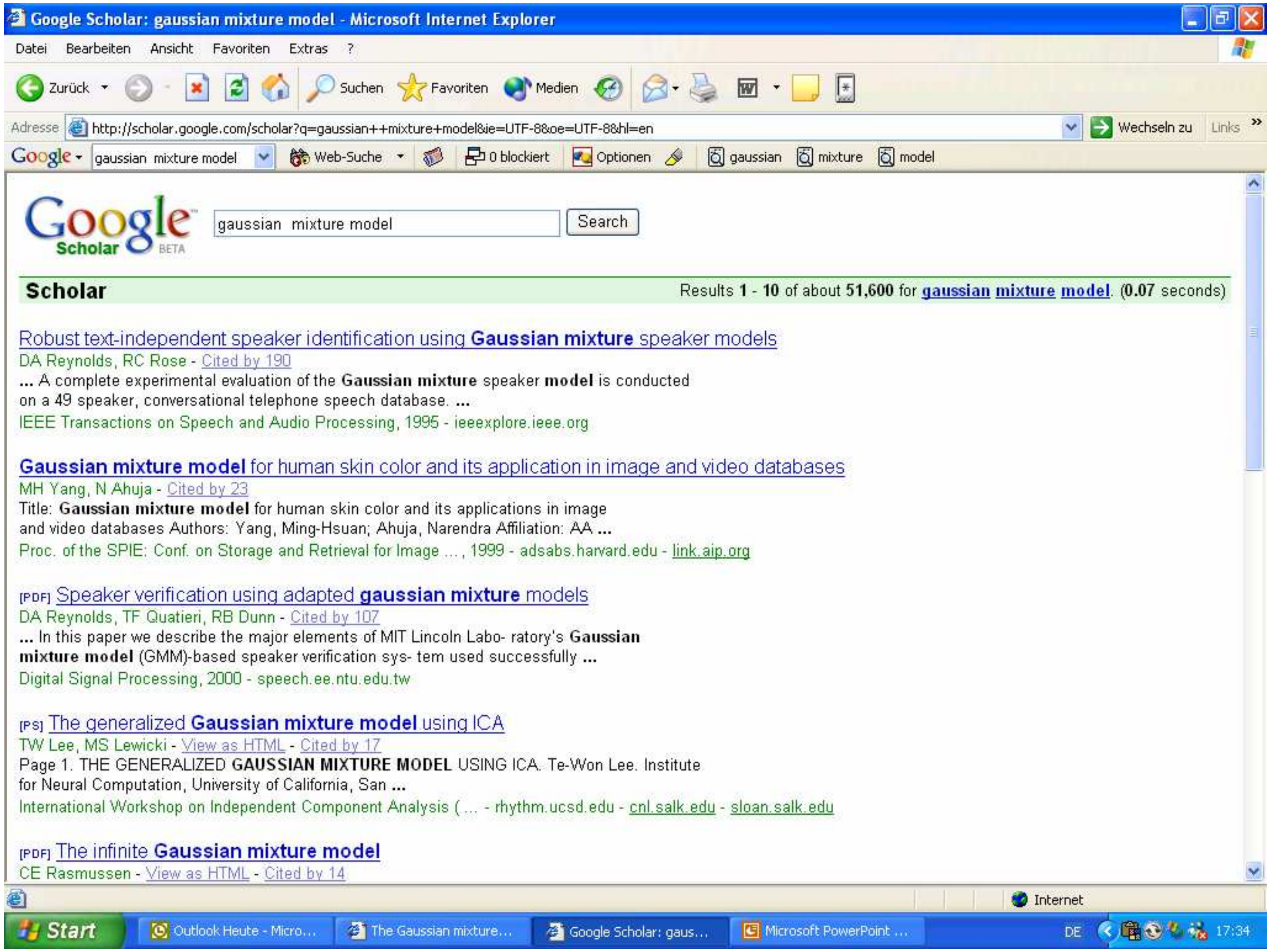
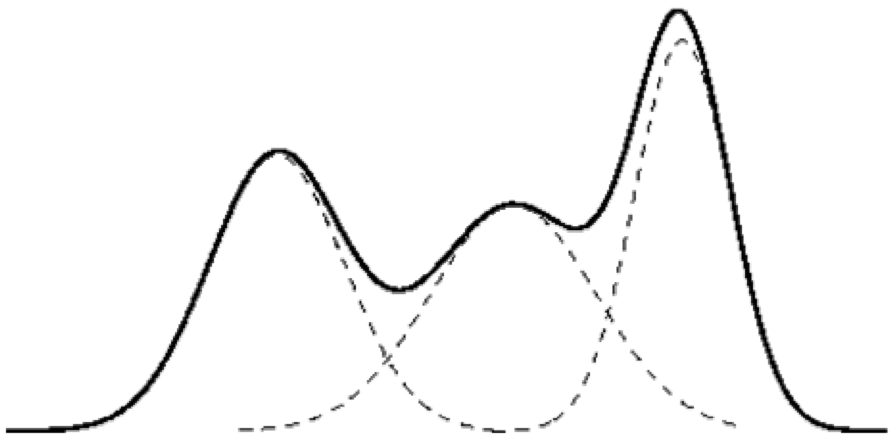
8. Gaussian Mixture Models and the EM-Algorithm

