Multi-view Warped Mixtures for Shape Correspondence Analysis

Hernán F. García and Mauricio A. Álvarez

February 18, 2019

Abstract

Probabilistic approach for the correspondence problem based on the Single View WMM[3].

1 Multiview Warped Mixture Models

We can use the single view problem of the warped mixture model from [3] in which they warp a latent mixture of Gaussians into nonparametric cluster shapes. The low-dimensional latent mixture model summarizes the properties of the high-dimensional density manifolds describing the data.

Our idea is to introduce a model which warps a multiview latent mixture of Gaussians (possibly MRD) to produce nonparametric cluster shapes. [1]

1.1 The model

Let as define a multi-view data set as $\mathcal{Y} = \{\mathbf{Y}^v\}_{v=1}^{V}$, where each view is defined as $\mathbf{Y}^v \in \mathbb{R}^{N_v \times D_v}$. This leads to the likelihood Given mixture assignments, likelihood has only two parts: GP-LVM and GMM

$$p\left(\mathbf{Y}^{\mathcal{V}}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}\right) = \prod_{v=1}^{V} p(\mathbf{Y}^v|\mathbf{X}, \boldsymbol{\theta}) \times \prod_{i} \sum_{c=1}^{\infty} \lambda_c \mathcal{N}\left(\mathbf{x}_i|\boldsymbol{\mu}_c, \mathbf{R}_c^{-1}\right), \quad \mathbf{x}_i \in \mathbf{Z}_c \tag{1}$$

Based on the iWMM (see [3]) our generative model generates multiple observations $\mathbf{Y}^{\mathcal{V}}$ according to the following generative process:

1. Draw mixture weights $\boldsymbol{\lambda} \sim \text{GEM}(\eta)$
2. For each cluster $c = 1, \cdots, \infty$
   (a) Draw precision $\mathbf{R}_c \sim \mathcal{W}(\mathbf{S}^{-1}, v)$
   (b) Draw mean $\boldsymbol{\mu}_c \sim \mathcal{N}(\mathbf{u}, (r\mathbf{R}_c)^{-1})$
3. For each view $v = 1, \cdots, \mathcal{V}$
   (a) For each observation $n = 1, \cdots, N_v$
      i. Draw latent assignment $z_{nv} \sim \text{Mult}(\boldsymbol{\lambda})$
      ii. Draw latent coordinates $\mathbf{x}_{nv} \sim \mathcal{N}(\boldsymbol{\mu}_{z_{nv}}, \mathbf{R}_{z_{nv}}^{-1})$
4. For each view $v = 1, \cdots, \mathcal{V}$
   (a) For each observed dimension $d = 1, \cdots, D_v$
      i. Draw function $\mathbf{f}_d^v \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}^v)$
5. For each view $v = 1, \cdots, \mathcal{V}$
   (a) For each observed dimension $d = 1, \cdots, D_v$
      i. Draw projection variable $w_d^v \sim \mathcal{N}\left(0, \rho_d^v\right)$

[1] The possibly low-dimensional latent mixture model allows us to summarize the properties of the high-dimensional clusters (or density manifolds) describing the data. The number of manifolds, as well as the shape and dimension of each manifold is automatically inferred.

ii. For each observation $n = 1, \cdots, N_v$

A. Draw feature $y_{nd}^v \sim \mathcal{N}\left(w_d^v f_d^v(\mathbf{x}_{nv}), \beta^{-1}\right)$

We define the dimensionalities of our variables as:

- K: real number of clusters
- Q: dimensionality of the Latent Space
- D_v: dimensionality of the input data in the v-th view
- $\boldsymbol{\lambda} \in \mathbb{R}^{K\times 1}$
- $\mathbf{R}_c \in \mathbb{R}^{Q\times Q}$
- $\boldsymbol{\mu}_c \in \mathbb{R}^{Q\times 1}$

