Tensor Factorized Density Estimation

Rasmus Tuxen, Simon A. Pedersen & Marc S. Bøg
Technical University of Denmark, Department of Applied Mathematics and Computer Science

Motivation

Estimating the densities of high dimensional datasets is a crucial role in unsupervised learning tasks. The purpose of this study is to investigate the use of tensor factorization methods on probabilistic mixture models for the purposes of density estimation.

Density estimation

The process of constructing a density function from observed data. Used in:

- Outlier detection
- Data imputation

Difficult to estimate density of high-dimensional datasets due to *"curse of dimensionality"*

Tensor decompositions

Tensors are multidimensional arrays, and their decomposition can help with uncovering underlying hidden low-dimensional structure in the larger multidimensional tensor.

- Efficient representation using a geometric approach
- Active research field with many methods

Tensor Factorized Density Estimation (TFDE) models

We adapt two decomposition methods into the language of mixture models:

Canonical Polyadic decomposition [2]:

$$p(x_1, \dots, x_m) = \sum_{k=1}^{K} p(k) \prod_{m}^{M} p(x_m|k)$$

Tensor train [4]:

$$p(x_1, \dots, x_M) = \sum_{k_0,\dots,k_M}^{K_0,\dots,K_M} p(k_0) \prod_{m=1}^{M} p(x_m, k_m | k_{m-1})$$

The models contain only a single hyperparameter K, that determines the number of distributions in each dimension.

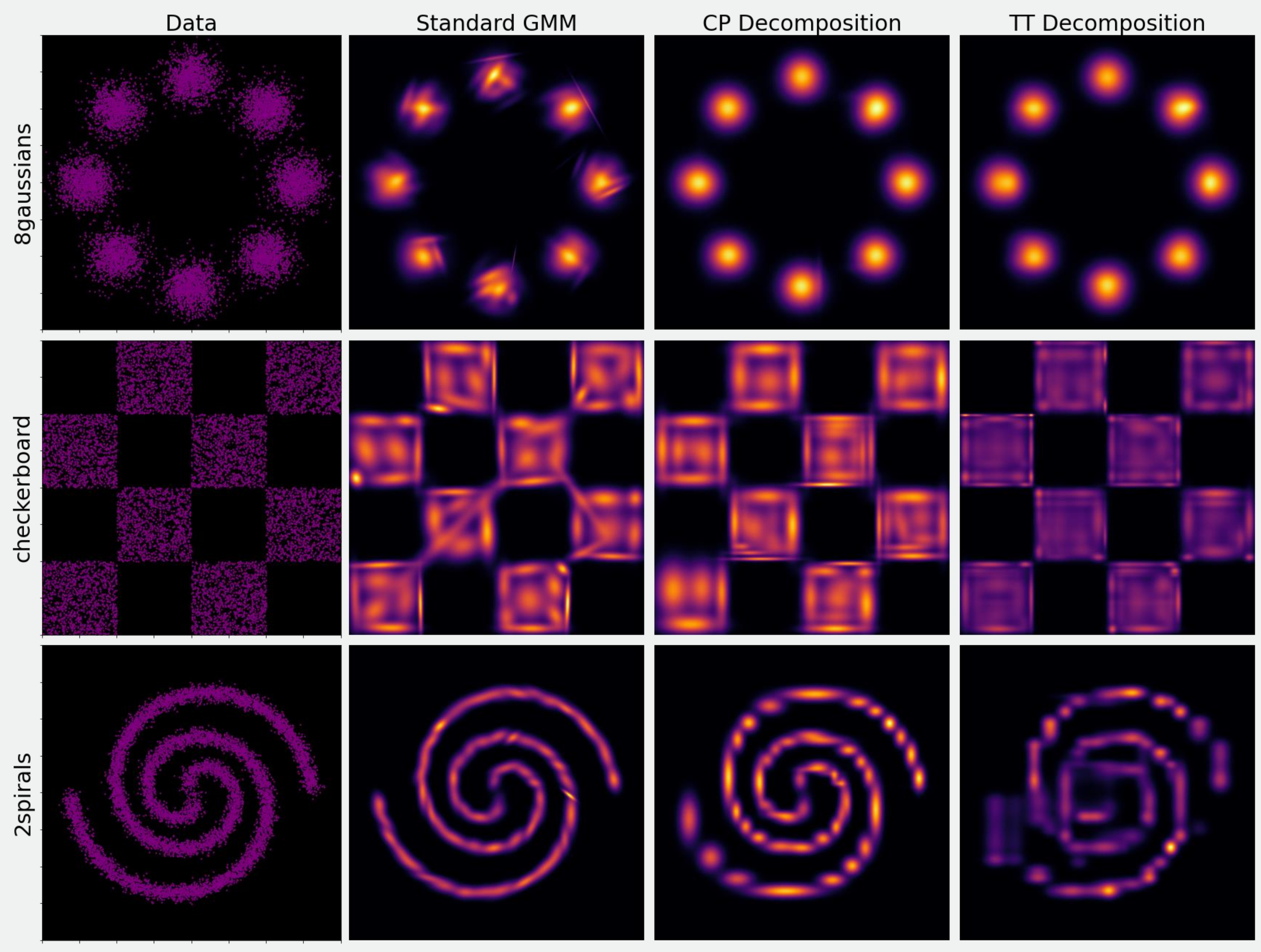

Figure 1: A standard Gaussian Mixture Model (GMM), The Tensor train (TT) model and the Canonical Polyadic (CP) model are fit on two dimensional toy datasets with gaussian distributions.

Advantages of TFDE models .

