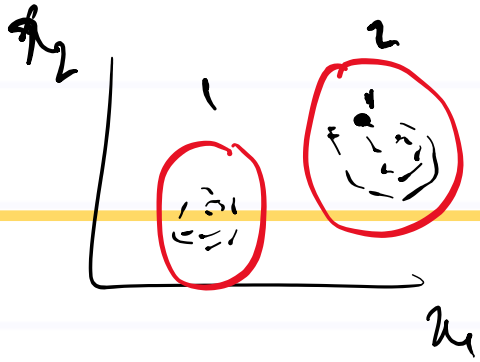
Gaussian Mixture Model

EE698V - Machine Learning for Signal Processing

Vipul Arora

K-means Clustering

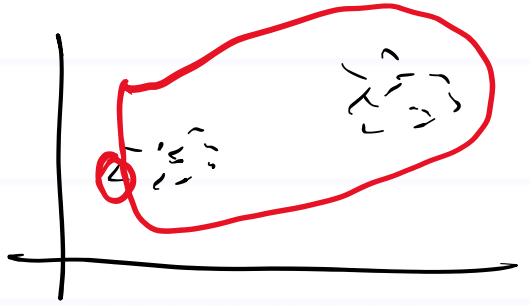


- Distortion measure: $\mathcal{L} = \sum_{n,k} r_{nk} \|x_k - \mu_k\|^2$
- k is the cluster index
- n is the sample index
- $r_{nk} = 1$ if sample n belongs to cluster k (hard clustering)
- $\mathcal{L}$ avoids very large (+ singular) clusters

