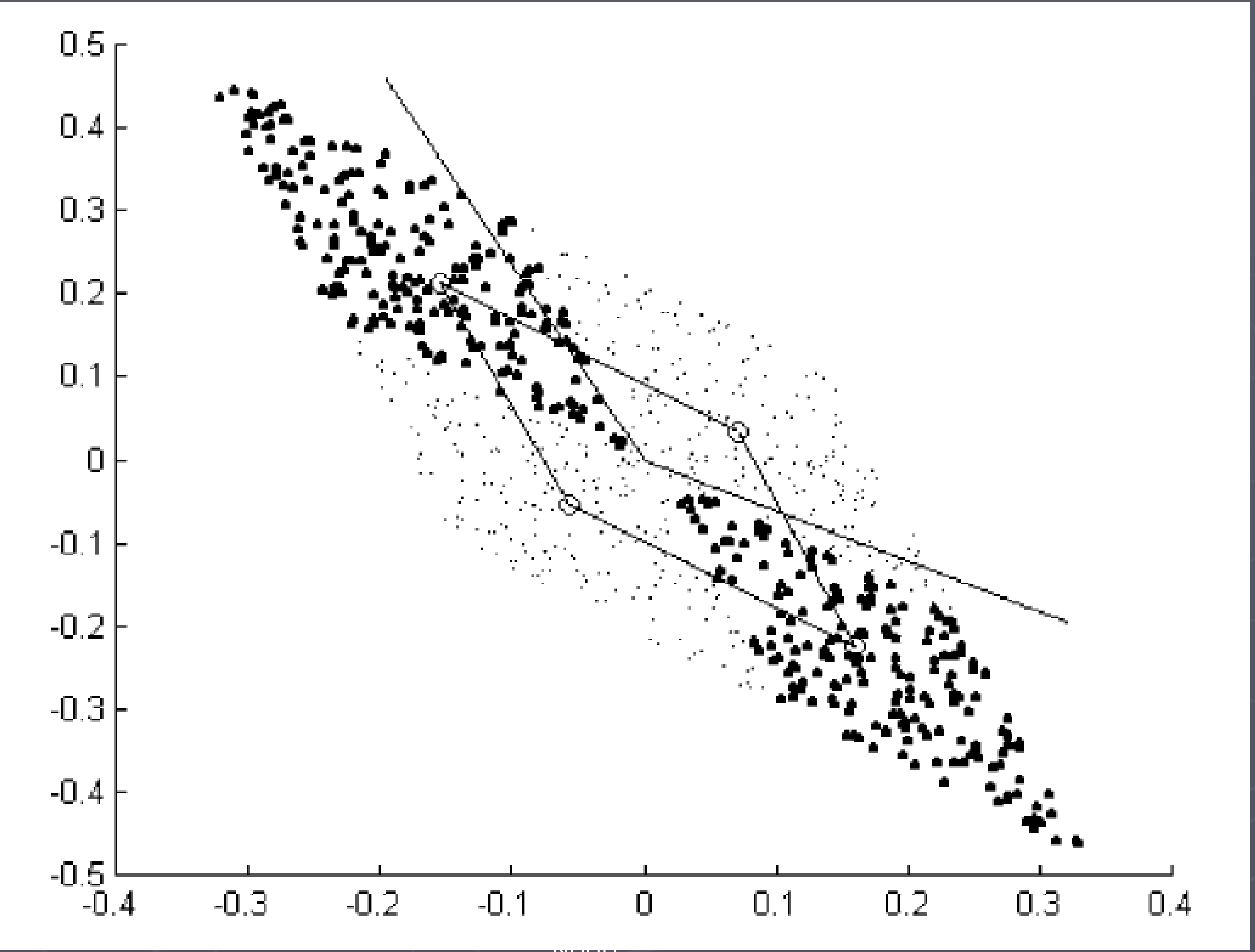
Numerical procedure

1. Initialize β as a sufficiently low value, $A = 0.01 \times I$ (identity matrix),

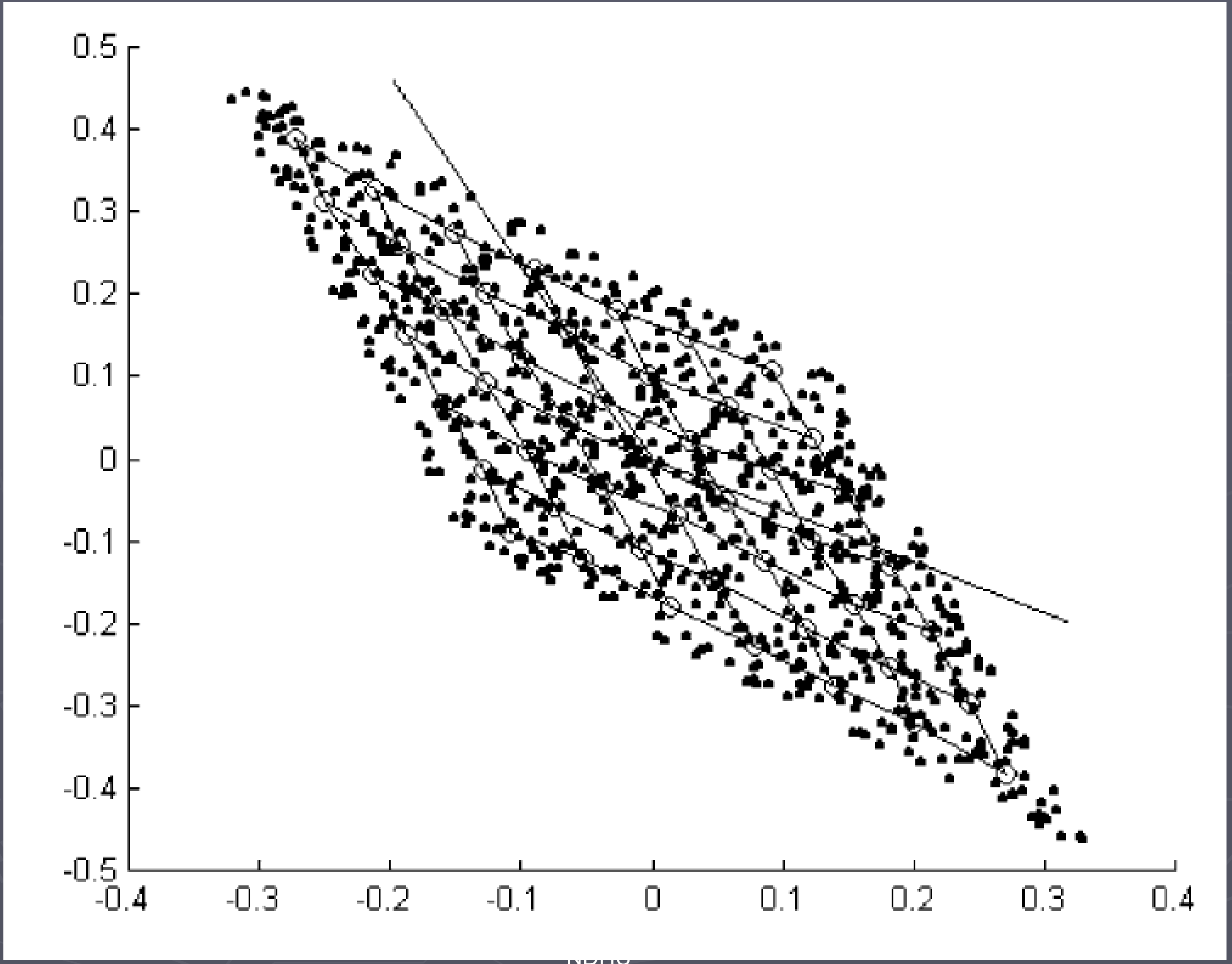
$$y_k \approx \frac{1}{N} \sum_i x_i, \quad \langle \delta_{ik} \rangle \approx \frac{1}{K}$$

2. Update $\{\langle \delta_{im} \rangle\}$ by Eqs. (12) and (13).
3. Update $\{y_m\}$by Eq. (15).
4. Update A by Eqs. (18) and (19).
5. If $\sum_i \sum_m \langle \delta_{im} \rangle^2 > \theta$ then halt, else $\beta \leftarrow \beta * (1/0.98)$, and *goto step* 2, where θ is a threshold, such as $\theta = 0.98^* N$.

