Maximum Likelihood for Gaussians on Graphs

Brijnesh J. Jain and Klaus Obermayer

Berlin University of Technology, Germany

May 20, 2011

Outline

Introduction

- **Gaussian distributions**
  - are often used as first approximation of random variables on vectors that cluster around a single mean
  - form basic building block for Gaussian mixture models
  - Gaussian mixtures + maximum likelihood method smoothly approximate arbitrarily shaped densities on vectors
- **Problem:**
  - Data is often represented by structures rather than by vectors
  - Gaussian distributions are undefined on structures
  - How can we approximate distributions on structures
    - that cluster around a single structure?
    - are of arbitrarily shaped form?

Introduction

- **Aim in this talk:**
  Adapt Gaussian distribution to attributed graphs such that parameters can be fitted by the maximum likelihood method in a feasible way.
- **Ansatz:**
  Orbifold framework
  - orbifold $\sim$ quotient of manifold by a finite group action
    $\Rightarrow$ orbifolds are locally like a manifold almost everywhere
    $\Rightarrow$ provides access to techniques from differential geometry
    $\Rightarrow$ induces probability space that regards graphs as events

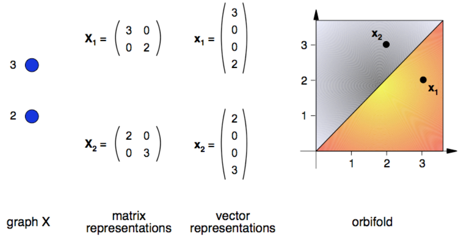
Orbifolds

- Regard graph X as equivalence class $[\mathbf{X}]$ of all matrices $\mathbf{X} \in \mathcal{X}$ that represent graph X
- Orbifold $\mathcal{X}_{\mathcal{G}}$ is set of all equivalence classes of matrices $\mathbf{X} \in \mathcal{X}$

  $\Rightarrow$ graph X is point in orbifold $\mathcal{X}_{\mathcal{G}}$

  $\Rightarrow$ orbifold is locally homeomorphic to Euclidean space almost everywhere
- **Note:**

  idea can be generalized to graphs
  - of arbitrary but bounded size
  - with arbitrary attributes



Orbifolds - Metric Structures

- **Motivation:**

  $\Rightarrow$ Gaussian is based on Euclidean metric

  $\Rightarrow$ construct metric on graphs related to Euclidean metric

- **Intrinsic metric:**

$$d(X, Y) = \min_{\mathbf{x} \in X, \mathbf{y} \in Y} \|\mathbf{x} - \mathbf{y}\|$$

- **Optimal alignment:**
  a pair $(\mathbf{x}, \mathbf{y}) \in X \times Y$ with

$$\|\mathbf{x} - \mathbf{y}\| = d(X, Y)$$

- **Note:**
  intrinsic metric is
  - widely used metric
  - NP hard

X

3 2

$\mathbf{X_1} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$ $\quad \mathbf{x_1} = \begin{pmatrix} 3 \\ 0 \\ 0 \\ 2 \end{pmatrix}$

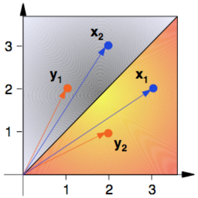
$\mathbf{X_2} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ $\quad \mathbf{x_2} = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 3 \end{pmatrix}$

Y

1 2

$\mathbf{Y_1} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ $\quad \mathbf{y_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 2 \end{pmatrix}$

$\mathbf{Y_2} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ $\quad \mathbf{y_2} = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 1 \end{pmatrix}$



$\mathbf{d(X, Y) = \min \|x_i - y_j\| = 2}$

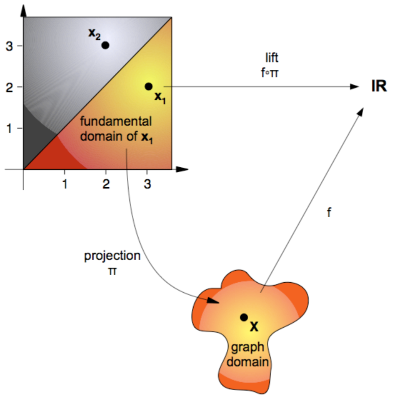
$\|x_1 - y_1\|^2 = (3\text{-}1)^2 + (2\text{-}2)^2 = 4$
$\|x_1 - y_2\|^2 = (3\text{-}2)^2 + (2\text{-}1)^2 = 2$
$\|x_2 - y_1\|^2 = (2\text{-}1)^2 + (3\text{-}2)^2 = 2$
$\|x_2 - y_2\|^2 = (2\text{-}2)^2 + (3\text{-}1)^2 = 4$

Orbifolds - Fundamental Domains

- **Dirichlet fundamental domain** of $\mathbf{x} \in \mathcal{X}$:

$$\mathcal{D}_{\mathbf{x}} = \{\mathbf{y} \in \mathcal{X} \;:\; \|\mathbf{x} - \mathbf{y}\| = d([\mathbf{x}], [\mathbf{y}])\} \;.$$

- **Fundamental observation:**
  Studying distributions f on graphs that cluster around center X can be reduced to studying lifts $\tilde{f}$ of f on a fundamental domain $\mathcal{D}_{\mathbf{x}}$ of an arbitrary vector representation $\mathbf{x} \in X$.



Quotient Gaussians

Quotient Gaussian distribution on $\mathcal{X}_{\mathcal{G}}$

$$f(X) = \frac{1}{a(C,\sigma)} \cdot \exp\left(-\frac{d(X,C)^2}{2\sigma^2}\right),$$

where

- C is the center graph
- σ is the width
- $a(C,\sigma)$ is the height that scales f to a density

$$a(C,\sigma) = \int_{X_G} \phi(X|C,\sigma)\lambda_{\mathcal{G}}(dX).$$

Quotient Gaussian

- quotient Gaussian f with center C can be lifted to Euclidean space $\mathcal{X}$ with centers $\mathbf{c}$.