Applications of Gaussian Mixture Models

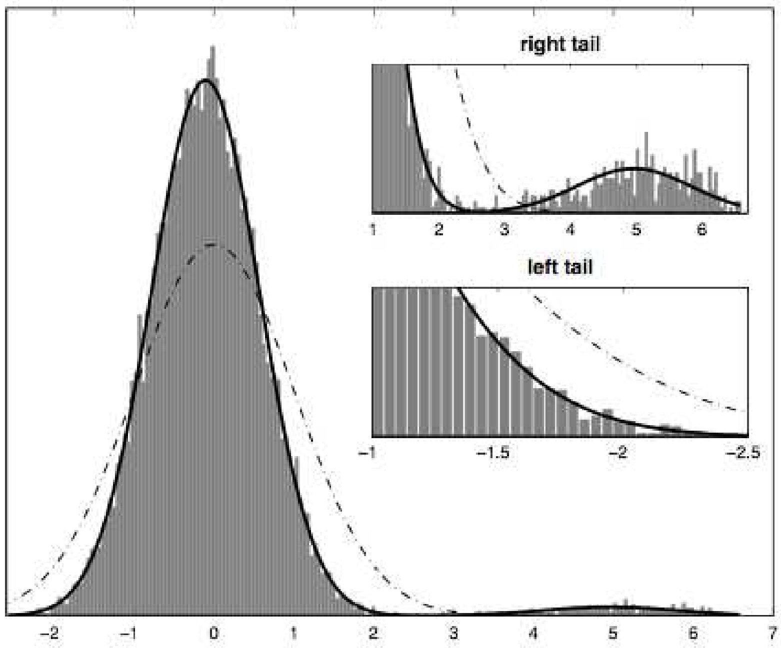


Scematic 1-D Example

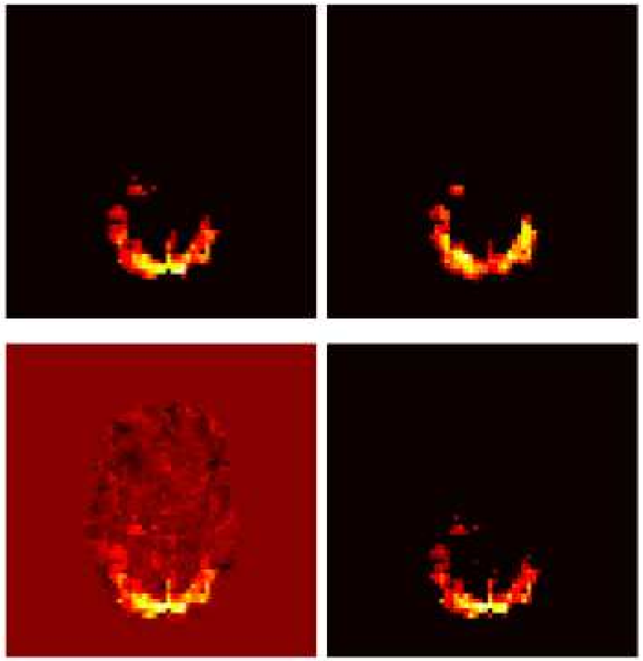


Application: Signals from Magnetic Resonance in Medicine

(www.fmrib.ox.ac.uk/.../ tr02cb1/node8.html)