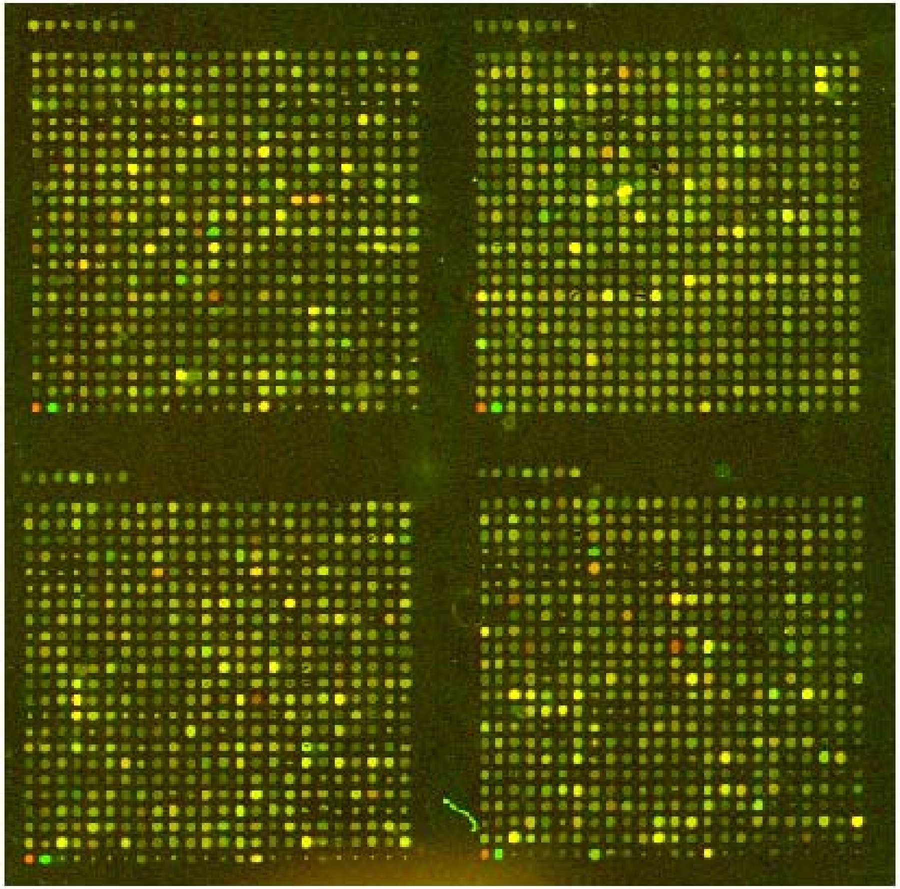
Numerical results



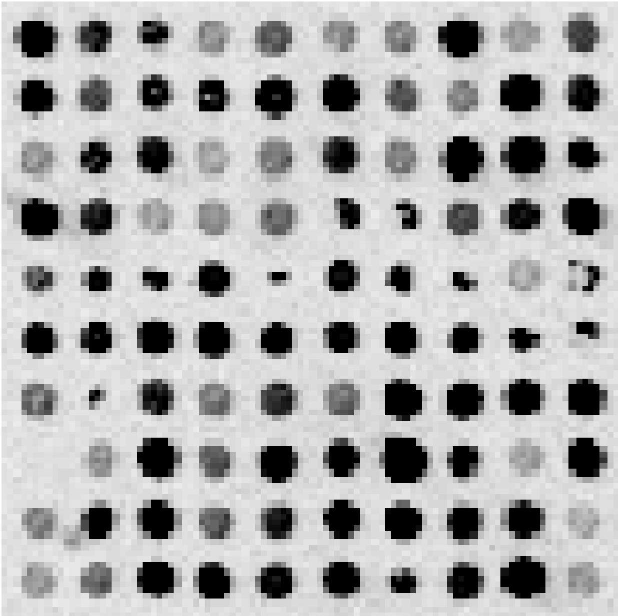
Numerical results

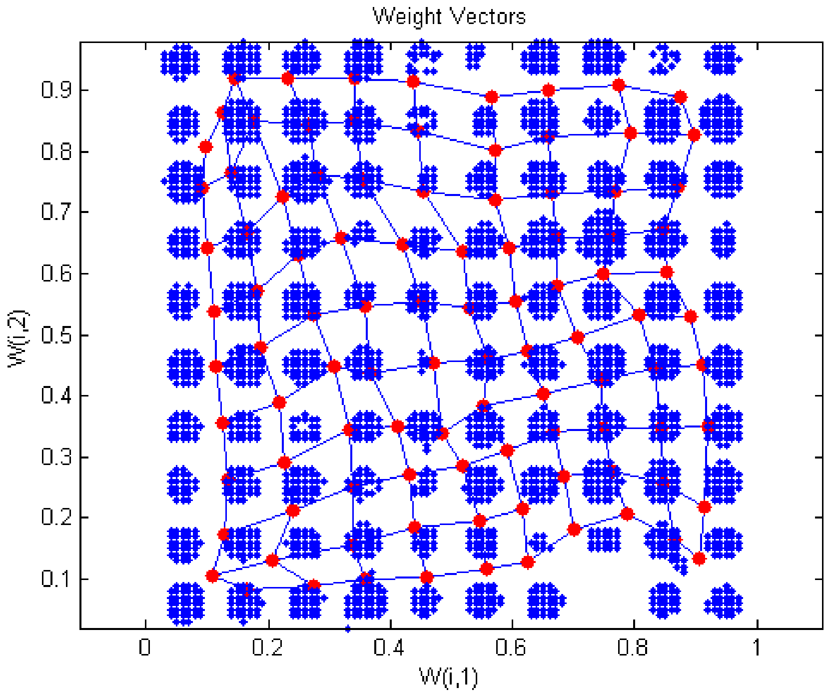


Micro-Array image

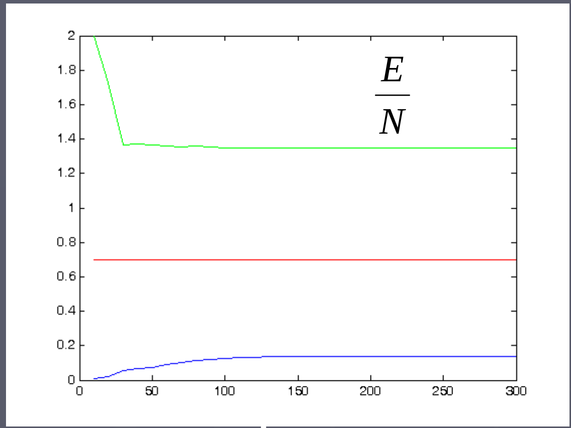


Griding

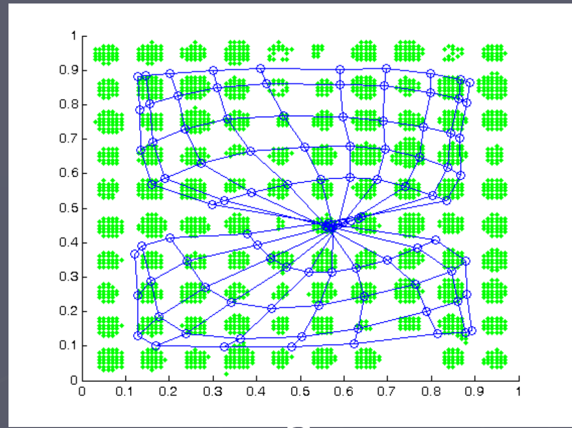




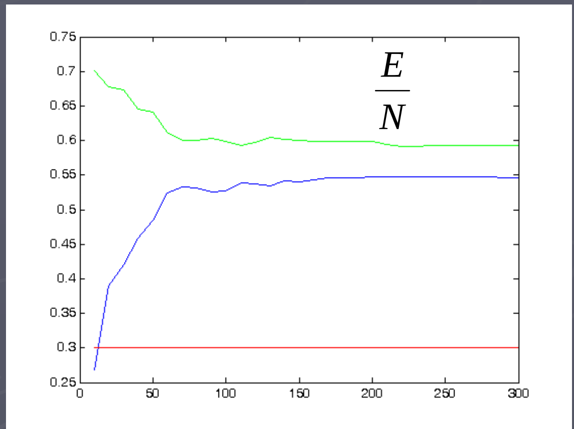
Kohonen SOM algorithm


$\frac{E}{N}$
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2
0 50 100 150 200 250 300