$r_{11} = 0$

$r_{12} = 1$

membership fn



Optimal value of r_{nk} and μ_k

$$\mathcal{L} = \sum_{n,k} r_{nk} \|x_n - \mu_k\|^2$$

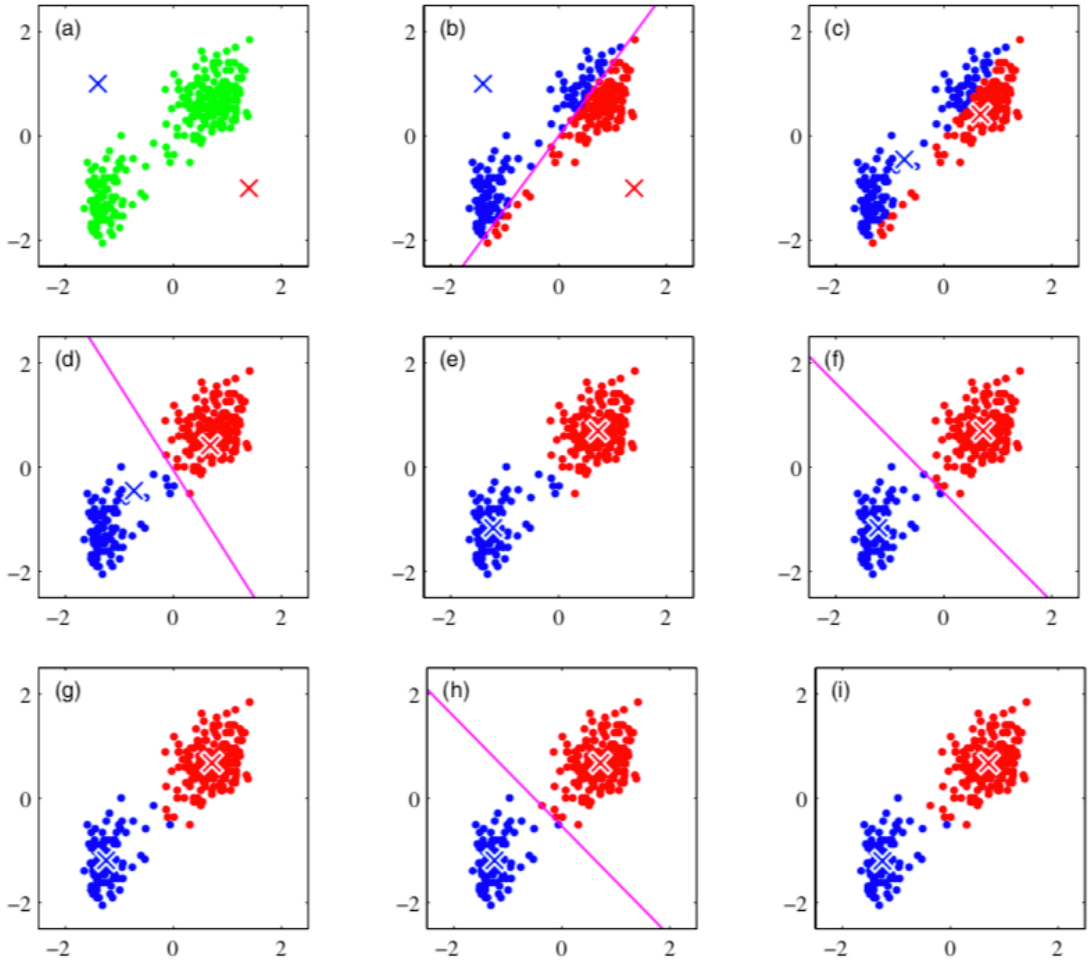
Given r_{nk}

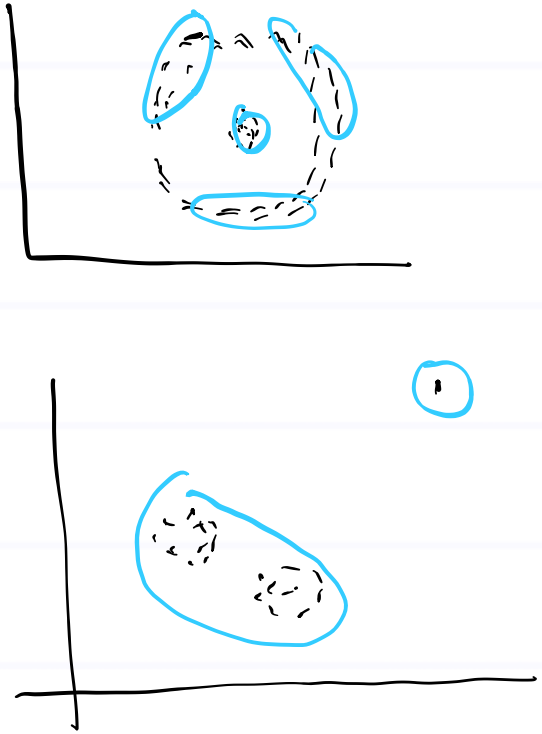
$$\frac{\partial \mathcal{L}}{\partial \mu_{k'}} = 2 \sum_n r_{nk'} (x_{k'} - \mu_{k'}) = 0 \quad \Rightarrow \quad \mu_{k'} = \frac{\sum_n r_{nk'} x_{k'}}{\sum_n r_{nk'}}$$

