Deep Gaussian Mixture Models (DGMMs)

Ciobanu Sebastian

Coordonator: Conf. dr. Liviu Ciortuz

Cercetare – semestrul 3

- Documentare:
  - Reducerea dimensionalității datelor (nesupervizat, clasic): PCA, FA, ICA etc.
  - AdaBoost (supervizat, clasic): clase de concepte învățabile în sens empiric gamma-slab
- **Mixturi profunde de distribuții normale** (nesupervizat, recent)

Scopul

- Semestrul trecut: procese gaussiene (model pentru serii de timp)
- Există varianta Procese gaussiene profunde: Deep *Gaussian Processes, Andreas C. Damianou and Neil D. Lawrence, 2013*
- Mi-am propus ca mai întâi să abordez un model de tip *profund* de dificultate medie

- Semestrul acesta: am început cu PCA, FA...

DGMMs

Intro

DGMMs – Distribuția normală multivariată

$$X = [X_1, \cdots, X_D]$$

$$X \sim \mathcal{N}(\mu, \Sigma)$$

$$\mu \in \mathbb{R}^D$$

$\Sigma \in \mathbb{R}^{D \times D}$ - simetrică, pozitiv definită

$$\mathcal{N}(x; \mu, \Sigma) = \frac{1}{(2\pi)^{D/2} \det(\Sigma)^{1/2}} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu))$$

- Având datele observabile, vrem să estimăm parametrii.
- Există formule analitice.

X1	X2	X3
1.1	2.33	3.43
1	2	3
…	…	…

$$D = \{x_1, \ldots, x_n\}$$

$$\mu_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\Sigma_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_{\text{MLE}})(x_i - \mu_{\text{MLE}})^T$$

DGMMs – Factor Analysis (FA)

- Idee: datele observabile sunt generate de (au în spate) niște cauze latente

X1	X2	X3=X1+X2
1.1	2.33	3.43

generează

Z1	Z2
1.1	2.33

- Exemple practice:
  - O întrebare la un examen de matematică poate testa dacă studentul are noțiuni din sfera algebrei, geometriei, probabilităților.
  - Un document poate fi despre economie, politică, educație, sport
  - O poză poate conține obiecte cunoscute: apus, copac, pisică.

DGMMs – Factor Analysis (FA)

- Având datele observabile, vrem să aflăm datele latente [de o dimensiune mai mică, de obicei] sau o **relație observabil-latent**

DGMMs – Factor Analysis (FA)

- Probabilist. Presupuneri:

$$\boxed{\begin{aligned} z &\sim \mathcal{N}(0, I) \\ x|z &\sim \mathcal{N}(\mu + \Lambda z, \Psi) \end{aligned}} \quad \Leftrightarrow \quad \boxed{\begin{aligned} z &\sim \mathcal{N}(0, I) \\ \epsilon &\sim \mathcal{N}(0, \Psi) \\ x &= \mu + \Lambda z + \epsilon. \end{aligned}}$$

ϵ, z - independente

$$z \in \mathbb{R}^d$$

$$\mu \in \mathbb{R}^D$$

$$\epsilon \in \mathbb{R}^D$$

$$\Lambda \in \mathbb{R}^{D \times d}$$

$\Psi \in \mathbb{R}^{D \times D}$ - diagonală

$$d \leq D$$

DGMMs – Factor Analysis (FA)

- Rezultat (din propr. distr. normale):

$$x \sim \mathcal{N}(\mu, \Lambda\Lambda^T + \Psi)$$

- Efect (secundar):

$$x \sim \mathcal{N}(\mu, \Sigma) \xrightarrow{\Sigma \approx \Lambda\Lambda^T + \Psi} x \sim \mathcal{N}(\mu, \Lambda\Lambda^T + \Psi)$$

(mai puțini parametri de estimat:

D=6, d=1: 21 vs 12

D=7, d=1: 28 vs 14 etc.)

DGMMs – Factor Analysis (FA)

- Cum estimăm parametrii? Nu există formule analitice
- Cu algoritmul EM (Expectation Maximization)
  - Pentru că lucrăm cu variabile latente...

EM/FA:

$$\mu^* = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Pasul E

$$\beta = \Lambda^T(\Psi + \Lambda\Lambda^T)^{-1}$$

$$E[z_i|x_i] = \beta(x_i - \mu^*),\ \forall i \in \{1,\ldots,n\}$$

$$E[z_i z_i^T|x_i] = I - \beta\Lambda + E[z_i|x_i]E[z_i^T|x_i],\ \forall i \in \{1,\ldots,n\}$$

Pasul M

$$\Lambda = \left(\sum_{i=1}^{n}(x_i - \mu^*)E[z_i^T|x_i]\right)\left(\sum_{i=1}^{n}E[z_i z_i^T|x_i]\right)^{-1}$$

$$\Psi = \frac{1}{n}\mathrm{diag}\left(\sum_{i=1}^{n} x_i x_i^T - \Lambda E[z_i|x_i]x_i^T\right)$$

Sursă: după *The EM Algorithm for Mixtures of Factor Analyzers*, Zoubin Ghahramani, Geoffrey E. Hinton, 1997

DGMMs – Factor Analysis (FA)

- Unicitate? Nu.

Fie μ^*, Λ^*, Ψ^* parametrii (furnizaţi de un algoritm de învăţare).

Atunci, dacă pe Λ^* îl înlocuim cu $\Lambda^* R$, unde $R \in \mathbb{R}^{d \times d}$ - matrice ortogonală ($RR^T = R^T R = I$), vom avea acelaşi efect (secundar):

$$\boxed{x \sim \mathcal{N}(\mu^*, \Lambda^*(\Lambda^*)^T + \Psi^*)}$$

$$\boxed{\Lambda^* = \Lambda^* R}$$

$$x \sim \mathcal{N}(\mu^*, \Lambda^* R(\Lambda^* R)^T + \Psi^*)$$
$$\Downarrow$$
$$x \sim \mathcal{N}(\mu^*, \Lambda^* R R^T(\Lambda^*)^T + \Psi^*)$$
$$\Downarrow$$
$$x \sim \mathcal{N}(\mu^*, \Lambda^*(\Lambda^*)^T + \Psi^*)$$

- Soluție: impunem constrângeri asupra parametrilor
  - De exemplu: impunem ca $\Lambda^T \Psi^{-1} \Lambda$ să fie diagonală

DGMMs – *Deep(er)* Factor Analysis

$$\boxed{\begin{aligned} z &\sim \mathcal{N}(\underline{A}, \underline{B}) \\ x|z &\sim \mathcal{N}(\mu + \Lambda z, \Psi) \end{aligned}} \quad \Leftrightarrow \quad \boxed{\begin{aligned} z &\sim \mathcal{N}(\underline{A}, \underline{B}) \\ \epsilon &\sim \mathcal{N}(0, \Psi) \\ x &= \mu + \Lambda z + \epsilon \end{aligned}}$$

$$\Downarrow$$

$$\boxed{x \sim \mathcal{N}(\mu + \Lambda \underline{A}, \Lambda \underline{B} \Lambda^T + \Psi)}$$

DGMMs – *Deep(er)* Factor Analysis

Adâncime = 2

$$\boxed{x = z^{(0)}}$$

$$z^{(2)} \sim \mathcal{N}(0, I)$$
$$z^{(1)}|z^{(2)} \sim \mathcal{N}(\mu^{(2)} + \Lambda^{(2)} z^{(2)}, \Psi^{(2)})$$
$$z^{(0)}|[z^{(2)},]z^{(1)} \sim \mathcal{N}(\mu^{(1)} + \Lambda^{(1)} z^{(1)}, \Psi^{(1)})$$

$$\Leftrightarrow$$

$$z^{(2)} \sim \mathcal{N}(0, I)$$
$$\epsilon^{(2)} \sim \mathcal{N}(0, \psi^{(2)})$$
$$z^{(1)} = \mu^{(2)} + \Lambda^{(2)} z^{(2)} + \epsilon^{(2)}$$
$$\epsilon^{(1)} \sim \mathcal{N}(0, \psi^{(1)})$$
$$z^{(0)} = \mu^{(1)} + \Lambda^{(1)} z^{(1)} + \epsilon^{(1)}$$

$$\Downarrow$$

$$z^{(1)} \sim \mathcal{N}(\underline{\mu^{(2)}}, \underline{\Lambda^{(2)}(\Lambda^{(2)})^T + \Psi^{(2)}})$$
$$z^{(0)} \sim \mathcal{N}(\underline{\mu^{(1)} + \Lambda^{(1)}}(\underline{\mu^{(2)}}), \underline{\Lambda^{(1)}}(\underline{\Lambda^{(2)}(\Lambda^{(2)})^T + \Psi^{(2)}})\underline{(\Lambda^{(1)})^T + \Psi^{(1)}})$$

DGMMs – *Deep(er)* Factor Analysis

Adâncime = h

$$\boxed{x = z^{(0)}}$$

…

$$z^{(0)}|z^{(1)} \sim \mathcal{N}(\mu^{(1)} + \Lambda^{(1)} z^{(1)}, \Psi^{(1)})$$

$$\Leftrightarrow$$

…

$$z^{(0)} = \mu^{(1)} + \Lambda^{(1)} z^{(1)} + \epsilon^{(1)}$$

$$\Downarrow$$

$$z^{(0)} \sim \mathcal{N}\left(\tilde{\mu}^{(1)}, \tilde{\Sigma}^{(1)}\right)$$

(t+1) (t+1)

$$\tilde{\mu}^{(1)} = \mu^{(1)} + \sum_{l=2}^{h} \left(\prod_{m=1}^{l-1} \Lambda^{(m)} \right) \mu^{(l)}$$

l = t+2 m = t+1

(t+1) (t+1)