1.2 Latent Multi-view Warped Mixture Model

Our model is set as a multi-view Gaussian Process Latent Variable model as [2]. First we assume that observations are generated by mapping the latent coordinates through a set of smooth functions, over which Gaussian process priors are placed. Under the GPLVM, the probability of observations given the latent coordinates, integrating out the mapping functions, is defined as

$$p\left(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}\right) = \prod_{v=1}^{V} p\left(\mathbf{Y}^v|\mathbf{X}^v, \boldsymbol{\theta}^v\right) \tag{2}$$

$$= \prod_{v=1}^{V}\prod_{d=1}^{D_v} p\left(\mathbf{y}_d^v|\mathbf{X}^v, \boldsymbol{\theta}^v\right), \tag{3}$$

where $\mathbf{y}_d^v$ represents the dth column of $\mathbf{Y}^v$ and

$$p\left(\mathbf{y}_d^v|\mathbf{X}^v, \boldsymbol{\theta}^v\right) = \mathcal{N}\left(\mathbf{y}_d^v|\mathbf{0}, \beta^{-1}\mathbf{I} + w_d^2\mathbf{K}^v\right). \tag{4}$$

Our multi-view iWMM assumes that the latent coordinates (per View) are generated from a Dirichlet process mixture model. In particular, we use the following infinite Gaussian mixture model,

$$p(\mathbf{x}^v|\lambda_c, \boldsymbol{\mu}_c, \mathbf{R}_c) = \sum_{c=1}^{\infty} \lambda_c \mathcal{N}\left(\mathbf{x}^v|\boldsymbol{\mu}_c, \mathbf{R}_c^{-1}\right), \tag{5}$$

where λ_c, $\boldsymbol{\mu}_c$ and $\mathbf{R}_c$ are the mixture weight, mean, and precision matrix of the cth mixture component.

As in the iWMM [3], we place Gaussian-Wishart priors on the Gaussian parameters $\{\boldsymbol{\mu}_c, \mathbf{R}_c\}$.

$$p\left(\boldsymbol{\mu}_c, \mathbf{R}_c\right) = \mathcal{N}\left(\boldsymbol{\mu}_c|\mathbf{u}, (r\mathbf{R}_c)^{-1}\right)\mathcal{W}\left(\mathbf{R}_c|\mathbf{S}^{-1}, \nu\right) \tag{6}$$

where $\mathbf{u}$ is the mean of $\boldsymbol{\mu}_c$, r is the relative precision of $\boldsymbol{\mu}_c$, $\mathbf{S}^{-1}$ is the scale matrix for $\mathbf{R}_c$, and ν is the number of degrees of freedom for $\mathbf{R}_c$.

By using conjugate Gaussian-Wishart priors for the parameters of the Gaussian mixture components, we can analytically integrate out those parameters, given the assignments of points to components. Let z_n^v be the latent assignment of the nth object in the vth view. The probability of latent coordinates $\mathbf{X}$ given latent assignments $\mathbf{Z}^v = (z_1, \ldots, z_{N_v})$ is obtained by integrating out the Gaussian parameters $\{\boldsymbol{\mu}_c, \mathbf{R}_c\}$ as follows:

$$p\left(\mathbf{X}|\mathbf{Z}, \mathbf{S}, \nu, r\right) = \prod_{c=1}^{\infty} \pi^{-\frac{\sum_v N_{vc}Q}{2}} \frac{r^{Q/2}|\mathbf{S}|^{\nu/2}}{r_c^{Q/2}|\mathbf{S}_c|^{\nu_c/2}} \prod_{q=1}^{Q} \frac{\Gamma\left(\frac{\nu_c+1-q}{2}\right)}{\Gamma\left(\frac{\nu+1-q}{2}\right)}, \tag{7}$$

where N_{vc} is the number of objects in the vth view assigned to the cth cluster, $\Gamma(\cdot)$ is the Gamma function and