Given $\mu_{k'}$

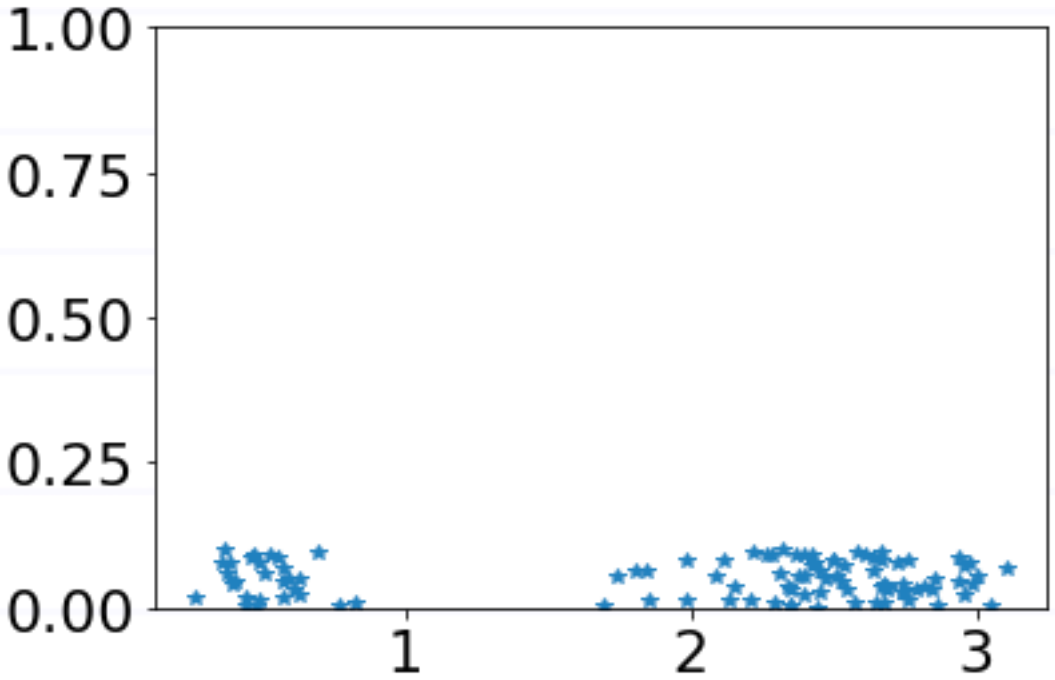
$$r_{nk} = \begin{cases} 1 & , \text{ if } k = \arg\min_{k'} \| x_n - \mu_{k'} \| \\ 0 & , \text{ o.w.} \end{cases}$$

K-means iteration





How would you model this data?

