Drifting Linear Dynamics

Tapani Raiko[1,2] **Alexander Ilin**[1,2] **Natalia Korsakova**[1,3] **Erkki Oja**[1] **Harri Valpola**[1,2]
[1]Aalto University, [2]ZenRobotics Ltd, [3]University of Eastern Finland

Abstract

We study a model where linear dynamics depend on a continuous valued hidden state as opposed to switching linear dynamics where the hidden state is discrete. We also provide efficient learning rules.

Consider a video sequence where translation can be described with linear dynamics. Different directions and speeds of movement each require a different linear mapping between the pixels of successive images. These linear mappings are not arbitrary, but they form a well-formed subset of all possible linear mappings. We propose a model that can find a subset (or a subspace) of such transformations from data.

1 Model Definition

Let us define a temporal model for a data set of T samples of K-dimensional observed vectors $\mathbf{x}_t$ and such that the linear mapping $\mathbf{B}_t$ is not a constant:

$$\mathbf{x}_{t+1} = \mathbf{B}_t \mathbf{x}_t + \boldsymbol{\epsilon}_t \tag{1}$$

$$p(\epsilon_{kt}) = \mathcal{N}\left(\epsilon_{kt}; 0, \sigma_x^2\right), \tag{2}$$

where $\mathbf{B}$ is a $K \times K$ auxiliary matrix and σ_x^2 parameterise the noise variance. Now let us collect the columns of the matrix $\mathbf{B}_t$ into a K^2-dimensional vector $\mathbf{b}_t = \mathbf{B}_t(:)$ and model it with a PCA-kind of a model:

$$\mathbf{b}_t = \mathbf{A}\mathbf{s}_t, \tag{3}$$

where $\mathbf{A}$ is a $K^2 \times L$ parameter matrix $\mathbf{s}_t$ are L-dimensional latent vectors. Instead of an explicit bias term, we assume that the last component of the state vectors $\mathbf{x}_t$ and the latent variables $\mathbf{s}_t$ are constants 1.

We include a temporal model for $\mathbf{s}_t$, that is:

$$\mathbf{s}_{t+1} = \mathbf{D}\mathbf{s}_t + \boldsymbol{\delta}_t \tag{4}$$

$$p(\delta_{lt}) = \mathcal{N}\left(\delta_{lt}; 0, \sigma_s^2\right), \tag{5}$$

Appearing in Proceedings of the 13^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna Resort, Sardinia, Italy. Volume 9 of JMLR: W&CP 9.

where $\mathbf{D}$ is an $L \times L$ parameter matrix.

Since the matrix $\mathbf{A}$ is so large ($K^2 \times L$), it is important to regularize it. We include a Gaussian prior for each element of $\mathbf{A}$ with a zero mean and a variance σ_a^2.

1.1 Learning

The learning algorithm is such that there is a gradient or conjugate gradient update rule for $\mathbf{S} = (\mathbf{s}_1 \dots \mathbf{s}_T)$, and for each proposed update of $\mathbf{S}$, we solve $\mathbf{A}$ and $\mathbf{D}$ before checking how good the proposed step was. Preliminary experiments indicate that this is faster than joint optimisation with the conjugate gradient.

Since solving $\mathbf{A}$ is nontrivial, let us use an index $m = 1 \dots M$ that enumerates all tuples (l, k) and thus $M = LK$. Now we can define $z_{mt} = s_{lt} x_{kt}$ and rearrange $\mathbf{A}'$ to be $\mathbf{A}$ as a $K \times KL$ matrix. The model equation (1) becomes $\mathbf{x}_{t+1} = \mathbf{A}'\mathbf{z}_t + \boldsymbol{\epsilon}_t$, which can be solved like linear regression.

2 Related Work

Previous works with changing dynamics are mostly about switching between a discrete number of possible dynamics. Even a paper on drifting dynamics [1] first finds a discrete set and only then models the drift from one state to the other.

The proposed model is closely related to Gated Boltzmann machines [2] where there is also a weight tensor that connects three different vectors. They can also be used to model image transformations.

References

[1] J. Kohlmorgen, K. R. Mueller, and K. Pawelzik. Analysis of drifting dynamics with neural network hidden Markov models. In *NIPS*, 1998.

[2] R. Memisevic and G. Hinton. Unsupervised learning of image transformations. Technical report, Department of Computer Science, University of Toronto, Toronto, Canada, 2006.