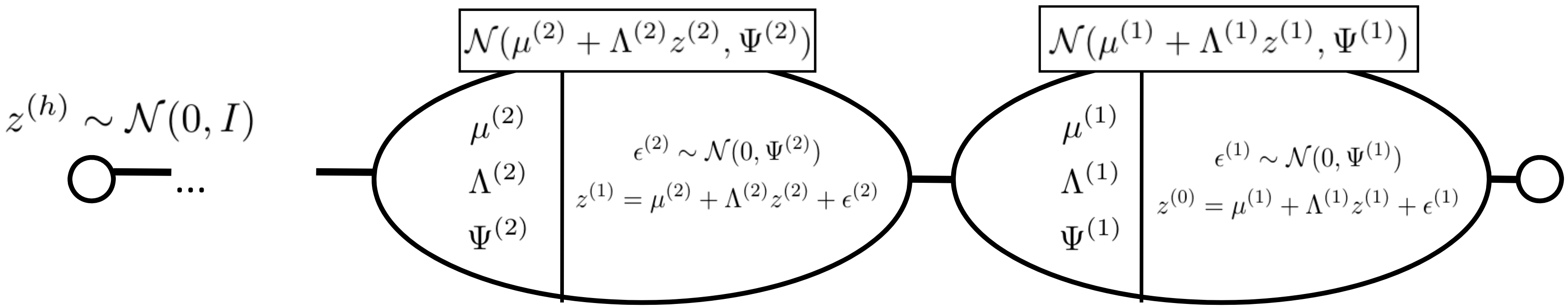
$$\tilde{\Sigma}^{(1)} = \Psi^{(1)} + \sum_{l=2}^{h} \left(\left(\prod_{m=1}^{l-1} \Lambda^{(m)} \right) \Psi^{(l)} \left(\prod_{m=1}^{l-1} \Lambda^{(m)} \right)^{T} \right) + \left(\prod_{m=1}^{h} \Lambda^{(m)} \right) \left(\prod_{m=1}^{h} \Lambda^{(m)} \right)^{T}$$

l=t+2 m=t+1 m=t+1 m=t+1 m=t+1

DGMMs – *Deep(er)* Factor Analysis

Adâncime = h

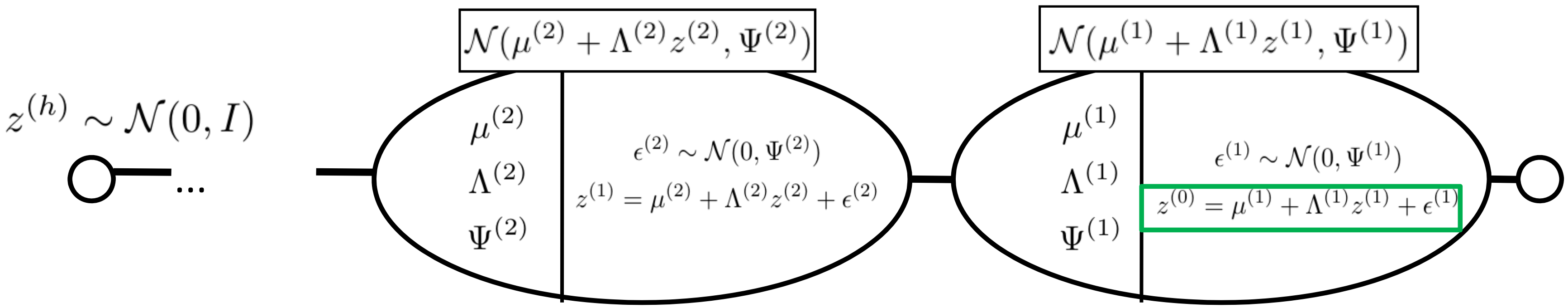
$x = z^{(0)}$



h+1	h	...	2	1	0
	-		...	$z^{(1)} \mid z^{(2)} \sim \mathcal{N}(\mu^{(2)} + \Lambda^{(2)} z^{(2)}, \Psi^{(2)})$	$z^{(0)} \mid z^{(1)} \sim \mathcal{N}(\mu^{(1)} + \Lambda^{(1)} z^{(1)}, \Psi^{(1)})$
	$z^{(h)} \sim \mathcal{N}(0, I)$			$z^{(1)} \sim \mathcal{N}\left(\tilde{\mu}^{(2)}, \tilde{\Sigma}^{(2)}\right)$	$z^{(0)} \sim \mathcal{N}\left(\tilde{\mu}^{(1)}, \tilde{\Sigma}^{(1)}\right)$

DGMMs – *Deep(er)* Factor Analysis

Adâncime = h $x = z^{(0)}$



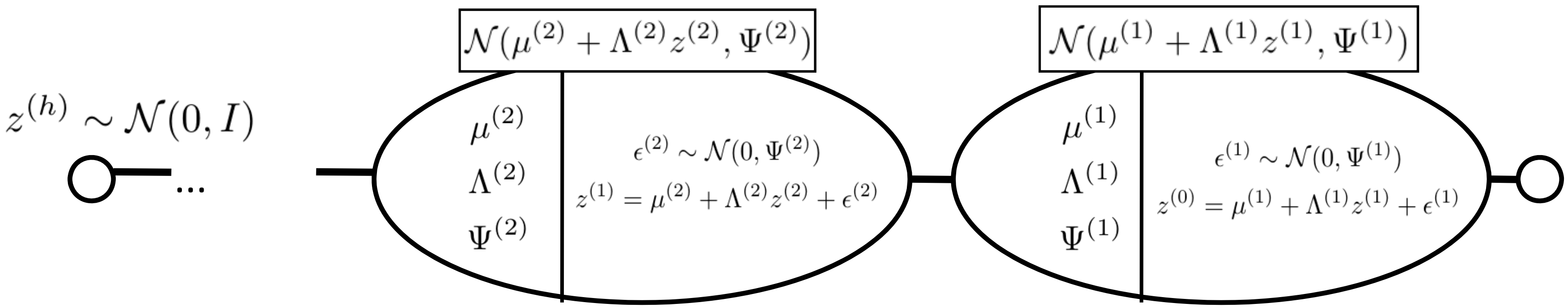
h+1	h	...	2	1	0
	-		...	$z^{(1)} \mid z^{(2)} \sim \mathcal{N}(\mu^{(2)} + \Lambda^{(2)} z^{(2)}, \Psi^{(2)})$	$z^{(0)} \mid z^{(1)} \sim \mathcal{N}(\mu^{(1)} + \Lambda^{(1)} z^{(1)}, \Psi^{(1)})$
	$z^{(h)} \sim \mathcal{N}(0, I)$			$z^{(1)} \sim \mathcal{N}\left(\tilde{\mu}^{(2)}, \tilde{\Sigma}^{(2)}\right)$	$z^{(0)} \sim \mathcal{N}\left(\tilde{\mu}^{(1)}, \tilde{\Sigma}^{(1)}\right)$

La fiecare nivel: a b c

DGMMs – *Deep(er)* Factor Analysis

Adâncime = h

$x = z^{(0)}$



h+1	h	...	2	1	0
	-		...	$z^{(1)} \mid z^{(2)} \sim \mathcal{N}(\mu^{(2)} + \Lambda^{(2)} z^{(2)}, \Psi^{(2)})$	$z^{(0)} \mid z^{(1)} \sim \mathcal{N}(\mu^{(1)} + \Lambda^{(1)} z^{(1)}, \Psi^{(1)})$
	$z^{(h)} \sim \mathcal{N}(0, I)$			$z^{(1)} \sim \mathcal{N}\left(\tilde{\mu}^{(2)}, \tilde{\Sigma}^{(2)}\right)$	$z^{(0)} \sim \mathcal{N}\left(\tilde{\mu}^{(1)}, \tilde{\Sigma}^{(1)}\right)$

Cum generăm o instanță?

DGMMs – *Deep(er)* Factor Analysis

- Idee de antrenare
  - Antrenare pe niveluri
  - Aplicăm algoritmul EM/FA (modificat)
  - De la nivelul 1 la nivelul h. De ce nu invers?

Pasul E

$$\beta = \Lambda^T(\Psi + \Lambda\Lambda^T)^{-1}$$

$$E[\overset{\boxed{z_i^{(1)}|z_i^{(0)}}}{\underline{z_i|x_i}}] = \beta(x_i - \mu^*),\ \forall i \in \{1, \ldots, n\}$$

$$E[z_i z_i^T|x_i] = I - \beta\Lambda + E[z_i|x_i]E[z_i^T|x_i],\ \forall i \in \{1, \ldots, n\}$$

De la nivelul 1 la nivelul 2

$$\boxed{z_i^{(1)} = E[z_i^{(1)}|z_i^{(0)}]}$$

DGMMs – Mixturi de distribuții pb.

- Definiție
  - Combinație convexă de mase/densități de pb.:

$$p(x;\pi,\theta_1,\ldots,\theta_k) = \pi_1 p_1(x;\theta_1) + \cdots + \pi_k p_k(x;\theta_k)$$
$$\pi_1 + \cdots + \pi_k = 1$$

  - Exemplu cu parametrii dați:

$$p(x) = 0.2\mathcal{N}(x;0,1) + 0.8\mathcal{N}(x;2,2^2)$$

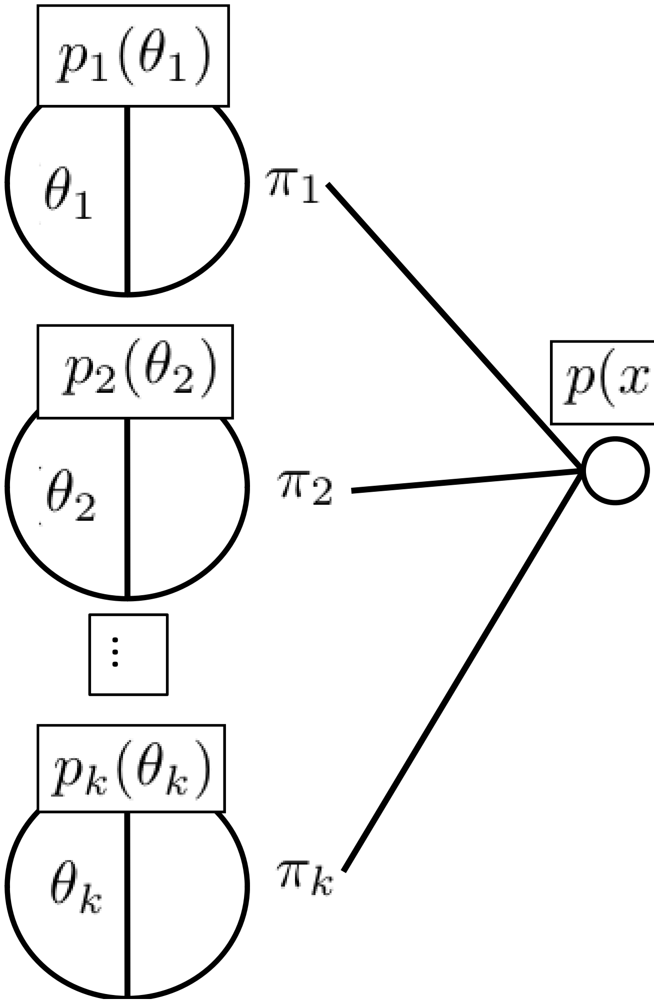
  - Exemplu fără parametrii dați:

$$p(x;\pi_1,\mu_1,\mu_2,\sigma_1,\sigma_2) = \pi_1\mathcal{N}(x;\mu_1,\sigma_1^2) + (1-\pi_1)\mathcal{N}(x;\mu_2,\sigma_2^2)$$

    - În acest ultim caz, parametrii ar trebui estimați

DGMMs – Mixturi de distribuții pb.

