DTU

Student IDs	Names
s173936	Simon Amtoft Pedersen
s173905	Marc Sun Bøg
s173910	Rasmus Tuxen

02460 - Advanced Machine Learning

Tensor Factorized Density Estimators

Density Estimation

- Unsupervised task in Machine Learning
- Using probabilistic models, e.g. the Gaussian Mixture Model
- Strengths: few parameters, probabilistic interpretation
- Weaknesses: initialization, convergence, overfitting issues

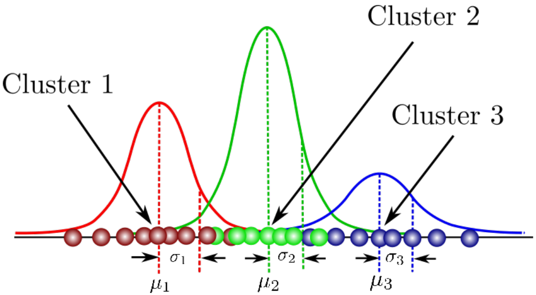

Figure: Illustration of the Gaussian Mixture Model [Conteras 2019]

Tensor Decompositions

- Tensors are multidimensional arrays
- Underlying hidden-low dimensional structures
- Efficient representation

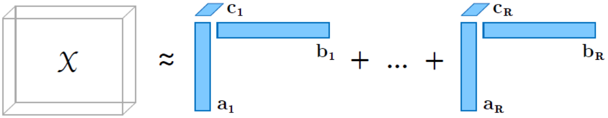

Figure: Illustration of the Canonical Polyadic decomposition [Diniz 2019]

Purpose of the project

The purpose of our project has been to *explore the use of tensor factorization methods to estimate probability densities of data, and make comparisons of these to the state-of-the-art methods.*

Factorization of Mixture Models

Canonical Polyadic (CP)

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

Tensor Train (TT)

$$p(x_1, \dots, x_M) = \sum_{k_0,\dots,k_M}^{K_0,\dots,K_M} p(k_0) \prod_{m=1}^{M} p(x_m, k_m|k_{m-1}) \tag{2}$$

- Decomposition of probability densities
- Enables heterogeneous modelling

Results (I)

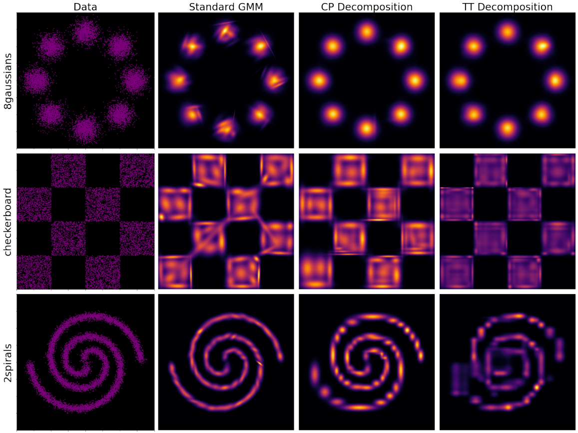

Figure: Density plots for the standard GMM model, and CP- and TT- decompositions models, after being fit to three different toy data distributions.

Results (II)

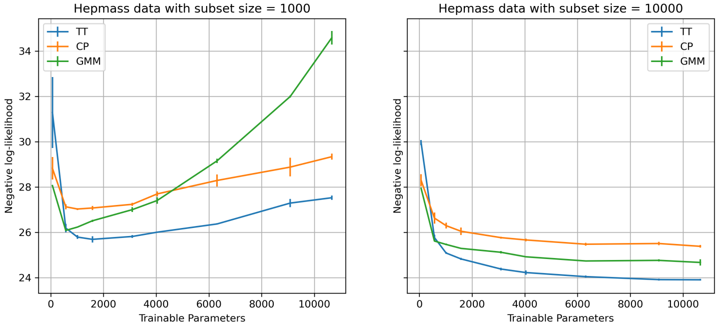

Figure: Learning rate for Tensor train, Canonical Polyadic and Gaussian mixture model with full covariance trained on different subsets of the Hepmass dataset (23 dimensions, and $\approx 700,000$ samples) with the same number of free parameters. The error is measured on the full test set from the Hepmass dataset.

