Lecture 2

# Graphical Models in Action

Prof. Dahua Lin
dhlin@ie.cuhk.edu.hk

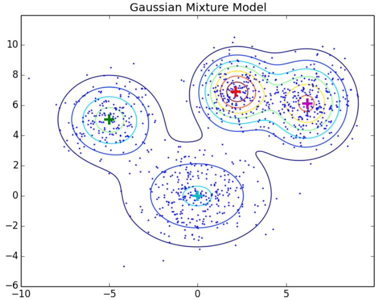
# Roadmap

1. Basic Procedure
2. Gaussian Mixture Models (GMM)
3. Extend GMM
4. Markov Random Fields for Images

# Formulate a Graphical Model

- Understand the **problem**
  - what kind of entities/factors are involved?
  - How do they interact with each other?
  - Any constraints to take into account?
- Formulate the **model**
  - Introduce variables
  - Specify relations among them
    - **make assumptions and modeling choices**
  - Formalize the graphical model
- Derive the inference & estimation **algorithms**.

# Gaussian Mixture Model: Motivation & Assumptions



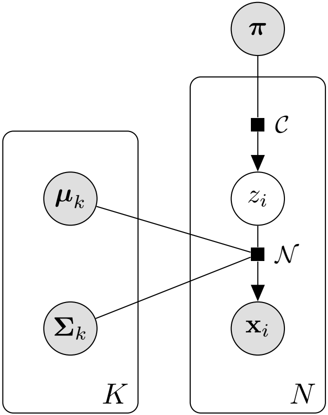

- Observation: clusters
- Assumptions:
  - latent **components**
  - independent generation of components
  - independent generation of points in each component.

# Formulate Gaussian Mixture Model

- Variables:
    - sample points: $\mathbf{x}_i$
    - component indicators: $z_i$
- Generative procedure: for each $i$,
    1. choose a component: $z_i \in \boldsymbol{\pi}$
    2. generate a point: $\mathbf{x}_i \in \mathcal{N}(\boldsymbol{\mu}_{z_i}, \boldsymbol{\Sigma}_{z_i})$.
- Model parameters:
    - component parameters: $\{(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\}_{k=1:K}$.
    - choice prior: $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$.
- Joint distribution:

$$p(X, Z \mid \Theta) = \prod_{i=1}^{N} p_{\mathcal{C}}(z_i|\boldsymbol{\pi}) p_{\mathcal{N}}(\mathbf{x}_i|\boldsymbol{\mu}_{z_i}, \boldsymbol{\Sigma}_{z_i}).$$

# Graphical Representations



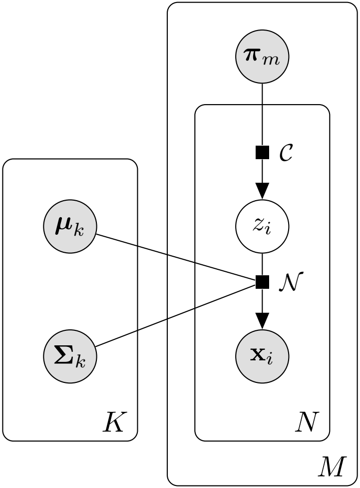

```
\usepackage{tikz}
\usetikzlibrary{bayesnet}

\newcommand{\vx}{\mathbf{x}}
\newcommand{\vpi}{\boldsymbol{\pi}}
\newcommand{\vmu}{\boldsymbol{\mu}}
\newcommand{\mSigma}{\boldsymbol{\Sigma}}
\newcommand{\cC}{\mathcal{C}}
\newcommand{\cN}{\mathcal{N}}

\begin{tikzpicture}
    % variables
    \node[obs]                                   (x)   {$\vx_i$};
    \node[latent, above=of x]                    (z)   {$z_i$};
    \node[obs,    above=of z, yshift=15pt]       (pi)  {$\vpi$};
    \node[obs,    left=50pt of z, yshift=-5pt]   (mu)  {$\vmu_k$};
    \node[obs,    below=of mu]                   (sig) {$\mSigma_k$};
    % factors
    \factor[above=15pt of z] {choose} {right:$\cC$} {pi}        {z};
    \factor[above=15pt of x] {gen}    {right:$\cN$} {z,mu,sig}  {x};
    % plates
    \plate[inner sep=15pt] {sample}    {(z)(x)(choose)} {};
    \plate[inner sep=15pt] {component} {(mu)(sig)}      {};
    \node[const, above left=5pt of sample.south east]    (N){$N$};
    \node[const, above left=5pt of component.south east] (K){$K$};
\end{tikzpicture}
```