- Vizualizare

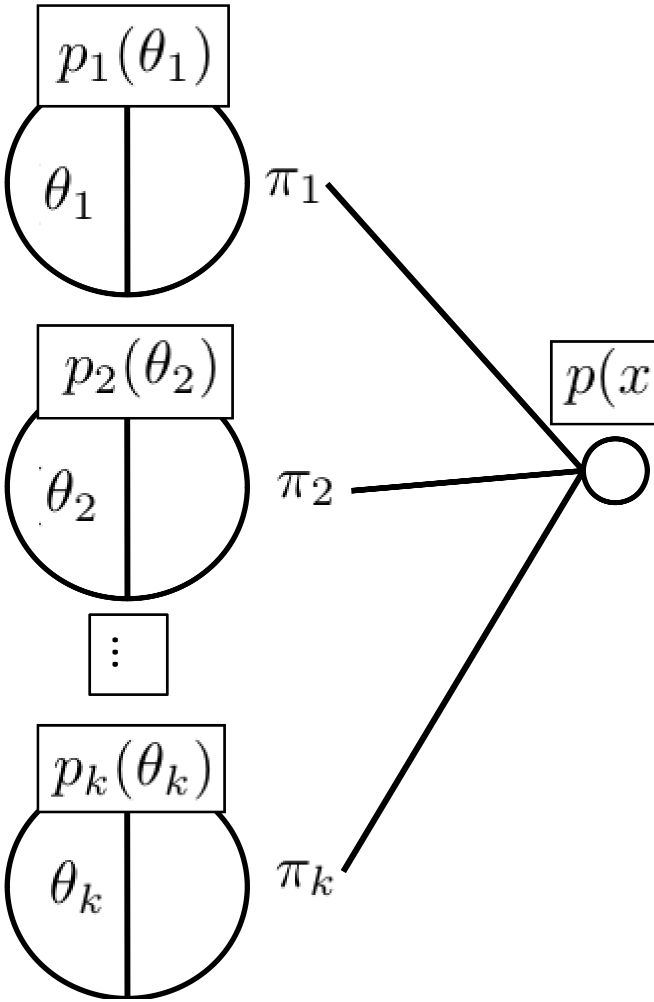


$$\boxed{p(x;\pi,\theta_1,\ldots,\theta_k) = \pi_1 p_1(x;\theta_1) + \cdots + \pi_k p_k(x;\theta_k)}$$

$$\pi_1 + \cdots + \pi_k = 1$$

DGMMs – Mixturi de distribuții pb.

- Vizualizare



$$\boxed{p(x;\pi,\theta_1,\ldots,\theta_k) = \pi_1 p_1(x;\theta_1) + \cdots + \pi_k p_k(x;\theta_k)}$$

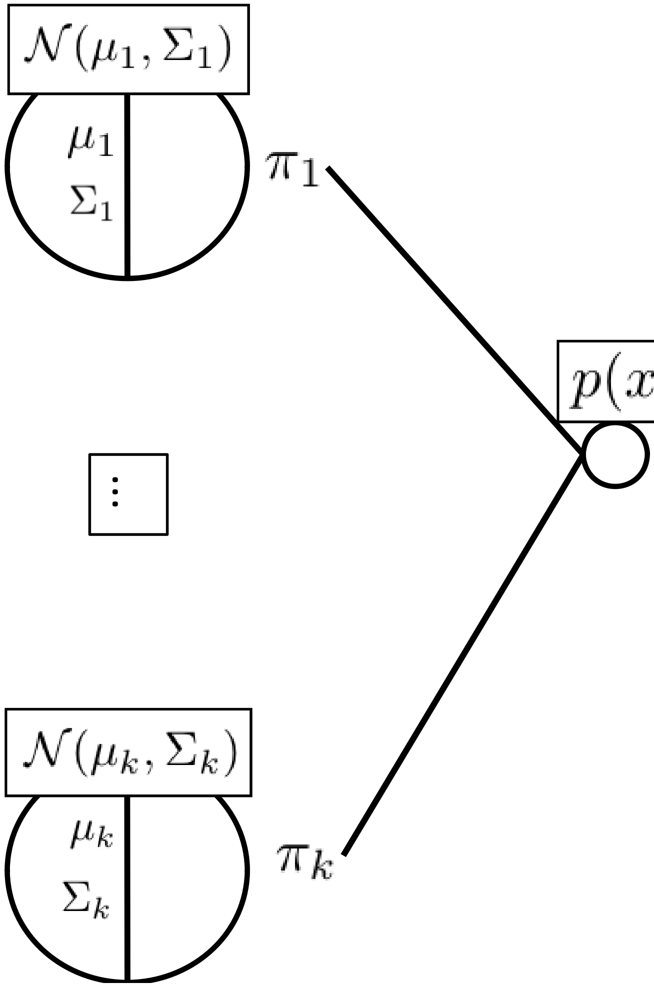
$$\pi_1 + \cdots + \pi_k = 1$$

Cum generăm o instanță?

DGMMs – Mixturi de distribuții pb.

- Cazul clasic: Mixturi de distribuții normale,

DAR...

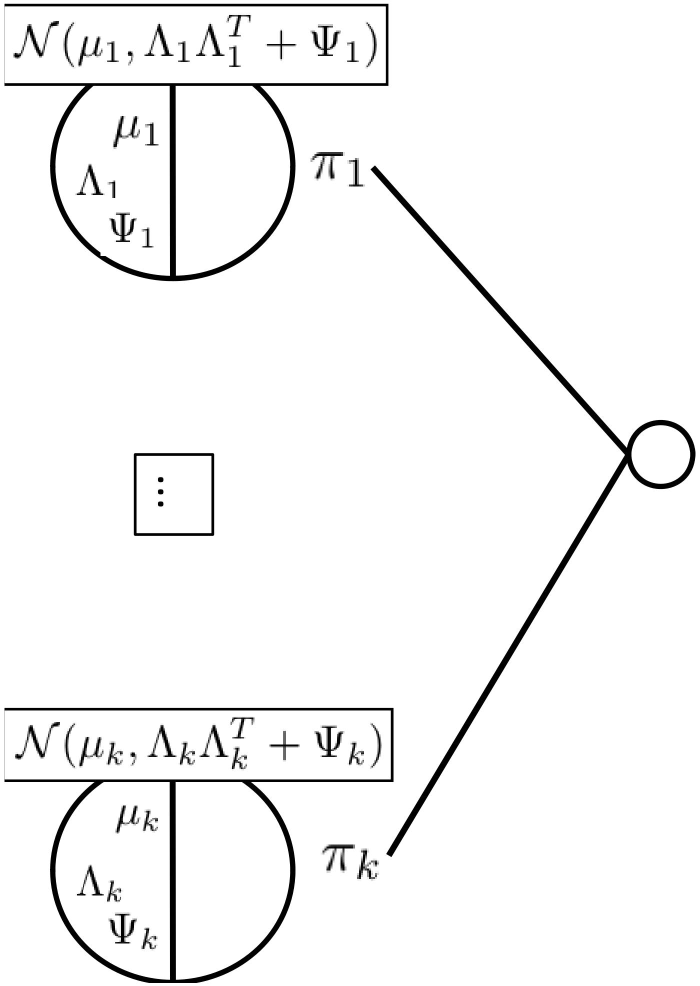


$$\boxed{p(x; \mu, \Sigma) = \pi_1 \mathcal{N}(x; \mu_1, \Sigma_1) + \cdots + \pi_k \mathcal{N}(x; \mu_k, \Sigma_k)}$$

$$\pi_1 + \cdots + \pi_k = 1$$

DGMMs – Mixturi de distribuții pb.