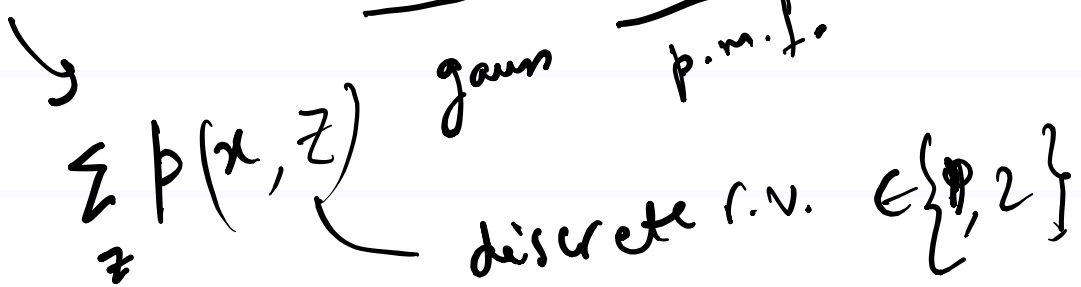


Latent Variable Models

Reference: PRML Chapter 9

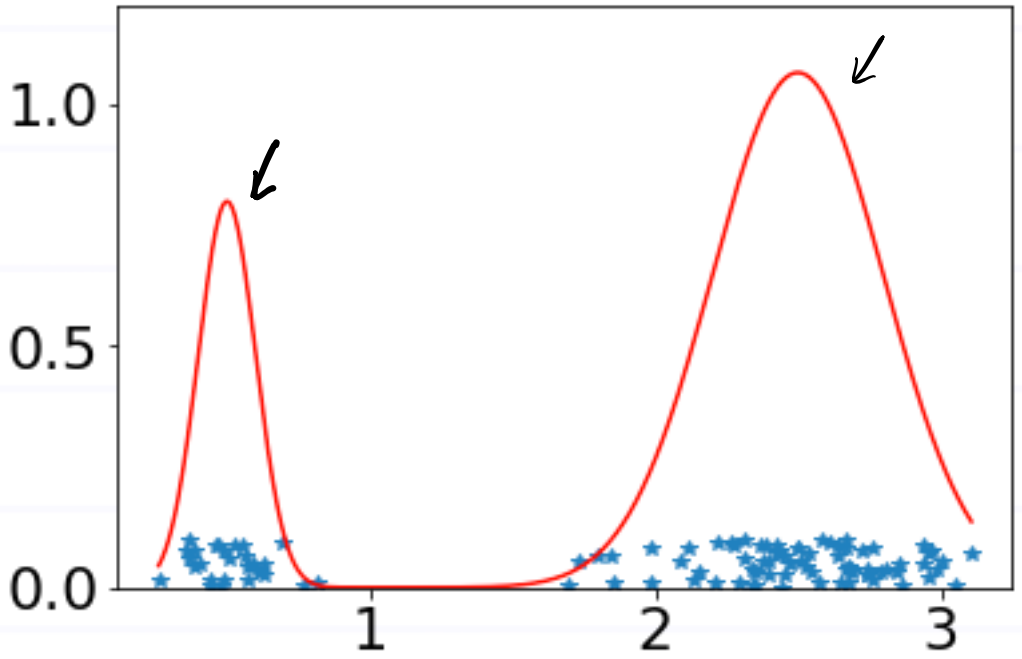
Latent Variables

- Latent variables help complicated distributions to be formed from simpler components
- $p(x) = p(x|z=1)p(z=1) + p(x|z=2)p(z=2)$
- $p(x) = \sum_z p(x|z)p(z)$



Gaussian Mixture Model

$$p(x) = \sum_k p(z_k)\mathcal{N}(x; \mu_k, \sigma_k)\,; z_k \in \{0,1\}, \sum_k z_k = 1$$

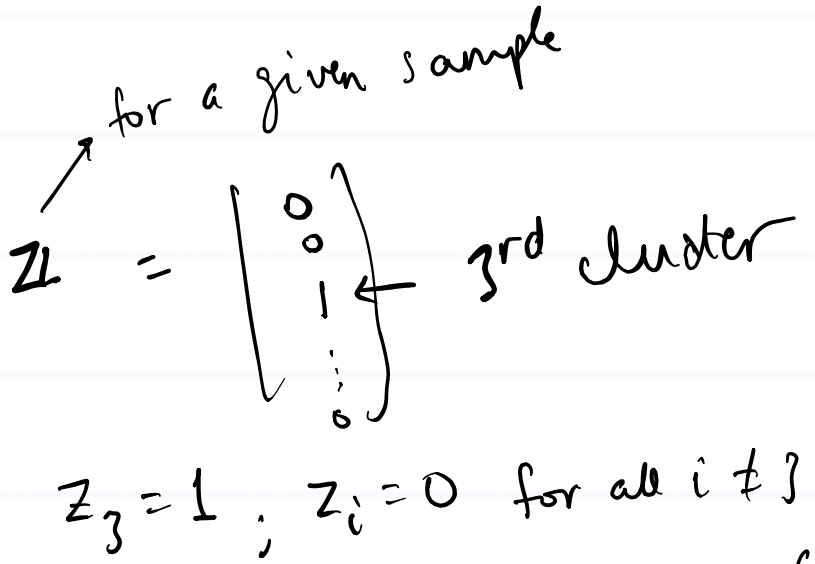


$z = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

one hot vector

Gaussian Mixture Model

- $p(\boldsymbol{x}) = \sum_z p(z)p(\boldsymbol{x}|z)$
  - $z \in \{1,2,\ldots,K\}$ is a discrete r.v.
  - $\boldsymbol{x} \in \mathbb{R}^D$



Equivalently,

- $p(\boldsymbol{x}) = \sum_k p(z_k = 1)p(\boldsymbol{x}|z_k)$
  - z_k is kth element of a 1-hot vector. It is 1 if $\boldsymbol{x}$ belongs to kth cluster $z_k = \delta_{k3}$
- $p(x) = \sum_k \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$