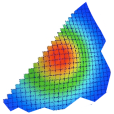
  $\Rightarrow$ pointwise maximum of a set $\mathcal{N}_{C,a,\sigma}$ of Gaussians on $\mathcal{X}$ with identical σ, but distinct centers $\mathbf{c} \in C$.

- choose arbitrary Gaussian on $\mathcal{X}$ from $\mathcal{N}_{C,a,\sigma}$ with center $\mathbf{c} \in C$

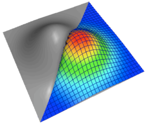
  $\Rightarrow$ quotient Gaussian f can be viewed as truncated Gaussian $\tilde{f}^t$ on $\mathcal{D}_{\mathbf{c}}$.

  $\Rightarrow$ height $a(\mathbf{c}, \sigma)$ can be viewed as probability of being in $\mathcal{D}_{\mathbf{c}}$.

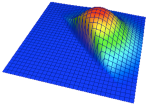
quotient Gaussian f in graph domain



liftet quotient Gaussian $\tilde{f}$ in Euclidean space



truncated Gaussian $\tilde{f}^t$ in Euclidean space



Quotient Gaussian

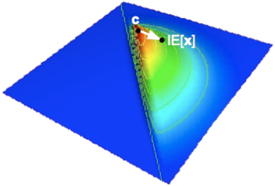
- **Central moments** $\mathbb{E}[\mathbf{x}]$ and $\mathbb{V}[\mathbf{x}]$:

$$\begin{aligned} \mathbb{E}[\mathbf{x}] &= \mathbf{c} + \delta_{\mathbb{E}}(\mathbf{c}, \sigma) \\ \mathbb{V}[\mathbf{x}] &= \sigma^2 + \delta_{\mathbb{V}}(\mathbf{c}, \sigma) \end{aligned}$$

  $\Rightarrow$ center $\mathbf{c}$ is not the expectation $\mathbb{E}[\mathbf{x}]$

  $\Rightarrow$ sq. width σ^2 is not the variance $\mathbb{V}[\mathbf{x}]$

- **Goal:**
  make inferences on $\mathbf{c}$ and σ rather than $\mathbb{E}[\mathbf{x}]$ and $\mathbb{V}[\mathbf{x}]$



Maximum Likelihood

- **Given:** Sample
$$\mathcal{S} = \{X_1, \ldots, X_N\} \subseteq \mathcal{X}_{\mathcal{G}}$$
of iid graphs drawn from some quotient Gaussian f_{C_*, σ_*^2}
- **Goal:** Estimate true but unknown parameters $\Theta_* = \left(C_*, \sigma_*^2\right)$
- **Approach:** Apply maximum likelihood method as follows:
  1. choose arbitrary $\mathbf{c} \in C$
  2. optimally align $\mathbf{x}_i \in X_i$ against $\mathbf{c}$ (graph matching)
  3. maximize log-likelihood of truncated Gaussian $\tilde{f}^t_{\mathbf{c},\sigma^2}$ on $\mathcal{D}_{\mathbf{c}}$

$$\tilde{\ell}(\mathbf{c}, \sigma^2) = \sum_{i=1}^{N} \ln \tilde{f}^t_{\mathbf{c},\sigma^2}\left(\mathbf{x}_i\right).$$

Maximizing the Log-Likelihood

- Setting the gradients of the log-likelihood to zero and solving yields

$$\mathbf{c} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i - \delta_{\mathbb{E}}(\mathbf{c}, \sigma) \tag{1}$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}_i - \mathbf{c}\|^2 - \delta_{\mathbb{V}}(\mathbf{c}, \sigma). \tag{2}$$

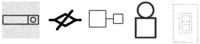
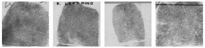
$\Rightarrow$ ML estimate of $\mathbf{c}$, σ^2 is estimate of $\mathbb{E}[\mathbf{x}]$, $\mathbb{V}[\mathbf{x}]$ plus adjustment

- Adjustments $\delta_{\mathbb{E}}(\mathbf{c}, \sigma)$ and $\delta_{\mathbb{V}}(\mathbf{c}, \sigma)$:
  - can be approximated using Monte Carlo integration
  - here: ignore adjustments for computational reasons
    $\Rightarrow$ estimate center by algorithm for sample mean of graphs

Experiments

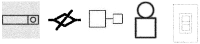
- **Aim:** Assess performance of ML method for quotient Gaussians in conjunction with Bayes classifier
- **Data:** Benchmark data of the IAM graph database repository

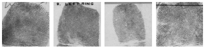
data set	graphs	classes	avg(nodes)	max(nodes)
letter	2250	15	4.7	8
grec	1100	22	11.5	24
fingerprint	2800	4	8.3	26

Classification Results

	Letter	GREC	Fingerprint
kNN	82.0	96.8	80.0
SK+SVM	92.9	92.4	83.1
LE+SVM	92.5	96.8	82.8
LGQ	81.7	86.9	79.9
LGQ2.1	86.3	92.6	81.6
RS-LGQ	87.3	97.4	84.1
ml+bayes	81.2	89.9	79.2





Conclusion

- Extension of Gaussian distribution to orbifolds
- Extension of ML method for (mixtures of) quotient Gaussians
- simulations indicate that approximation works
- orbifold framework turns out to be a versatile alternative for bridging the gap between statistical and structural methods
- Future work: Extend ML method for other distributions