- Cazul alternativ: Mixturi de FA (1)

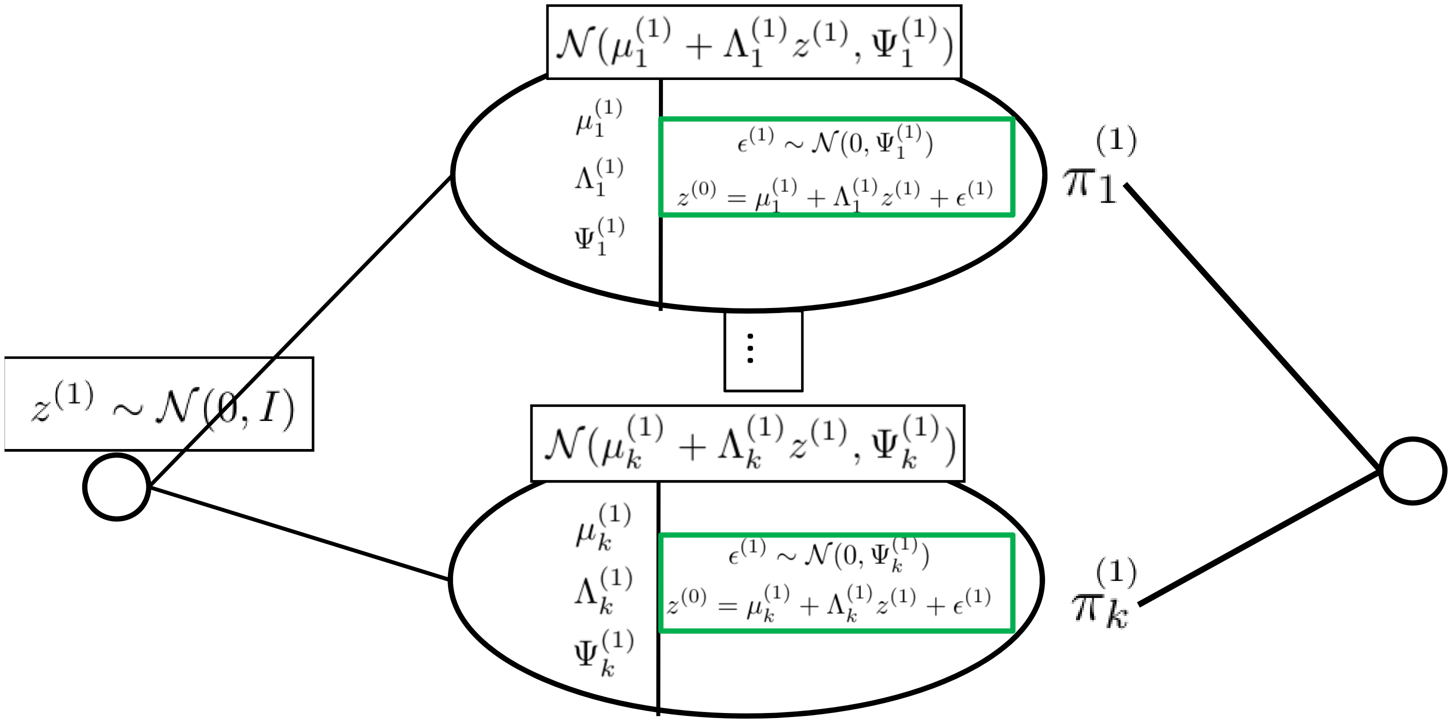


$$\begin{aligned} p(x;\mu,\Lambda,\Psi) &= \pi_1\mathcal{N}(x;\mu_1,\Lambda_1\Lambda_1^T+\Psi_1) + \\ &+ \cdots + \\ &+ \pi_k\mathcal{N}(x;\mu_k,\Lambda_k\Lambda_k^T+\Psi_k) \end{aligned}$$

$$\pi_1 + \cdots + \pi_k = 1$$

DGMMs – Mixturi de distribuții pb.

- Cazul alternativ: Mixturi de FA (2)



2 1 0

-

$z^{(1)} \sim \mathcal{N}(0, I)$

$$z^{(0)} | z^{(1)} \sim \pi_1^{(1)} \mathcal{N}(\mu_1^{(1)} + \Lambda_1^{(1)} z^{(1)}, \Psi_1^{(1)}) + \cdots + \pi_k^{(1)} \mathcal{N}(\mu_k^{(1)} + \Lambda_k^{(1)} z^{(1)}, \Psi_k^{(1)})$$

$$z^{(0)} \sim \pi_1^{(1)} \mathcal{N}(\mu_1^{(1)}, \Lambda_1^{(1)} (\Lambda_1^{(1)})^T + \Psi_1^{(1)}) + \cdots + \pi_k^{(1)} \mathcal{N}(\mu_k^{(1)}, \Lambda_k^{(1)} (\Lambda_k^{(1)})^T + \Psi_k^{(1)})$$

DGMMs – Mixturi de distribuții pb.

- Cum estimăm parametrii? Nu există formule analitice
- Cu algoritmul EM (Expectation Maximization)
  - Pentru că lucrăm cu variabile latente...

EM/MFA

$\omega_{ij} = 1 \Leftrightarrow x_i$ - generat de componenta j

Pasul E:

$$h_{ij} = E[\omega_{ij}|x_i] = \frac{\pi_j \mathcal{N}(x_i; \mu_j, \Lambda_j \Lambda_j^T + \Psi_j)}{\pi_1 \mathcal{N}(x_i; \mu_1, \Lambda_1 \Lambda_1^T + \Psi_1) + \cdots + \pi_k \mathcal{N}(x_i; \mu_k, \Lambda_k \Lambda_k^T + \Psi_k)}$$

$$\beta_j = \Lambda_j^T (\Psi_j + \Lambda_j \Lambda_j^T)^{-1}$$

$$E[z_i|x_i, \omega_{ij} = 1] = \beta_j (x_i - \mu_j), \forall i \in \{1, \ldots, n\}, \forall j \in \{1, \ldots, k\}$$

$$E[z_i z_i^T|x_i, \omega_{ij} = 1] = I - \beta_j \Lambda_j + E[z_i|x_i, \omega_{ij} = 1] E[z_i^T|x_i, \omega_{ij} = 1], \forall i \in \{1, \ldots, n\}, \forall j \in \{1, \ldots, k\}$$

Sursă: după *The EM Algorithm for Mixtures of Factor Analyzers*, Zoubin Ghahramani, Geoffrey E. Hinton, 1997

DGMMs – Mixturi de distribuții pb.

Pasul M:

$$\tilde{z}_i = \begin{bmatrix} z_i \\ 1 \end{bmatrix}$$

$$E[\tilde{z}_i | x_i, \omega_{ij} = 1] = \begin{bmatrix} E[z_i | x_i, \omega_{ij} = 1] \\ 1 \end{bmatrix}$$

$$E[\tilde{z}_i \tilde{z}_i{}^T | x_i, \omega_{ij} = 1] = \begin{bmatrix} E[z_i z_i^T | x_i, \omega_{ij} = 1] & E[z_i | x_i, \omega_{ij} = 1] \\ E[z_i^T | x_i, \omega_{ij} = 1] & 1 \end{bmatrix}$$

$$[\Lambda_j, \mu_j] = \tilde{\Lambda}_j = \left(\sum_{i=1}^{n} h_{ij} x_i E[z_i^T | x_i, \omega_{ij} = 1] \right) \left(\sum_{i=1}^{n} h_{ij} E[z_i z_i^T | x_i, \omega_{ij} = 1] \right)^{-1} \quad \forall j \in \{1, \ldots, k\}$$

$$\Psi_j = \frac{1}{n} \mathrm{diag} \left(\sum_{i=1}^{n} h_{ij} (x_i x_i^T - \tilde{\Lambda}_j E[z_i | x_i, \omega_{ij} = 1] x_i^T) \right) \quad \forall j \in \{1, \ldots, k\}$$

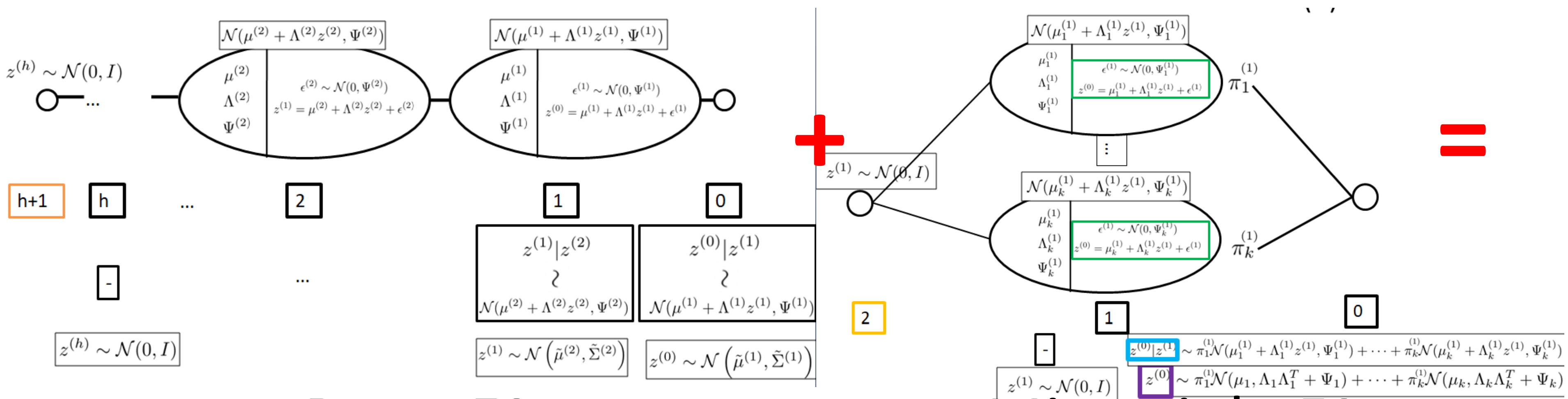
$$\pi_j = \frac{1}{n} \sum_{i=1}^{n} h_{ij} \qquad \forall j \in \{1, \ldots, k\}$$

- <u>Seamănă foarte mult cu EM/FA</u>

DGMMs

Deep Gaussian Mixture Models,
Cinzia Viroli, Geoffrey J. McLachlan,
2017

Deep FA + MFA = ...