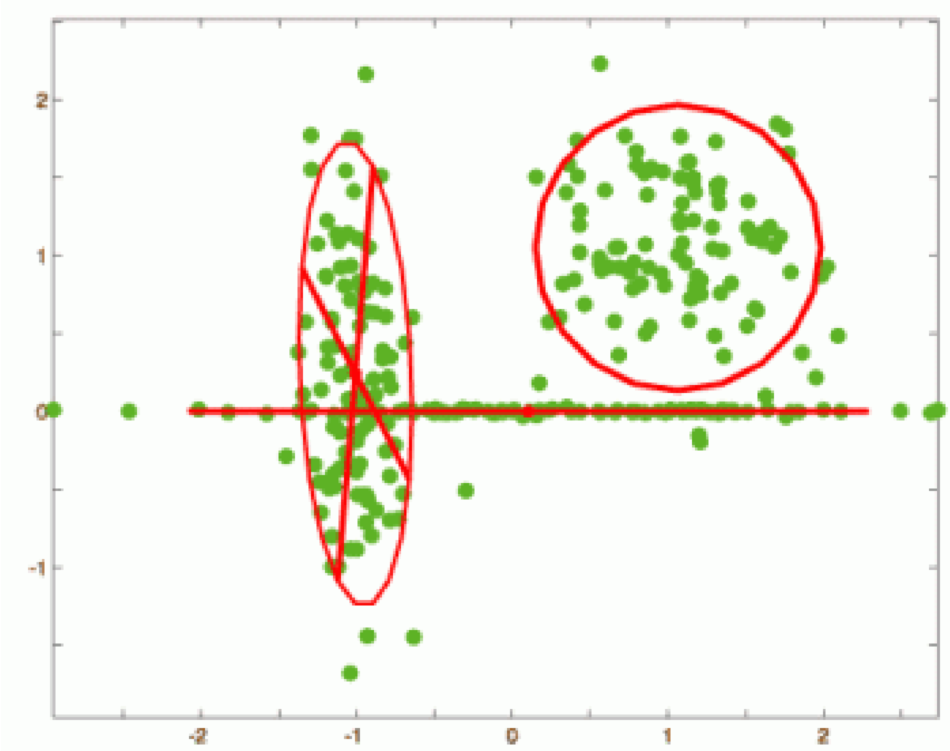


(i)



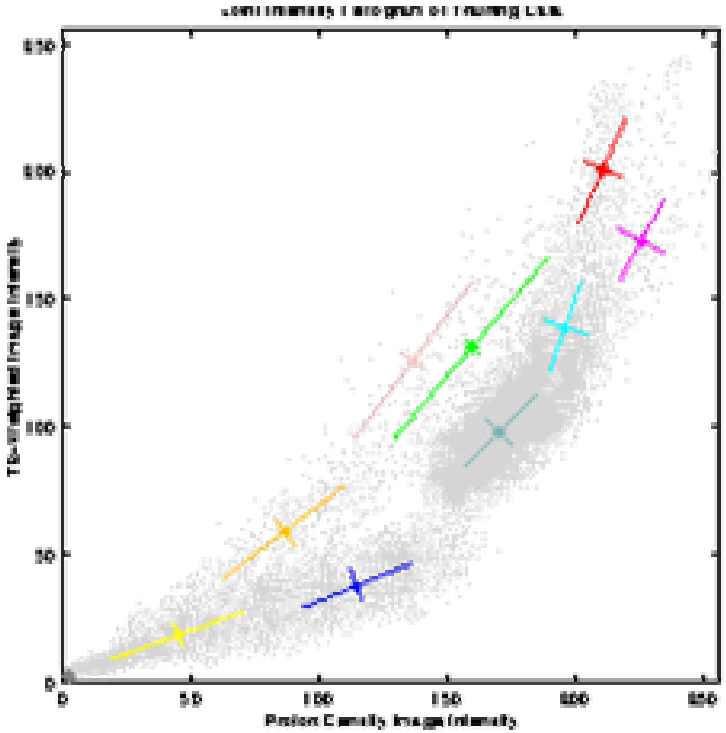
(ii)

