Code for learning a manifold model using the non–parametric MFA: short documentation

1 Usage

Call the model learning procedure using the following syntax:

spl = MFA_DP($\boldsymbol{X}$,param)

where $\boldsymbol{X}$ is a $p \times n$ data matrix of p–dimensional vectors arranged in n columns, and spl and param are output and input structures, respectively. The structure contents are detailed in the following table. Many of the structure members below are cell arrays, where each cell corresponds to one cluster. There are T clusters. The maximum dimensionality for cluster t is K_t. We usually set all K_t to a constant value K and use $T = K = 50$ or some adequately large number.

Inputs	Outputs
param.k: vector of maximum dimensionalities	spl.A: cell array of posterior factor loadings $\tilde{\boldsymbol{A}}_t$
param.cet: initial value for the Dirichlet process concentration parameter (default=1)	spl.z: cell array of sparse binary vectors $\boldsymbol{z}_t$ multiplying the factor scores
param.maxit: Total # of MCMC iterations (burn–in+collection)	spl.w: cell array of posterior factor weights $\boldsymbol{\Delta}_t$
param.num: # of MCMC collection iterations	spl.mu: cell array of cluster means $\boldsymbol{\mu}_t$
	spl.Phi: cell array with vectors of noise variances $[\alpha_t \cdots \alpha_t]$
	spl.pai: cell array with vectors of Bernoulli probabilities $\boldsymbol{\pi}_t$
	spl.H: $T \times n$ matrix of cluster occupancies
	spl.qai: vector $\boldsymbol{\lambda}$ of cluster weights
	spl.S1: cell array with vectors $\boldsymbol{\xi}_t = \frac{1}{M} \sum_{m=1}^{M} \left(\sum_{i:t(i)=t} \hat{\boldsymbol{w}}_t / \sum_{i:t(i)=t} 1 \right)$ (see section II-D in the paper)
	spl.S2: cell array with matrices $\boldsymbol{\Lambda}_t + \boldsymbol{\xi}_t \boldsymbol{\xi}_t^\top$ (see section II-D in the paper)
	spl.X_hat: matrix of size $p \times n$ containing the reconstructed training points, according to the model
	mcmc.Z, mcmc.W: cell arrays containing the Gibbs samples of $\boldsymbol{z}_t$ and $\boldsymbol{\Delta}_t$ for each t

Note that T=size(param.k,1) and that K_t=param.k(t).

2   Interpretation of the output

2.1   Clustering

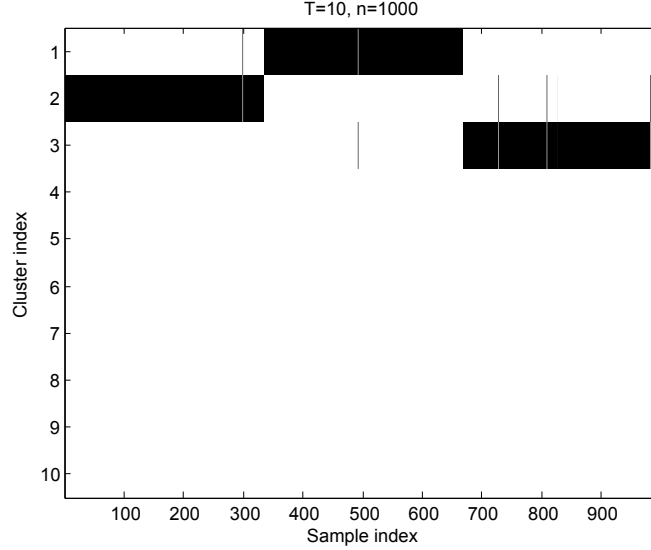The matrix spl.H contains cluster membership information for each training point, as illustrated in Figure 1.

Figure 1: Example for matrix spl.H, which contains the cluster index for each training point. It is possible to identify three clusters in this example, where we have set $T = 10$ and $n = 1000$.