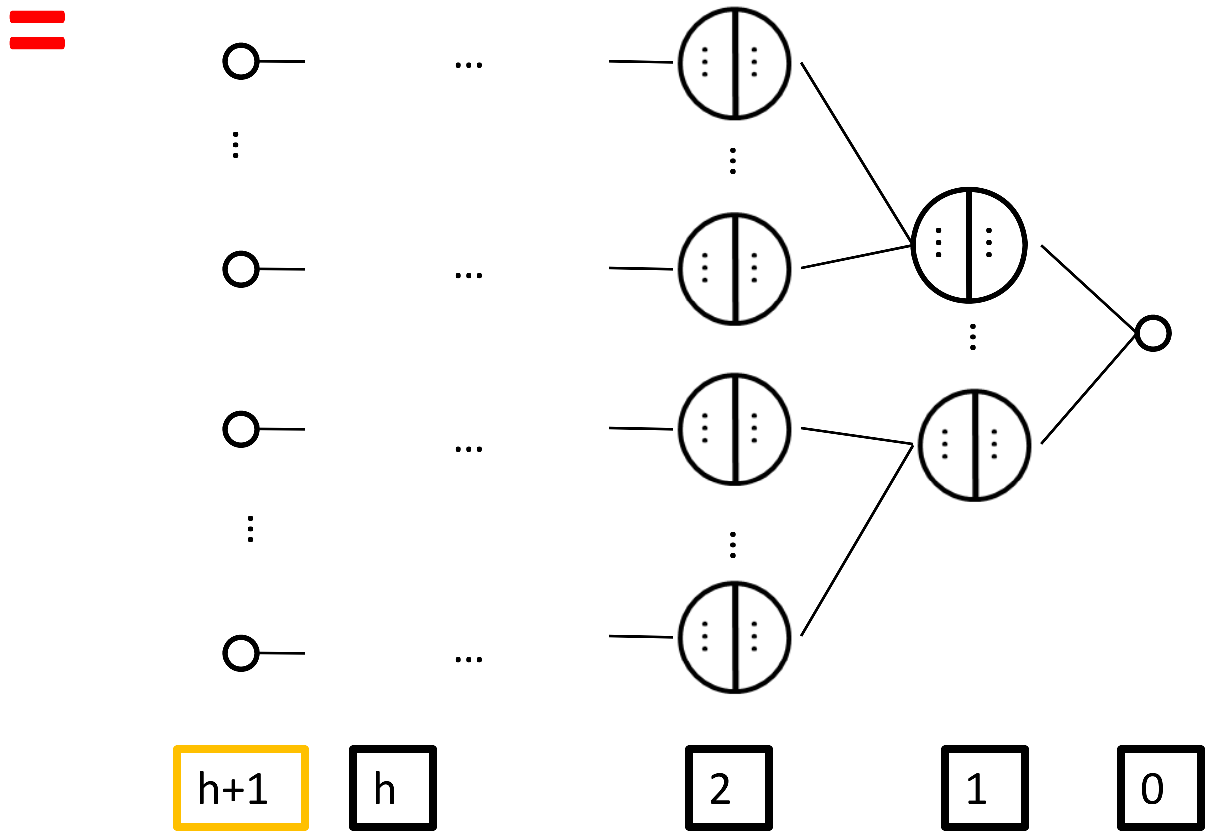


Deep FA + Mixturi de FA =

Idee cheie 1: De ce nu mixturi de distr. normale?

Pierdem proprietatea de adâncime...

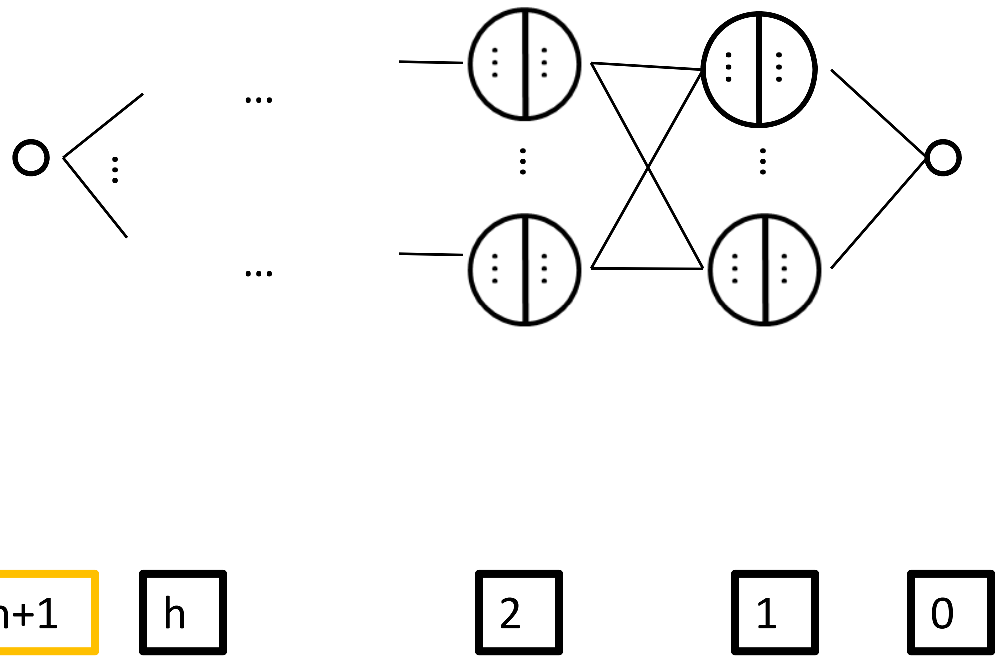
… = DGMM?



Dar astfel vom avea foarte mulți parametri... O alternativă?

