02460 - Advanced Machine Learning LOG-BOOK

Simon Amtoft Pedersen, s173936@student.dtu.dk
Marc Sun Bøg, 173905@student.dtu.dk
Rasmus Tuxen, s173910@student.dtu.dk

The main purpose of the logbook is that it serves as a tool for you to organize the project. Further, it serves as a way to collecting information related to the learning objectives:

- Presentation of methods and results at meetings with project supervisor and fellow students
- Plan and carry out the course of the project in collaboration with the project supervisor
- Organize and coordinate the work in the project group

Overall Project Goals

Define own learning objectives for the project

- Understand the parts of the theory behind tensor factorization methods, such as CP, TT and TR decompositions.
- Understand the necessary theory of tensor factorization, to lay the foundation for implementing and training a TFDE model.
- Implement and train a TFDE model using one of the probabilistic programming tools, TensorFlow Probability (TFP) or Pyro.
- Analyze results of different implementations of the TFDE model.
- Assess if the trained model gives desired results.
- Compare results of own model to state-of-the-art e.g., FFJORD.

Carry out a well-founded delimitation of the project and formulate specific hypotheses and aims

In this project we aim to explore the use of tensor factorization methods to estimate probability densities of data and make comparisons to the state-of-the-art methods. Additionally, we want to investigate the advantages of tensor factorizations, such as being able to tailor density estimators to datasets consisting of arbitrary distributions in each dimension such as continuous and discrete features.

Our hypothesis is that the tensor factorization methods can be applied to density estimation problems and provide reasonable density estimations even with relatively few parameters.

As this field is more or less uncharted territory, we will focus on TT- and CP- decomposition modelling, and not looking into hyperparameter optimization, besides exploring the impact of the number of components in the models.

Project Meetings

Week 0: 18.02.21 - 24.02.21

Reading: Three papers regarding CP-, tensor train, tensor ring decomposition:
- Oseledets, Ivan V. "Tensor-train decomposition." SIAM Journal on Scientific Computing 33.5 (2011): 2295-2317
- Zhao, Qibin, et al. "Tensor ring decomposition." arXiv preprint arXiv:1606.05535 (2016).
- FFJORD paper https://openreview.net/pdf?id=rJxgknCcK7

Week 1: 25.02.21 - 03.03.21

Reading: TensorFlow Probability documentation. Literature study to find papers

Implementation: Everyone looked into TensorFlow Probability, and how to make a diagonal GMM distribution. Also implement generation of toy data as done in the FFJORD paper.

Results: Functions for toy data generation. Creation of gaussian distributions in TFP.

Decisions: Decided to use TensorFlow Probability as our probabilistic programming tool.

Week 2: 04.03.21 - 10.03.21

Reading: TensorFlow Probability documentation. Literature study to find papers

Implementation: Create distributions with TFP such that each cluster has a 2-D distribution consisting of univariate gaussians in each dimension.

Results: A fit of the gaussian TFD distributions to the toy data 8gaussians from FFJORD succeeded.

Decisions: We shouldn't use the EM algorithm in the first place. Try using negative log likelihood with Adam optimizer and use logits instead of probabilities to model Categorical distributions of the Mixture components to avoid issue of weights not summing to 1.

Week 3: 11.03.21 - 17.03.21

Implementation: CP Model trainable with standard negative log likelihood loss. Initial naive implementation of Tensor Train model.

Results: Tensor Train model is absurdly slow. CP Model trains and fits data well using - log p(x). Made a fit on 2-D toy data using the Tensor Train model, which is able to model more complex data using combinations of gaussians.

Decisions: Remove loops, use matrix operations for Tensor Train in order to utilize parallelization.

Week 4: 18.03.21 - 24.03.21

Questions: Should we use different mu and sigma matrix trainable variables for each dimension, or use the same for all dimensions? Should the range of ki be different for each i, or should they stay the same? What is NATS from the FFJORD paper?

Implementation: Vectorize TT model to achieve a speed-up to be able to train faster on many more epochs.

Results: TT model was vectorized, resulting in a great speed-up and better fit to the data, especially for large batch sizes.

Decisions: For now we keep using only Gaussian distributions with the same dimensions of mu and sigma, such that all ki = [0, ..., K].

Week 5: 25.03.21 - 07.04.21 (easter off)

Implementation: Make a fair comparison between GMM, CP and TT on toy data. Investigation of the real FFJORD data sets.

Results: Fair-ish comparisons of CP, TT and GMM on both Toy and (some) Real data.

Decisions: Our ‘extra contribution’ is to compare the CP and TT models using not only Gaussian distributions, but also other distributions, e.g. Poisson distributions. Create heterogeneous models. Not continue looking into Tensor Ring Decomposition.
Fix data loaders.

Week 6: 08.04.21 - 14.04.21

Questions: How to make integration tests for heterogeneous models (with categorical distributions)?

Implementation: Initial implementation of heterogeneous model.

Results: Made initial heterogeneous model. Changes Tensortrain to compute in log-domain (for numerical stability in higher dimensions)

Decisions: Keep working on heterogeneous models. On toy data, try categorical distribution of data in 3D where the third dimension is the categories, where there are some kind of similarity between the categories, but also some differences. Make hold-out Cross validation for hyperparameter selection.