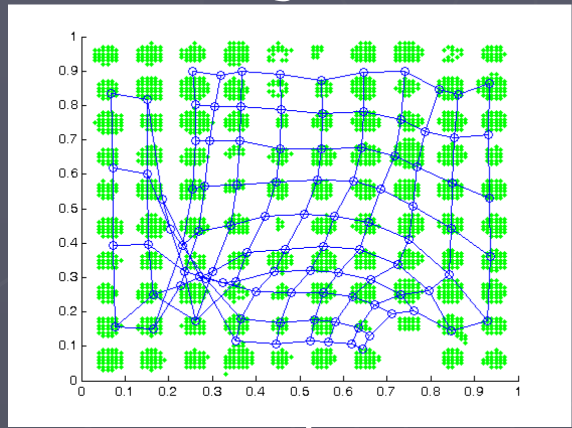
a


0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

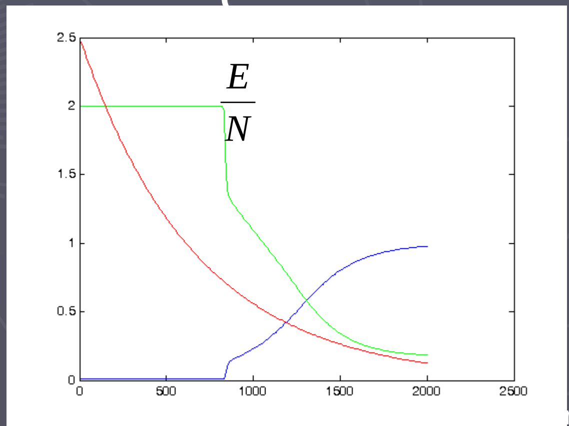
b


$\frac{E}{N}$
0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75
0 50 100 150 200 250 300