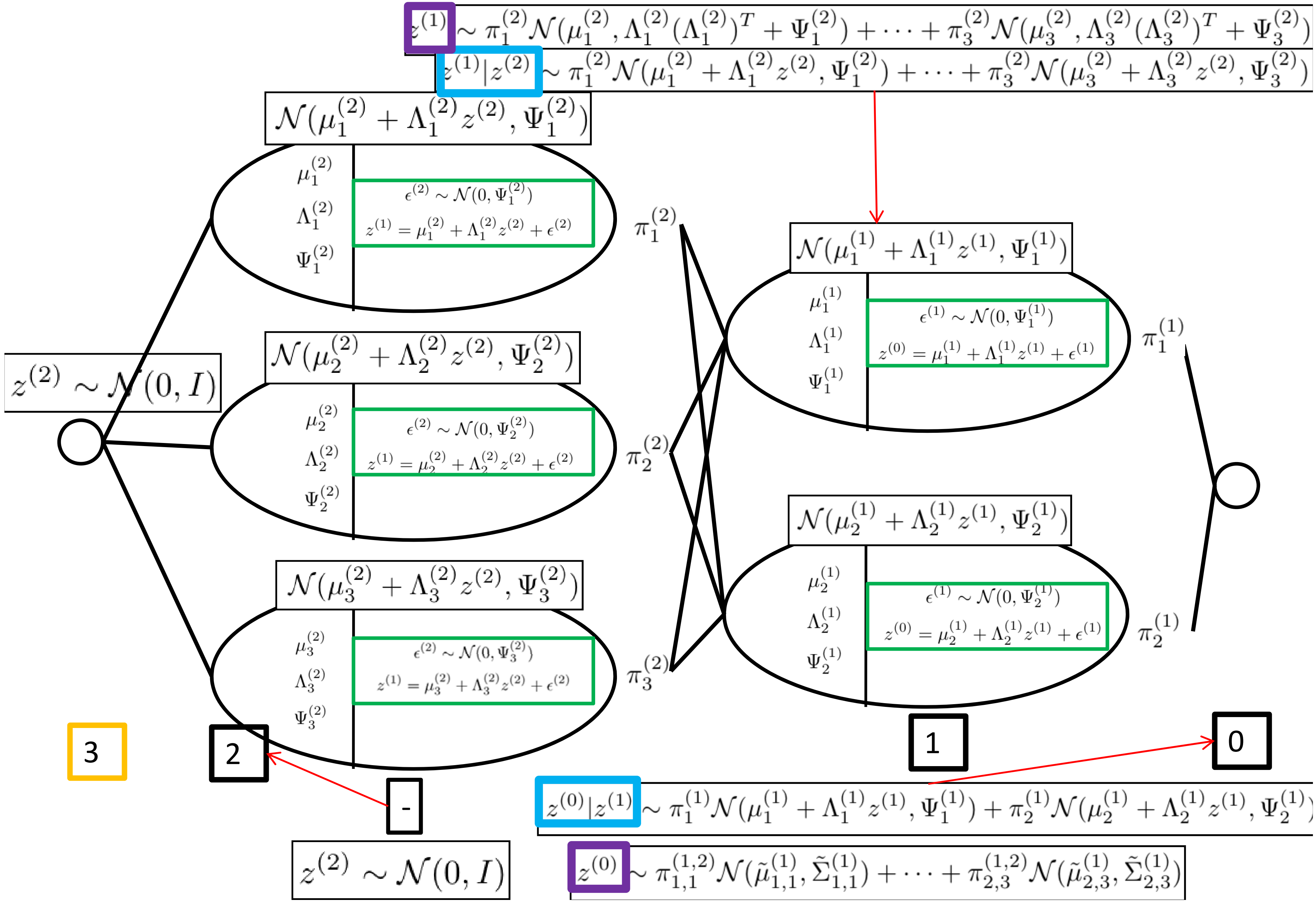
... = DGMM

=

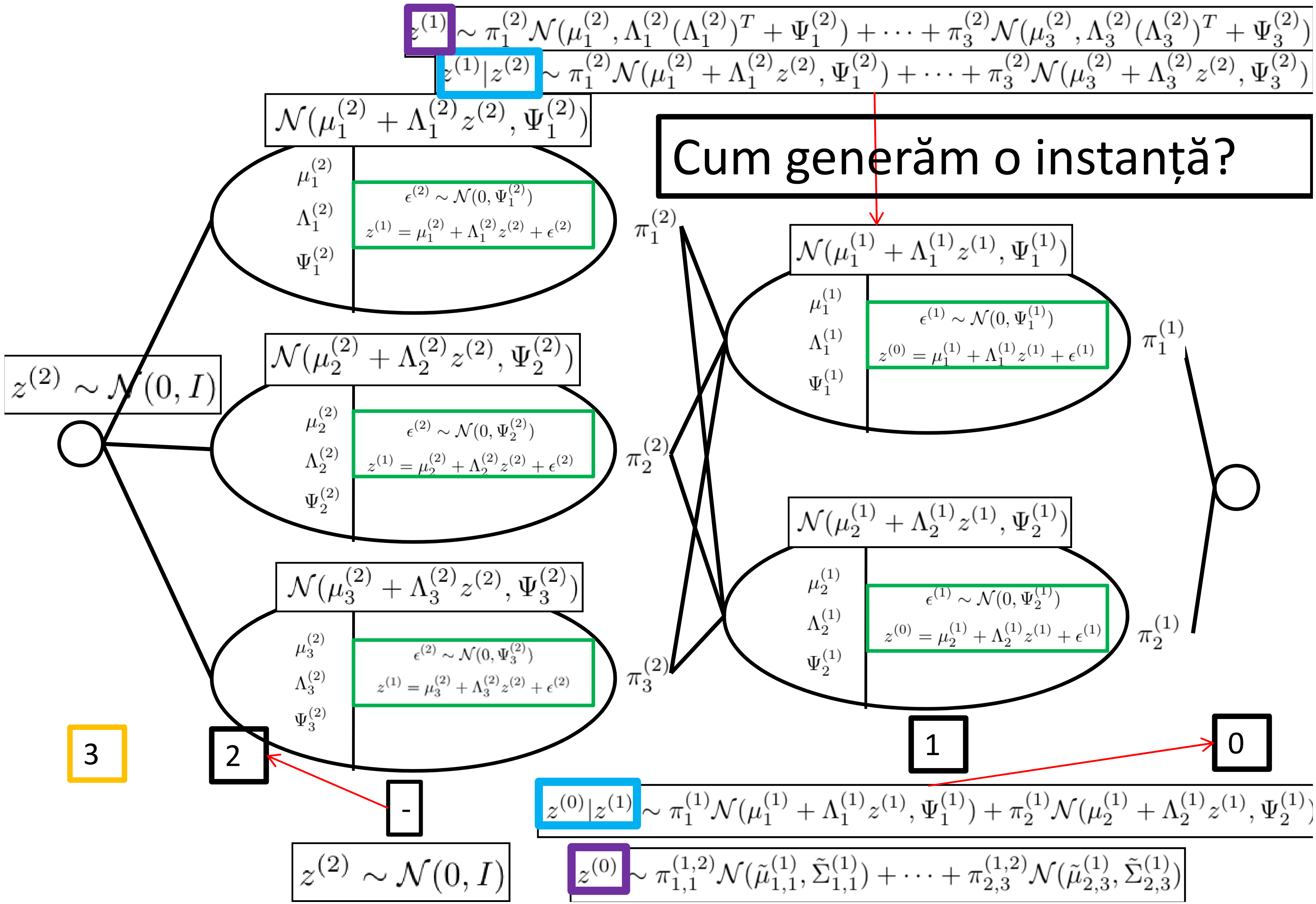


Idee cheie 2: parametri comuni

Exemplu DGMM



Exemplu DGMM



DGMMs

- Definiție

$$
\begin{aligned}
&(1) \qquad \mathbf{y}_i = \eta^{(1)}_{s_1} + \Lambda^{(1)}_{s_1}\mathbf{z}^{(1)}_i + \mathbf{u}^{(1)}_i \text{ with prob. } \pi^{(1)}_{s_1},\ s_1 = 1,\dots,k_1,\\
&(2) \qquad \mathbf{z}^{(1)}_i = \eta^{(2)}_{s_2} + \Lambda^{(2)}_{s_2}\mathbf{z}^{(2)}_i + \mathbf{u}^{(2)}_i \text{ with prob. } \pi^{(2)}_{s_2},\ s_2 = 1,\dots,k_2,\\
&\qquad\qquad \dots\\
&(h) \qquad \mathbf{z}^{(h-1)}_i = \eta^{(h)}_{s_h} + \Lambda^{(h)}_{s_h}\mathbf{z}^{(h)}_i + \mathbf{u}^{(h)}_i \text{ with prob. } \pi^{(h)}_{s_h}, s_h = 1,\dots,k_h,
\end{aligned}
$$

- Dimensiunea de la fiecare nivel este aceeași!

(dimensiunea datelor observabile)

DGMMs - probleme

- Există soluții simetrice:
  - Din cauza aceleiași dimensiuni de la fiecare nivel
    - Soluție: impunem ca dimensiunile să descrească strict pe măsură ce crește indexul nivelului
  - Din cauza non-unicității FA
    - Soluție: menționată la FA

DGMMs – învățare

- Cum estimăm parametrii? Nu există formule analitice
- Cu algoritmul EM (Expectation Maximization)
  - Pentru că lucrăm cu variabile latente...
- Folosesc algoritmul *Stochastic* EM
  - Pasul S: generăm date
  - Pasul E: calculăm E-urile din datele generate
  - Pasul M

DGMMs – învățare

- Idee de antrenare
  - Antrenare pe niveluri
  - Aplicăm algoritmul EM/MFA (modificat)
  - De la nivelul 1 la nivelul h.

DGMMs – învățare

For $l = 1, \ldots, h$

- S-STEP ($\mathbf{z}_i^{(l-1)}$ is known)

  Generate M replicates $\mathbf{z}_{i,m}^{(l)}$ from $f(\mathbf{z}_i^{(l)}|\mathbf{z}_i^{(l-1)}, s; \mathbf{\Theta}')$.

- E-STEP - Approximate:

$$E[\mathbf{z}_i^{(l)}|\mathbf{z}_i^{(l-1)}, s; \mathbf{\Theta}'] \cong \frac{\sum_{m=1}^{M} \mathbf{z}_{i,m}^{(l)}}{M}$$

and

$$E[\mathbf{z}_i^{(l)}\mathbf{z}_i^{(l)\top}|\mathbf{z}_i^{(l-1)}, s; \mathbf{\Theta}'] \cong \frac{\sum_{m=1}^{M} \mathbf{z}_{i,m}^{(l)}\mathbf{z}_{i,m}^{(l)\top}}{M}.$$

DGMMs – învățare

- M-STEP - Compute:

$$
\begin{aligned}
\hat{\Lambda}_{s_l}^{(l)} &= \frac{\sum_{i=1}^{n} p(s|\mathbf{z}_i^{(l-1)})(\mathbf{z}_i^{(l-1)} - \eta_{s_l}^{(l)})E[\mathbf{z}_i^{(l)\top}|\mathbf{z}_i^{(l-1)}, s]E[\mathbf{z}_i^{(l)}\mathbf{z}_i^{(l)\top}|\mathbf{z}_i^{(l-1)}, s]^{-1}}{\sum_{i=1}^{n} p(s|\mathbf{z}_i^{(l-1)})}, \\
\hat{\Psi}_{s_l}^{(l)} &= \frac{\sum_{i=1}^{n} p(s|\mathbf{z}_i^{(l-1)}) \left[(\mathbf{z}_i^{(l-1)} - \eta_{s_l})(\mathbf{z}_i^{(l-1)} - \eta_{s_l})^\top - (\mathbf{z}_i^{(l-1)} - \eta_{s_l})E[\mathbf{z}_i^{(l)\top}|\mathbf{z}_i^{(l-1)}, s]\hat{\Lambda}_{s_l}^\top\right]}{\sum_{i=1}^{n} p(s|\mathbf{z}_i^{(l-1)})}, \\
\hat{\eta}_{s_l}^{(l)} &= \frac{\sum_{i=1}^{n} p(s|\mathbf{z}_i^{(l-1)}) \left[\mathbf{z}_i^{(l-1)} - \Lambda_{s_l}E[\mathbf{z}_i^{(l)\top}|\mathbf{z}_i^{(l-1)}, s]\right]}{\sum_{i=1}^{n} p(s|\mathbf{z}_i^{(l-1)})}, \\
\hat{\pi}_s^{(l)} &= \sum_{i=1}^{n} f\left(s_l|\mathbf{y}_i\right),
\end{aligned}
$$

DGMMs – aplicații

- Clusterizare
  - $p(z^{(0)}|z^{(1)})$ este o mixtură
  - $p(z^{(0)})$ este o mixtură
- Acest model înglobează mai multe modele
  - FA
  - MFA
  - GMM...
  - Deep MFA
  - ...
    - => poate fi folosit și în alte contexte:
      - de ex. Estimarea densității (density estimation)

Modelling high-dimensional data by mixtures of factor analyzers, McLachlan, G., D. Peel, and R. Bean, 2003
Deep Mixtures of Factor Analysers, Yichuan Tang, Ruslan Salakhutdinov, Georey Hinton, 2012

Rezultatele mele

- **Probleme**
  - În articol:
    - Clusterizare după $p(z^{(0)}|z^{(1)})$, dar nu știm $z^{(1)}$
      - Soluția 1: $\boxed{z^{(0)} \sim \pi_1^{(1)}\mathcal{N}(\mu_1^{(1)}, \Lambda_1^{(1)}(\Lambda_1^{(1)})^T + \Psi_1^{(1)}) + \cdots + \pi_k^{(1)}\mathcal{N}(\mu_k^{(1)}, \Lambda_k^{(1)}(\Lambda_k^{(1)})^T + \Psi_k^{(1)})}$
      - Soluția 2: $\boxed{z^{(0)} = \sum_{\text{path}\in\text{paths}} \pi_{\text{path}}^{(1,\ldots,h)}\mathcal{N}(\tilde{\mu}_{\text{path}}^{(1)}, \tilde{\Sigma}_{\text{path}}^{(1)})}$
    - $\pi_{\text{path}}^{(1,\ldots,h)}$ nu este cunoscut
      - Soluție: presupunem independența

$$\boxed{\pi_{\text{path}}^{(1,\ldots,h)} = \prod_{l=1}^{h} \pi_{\text{path}_l}^{(l)}}$$

Rezultatele mele

- **Probleme**
  - În articol:
    - Nu se specifică explicit cum se calculează $f\left(s_l|\mathbf{y}_i\right)$
    - Nu se specifică explicit cum se calculează $p(s|\mathbf{z}_i^{(l-1)})$
      - Soluție: le-am calculat *clasic*
      - Problemă: probabilitățile π >> 1
        - » Soluție: un fel de normalizare
        - » Problemă: la un nivel suma π -urilor $\neq$ 1
          - Soluție: nu mai actualizăm π -urile