# Group-wise GMM: First Attempt

- Suppose we have multiple groups of data, each governed by a different distribution of components.
- Formulation
  - Each group $G_m$ has a prior $\boldsymbol{\pi}_m$
  - Generate the $i$-th point in $G_m$:

$$z_i \sim \boldsymbol{\pi}_m$$
$$\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

- **How to generate new groups?**
  - We need a prior over $\boldsymbol{\pi}_m$.



# Common Choices of Priors

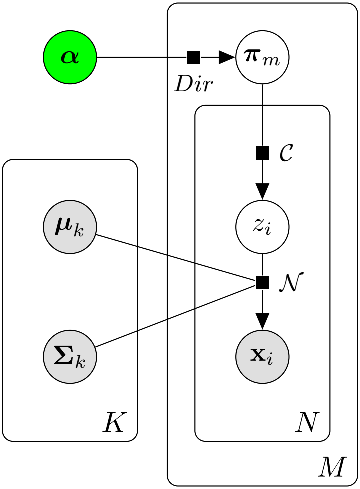
| Parameters | Prior |
|---|---|
| Probability value (e.g. *Bernoulli*) | $Beta(\alpha, \beta)$ |
| Probability vector (e.g. *Categorical*) | $Dirichlet(\boldsymbol{\alpha})$ |
| Mean vector (e.g. *Normal*) | $Normal(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ |
| Variance (e.g. *Normal*) | $InvGamma(\alpha, \beta)$ |
| Rate (e.g. *Poisson*, *Exponential*) | $Gamma(\alpha, \beta)$ |

# Group-wise GMM: Generalizable to New Groups

- Introduce a Dirichlet Prior over $\boldsymbol{\pi}_m$ to allow the generation of new groups.
- Formulation
  - For each group $G_m$: $\boldsymbol{\pi}_m \sim \mathrm{Dir}(\boldsymbol{\alpha})$.
  - Generate the $i$-th point in $G_m$:

$$z_i \sim \boldsymbol{\pi}_m$$
$$\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

  - **Note:** $\boldsymbol{\pi}_m$ is now a latent variable.



# Temporal Structures

- **Temporal structures** are ubiquitous in the real world.
- Three ways to incorporate dynamics
  - Dynamics on $\mathbf{x}_i$
  - Dynamics on $z_i$
  - Dynamics on $\boldsymbol{\pi}$

# GMM with Dynamics on $\mathbf{x}_i$

- Suppose the motion of each point respectively follows a Markovian motion model $p_S$.
- Formulation
  - Initial distribution ($t = 0$) is a GMM.
  - At time step $t > 0$, $\mathbf{x}_i^{(t)}$ is generated *independently* conditioned on $\mathbf{x}_i^{(t-1)}$ for each point $i$, following $p_S(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}; z_i)$.