Schematic 2 dimensional Example



Segmentation of Images

(**www.ai.mit.edu/.../ 9810-MICCAI-Reg/node6.html**)



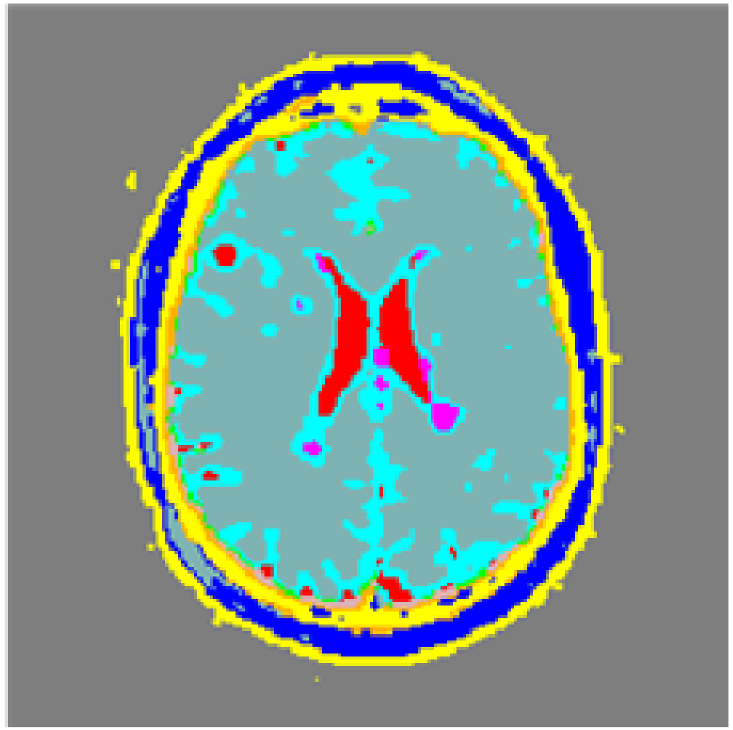
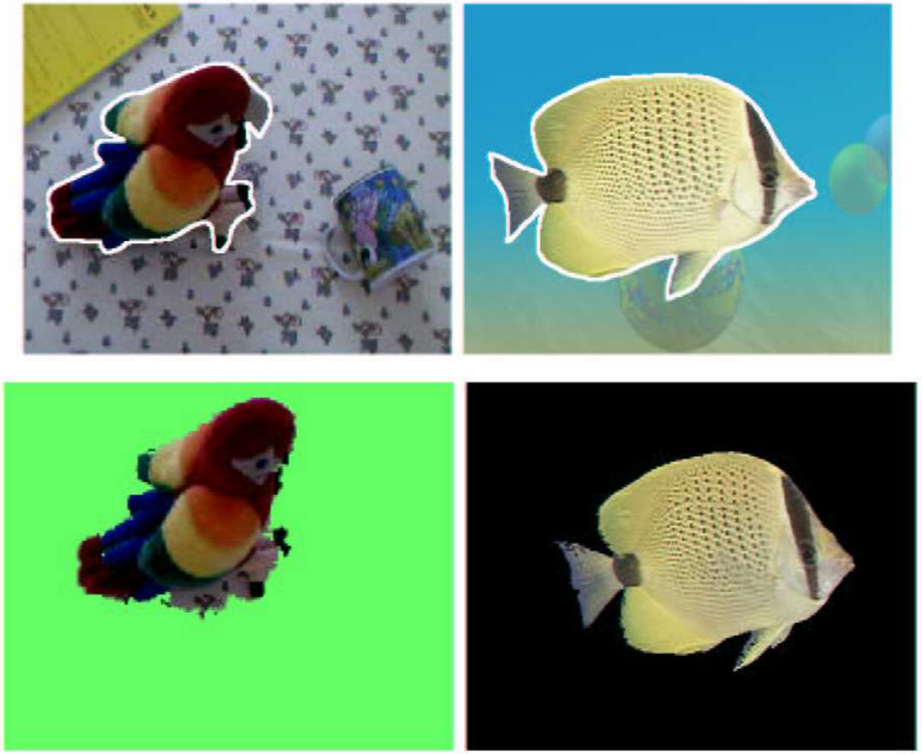


