University of California
Berkeley

College of Engineering
Department of Electrical Engineering
and Computer Sciences

Professors : N.Morgan / B.Gold

EE225D

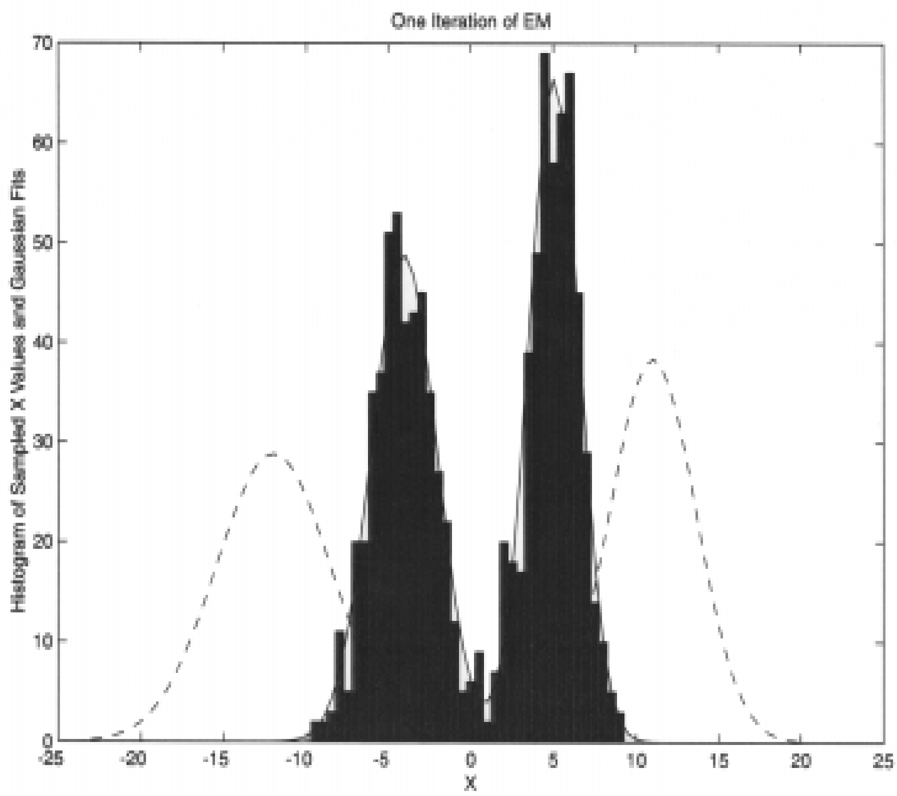
Spring,1999

Statistical Pattern Recognition

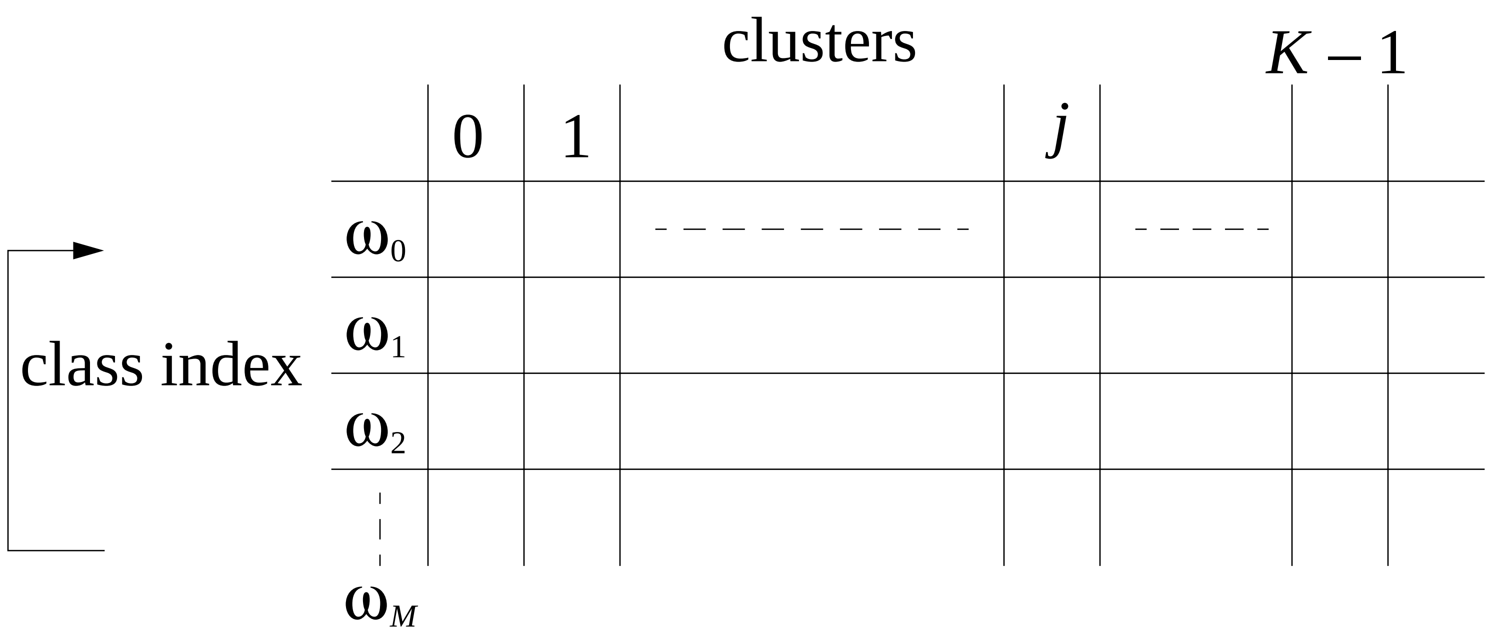
Lecture 10

Last Time

$$\log p(x|\omega_i) + \log p(\omega_i)$$


One Iteration of EM
Histogram of Sampled X Values and Gaussian Fits
X
0
10
20
30
40
50
60
70
-25
-20
-15
-10
-5
0
5
10
15
20
25

Discrete Density Estimation



$$p(\omega_i) = \frac{\sum_j n_{ij}}{\sum_{i,j} n_{ij}} = \frac{\text{row total}}{\text{total}}$$

$$p(\omega_i|x) \approx p(\omega_i|y_j) = \frac{p(\omega_i, y_j)}{p(y_j)} = \frac{n_{ij}}{\sum_i n_{ij}}$$