# Stochastic Dynamics

- *Stochastic dynamics* of real vectors can often be characterized by a **Stochastic Differential Equation (SDE)**:

$$\underbrace{dX_t}_{\text{actual movement}} = \underbrace{\mu(X_t, t)dt}_{\text{deterministic part}} + \underbrace{\sigma(X_t, t)dB_t}_{\text{random perturbation}}$$

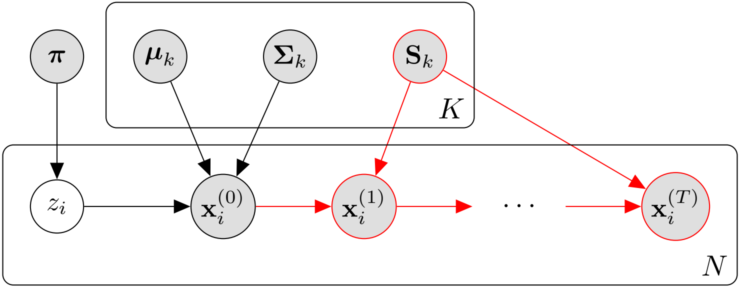
- Discretization of an SDE into a Markov chain

$$\mathbf{x}_t \mid \mathbf{x}_{t-1} \sim \mathcal{N}\left(\mathbf{A}\mathbf{x}_{t-1} + \mathbf{b}, \boldsymbol{\Lambda}\right)$$

  - We use a *linear approximation*, assuming very small intervals.
  - This is a special case of the **Kalman filter model**, with *zero input*.
  - $\mathbf{x}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) \implies \mathbf{x}_{t+1} \sim \mathcal{N}(\mathbf{x}_t + \mathbf{A}\mathbf{x}_t + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^T + \boldsymbol{\Lambda})$

# GMM with Dynamics on $\mathbf{x}_i$ (graph)

$$\prod_{i=1}^{N} p_{\mathcal{C}}(z_i|\boldsymbol{\pi}) p_{\mathcal{N}}(\mathbf{x}_i^{(0)}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \underbrace{\prod_{t=1}^{T_i} p_S(\mathbf{x}_i^{(t)}|\mathbf{x}_i^{(t-1)}; \mathbf{S}_k)}_{\text{dynamics of } \mathbf{x}_i}, \text{ with } \mathbf{S}_k = (\mathbf{A}_k, \mathbf{b}_k, \boldsymbol{\Lambda}_k)$$

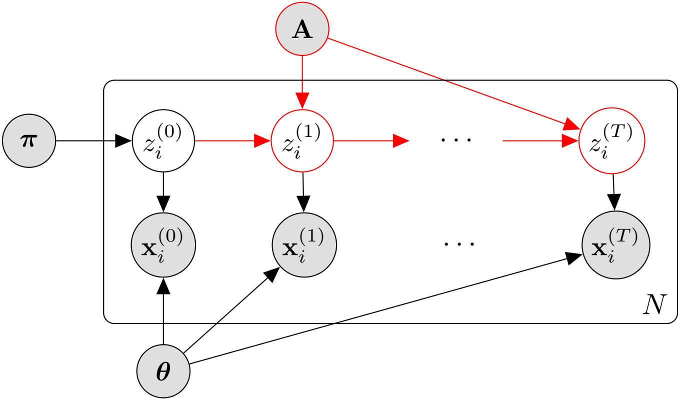


# GMM with Dynamics on $z_i$: HMM

- Assumptions:
  - Along a sequence, the latent component may switch from one to another at each time step, following a Markov chain.
  - The observed samples are generated independently conditioned on the component choices.
- Formulation, for the $i$-th sequence:
  - Draw the initial component $z_0 \sim \boldsymbol{\pi}$.
  - At each time step $t > 0$, $z_t$ is generated conditioned on $z_{t-1}$ following a Markov chain $p(z_t | z_{t-1}; \mathbf{A})$.
  - $\mathbf{x}_i^{(t)}$ is generated independently conditioned on $z_i^{(t)}$, for each $t \geq 0$.
- This formulation is known as the **Hidden Markov Model (HMM)**.