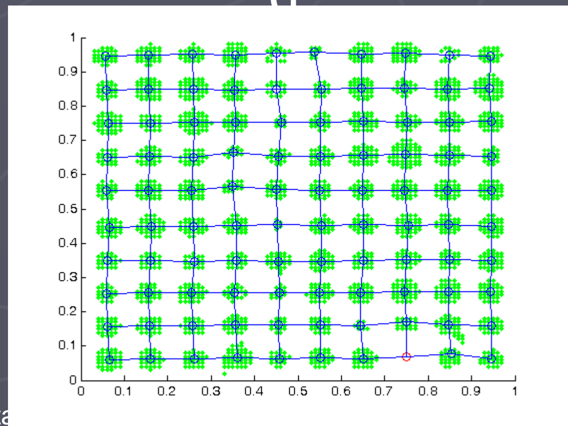
c


0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

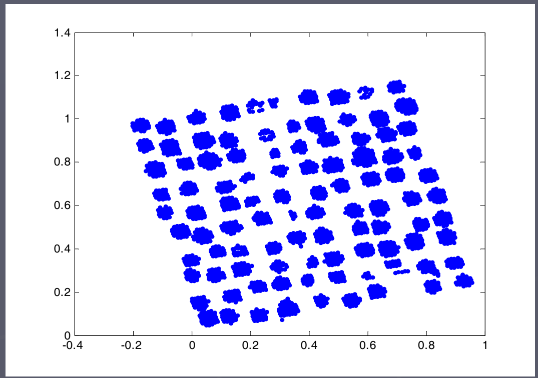
d

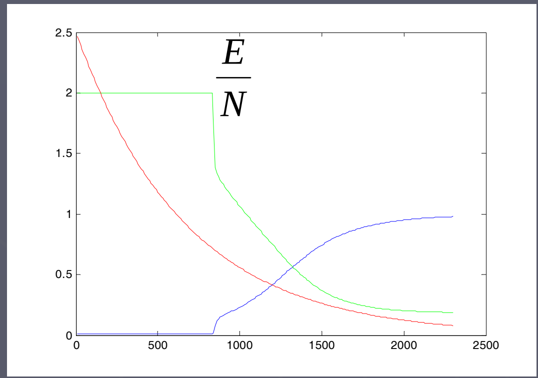

$\frac{E}{N}$
0 0.5 1 1.5 2 2.5
0 500 1000 1500 2000 2500

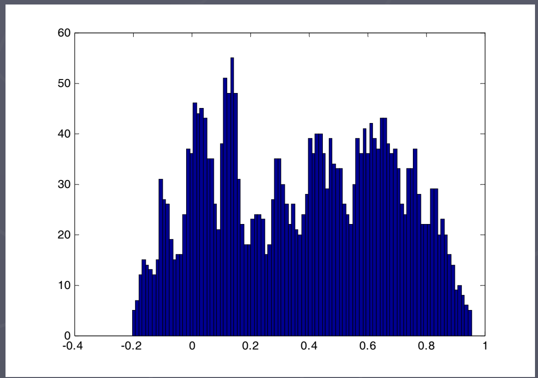
e

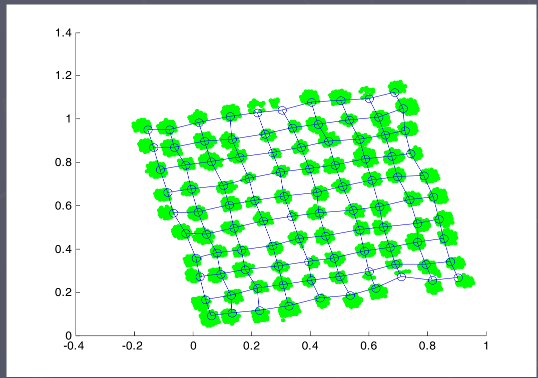

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

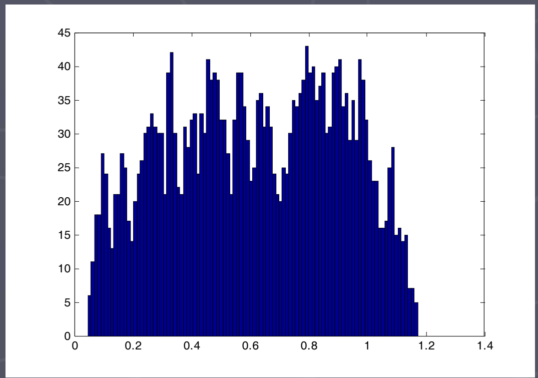
f

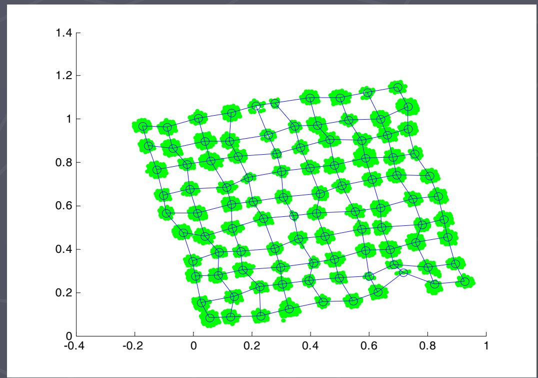



E
N

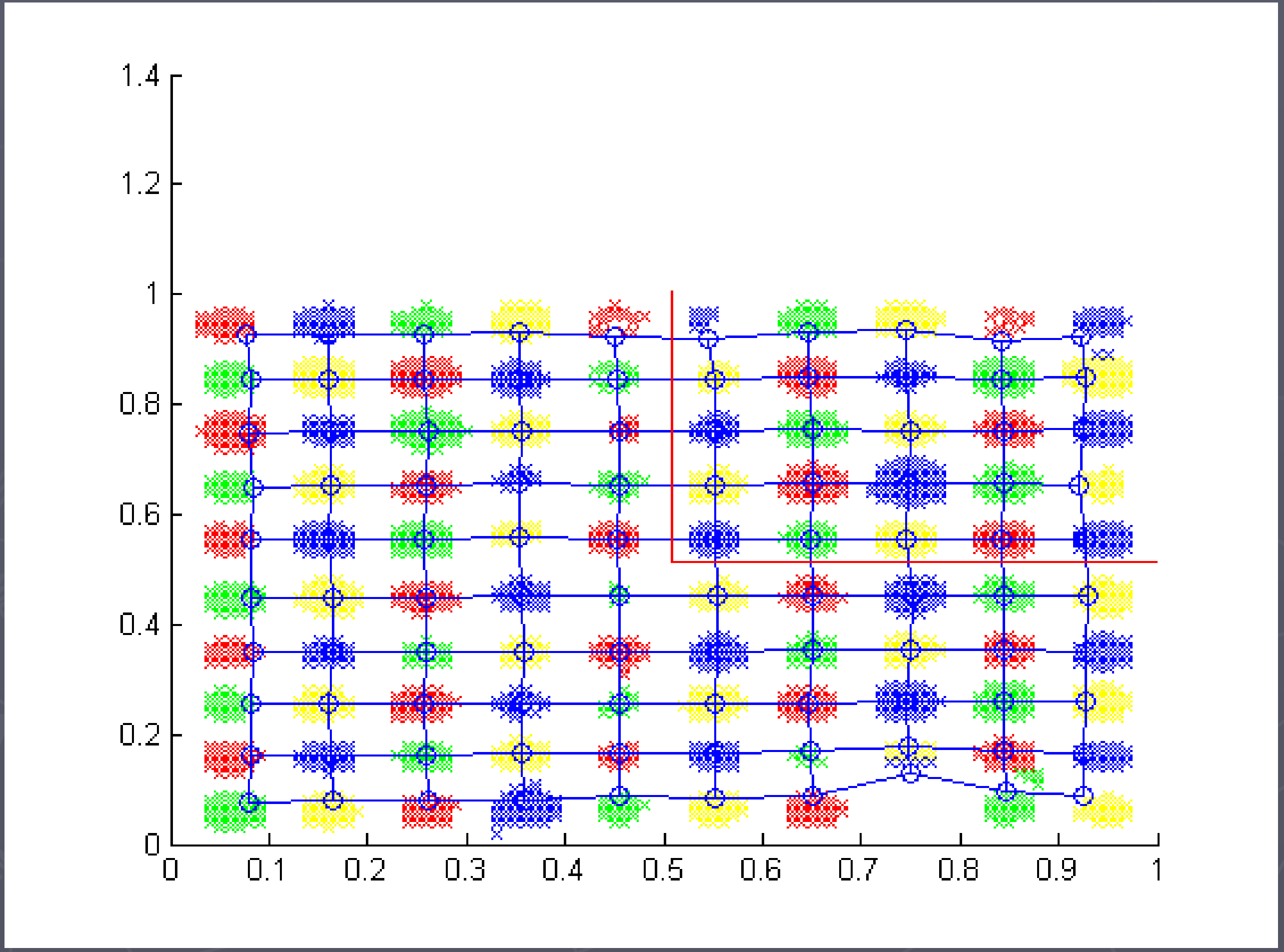




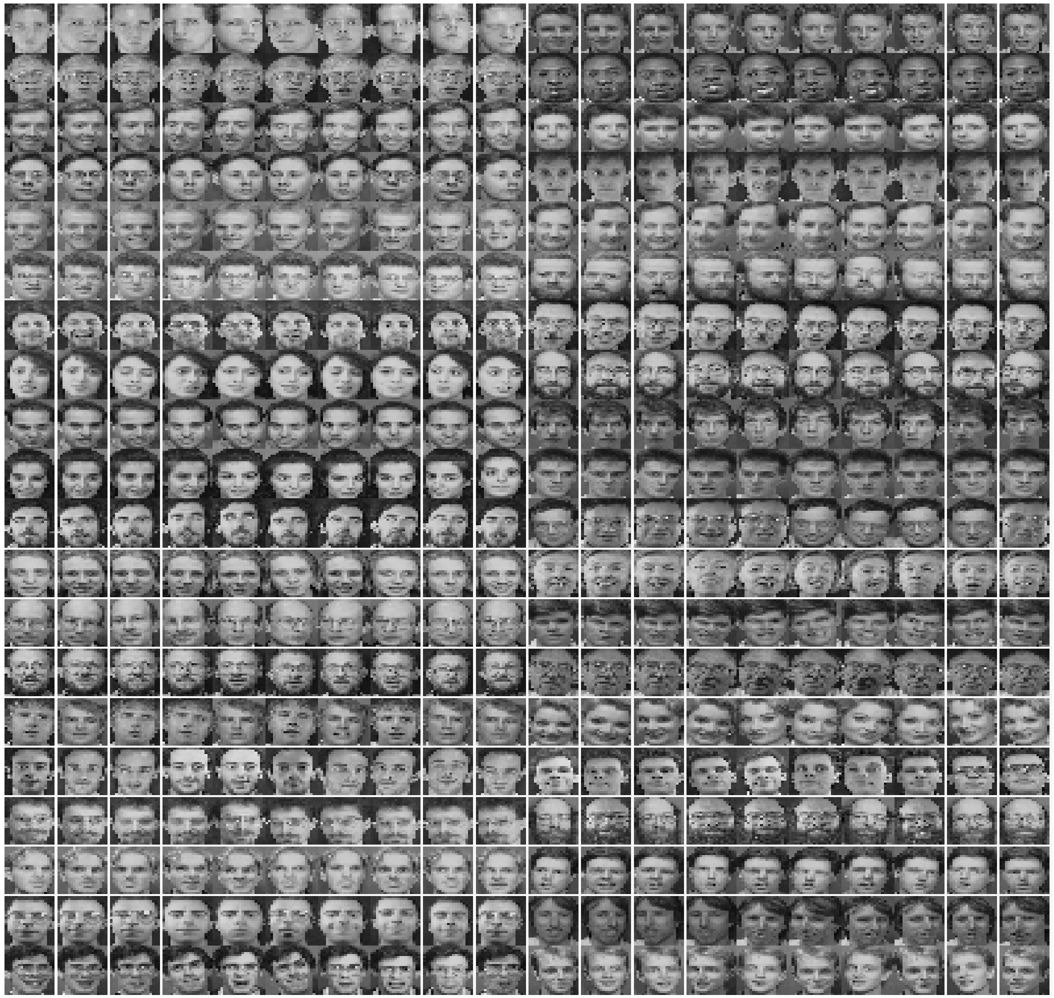




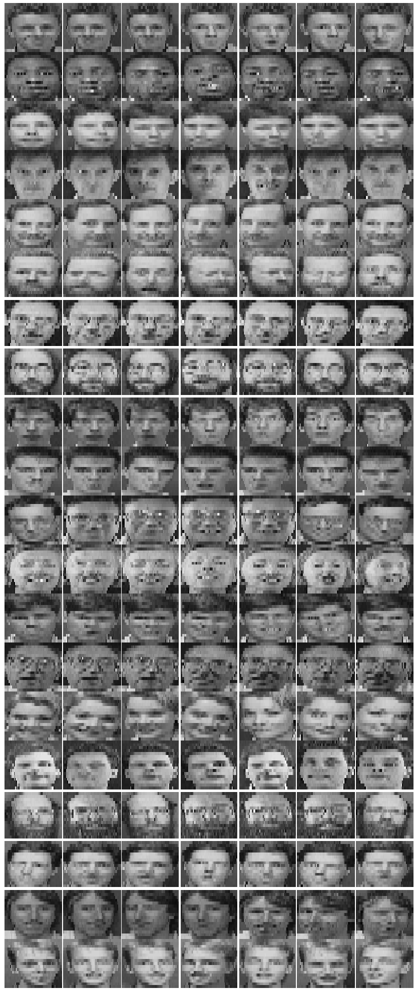