K-means Clustering

1. Choose N centers
2. Assign paths to nearest
3. Compute centers
4. Assess
5. Write “codebook”

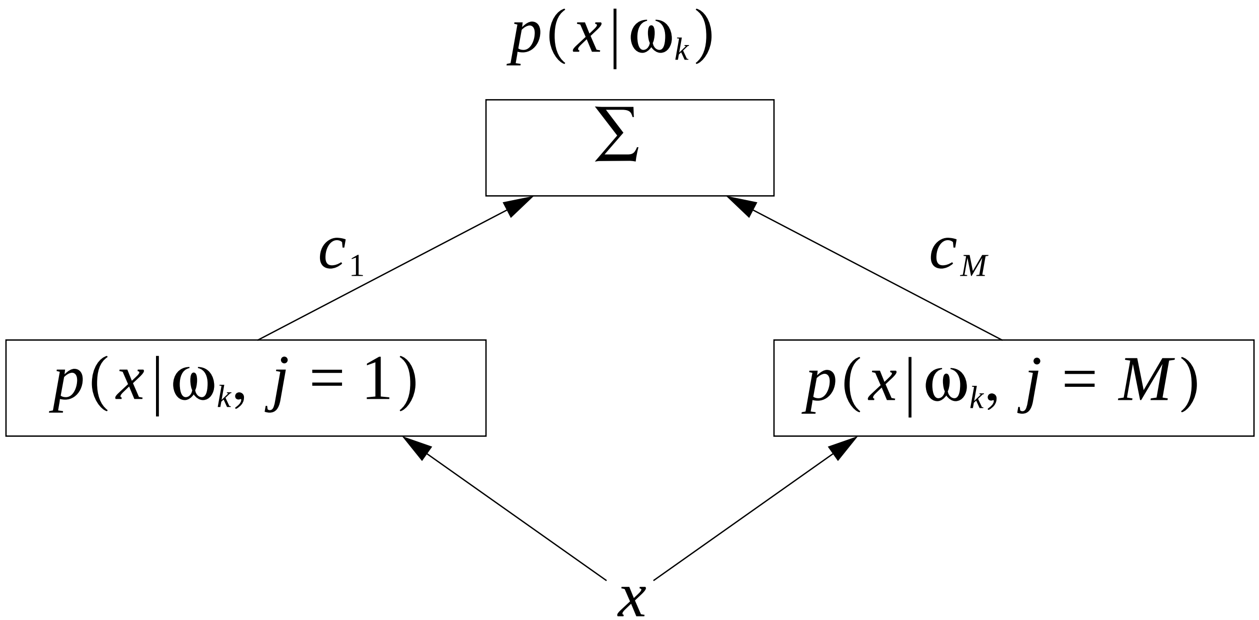
Example: speech frame classification

- 256 pt DFT, 128 spec vals, take log power
- Use K-means, find 64 centers, make table
- Assign each spectrum to a codebook entry, count co-occurences with phoneme labels, get probs

Estimators requiring iterative training

- Gaussian mixtures
- Neural networks

Gaussian Mixtures



$$p(x|\omega_k) = \sum_{j=1}^{M} p(j, x|\omega_k) = \sum_{j=1}^{M} \underbrace{p(j|\omega_k)}_{c_j} p(x|\omega_k, j)$$

c_j = prob x originated from dist j

Expectation Maximization

(Also sometimes called Estimate-and-Maximize)