$$r_c = r + \sum_v N_{vc}, \quad \nu_c = \nu + \sum_v N_{vc},$$
$$\mathbf{u}_c = \frac{r\mathbf{u} + \sum_{v=1}^V \sum_{n:z_{nv}=c} \mathbf{x}_{nv}}{r + \sum_v N_{vc}},$$
$$\mathbf{S}_c = \mathbf{S} + \sum_{v=1}^V \sum_{n:z_{nv}=c} \mathbf{x}_{nv}\mathbf{x}_{nv}^\top + r\mathbf{u}\mathbf{u}^\top - r_c\mathbf{u}_c\mathbf{u}_c^\top,$$

are the posterior Gaussian-Wishart parameters of the cth component. As in [3], we use a Dirichlet process with concentration parameter η for infinite mixture modeling in the latent space.

The probability of $\mathbf{Z}$ is given as follows

$$p(\mathbf{Z}|\eta) = \prod_{v=1}^V \frac{\eta^C \prod_{c=1}^C (N_{vc}-1)!}{\eta(\eta+1)\cdots(\eta+N_v-1)}, \tag{8}$$

where C is the number of components for which $N_{vc} > 0$. The joint distribution is given by

$$p(\mathbf{Y}, \mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}, \nu, \mathbf{u}, r\eta) = \prod_{v=1}^V p(\mathbf{Y}^v|\mathbf{X}^v, \boldsymbol{\theta}^v)\, p(\mathbf{X}^v|\mathbf{Z}_v, \mathbf{S}_v, \nu, \mathbf{u}, r)\, p(\mathbf{Z}_v|\eta). \tag{9}$$

1.3 Inference

We infer the posterior distribution of the latent coordinates $\mathbf{X} = \{\mathbf{X}^v\}_v^V$ and cluster assignments $\mathbf{Z}^v$ using Markov chain Monte Carlo (MCMC). In particular, we alternate collapsed Gibbs sampling of $\mathbf{Z}$, and hybrid Monte Carlo sampling of $\mathbf{X}$. Given $\mathbf{X}$, we can efficiently sample $\mathbf{Z}^v$ using collapsed Gibbs sampling, integrating out the mixture parameters. Given $\mathbf{Z}^v$, we can calculate the gradient of the unnormalized posterior distribution of $\mathbf{X}$, integrating over warping functions. This gradient allows us to sample $\mathbf{X}$ using hybrid Monte Carlo.

First, we explain collapsed Gibbs sampling for Z. Given a sample of X, $p(\mathbf{Z}|\mathbf{X}, \mathbf{S}, v, \mathbf{u}, r, \eta)$ does not depend on Y. This lets resample cluster assignments, integrating out the iGMM likelihood in close form. Given the current state of all but one latent component zn, a new value for zn is sampled from the following probability:

$$p\left(\mathbf{z}_{nv} = c|\mathbf{X}, \mathbf{Z}_{\setminus nv}, \mathbf{S}, v, \mathbf{u}, r, \eta\right) \propto \begin{cases} N_{vc\setminus nv} \cdot p(\mathbf{x}_{nv}|\mathbf{X}_c, \mathbf{S}, v, \mathbf{u}, r) & \text{existing components} \\ \eta \cdot p(\mathbf{x}_{nv}|\mathbf{S}, v, \mathbf{u}, r) & \text{a new cluster} \end{cases}$$

where $\mathbf{X}_c = \{\mathbf{x}_{nv}|\mathbf{z}_{nv} = c\}$ is the set of latent coordinates assigned to the c^{th} component, and $\setminus nv$ represents the value or set when excluding the n-th observation in the v-th view. We can analytically calculate $p(\mathbf{x}_{nv}|\mathbf{X}_c, \mathbf{S}, v, \mathbf{u}, r)$ as follows:

$$p(\mathbf{x}_{nv}|\mathbf{X}_c, \mathbf{S}, v, \mathbf{u}, r) = \pi^{-\frac{N_{vc\setminus n}Q}{2}} \frac{r^{Q/2}|\mathbf{S}|^{\nu/2}}{r_c^{Q/2}|\mathbf{S}_c|^{\nu_c/2}} \prod_{q=1}^Q \frac{\Gamma\left(\frac{\nu_c+1-q}{2}\right)}{\Gamma\left(\frac{\nu+1-q}{2}\right)}$$

2 Results

3 Clustering performance on real datasets

3.1 Comparison with linear approaches

First, we test the performance of our approaches (both NL-UCM and MV-WMM) regarding the adjusted Rand index (we report both average and standard deviation), to quantify the similarity between the

Table 1: Average Rand index for evaluating clustering performance.