2.2   Intrinsic dimension estimate

Strictly speaking, the number of nonzero coordinates of $\boldsymbol{z}_t$, *i.e.*, $\|\boldsymbol{z}_t\|_0$, yields an estimate of the true intrinsic dimensionality of the manifold, at least in the local region covered by cluster t. However, it is more informative to take into account the posterior weights $\boldsymbol{\Delta}_t$ multiplying each factor - this allows us to gauge the relative importance of each factor that has non-zero entry in $\boldsymbol{z}_t$. Thus, in order to obtain an estimate of the intrinsic dimensionality, we suggest thresholding

abs(spl.w{t} .* spl.z{t})

at an appropriate level (for instance, 10% of the maximum component of abs(spl.w{t} .* spl.z{t})), counting the number of components above the threshold, and then averaging over all t. Note that abs(spl.w{t} .* spl.z{t}) corresponds to $|\boldsymbol{\Delta}_t \circ \boldsymbol{z}_t|$. As explained in section II-D in the paper, this adequately accounts for posterior information.

A more complete picture can be obtained by looking at the posterior factor structure of each Gaussian mixture component (see section II-D in the paper). This can be obtained as follows:

[A1,mu1] = get_posterior_mfa(spl)

where A1 is a cell structure containing the posterior factor loadings and mu1 is another cell structure of posterior means. The relevant factors for cluster t correspond to nonzero columns in the $D \times K$ matrix

2

A1$\{t\}$. This is illustrated in Figure 2, where we return to the example of the previous subsection. There are three clusters, each having one significant factor.
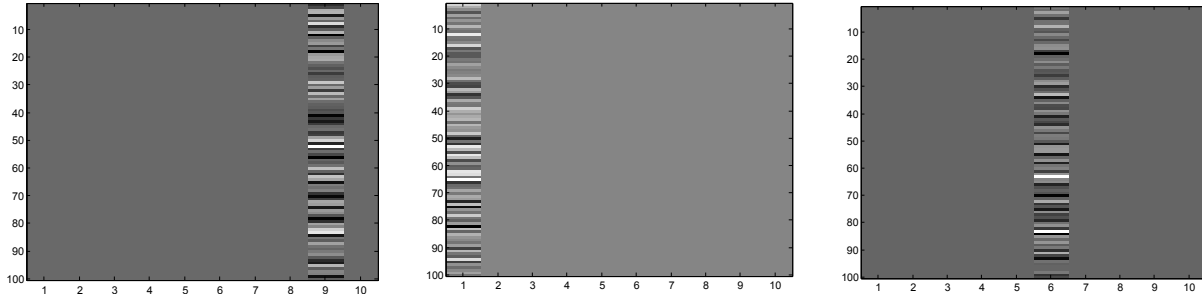

Figure 2: Example for matrices A1$\{1\}$–A$\{3\}$, corresponding to the three clusters identified in Figure 1. There is one significant factor for each cluster, suggesting an intrinsic dimension of one.

3   Drawing from the MFA model

One of the key advantages of the MFA statistical approach is that we obtain a generative model and therefore can generate new data. The following command returns m new vectors drawn from the learned mixture model, in the columns of the $p \times m$ matrix $\boldsymbol{Z}$:

$$\boldsymbol{Z} = \text{draw\_mfa}(\text{spl}, m)$$

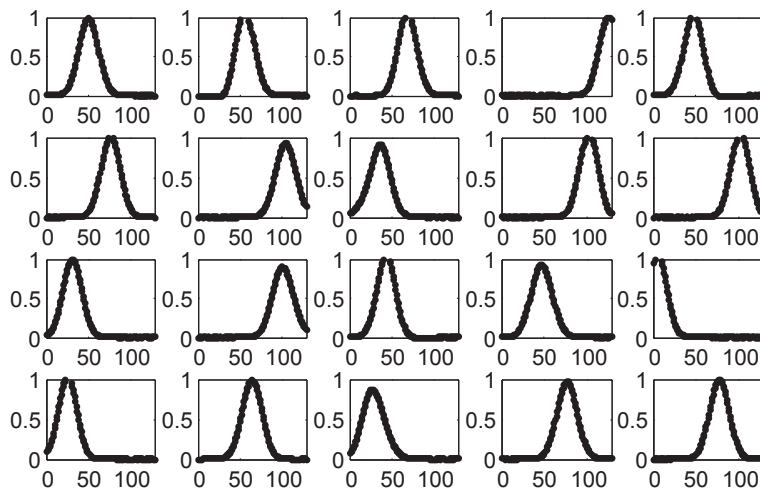As an example, using the MFA learned from shifted Gaussian data, Figure 3 shows new shifted Gaussians drawn from the generative model.

Figure 3: Shifted Gaussians drawn from the learned MFA model.

3