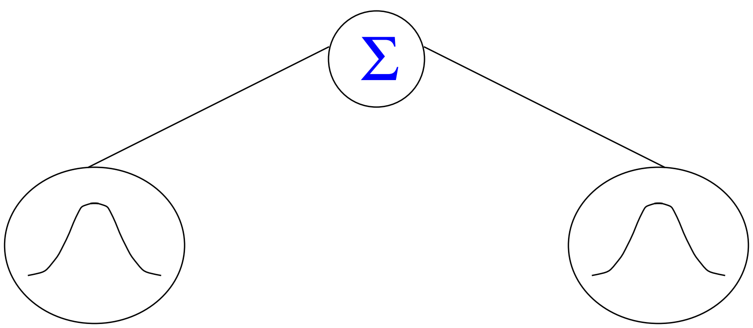
- Potentially quite general
- Cannot analytically determine parameters
- E step: Conditional Expectation of unknown variable given what is known
- M step: Choose parameters to maximize E

$p(\omega_i|x)$ $p(x|\omega_i)$

$p(\omega_i|x, \theta)$ $p(x|\omega_i, \theta)$

for class ω_i $p(x|\theta)$ $\rightarrow$ ML

$$\arg_\theta \max p(x|\theta)$$



Let k be hidden variables

x be observed

θ be params

θ_{old} old params

$$E\{\log p(k, x|\theta)\} = \sum_k p(k|x, \theta_{old}) \log[p(k, x|\theta)]$$

$$= \sum_k p(k|x, \theta_{old}) \log[p(k|x, \theta) p(x|\theta)]$$

$$Q(\theta, \theta_{old}) = \sum_k p(k|x, \theta_{old}) \log[p(k|(x, \theta))]$$

Sums to 1

$$+ \sum_k p(k|x, \theta_{old}) \underbrace{\log[p(x|\theta)]}_{\text{Indep. of k}}$$

$$Q(\theta_{old}, \theta_{old}) = \sum_k p(K|x, \theta_{old}) \log[p(k|x, \theta_{old})] + \log p(x|\theta_{old})$$

$$\text{diff} = \log p(x|\theta) - \log p(x|\theta_{old})$$

$$-\sum_k p(k|x, \theta_{old}) \log \frac{p(k|x, \theta_{old})}{p(k|x, \theta)}$$

Gaussian Mixture

$$p(x|\theta) = \sum_{k=1}^{K} p(x, k|\theta) = \sum_{k=1}^{K} p(k|\theta)p(x|k, \theta)$$

Log Joint Density

$$\log p(x, k|\theta) = \log[p(k|\theta)p(x|k, \theta)]$$

$$Q = \sum_{k=1}^{K} \sum_{n=1}^{N} p(k|x_n, \theta_{old})\log[p(k|\theta)p(x_n|k, \theta)]$$

$$= \sum_{k=1}^{K} \sum_{n=1}^{N} p(k|x_n, \theta_{old})\log p(k|\theta)$$

$$\sum_{k=1}^{K} \sum_{n=1}^{N} p(k|x_n, \theta_{old})\log p(x_n|k, \theta)$$

$$\text{Let } p(x_n|k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}\exp\left[-\frac{1}{2}\left(\frac{x_n-\mu_k}{\sigma_k}\right)^2\right]$$

$$Q = \sum_{k=1}^{K}\sum_{n=1}^{N} p(k|x_n, \theta_{old})\log p(k|\theta)$$

$$+ \sum_{k=1}^{K}\sum_{n=1}^{N} p(k|x_n, \theta_{old})\left(-\log\sigma_n + C + \frac{-(x_n-\mu_k)^2}{2\sigma_k^2}\right)$$

$$\frac{\partial Q}{\partial \mu_j} = 0$$

$$\Rightarrow \sum_{n=1}^{N} p(j|x_n, \theta_{old})\left(\frac{x_n}{\sigma_j^2} - \frac{\mu_j}{\sigma_j^2}\right) = 0$$

$$\Rightarrow \sum_{n=1}^{N} p(j|x_n, \theta_{old})x_n = \sum_{n=1}^{N} p(j|x_n, \theta_{old})\mu_j$$

$$\Rightarrow \mu_j = \frac{\sum_{n=1}^{N} p(j|x_n, \theta_{old})x_n}{\sum_{n=1}^{N} p(j|x_n, \theta_{old})}$$

EM Summary

- Choose parametric form
- Choose initial values
- Compute posterior estimates for hidden variables
- Choose parameters to maximize expectation of joint density (observed, hidden)
- Assess goodness of fit
- Good enough — Yes → Stop

No (→ back to "Compute posterior estimates for hidden variables")