Approach	Database					
	Wine	2-curve	3-semi	2-circle	Pinwheel	Vowel
MV-WMM($Q = 2$)	0.68 ± 0.03	0.83 ± 0.02	0.83 ± 0.01	0.88 ± 0.02	0.87 ± 0.02	0.65 ± 0.01
MV-WMM($Q = D$)	0.85 ± 0.02	0.83 ± 0.02	0.83 ± 0.01	0.88 ± 0.02	0.87 ± 0.02	0.73 ± 0.02

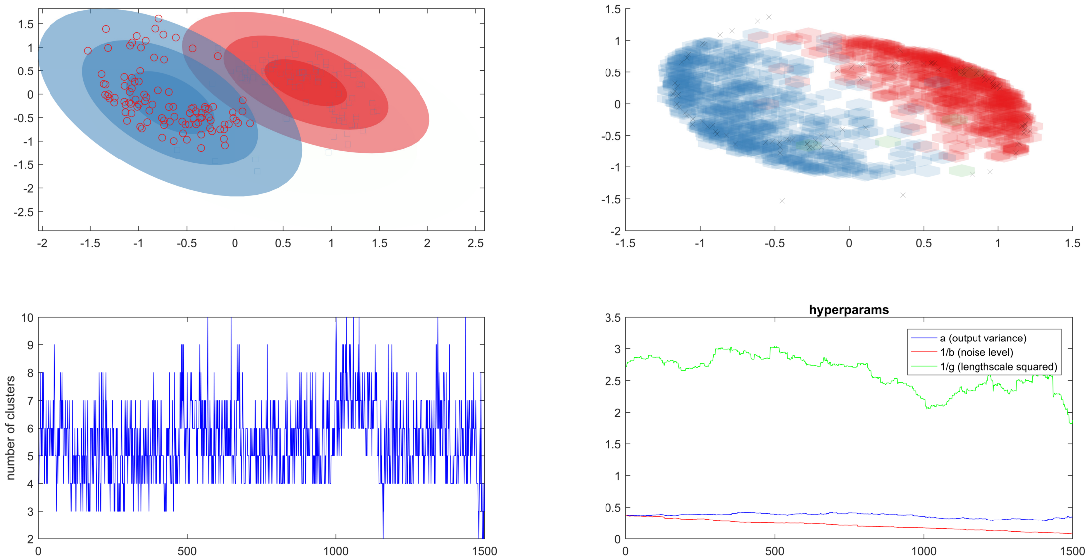

Figure 1: Experimental results for the Pinwheel dataset. Current mixture parameters, along with the latent positions (Top left). Original data and predicted assignments (Top right).

inferred clusters [1] and the true labels. For comparison, we use unsupervised clustering matching (UCM) [1], k-means (KM), and convex kernelized sorting (CKS) [**?**]. Table 2 shows that our approaches outperforms the state-of-the-art methods for unsupervised clustering for the three databases. The results also show that by mapping the observed data through random feature expansions, the model can handle real-world datasets with better performance than linear approaches as in the case of NL-UCM (i.e., 0.17 for the MNIST dataset against 0.085 obtained from the UCM method).

Table 2: Adjusted Rand index of the proposed method against the state-of-the-art methods for unsupervised clustering.