Estimating the parameters

- Given the samples $\{\boldsymbol{s}_n; n = 1, \ldots, N\}$
  - the learnable parameters are $\boldsymbol{\theta} = \{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, ; k = 1, \ldots, K\}$
  - K is a hyperparameter

MLE for μ_k

$$\ln \mathcal{L} = \sum_n \ln p(s_n|\theta) = \sum_n \ln \left(\sum_k \pi_k \mathcal{N}(s_n|\mu_k, \Sigma_k) \right)$$

$$\mathcal{N}(s_n|\mu_k, \Sigma_k) = \frac{1}{\sqrt{2\pi}^D |\Sigma|} e^{-\frac{1}{2}(s_n-\mu_k)^T \Sigma^{-1} (s_n-\mu_k)}$$

Differentiate w.r.t. $\mu_{k'}$:

$$\frac{\partial \ln \mathcal{L}}{\partial \mu_{k'}} = \sum_n \frac{\pi_{k'} \mathcal{N}(s_n|\mu_{k'}, \Sigma_{k'})}{\sum_k \pi_k \mathcal{N}(s_k|\mu_k, \Sigma_k)} \cdot \left(-\Sigma_{k'}^{-1} (s_n - \mu_{k'})\right) = 0$$

$$= \gamma_{nk'} \rightarrow p(z_{k'}=1) \text{ for } s_n$$

- prob. that s_n belongs to the k^{th} cluster (soft assignment)
- k-means does hard assignment

MLE for $\boldsymbol{\mu_k}$

$$\Rightarrow \sum_n \gamma_{nk'} \mu_{k'} = \sum_n \gamma_{nk'} s_n$$

$$\Rightarrow \boxed{\mu_{k'} = \frac{\sum_n \gamma_{nk'} s_n}{N_{k'}}}$$

$N_k = \sum_n \gamma_{nk}$: effective # samples assigned to cluster k

MLE for $\boldsymbol{\Sigma_k}$

Differentiating $\ln L$ w.r.t $\Sigma_{k'}$

$$\frac{\partial \ln L}{\partial \Sigma_{k'}} = \sum_n \gamma_{nk'} \left(-\frac{1}{2|\Sigma|_{k'}^2} \frac{\partial |\Sigma_{k'}|}{\partial \Sigma_{k'}} + \frac{1}{|\Sigma|} \frac{\partial}{\partial \Sigma} \left(-\frac{(s_n - \mu_{k'})^T \Sigma_{k'}^{-1} (s_n - \mu_{k'})}{2} \right) \right) = 0$$

$$\Rightarrow$$

$$|\Sigma_{k'}| (\Sigma_{k'}^{-1})^T$$

$$= \text{tr} \left\{ (s_n - \mu_{k'})(s_n - \mu_{k'})^T \Sigma_{k'}^{-1} \right\}$$

$$-(\Sigma_{k'}^{-1})^T (s_n - \mu_{k'})(s_n - \mu_{k'})^T (\Sigma_{k'}^{-1})^T$$

$$\Rightarrow \sum_n \gamma_{nk'} \left[(\Sigma_{k'}^{-1})^T - (\Sigma_{k'}^{-1})^T (s_n - \mu_{k'})(s_n - \mu_{k'})^T (\Sigma_{k'}^{-1})^T \right] = 0$$

$$\Rightarrow \quad \boxed{\Sigma_{k'} = \frac{1}{N_{k'}} \sum_n \gamma_{nk'} (s_n - \mu_{k'})(s_n - \mu_{k'})^T}$$

$$\sigma^2_{ML} = \frac{\sum_n (x_n - \mu)^2}{N}$$

MLE for π_k

Differentiate $\left[\ln \mathcal{L} + \lambda\left(\sum_k \pi_k - 1\right)\right]$ w.r.t. $\pi_{k'}$,

$$\frac{\partial [\ldots]}{\partial \pi_{k'}} = \sum_n \frac{\mathcal{N}(s_n | \mu_{k'}, \Sigma_{k'})}{\sum_k \pi_k \mathcal{N}(s_n | \mu_k, \Sigma_k)} + \lambda = 0 \qquad \ldots (1)$$