Results (III)

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: Negative log-likehood on test data for density estimation models; **lower is better**. In nats for tabular data and bits/dim. FFJORD results are from Grathwohl et al. 2018.

Discussion

- Singularity issues for single point clusters
  - Relative point probabilities skyrocket
- Explore initialization
  - Parameters drawn randomly from a specified interval
- High dimensional problems
  - Log-domain calculations
  - TensorFlow loop unrolling
- Ordering of Tensor Train

Conclusion

- TFDE models shows versatility for high dimensional data
- Good results compared to state-of-the-art
  - Generally TT performed better than CP.
- Heterogeneous modelling
  - Enables a single models to work both in a mix of continuous and non-continuous spaces

References

C. O. Conteras. Gaussian mixture models explained, 2019.

F. B. Diniz. A fast implementation for the canonical polyadic decomposition, 2019.

W. Grathwohl, R. T. Q. Chen, J. Bettencourt, I. Sutskever, and D. Duvenaud. FFJORD: free-form continuous dynamics for scalable reversible generative models. *CoRR*, abs/1810.01367, 2018. URL `http://arxiv.org/abs/1810.01367`.

Heterogeneous 2D -> 3D

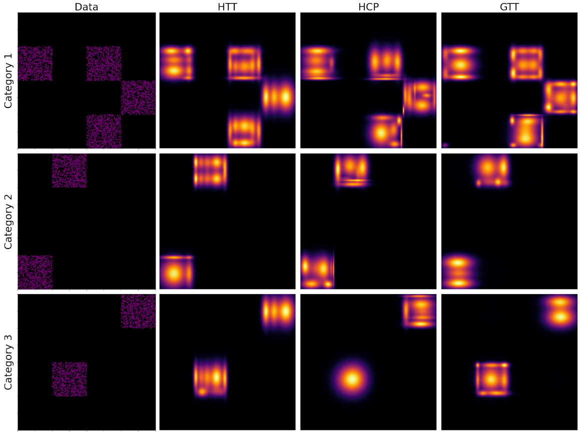

Figure: A 3D plot of the checkerboard 2D toy data split intothree categories, with each row representing a different cate-gory.

Sampling from Heterogenous Model

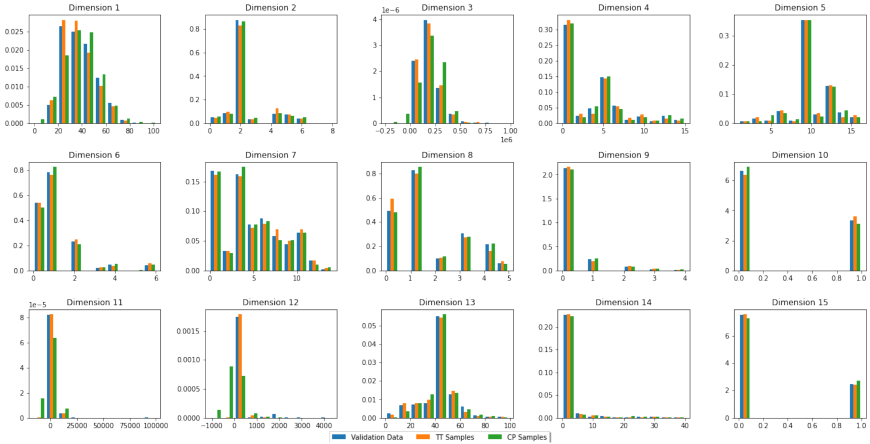

Figure: Sampling of CP and TT heterogeneous TFDE models trained on the Adult dataset, which contains 14 features and a label.

Results of Bayesian Classifier

Training heterogeneous models as a Bayesian Classifier

	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: Results from classification task on test set from Adult data for the heterogeneous TT and CP model and a simple classification tree. The prevalence of negative is 75.2 %.