Unsupervised MFA Learning



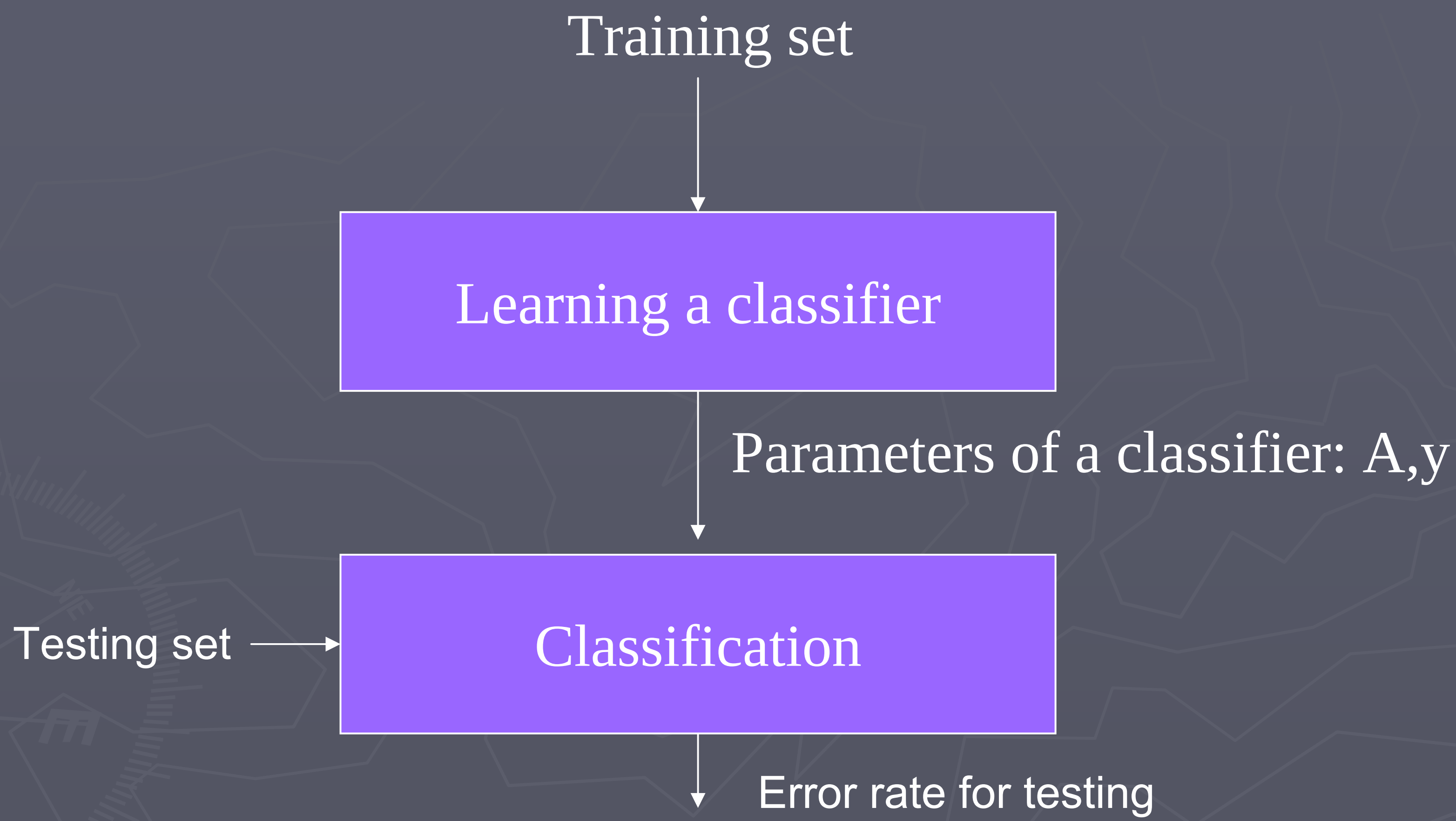
Face classification



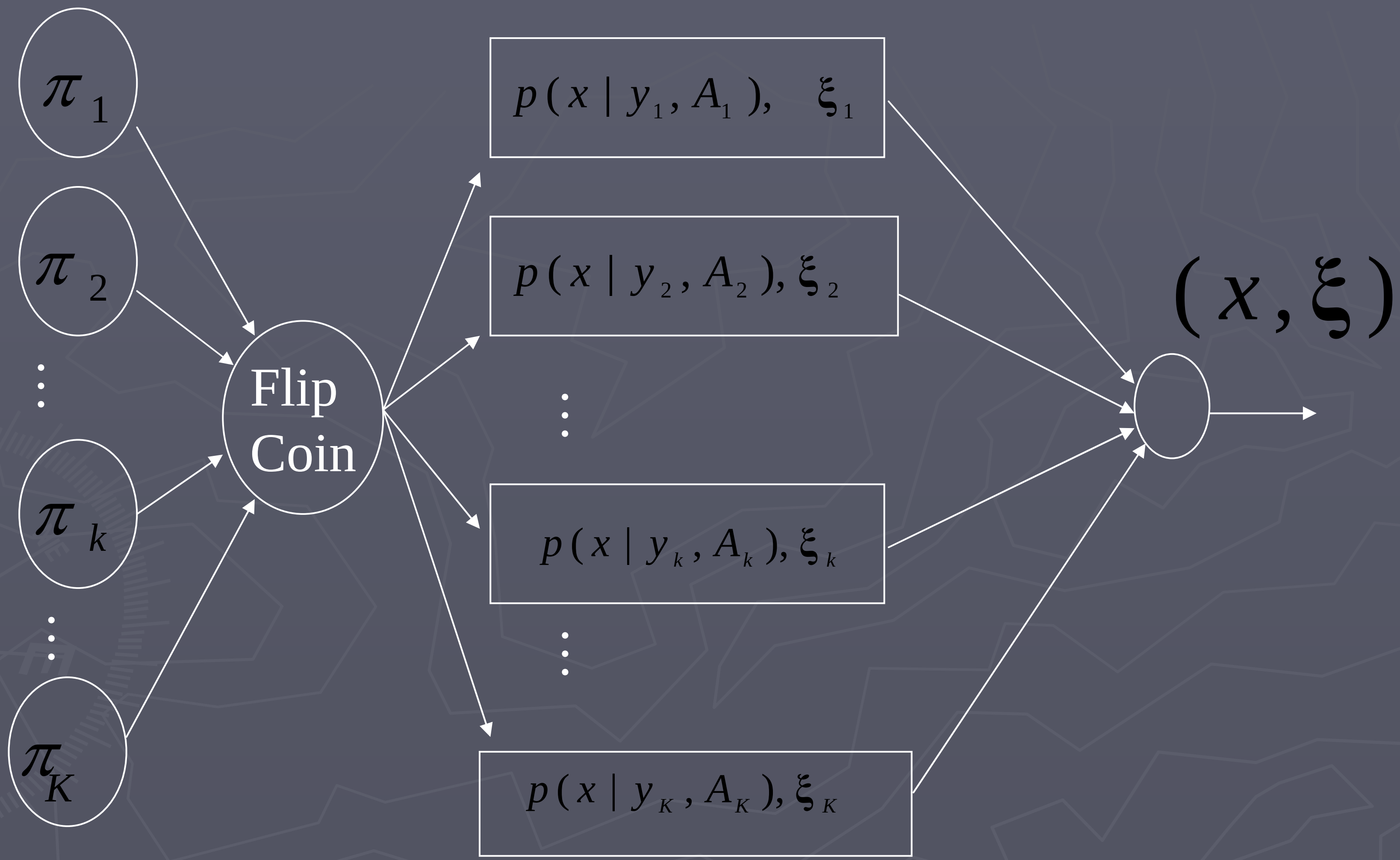
Data sets



Data flow


Training set
Learning a classifier
Parameters of a classifier: A,y
Testing set
Classification
Error rate for testing

A generative model



Classification

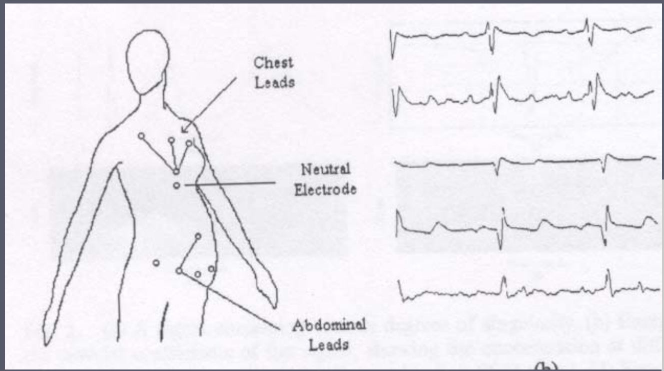
Minimize

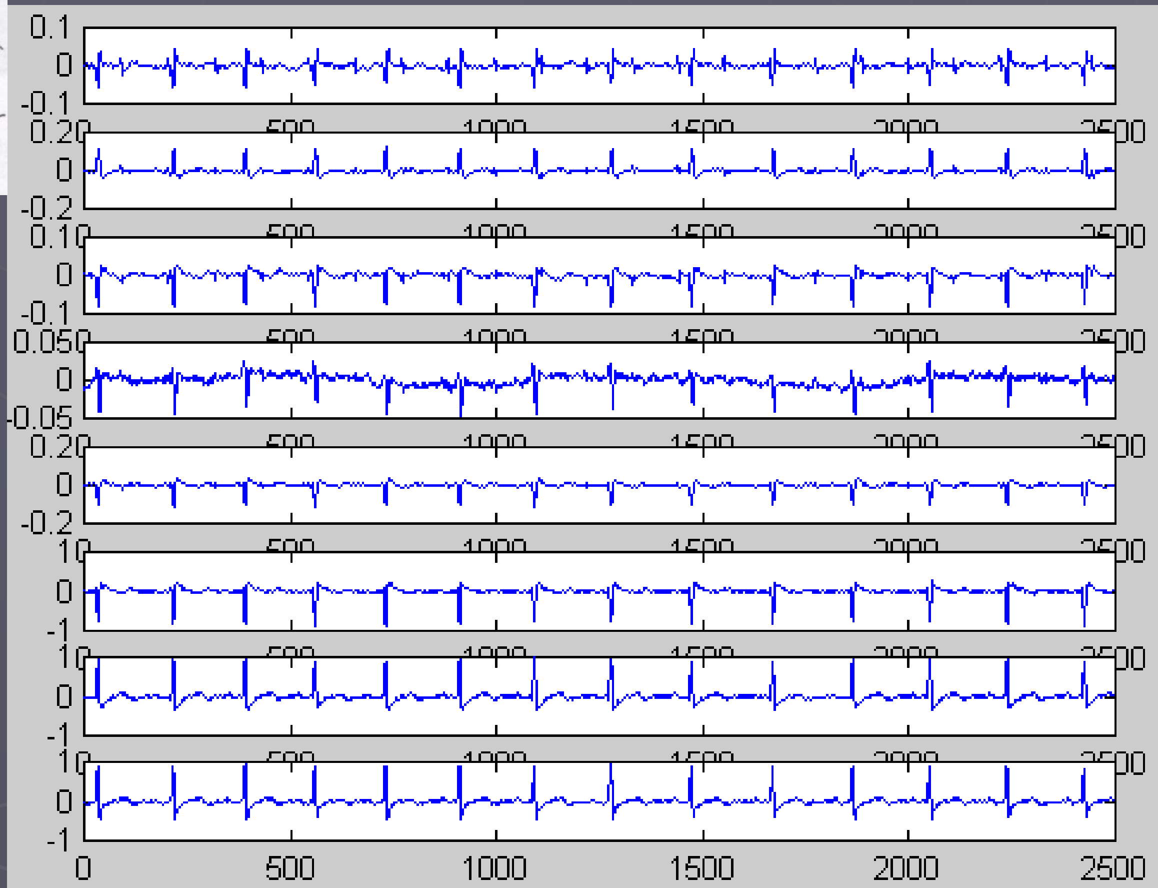
$$E(\delta, \xi, y, A) = E_1 + cE_2$$
$$= \frac{1}{2}\sum_i \sum_k \delta_{ik}(x_i - y_k)' A (x_i - y_k)$$
$$- \frac{N}{2}\log\det(A) + \frac{c}{2}\sum_i ||q_i - \Lambda\delta_i||^2,$$