Multiply with $\pi_{k'}$ and sum over k'

$$\Rightarrow \sum_n 1 + \lambda = 0 \Rightarrow \lambda = -N$$

Substituting in (1),

$$\sum_n \frac{\gamma_{nk'}}{\pi_{k'}} + (-N) = 0$$

$$\Rightarrow \boxed{\pi_{k'} = \frac{N_{k'}}{N}}$$

MLE for GMM

- Now none of the estimates is independent of each other
- So follow iterative approach, as in K-means
  1. Initialize all parameters
  2. ASSIGN: calculate γ_k
  3. UPDATE $\boldsymbol{\theta}$
  4. Stop if converged, else go to step 2

Given L, find optimal θ :-

- $\frac{\partial L}{\partial \theta} = 0$, solve for θ
- if θ's are interdependent, use iterative update (in tandem). ↗ possible if $\frac{\partial L}{\partial \theta} = 0$ is solvable for θ
- If $\frac{\partial L}{\partial \theta}$ not solvable, use Grad. desc.

θ_{ML}

Expectation Maximization Algorithm

Reference: PRML Section 9.3

EM Algo

all the data samples

all trainable parameters

latent vars

$$\ln p(S|\theta) = \ln \left\{ \sum_Z p(S, Z|\theta) \right\}$$

$$= \ln \left\{ \sum_Z p(Z|\theta)\ p(S|Z,\theta) \right\}$$

$$\ln p(S|\theta) = \sum_n \ln \left\{ \sum_k \pi_k \mathcal{N}(s_n|\mu_k, \Sigma_k) \right\}$$

exp, but ln can't act on it $\because$ of $\sum_k$

How to get rid of sum inside log?

EM Algo

- $p(S, Z|\theta)$, without marginalizing over Z , is exp.
- How to avoid marginalization?
- Assume a value of Z, and later average over all Z.

$$Q = \sum_Z \underbrace{[p(Z|S,\theta)^{old}}_{\text{arg.}} \underbrace{\ln p[S,Z|\theta)]}_{\text{exp.}} \qquad \underset{\theta}{\operatorname{argmax}}\, Q(\theta; \underset{=}{\theta}^{old})$$

But again $\frac{\partial Q}{\partial \theta}$ will include difficult terms ∴ take θ^{old} outside ln.

$$Q = E_{Z|S,\theta^{old}}[\ln p(S,Z|\theta)]$$

EM Algo

latent variable for cluster assignment

$$p(S, Z \mid \theta) = \prod_n \prod_k \left\{ \pi_k \, \mathcal{N}(s_n \mid \mu_k, \Sigma_k) \right\}^{z_{nk}}$$

$$\ln p(S, Z \mid \theta) = \sum_n \sum_k z_{nk} \ln \left\{ \pi_k \, \mathcal{N}(s_n \mid \mu_k, \Sigma_k) \right\}$$

$$\therefore Q = \sum_Z p(Z \mid S, \theta^{old}) \ln p(S, Z \mid \theta)$$

$$= \sum_n \sum_k \gamma_{nk} \ln \left\{ \pi_k \, \mathcal{N}(s_n \mid \mu_k, \Sigma_k) \right\}$$

EM Algo

Iterative Algorithm:

1. Initialize θ^{old}
2. ASSIGN: Evaluate $p(Z|X, \theta^{old})$, i.e., γ_{nk} E-step
3. UPDATE PARAMETERS: μ_k, Σ_k, π_k by maximizing Q M-step
4. Stop if converged, else go to Step 2

Further Reading (Not mandatory)

- PRML Section 9.4