Gaussian Mixtures with Missing Data: an Efficient EM Training Algorithm

Olivier Delalleau, Aaron Courville and Yoshua Bengio

Dept. IRO, Université de Montréal
C.P. 6128, Montreal, Qc, H3C 3J7, Canada
{*delallea,courvila,bengioy*}*@iro.umontreal.ca*
http://www.iro.umontreal.ca/∼lisa

In data-mining applications, we are frequently faced with a large fraction of missing entries in the data matrix, which is problematic for most discriminant machine learning algorithms. A solution that we explore here is the use of a generative model (a mixture of Gaussians with full covariances) to learn the underlying data distribution and replace missing values by their conditional expectation given the observed variables. Since training a Gaussian mixture with many different patterns of missing values can be computationally very expensive, we introduce a spanning-tree based algorithm that significantly speeds up training in these conditions. Such mixtures of Gaussians can be applied directly to supervised problems (Ghahramani and Jordan, 1994), but we observe that using them for missing value imputation before applying a separate discriminant learning algorithm yields better results.

Our contributions are two-fold:

1. We explain why the basic EM training algorithm is not practical in large-dimensional applications in the presence of missing values, and we propose a novel training algorithm that significantly speeds up training by EM. The algorithm we propose relies on the idea to re-use the computations performed on one training sample as a basis for the next sample, in order to obtain the quantities required by the EM update equations. We show how these computations can be minimized by ordering samples in such a way that two consecutive samples have similar "missing patterns", i.e. share missing values for similar variables. On 28x28 images with random squares of 5x5 pixels being forced to missing values, we obtain a speed-up on the order of 8 compared to standard EM training.

2. We show, both by visual inspection on image data (figure 1) and by feeding the imputed values to another classification algorithm (figure 2), how a mixture of Gaussians can model the data distribution so as to provide a valuable tool for missing values imputation.

References

Ghahramani, Z. and Jordan, M. I. (1994). Supervised learning from incomplete data via an EM approach. In Cowan, D., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems 6*, San Mateo, CA. Morgan Kaufmann.

Topic: learning algorithms
Preference: poster

Figure 1: Imputation of missing values in images (on the left, two digits with missing values in grey, and on the right the corresponding imputation predicted by the mixture of Gaussians).

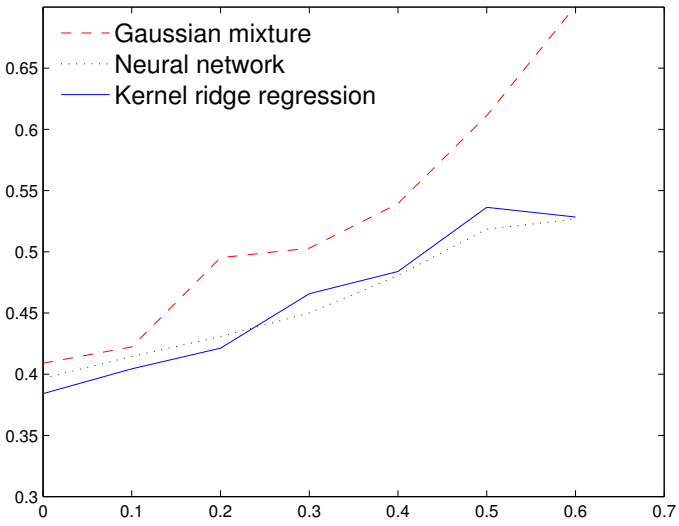

Figure 2: Test mean-squared error (Abalone) when the proportion of missing values increases. "Gaussian mixture" is the mixture alone, used as a classifier. Both "Neural network" and "Kernel ridge regression" have been trained using the conditional mean imputation of missing values provided by the same Gaussian mixture. We see that combining a discriminant algorithm with the generative Gaussian mixture model works better than the generative model alone.