subject to

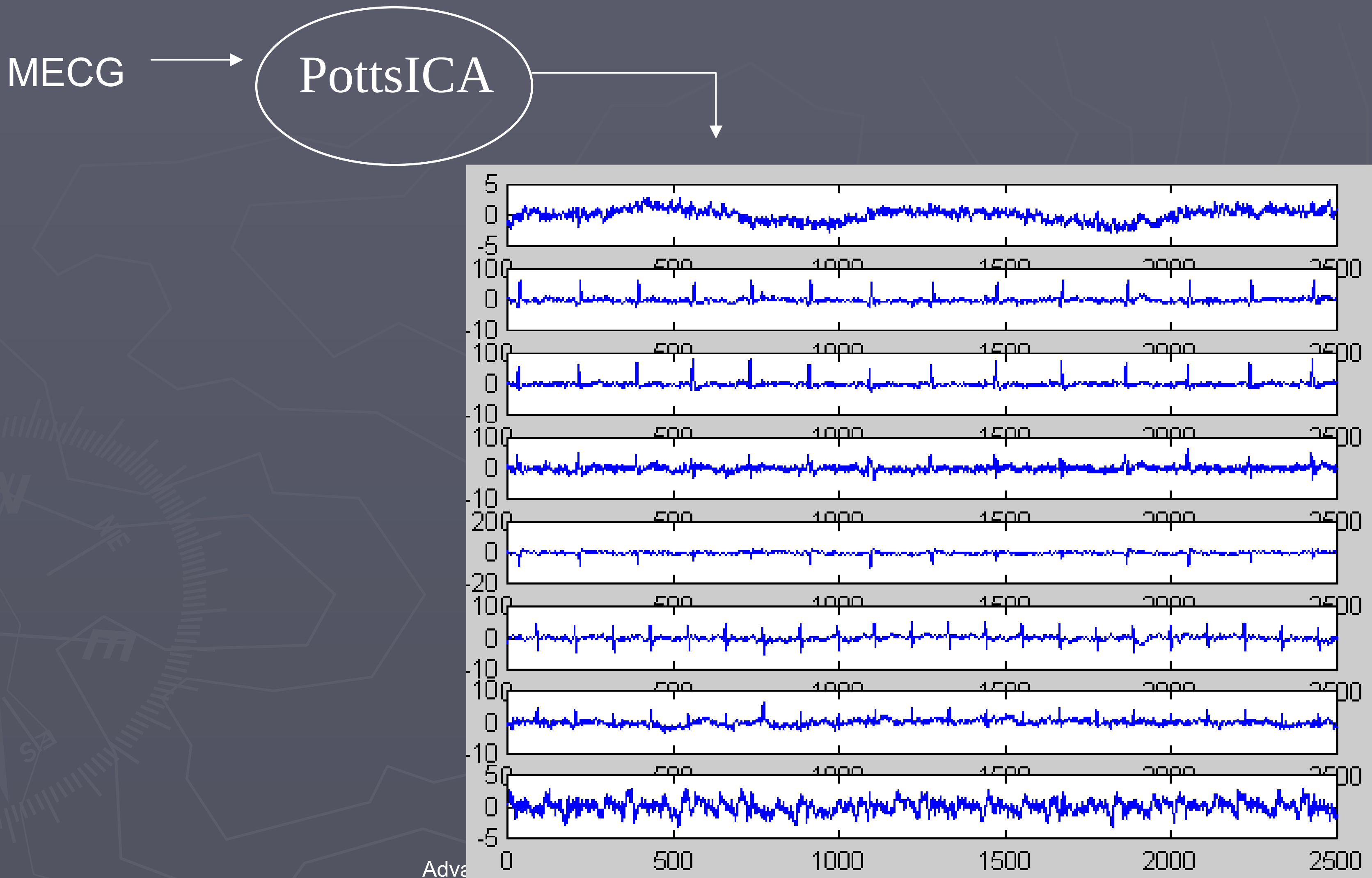
$$\delta_{ik} \in \{0, 1\}, \text{ for all } i, k$$
$$\sum_k \delta_{ik} = 1, \text{ for all } i$$
$$\xi_{km} \in \{0, 1\}, \text{ for all } k, m$$
$$\sum_m \delta_{km} = 1, \text{ for all } k,$$

Maternal ECG



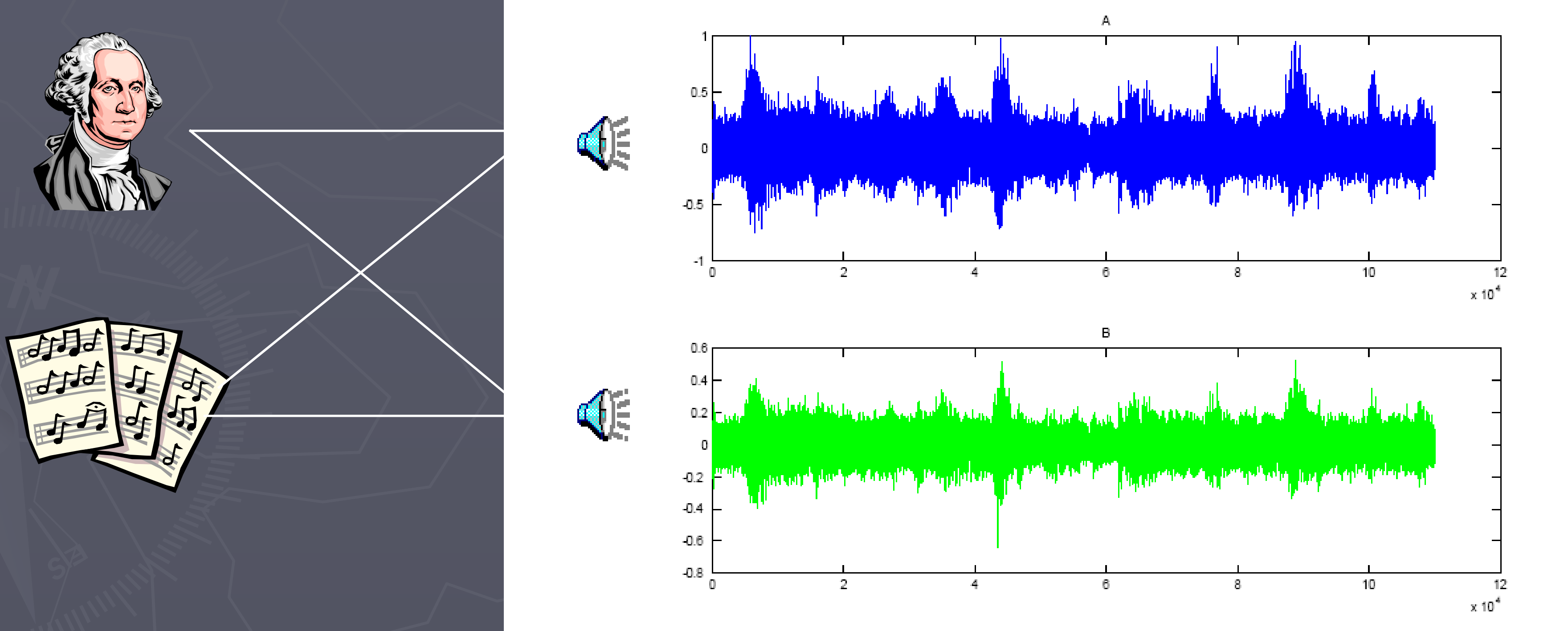


Fetal ECG extraction



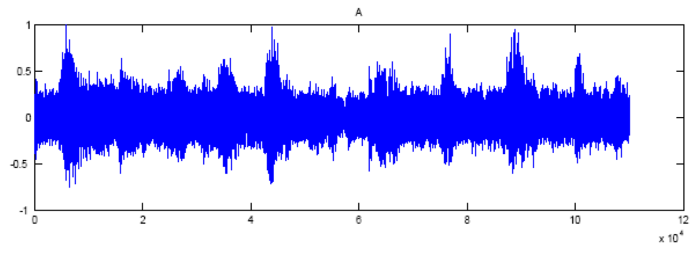
Blind source separation

The recordings of two microphones are shown in the following figure.