# Hidden Markov Model (graph)

$$p(z_i^{(0)}|\boldsymbol{\pi}) \prod_{t=1}^{T_i} p(z_i^{(t)}|z_i^{(t-1)}; \mathbf{A}) \cdot \prod_{t=0}^{T_i} p(\mathbf{x}_i^{(t)}|z_i^{(t)}; \boldsymbol{\theta})$$



# GMM with Dynamics on $\boldsymbol{\pi}$

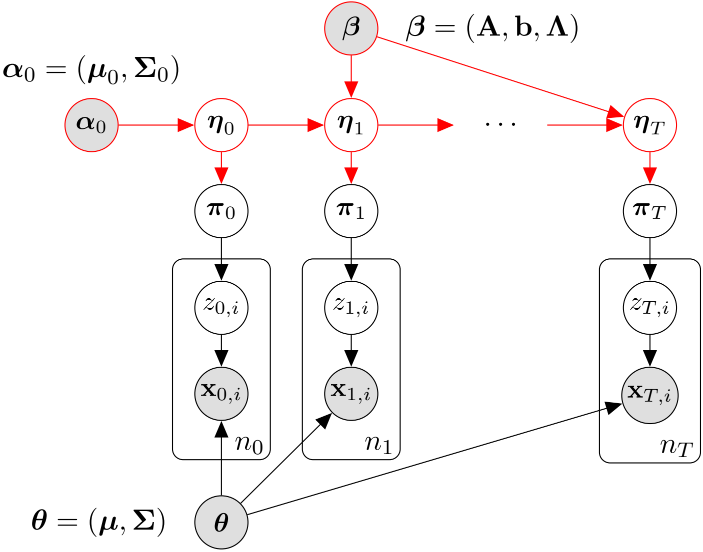
- The *proportions* among components would change over time, following $p(\boldsymbol{\pi}_t|\boldsymbol{\pi}_{t-1})$.
- The sample points at each time step $t$ are generated independently conditioned on $\boldsymbol{\pi}_t$.
- **Challenge:** how to formulate $p(\boldsymbol{\pi}_t|\boldsymbol{\pi}_{t-1})$?
  - **Solution:** Warped Gaussians.

# Warped Gaussians

- *Gaussian distribution* provides a convenient way to express **correlations** among real-valued variables.
- When $(\mathbf{x}_1, \mathbf{x}_2)$ are jointly Gaussian, then $p(\mathbf{x}_2|\mathbf{x}_1)$ is Gaussian: $\mathbf{x}_2 \sim \mathcal{N}(\mathbf{A}\mathbf{x}_1 + \mathbf{b}, \mathbf{\Lambda})$ for some $\mathbf{A}$, $\mathbf{b}$, and $\mathbf{\Lambda}$.
- To express the dependency between $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$, we can introduce latent Gaussian variables $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$:

$$\boldsymbol{\eta}_2|\boldsymbol{\eta}_1 \sim \mathcal{N}(\mathbf{A}\boldsymbol{\eta}_1 + \mathbf{b}, \mathbf{\Lambda})$$
$$\boldsymbol{\pi}_2 = \mathrm{Softmax}(\boldsymbol{\eta}_2)$$

# GMM with Dynamics on $\boldsymbol{\pi}$ (graph)



# MRF for Image Modeling

- **(Smoothness assumption):** neighboring pixels are similar.
- Formulation:

$$p(\mathbf{x}|\alpha) = \frac{1}{Z(\alpha)} \exp\left(-\frac{\alpha}{2} \sum_{(i,j)\in\mathcal{E}} (x_i - x_j)^2\right)$$

Here, $\mathcal{E}$ is the set of all neighboring pairs.

- **Are there any limitations?**

# Handle Boundaries: Gated MRF

- It is desirable to lift the smoothness constraint across object or region boundaries.
- Formulation:
  - Introduce an indicator $z_{ij}$ for each $(i, j) \in \mathcal{E}$ as a boundary indicator: $z_{ij} \in \mathrm{Ber}(\pi)$.
  - Condition on the boundaries:

$$p(\mathbf{x}|\mathbf{z}, \alpha_0, \alpha_1) \propto \exp\left(-\frac{1}{2}\sum_{(i,j)\in\mathcal{E}} \alpha_{z_{ij}}(x_i - x_j)^2\right)$$

# Handle Textures: Fields of Experts

- Characterize local patterns by modeling the distribution of the responses to *linear filters*.