Image Segmentation Using GMMs

(http://dircweb.king.ac.uk/papers/Thirde_D.J.2004_617535/1757_Thirde_D.pdf)



Gaussian Mixture Model

Mixture model:

$$P(x \mid \theta_1 \ldots \theta_K\, p_1 \ldots p_K) = \sum_{k=1}^{K} p_k\, p(x \mid \theta_k)$$

With:

mixture weights p_k

Gaussian distributions:

$$p(x \mid \theta_k) = p(x \mid \mu_k \Sigma_k)$$

$$= \frac{1}{(2\pi)^{d/2} \mid \Sigma \mid^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^t \Sigma^{-1} (\vec{x}-\vec{\mu})}$$

Gaussian Mixture Model

- Abbreviation: GMM
- GMM: weighted average of Gaussians
- Each Gaussian has its own mean and covariance matrix that has to be estimated separately
- Unlike in the case with just one Gaussian, you do not know which training sample contributes to which Gaussian and hence the existing formulas for mean and covariance matrix are not applicable

EM-Algorithm

- Iterative optimization of the likelihood
- Applications:
  - Mixture models
  - Grammars
- Auxiliary objective function

$$Q(\theta^j, \theta^{j+1}) = \sum_{i=1}^{N} \sum_{y} P(y \mid x_i, \theta^j) \log P(y, x_i \mid \theta^{j+1})$$

y : hidden property (e.g. assignement of training samples to Gaussians)

x_i : training samples

EM-Algorithm

- Initialize Θ^0
- j=0
- Iterate until converged:
  - Calculate $Q(\Theta^j, \Theta^{j+1})$
  - Update parameters $\theta^{j+1} = \arg\max_{\theta^{j+1}} Q(\theta^j, \theta^{j+1})$

Warning: actual applications of the EM algorithm can become quite messy!

Comments EM

- EM-algorithm can be derived from the maximum likelihood principle
- Formula for Q function depends on the specific problem to be solved
- For convergence, Q function has to satisfy a couple of properties (e.g. being bounded)
- Best being understood using an example

Determine necessary parts for Q

$$Q(\theta^j, \theta^{j+1}) = \sum_{i=1}^{N} \sum_{Y} \underline{P(y \mid x_i, \theta^j)} \log P(y, x_i \mid \theta^{j+1})$$

$$P(y \mid x_i, \theta^j) = \frac{P(y, x_i \mid \theta^j)}{\sum_{x_i} P(y, x_i \mid \theta^j)}$$

$$= \frac{P(y \mid \theta^j) P(x_i \mid y, \theta^j)}{\sum_{y} P(y \mid \theta^j) P(x_i \mid y, \theta^j)}$$

Denominator:
calculated directly from normalization using numerator

Application to a Mixture Model

- The hidden property is the id of the Gaussian distribution responsible for generating a particular piece of data
- y=k
- Notational simplification: drop j index wherever possible

Application to a Mixture Model: Identify parts of GMM

Mixture weights: $P(y \mid \theta) \rightarrow P(k \mid \theta) = p_k$

Gaussian distributions: $P(x_i \mid y, \theta) \rightarrow P(x_i \mid k, \theta) = P(x_i \mid \theta_k)$

First part of Q:

$$P(y \mid x_i, \theta) = \frac{P(y \mid \theta) P(x_i \mid y, \theta)}{\sum_y P(y \mid \theta) P(x_i \mid y, \theta)}$$

$$\rightarrow \frac{p_k P(x_i \mid \theta_k)}{\sum_y p_k P(x_i \mid \theta_k)}$$

Now consider second part of Q function

$$Q(\theta^j, \theta^{j+1}) = \sum_{i=1}^{N} \sum_{Y} P(y \mid x_i, \theta^j) \log \underline{P(y, x_i \mid \theta^{j+1})}$$

$$P(y, x_i \mid \theta^{j+1}) = P(y \mid \theta^{j+1}) P(x_i \mid y, \theta^{j+1})$$

$$\rightarrow p_k^{\,j+1} P(x_i \mid \theta_k^{\,j+1})$$

EM-Objective Function Q for Mixture Model

Generic Q

$$Q(\theta^j, \theta^{j+1}) = \sum_{i=1}^{N} \sum_{y} P(y \mid x_i, \theta^j) \log P(y, x_i \mid \theta^{j+1})$$

Specific Q for mixture model:

$$Q(\theta^j, \theta^{j+1}) = \sum_{i=1}^{N} \sum_{k} \frac{p_k{}^j P(x_i \mid \theta_k{}^j)}{\sum_{k'} p_{k'}{}^j P(x_i \mid \theta_{k'}{}^j)} \log\left(p_k{}^{j+1} P(x_i \mid \theta_k{}^{j+1})\right)$$

Ready for direct optimization?