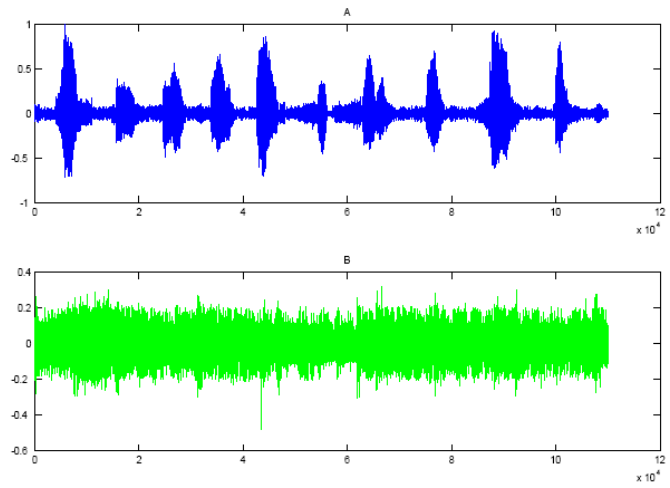
BSS

The recordings of two microphones are shown in the following figure.



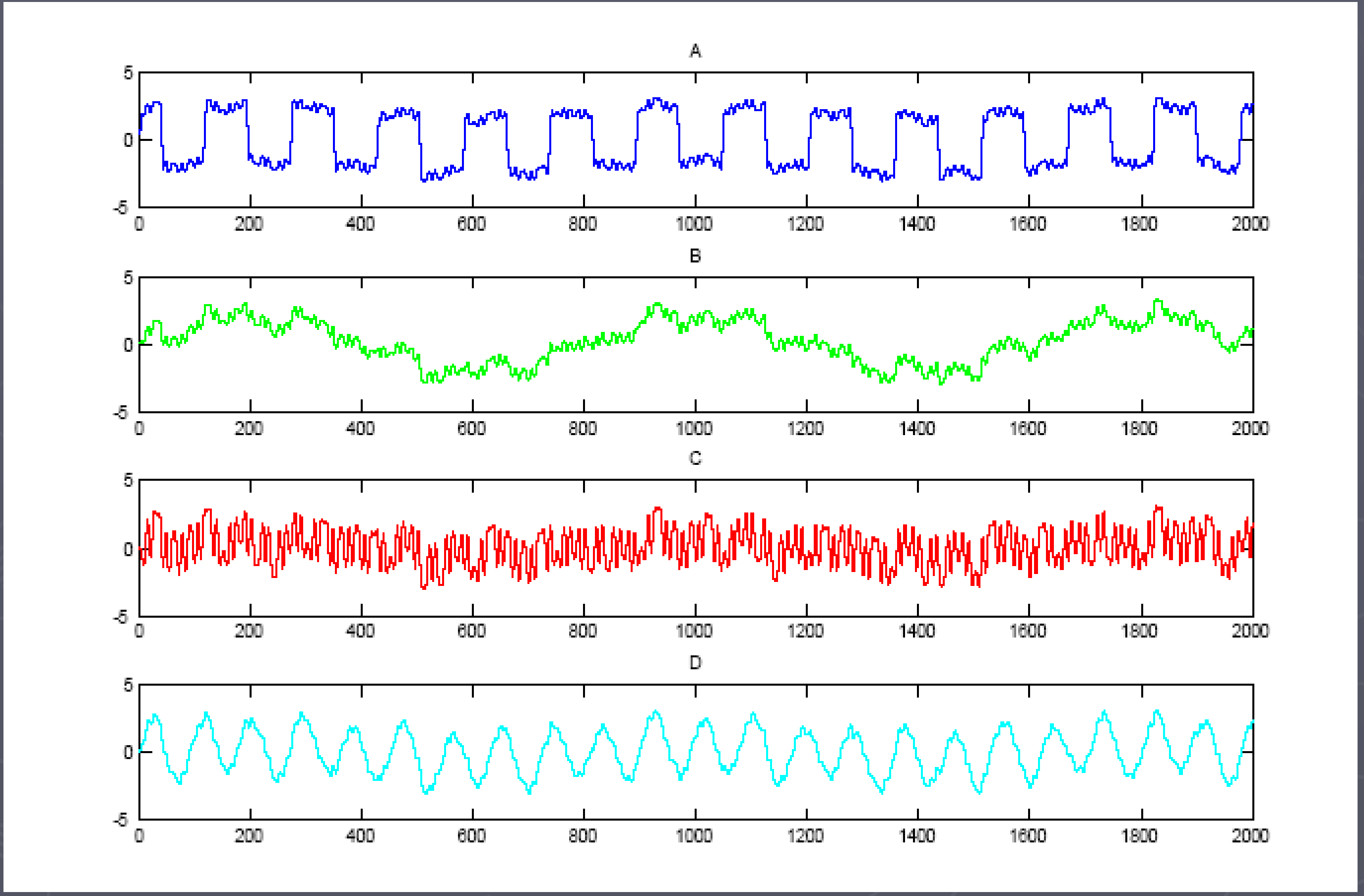
Convolutive ICA

The blind separation of music and speech are shown in the following figure.