Rezultatele mele

- **Probleme**
  - În articol:
    - La MFA nu folosim SEM, dar în articol da
    - Formulele din algoritm nu sunt exact cele de la MFA
      - Soluție: am implementat modele mai mici înglobate în DGMM după formulele clasice și după formulele din articol
      - Rezultat: cam același comportament la funcția de optimizat
      - Problemă: funcția de optimizat mai și descrește

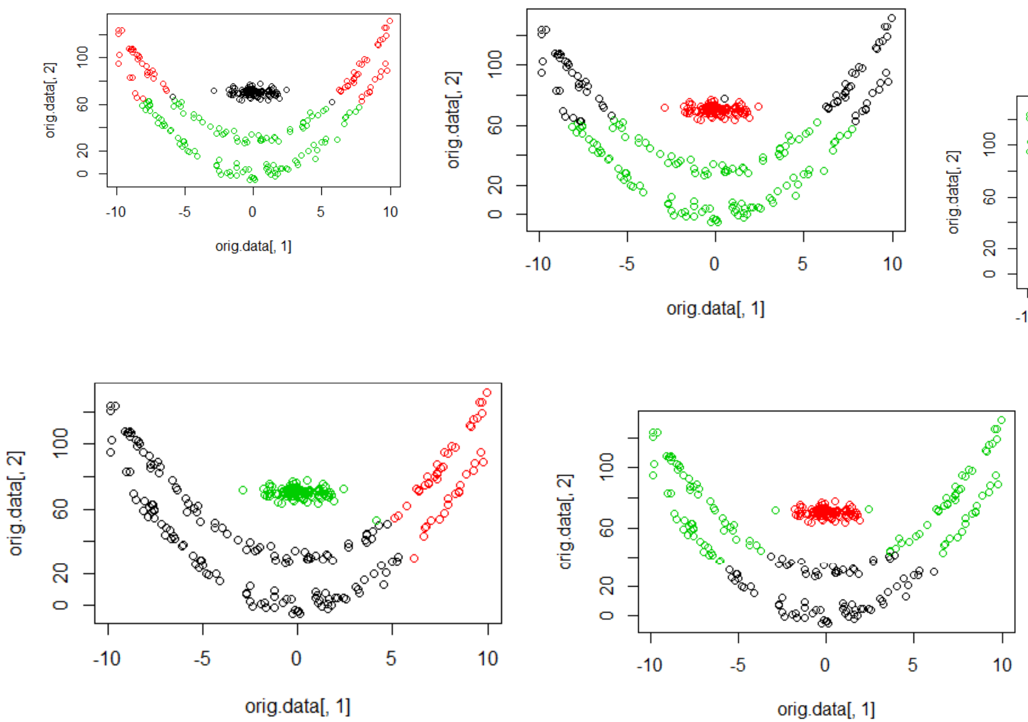
Rezultatele mele

- **Probleme**
  - Formulele de actualizare: inversa unei matrici
    - Problemă: matricea nu are inversă
    - Soluție: folosim pseudoinversa
  - Timpul de rulare (pentru imagini):
    - **FA**
    - 8 iterații
    - setul de date MNIST – cifra 8: 5851 poze de 28x28 (19 MB)
    - Timp: **2 zile**
  - ...

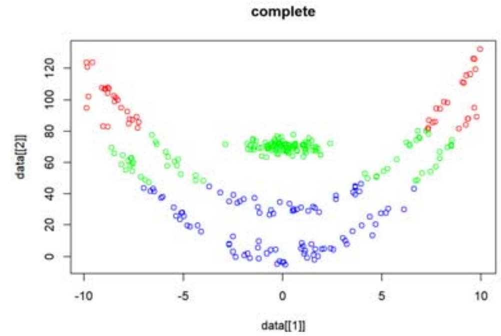
Rezultatele mele

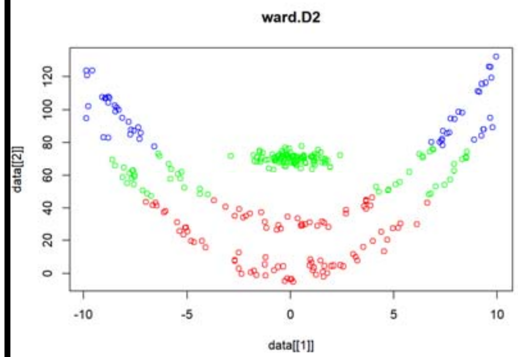
- Seturi de date clasice 2D
- Comparație vizuală cu metode clasice de clusterizare
- Mai multe rulări (numere aleatorii)
- Nu doar o singură arhitectură
- Clusterizare prin $p(z^{(0)}|z^{(1)})$ sau $p(z^{(0)})$