Optimize for mixture weights

Mixture weights are normalized:

$$\sum_k p_k{}^{j+1} = 1$$

Introduce Lagrange multiplier:

$$Q(\theta^j, \theta^{j+1}) = \sum_{i=1}^{N} \sum_k \frac{p_k{}^j P(x_i \mid \theta_k{}^j)}{\sum_{k'} p_{k'}{}^j P(x_i \mid \theta_{k'}{}^j)} \log\left(p_k{}^{j+1} P(x_i \mid \theta_k{}^{j+1})\right)$$
$$+ \mu(1 - \sum_k p_k{}^{j+1})$$

Comments Lagrange multiplier

- General “trick” to make optimization of a function simple, when constraints have to be taken into account
- Details (proof) can be found in good books on calculus

Taking first derivative

$$\frac{\partial \hat{Q}(\theta^j, \theta^{j+1})}{\partial p_k{}^{j+1}} = \sum_{i=1}^{N} \frac{p_k{}^j P(x_i \mid \theta_k{}^j)}{\sum_{k'} p_{k'}{}^j P(x_i \mid \theta_{k'}{}^j)} \frac{1}{p_k{}^{j+1}} - \mu = 0$$

Solve for new mixture weight:

$$p_k{}^{j+1} = \frac{1}{\mu} \sum_{i=1}^{N} \frac{p_k{}^j P(x_i \mid \theta_k{}^j)}{\sum_{k'} p_{k'}{}^j P(x_i \mid \theta_{k'}{}^j)}$$

What´s the problem
left to be solved?

Determine the Lagrange Multiplier

$$1 = \sum_k p_k{}^{j+1}$$

Normalization

$$= \frac{1}{\mu} \sum_{i=1}^{N} \frac{\sum_k p_k{}^j P(x_i \mid \theta_k{}^j)}{\sum_{k'} p_{k'}{}^j P(x_i \mid \theta_{k'}{}^j)}$$

Plug in result
from optimization

$$= \frac{1}{\mu} \sum_{i=1}^{N} 1 = \frac{N}{\mu}$$

$$\Rightarrow \qquad \mu = N$$

Iteration scheme for mixture weights

$$p_k^{\;j+1} = \frac{1}{N}\sum_{i=1}^{N}\frac{p_k^{\;j}P(x_i \mid \theta_k^{\;j})}{\sum_{k'} p_{k'}^{\;j}P(x_i \mid \theta_{k'}^{\;j})}$$

Usually a few iterations are sufficient

Update of Means and Covariance Matrix

Define auxiliary function $$\gamma^{j}{}_{i,k} = \frac{p_k{}^{j} P(x_i \mid \theta_k{}^{j})}{\sum_{k'} p_{k'}{}^{j} P(x_i \mid \theta_{k'}{}^{j})}$$

Auxiliary function is a measure for the contribution of training sample i to the Gaussian k

Update of Means and Covariance Matrix

Means: $$\mu_k^{\,j+1} = \frac{1}{\sum_{i=1}^{N} \gamma^j{}_{i,k}} \sum_{i=1}^{N} \gamma^j{}_{i,k} x_i$$

Covariances: $$\Sigma_k^{\,j+1} = \frac{1}{\sum_{i=1}^{N} \gamma^j{}_{i,k}} \sum_{i=1}^{N} \gamma^j{}_{i,k} (x_i - \mu_k)(x_i - \mu_k)^t$$

What are the changes
as compared to the maximum
likelihood estimate

Summary

- Applications of Gaussian Mixture Models (GMMs)
  - Speaker Identification
  - Image Segmentation
  - …
- Training of GMMs:
  - EM-Algorithm