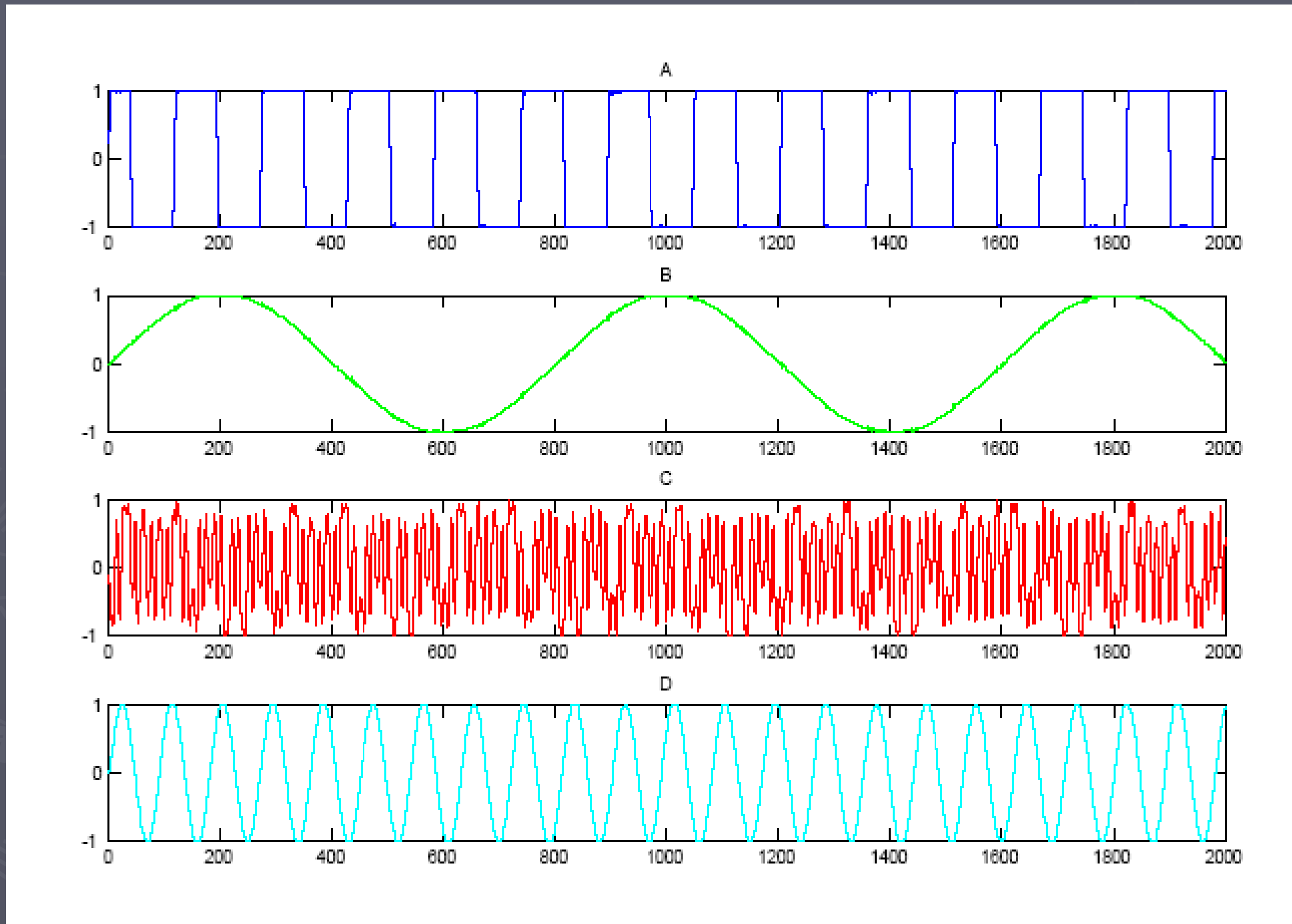


Convolutive mixtures

$$\tau = 5$$



Recovered sources

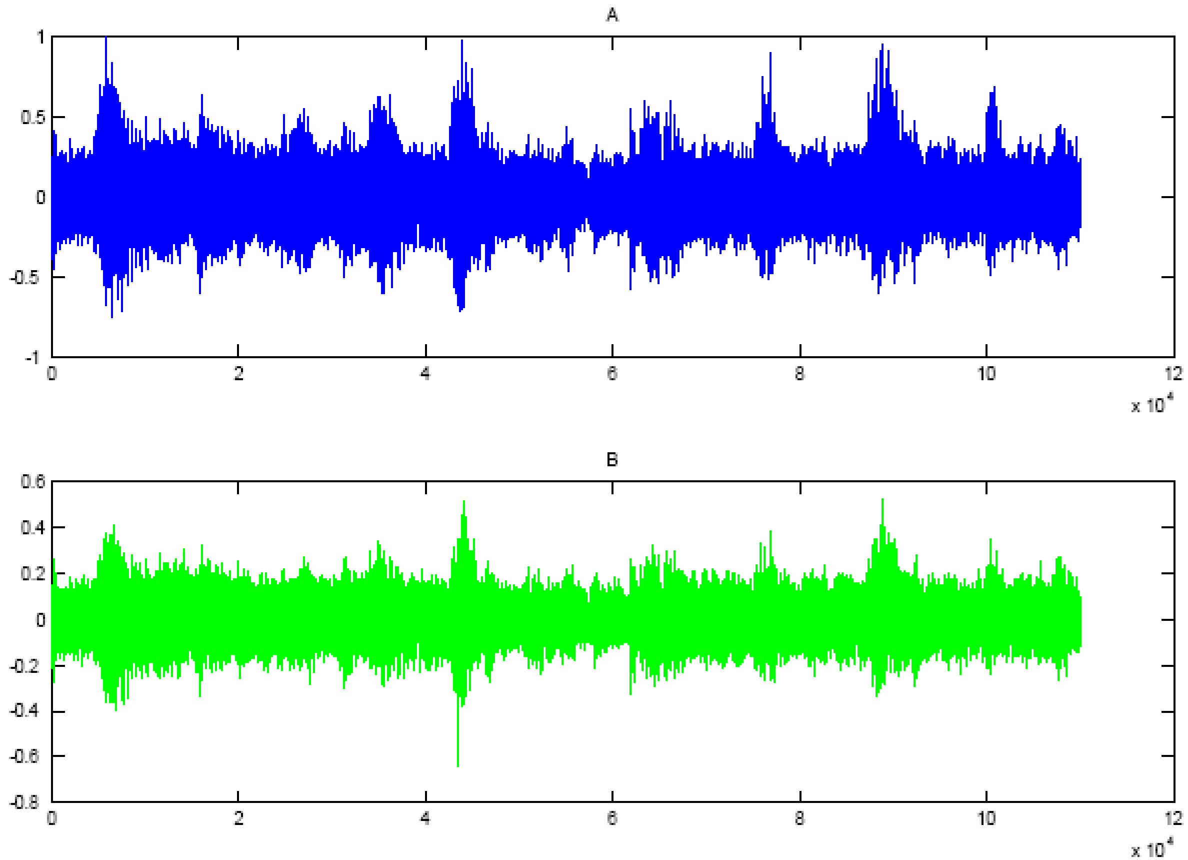


Blind separation of real world signals

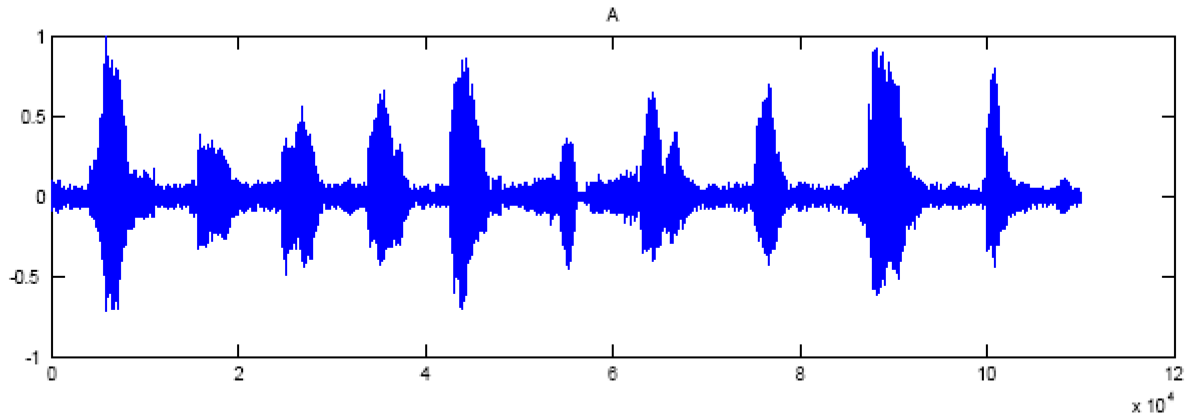
The experiment results as follows:

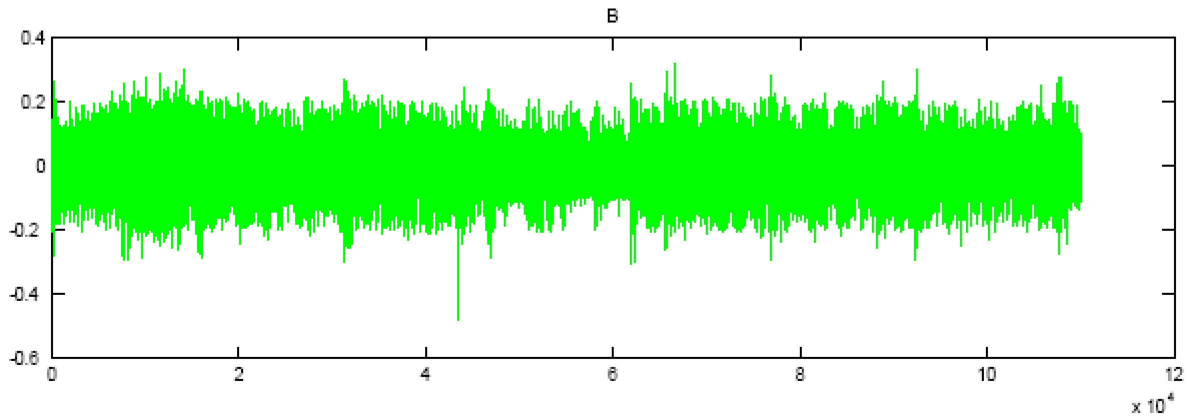
- Two-microphone recordings of music and speech.
  - Channel-1 Sound
  - Channel-2 Sound
- Blind separation of recordings of music and speech.
  - Channel-1 Sound
  - Channel-2 Sound

The recordings of two microphones are shown in the following figure.



The blind separation of music and speech are shown in the following figure.





ICA of mixed facial images

ICA

$$L' = \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K}\delta_{itk}\left\|\mathbf{W}_i\mathbf{x}(t) - h_{ik}\right\|^2$$
$$+ C_1(-\log|\det(\mathbf{W})|)$$
$$+ C_2\left(-\sum_{i=1}^{N}\sum_{k=1}^{K}p_{ik}\log p_{ik}\right)$$

subject to

$$\sum_{k=1}^{K}\delta_{itk} = 1,\ 1 \leqslant i \leqslant N,\ 1 \leqslant t \leqslant T$$
$$\delta_{itk} \in \{0,\ 1\},\ \text{for all } i\ t,\ k$$
$$p_{ik} = \frac{1}{T}\sum_{t=1}^{T}\delta_{itk},\ 1 \leqslant i \leqslant N,\ 1 \leqslant k \leqslant K$$

Demo

- Demo_face_nda
- Jade_K5
- Microarray_nen

Conclusions

- Successfully develop MILP modeling for self-organization, independent component analysis, classification
- Well verify effectiveness of AEM for relevant bio-signal analysis

References

- Jiann-Ming Wu & Lin, Neural Networks, Vol 15, No. 3(2002)
- Jiann-Ming Wu, Neural Computation,Vol. 14, No. 3(2002)
- Jiann-Ming Wu & Chiu, IEEE Trans. On Neural Networks, Vol. 12, No. 2, March (2001)