	Approach				
Database	UCM	KM	KM-CKS	**NL-UCM**	**MV-WMM**
Iris	0.383 ± 0.189	0.224 ± 0.0910	0.254 ± 0.154	$\mathbf{0.546 \pm 0.080}$	$\mathbf{0.498 \pm 0.001}$
Glass	0.160 ± 0.020	0.050 ± 0.008	0.052 ± 0.011	$\mathbf{0.378 \pm 0.045}$	$\mathbf{0.384 \pm 0.003}$
MNIST	0.085 ± 0.016	0.030 ± 0.007	0.037 ± 0.008	$\mathbf{0.167 \pm 0.013}$	$\mathbf{0.133 \pm 0.004}$

3.2 Non-rigid 3D shape datasets

References

[1] Tomoharu Iwata, Tsutomu Hirao, and Naonori Ueda. Probabilistic latent variable models for unsupervised many-to-many object matching. *Information Processing and Management*, 52(4):682 – 697, 2016.

[2] Neil Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *In NIPS*, page 2004, 2003.

[3] Zoubin Ghahramani Tomoharu Iwata, David Duvenaud. Warped mixtures for nonparametric cluster shapes. In *29th Conference on Uncertainty in Artificial Intelligence*, pages 311–319, 2013.

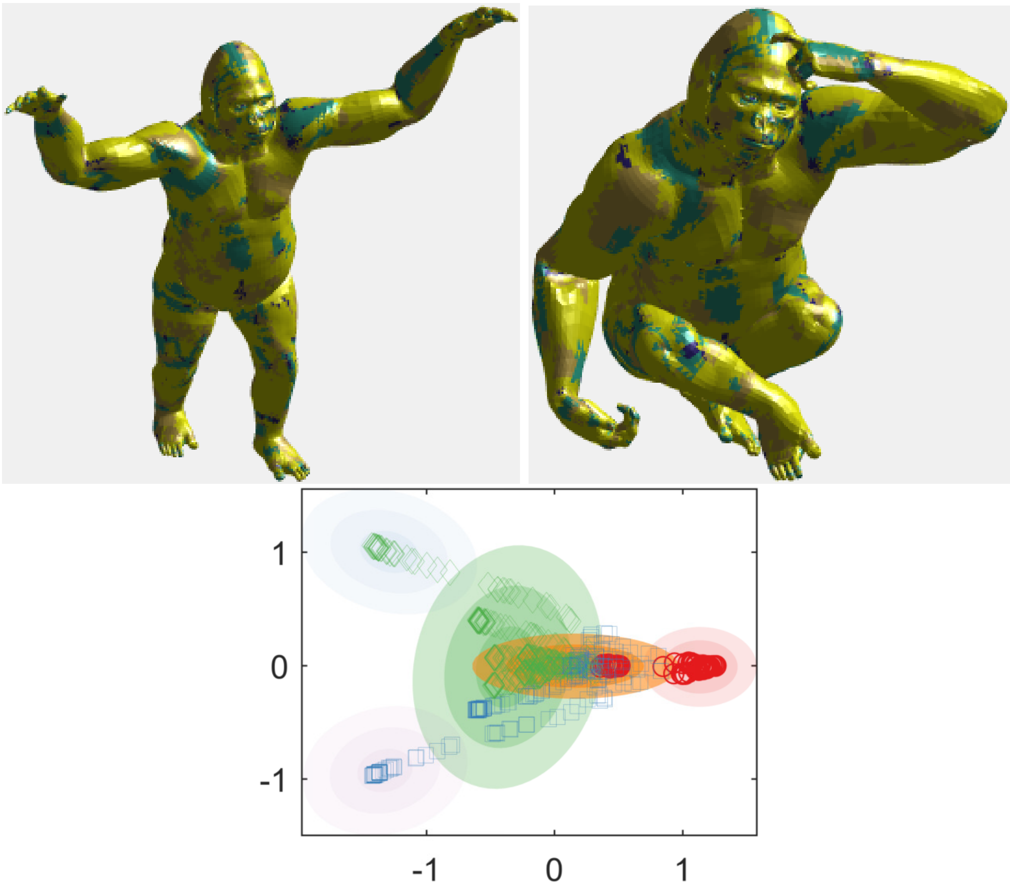

Figure 2: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.5857