- The Tensor train (TT) and Canonical Polyadic (CP) models are trained with gradient descent on high dimensional datasets and compared to a state-of-the-art method [6] for density estimation.
- The TFDE models are found to be better at utilizing the trainable parameters in high dimensional space than standard mixture models.
- The TFDE models can be fit directly on full datasets containing both continuous and discrete variables since each dimension can use a unique distribution.

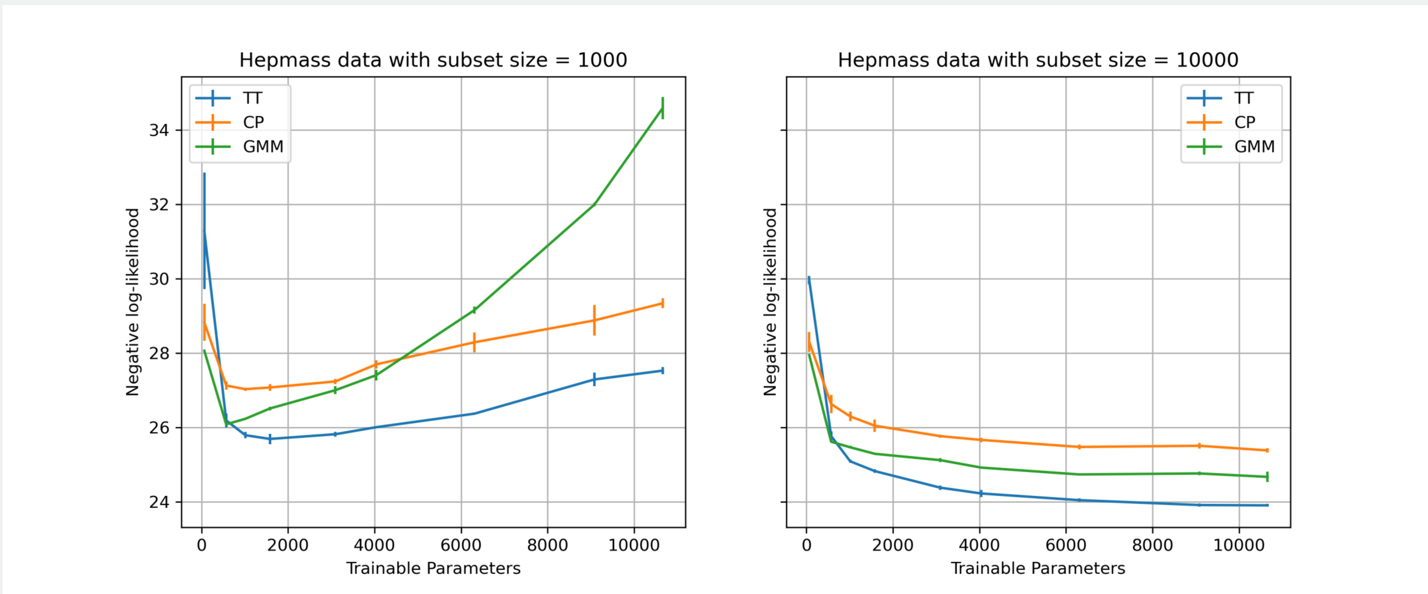

Figure 2: Learning rate for TT, CP and GMM trained on subsets of the HEPMASS datasets (23 dimensions and ~700.000 samples) with the same number of free parameters. The error is measured on the full test set of the HEPMASS dataset.

	TT	CP	CART
True Positive Rate	54.27%	37.69%	59.57%
True Negative Rate	91.46%	94.56%	87.36%
Accuracy	82.51%	80.87%	80.67%

Table 1: Results from a classification task on the 'Adult' dataset containing a mix of continuous and discrete variables for the heterogeneous TT and CP models and a simple classification tree. The prevalence of negative is 75.2%.

Dataset	TFDE [TT]	TFDE [CP]	FFJORD
POWER	-0.02	0.01	-0.46
GAS	-3.95	-5.44	-8.59
HEPMASS	22.38	23.60	14.92
MINIBOONE	33.29	41.43	10.43
BSDS300	-130.31	-127.47	-157.40
MNIST	0.06	2.57	1.05
CIFAR10	N/A	N/A	3.40

Table 2 : Negative log-likelihood on test data for density estimations models; **lower is better**. In nats for tabular data and bits/dim for MNIST and CIFAR10. FFJORD results are from [6]. The datasets are listed in order of greater dimensions. Hyperparameter was selected using holdout cross-validation. The CIFAR10 dataset was not evaluated due to a GPU implementation specific issue.

Conclusion

- TFDE models show versatility for high dimensional data
- Good results compared to state-of-the-art
  - Generally, TT performed better than CP
- Heterogeneous modelling
  - Enables single models to work both in a mix of continuous and non-continuous spaces
  - The accuracy is not that great, but the modelling on this type of dataset is possible.

References

[1] Stephan Rabanser, Oleksandr Shchur, and Stephan Günnemann, "Introduction to tensor decompositions and their applications in machine learning," 2017.
[2] Felipe Bottega Diniz, "A fast implementation for the canonical polyadic decomposition," 2019.
[3] Johan Håstad, Tensor rank is NP-complete, vol. 11, pp. 451–460, Journal of Algorithms, 04 2006.
[4] I. V. Oseledets, "Tensor-train decomposition," Society for Industrial and Applied Mathematics, 2009.
[5] Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud, "FFJORD: freeform continuous dynamics for scalable reversible generative models," CoRR, vol. abs/1810.01367, 2018.
[6] George Papamakarios, Theo Pavlakou, and Iain Murray, "Masked autoregressive flow for density estimation," 2018.

DTU Compute
Department of Applied Mathematics and Computer Science
DTU Compute
Department of Applied Mathematics and Computer Science