- *"Responses of linear filters applied to natural images typically exhibit highly kurtotic marginal distributions that resemble a Student-t distribution."* – Roth & Black.

- For a single patch $\mathbf{u}$, we use a **product of experts**:

$$p(\mathbf{u}) \propto \prod_{k=1}^{K} \phi_k(\mathbf{w}_k^T \mathbf{u}; \alpha_k)$$

  with

$$\phi_k(y; \alpha_k) = \left(1 + \frac{y^2}{2}\right)^{-\alpha_k}$$

# Field of Experts (cont'd)

- For an image $\mathbf{x}$:

$$p(\mathbf{x}|\boldsymbol{\alpha}) = \frac{1}{Z(\boldsymbol{\alpha})} \prod_{C \in cC} \prod_{k=1}^{K} \phi_k(\mathbf{w}_k^T \mathbf{x}_C; \alpha_k)$$

Here, $\mathcal{C}$ is the set of all involved patches. Patches can overlap.

- **Gibbs Form**:

$$p(\mathbf{x}|\boldsymbol{\alpha}) = \frac{1}{Z(\boldsymbol{\alpha})} \exp\left(-E(\mathbf{x}; \boldsymbol{\alpha})\right)$$

where

$$E(\mathbf{x}; \boldsymbol{\alpha}) = -\sum_{C \in \mathcal{C}} \sum_{k=1}^{K} \alpha_k \log\left(1 + \frac{1}{2}(\mathbf{w}_k^T \mathbf{x}_C)^2\right)$$

# Low-level Vision based on Image Prior

- **Image denoising**
  - Model:

$$\mathbf{x} \sim p_I(\mathbf{x}); \quad \tilde{\mathbf{x}} \sim \mathcal{N}(\mathbf{x}; \sigma^2 \mathbf{I})$$

  - Solution:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}} \; p(\mathbf{x}|\tilde{\mathbf{x}})$$

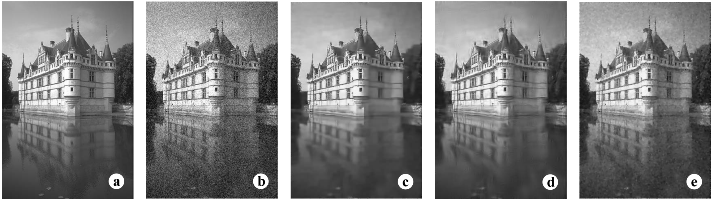


Figure 4. **Denoising results.** **(a)** Original noiseless image. **(b)** Image with additive Gaussian noise ($\sigma = 25$); PSNR = 20.29dB. **(c)** Denoised image using a Field of Experts; PSNR = 28.72dB. **(d)** Denoised image using the approach from [20]; PSNR = 28.90dB. **(e)** Denoised image using standard non-linear diffusion; PSNR = 27.18dB.

# Low-level Vision based on Image Prior (cont'd)

- **Image Inpainting**
  - Model: let $C$ be the missing part.

$$\mathbf{x} \sim p_I(\mathbf{x}); \quad \mathbf{x}(C) = 0$$

  - Solution:

$$\hat{\mathbf{x}}_C = \underset{\mathbf{x}_C}{\operatorname{argmax}} \; p(\mathbf{x}_C | \mathbf{x}_{D \setminus C})$$


Figure 6. **Inpainting with a Field of Experts.** **(a)** Original image with overlaid text. **(b)** Inpainting result from diffusion using the FoE prior. **(c)** Close-up comparison between **a** (left), **b** (middle), and the results from [3] (right).

## Summary

- Understanding/Observations $\Longrightarrow$ Assumptions
- Formulate the basic structures (*i.e.* nodes and edges).
- Choose a specific distribution for each factor.
- Obtain the joint distribution.
- Derive inference/estimation algorithms *(discuss later)*.