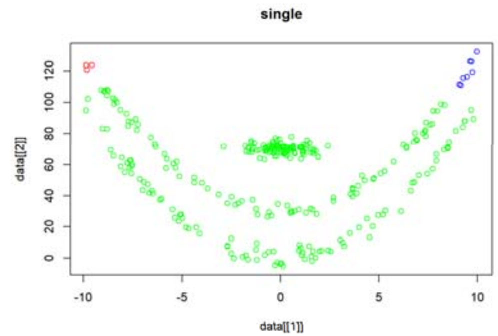
Rezultatele mele

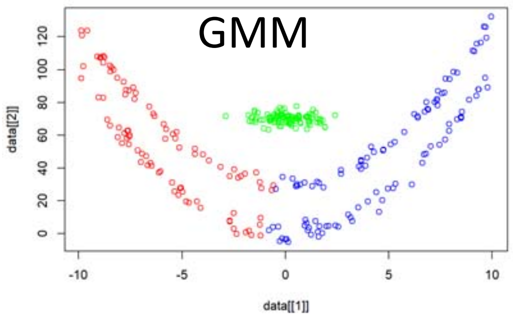


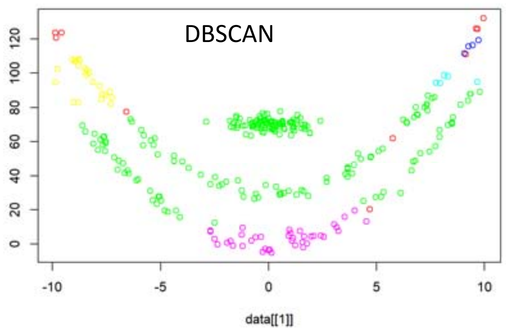
DGMM

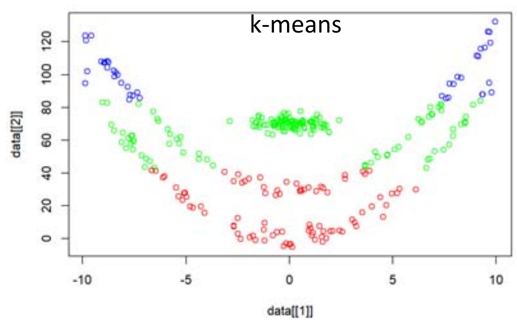


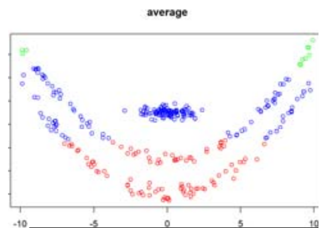






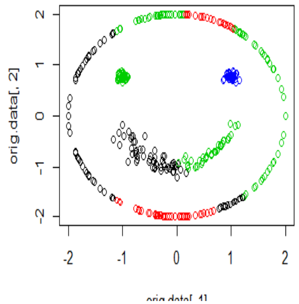


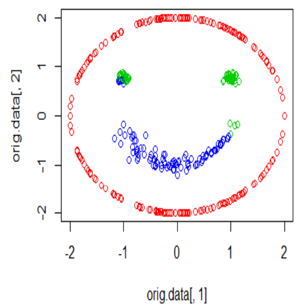


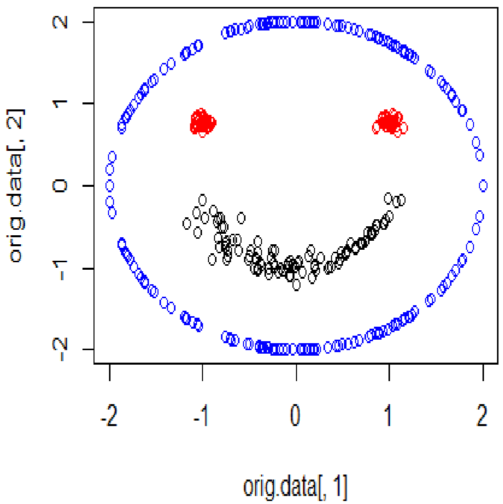


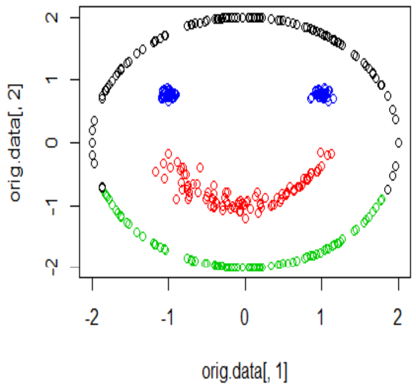
Alte metode

Rezultatele mele

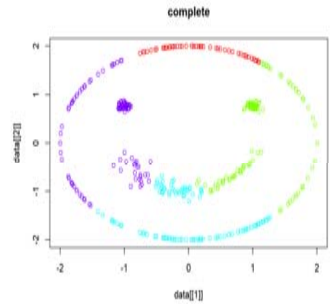


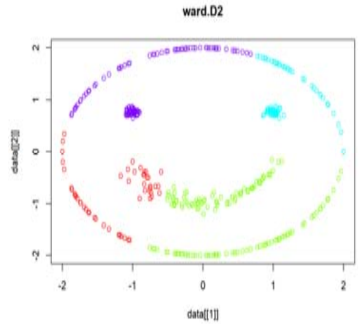


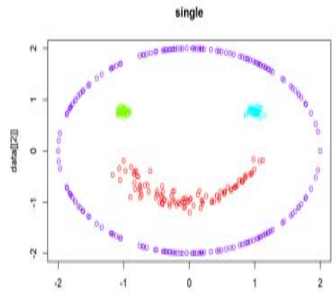




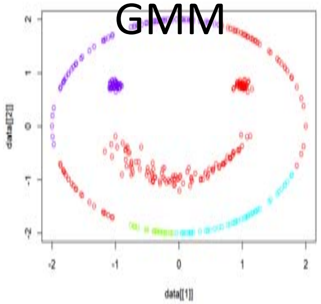
DGMM



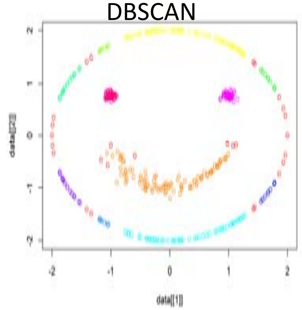




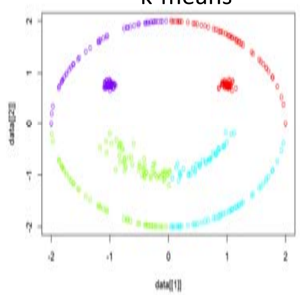
GMM

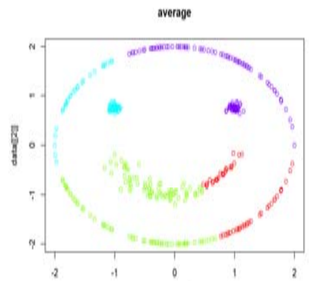


DBSCAN



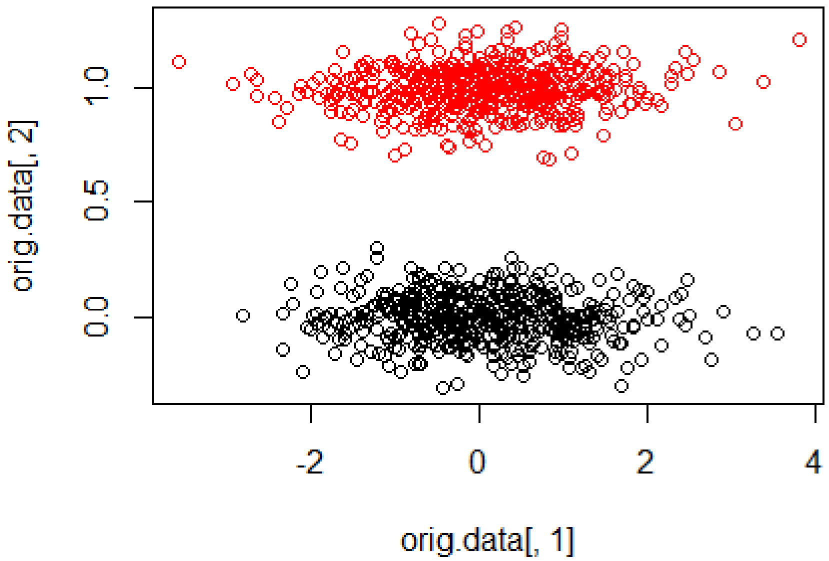
k-means



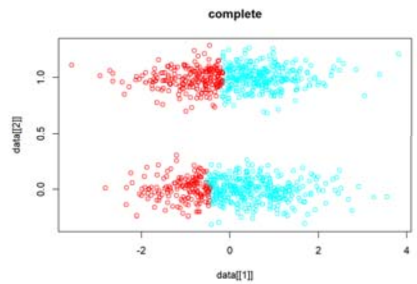


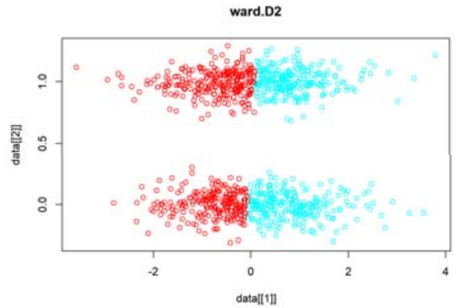
Alte metode

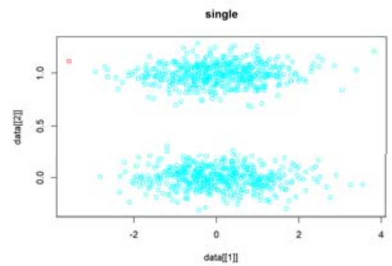
Rezultatele mele

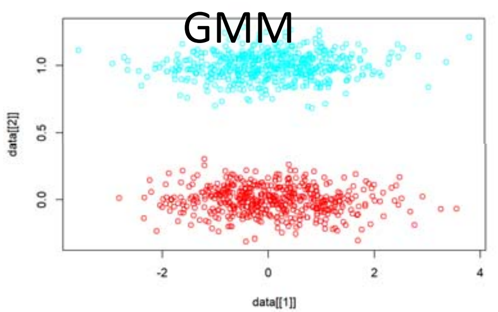


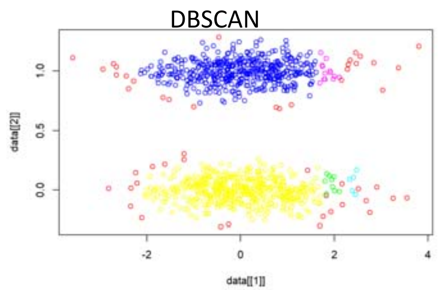
DGMM

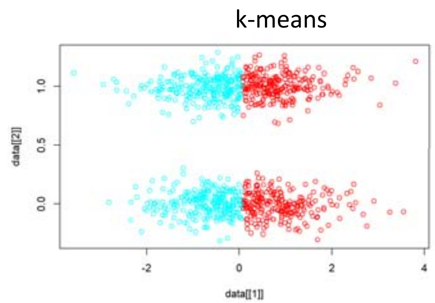


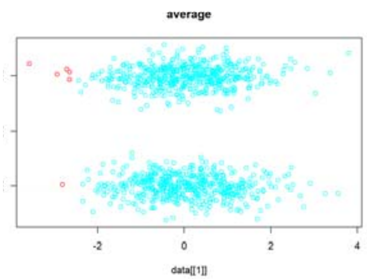












Alte metode

Rezultatele lor

Deep Gaussian Mixture Models, Cinzia Viroli, Geoffrey J. McLachlan, 2017

- **Clusterizare**
  - Date simulate
    - Generate de 100 de ori: *media (er. std.)*
    - Arhitecturi: $h = 2, d_1 = 2, d_2 = 1, k_1 = 4, k_2 = 1 - 5$
    - Alegerea modelului:
      - BIC (Bayesian Information Criteria) = f(#params, logLikelihood)

	Method	ARI		m.r.	
	k-means	0.661	(0.003)	0.134	(0.001)
Partition Around Medoids	PAM	0.667	(0.004)	0.132	(0.001)
Ward	Hclust	0.672	(0.013)	0.141	(0.006)
	GMM	0.653	(0.008)	0.178	(0.006)
Skew-normal mixture model	SNmm	0.535	(0.006)	0.251	(0.006)
Skew-t mixture model	STmm	0.566	(0.006)	0.236	(0.004)
	DGMM	0.788	(0.005)	0.087	(0.002)

ARI = Adjusted Rand Index
= sim(real_part,
clust_part)

m.r. = misclassification rate
=! acuratețe

Rezultatele lor

Deep Gaussian Mixture Models, Cinzia Viroli, Geoffrey J. McLachlan, 2017

- **Clusterizare**
  - Date reale
    - Arhitecturi: $h = 2 - 3$

$$D > d_1 > d_2 > d_3 \geq 1$$

$$k_1 = k^*$$

$$k_2 = 1 - 5$$

$$k_3 = 1 - 5$$

    - Alegerea modelului: BIC
    - 10 rulări: inițializări diferite

Rezultatele lor

Deep Gaussian Mixture Models, Cinzia Viroli, Geoffrey J. McLachlan, 2017

- **Clusterizare**
  - Date reale

Dataset	*Wine*		*Olive*		*Ecoli*		*Vehicle*		*Satellite*	
	ARI	m.r.	ARI	m.r.	ARI	m.r.	ARI	m.r.	ARI	m.r.
k-means	0.930	0.022	0.448	0.234	0.548	0.298	0.071	0.629	0.529	0.277
PAM	0.863	0.045	0.725	0.107	0.507	0.330	0.073	0.619	0.531	0.292
Hclust	0.865	0.045	0.493	0.215	0.518	0.330	0.092	0.623	0.446	0.337
GMM	0.917	0.028	0.535	0.195	0.395	0.414	0.089	0.621	0.461	0.374
SNmm	0.964	0.011	0.816	0.168	-	-	0.125	0.566	0.440	0.390
STmm	0.085	0.511	0.811	0.171	-	-	0.171	0.587	0.463	0.390
FMA	0.361	0.303	0.706	0.213	0.222	0.586	0.093	0.595	0.367	0.426
MFA	0.983	0.006	0.914	0.052	0.525	0.330	0.090	0.626	0.589	0.243
DGMM	0.983	0.006	0.997	0.002	0.749	0.187	0.191	0.481	0.604	0.249

FMA – Factor Mixture Analysis

MFA – Mixture of Factor Analyzers

Rezultatele lor

Deep Mixtures of Factor Analysers, Yichuan Tang, Ruslan Salakhutdinov, Georey Hinton, 2012

- **Estimarea densității**
  - Setul de date: Toronto Faces Dataset
  - Din setul de date:

Rezultatele lor

Deep Mixtures of Factor Analysers, Yichuan Tang, Ruslan Salakhutdinov, Georey Hinton, 2012

- **Estimarea densității** **(Distribuția învățată chiar generează fețe?)**
  – Fețe generate după estimarea densității conform:

MLE distribuția normală

MFA

Deep MFA
(înglobat de DGMM)

Concluzii

- Deep (next: Deep Gaussian Processes)
- Nou (pe lângă DGMM):
  - Stochastic EM
  - BIC
  - ARI
  - Estimarea densității prin distribuții normale: aplicare pe fețe
  - Ideea articolului:
    - clasic (FA, mixturi) -> modern (DGMM)