Week 7: 15.04.21 - 21.04.21

Questions: Would it be fine to train on a subset of the datasets?

Implementation: Implemented heterogenous TFDE models.

Results: Can fit heterogenous TFDE models to both toy and real data, with e.g. categorical and gaussian distributions. Get results on real FFJORD data for TFDE models.

Decisions: Only train and validate on subsets on the real datasets, since our models fit very slowly due to computational time. (Perhaps) Use ideal parameters from the subset and fit the whole dataset for final results in paper.

Week 8: 22.04.21 - 29.04.21

Questions: Classification with heterogeneous models on the Adult data.

Implementation: Re-implement log domain for TT.

Results: Computed results on ffjord data on HPC cluster for more iterations and on full dataset, which reduced the error of the results.

Decisions: Finalize paper.

Week 9: 30.04.21 - 06.05.21

Finish report and hand-in.

Supervisor Meetings

Week 1: 25.02.21

Presentation of results since last week

- Tensor-decomposition methods and their relationship to gaussian mixture models.

Action points for next week

- Implement a GMM using tensorflow/pyro framework
- Extract data from Fjord and use for training/visualization
- Find relevant papers

Week 2: 05.03.21

Presentation of results since last week

- Including visualization of gmm performance

Action points for next week

- Implement a GMM using CP-decomposition
- Find relevant papers and make ready to present them

Week 3: 11.03.21

Presentation of results since last week

- Presentation of CP-decomposition result and relevant papers.

Action points for next week

- Finish CP-decomposition presentation
- Make unit-test to verify the performance of densities
- Implement gmm using tensor-train decomposition in 2D data
- Test spiral, checkerboard and 8gaussians data for K=2,3,4 components. Visualize the results for the different components

Week 4: 18.03.21

Presentation of results since last meeting

- Density plots that can fit gaussians and squares.
- Prob. density integrates to 1

Action points for next week

- Generalize TT-implementation to admit M-dimensional data
- Speed up computation time for call().
- Make a softplus-constraint on sigma
- Set up real data from FFJORD-paper (download and make functions)
- Make a comparison between GMM w diag and full covariance (sklearn), CP and TT on synthetic data
- Investigate Tensor-Ring implementation
- Start small writing of report (introduction and method)
- Use TensorFlow classes instead?

Week 5: 25.03.21

Presentation of results since last meeting

- Show improved plots for Toy Data, by using the vectorized implementation of TT, and the difference in the negative log likelihood between CP and TT implementation.

Action points for next week

- Implement CV-testing for toy data
- Make fair comparison of models (gmm w. Full cov, CP, TT), that is; with same number of adjustable parameters
  - For CP, K should be K^2 of TT.
- Finish report section about Toy Data fitting.
  - With proper split of the data into training, evaluation and test sets.
    - Use hold-out sets, since the dataset is large.
  - Hyper parameter tuning
    - Check lecture on testing
- First iteration of a problem formulation for what we want to investigate.
- Investigate Tensor Ring (TR)
- Investigate the effects of initialisation of mu and sigma

Week 6: 08.04.21

Presentation of results since last meeting

- Fair comparison of models

Action points for next week

- Implement intelligent initialization scheme for TT method
- Implement CP model with full covariance (For comparison in the future.)
- Create comparison results for the Toy Data that can be put into the report.
- Implement correct pre-processing and reading of the miniboone data from FFJORD paper
- Implement heterogeneous CP and TT models.

Week 7: 15.04.21

Presentation of results since last meeting

- Show that heterogeneous models work on a simple level.

Action points for next week

- Make sure TT in log-domain is stable and correct
- Make synthetic data with M=3 where the 3rd dimension is categorical and the re-maining is slightly changed with each category.
    - E.g. Categorical with 3categories, [0.3, 0.3, 0.4]
    - 30% samples from Dist 1
    - 30% samples from Dist 2 etc.
    - Stack em, shuffle
- Test on categorical synth data (M=3) and make visualization of heterogeneous model vs GMM (Marc)
    - Train on 3D data
    - Fix category to e.g. 1/2/3, plot 2D contours of p(x1, x2, 1/2/3)
    - Compare Heterogeneous vs Gaussian visually (not log_prob)
- Make hold-out Cross validation scheme and train on : ffjord data/synthetic data
- Using optimal K from hold-out Cross-validation find loss on test data (ffjord) and make a table of achieved loss in report.
- Make a learning curve on a high-dimensional dataset where the models are only trained on a small subset of the data.
    - Plots with error bars (a couple initializations with N_init=1)

Week 8: 22.04.21

Presentation of results since last meeting

- Fit of heterogeneous models on the 8gaussians toy dataset
    - Samples, loss curves, densities
    - CP is approximately the same as TT for most Ks and different data sizes.
- Showcase of heterogeneous models on the “Adult” dataset
    - CP is approximately the same as TT for most Ks and different data sizes.

Action points for next week

- Make visualizations of TFDE models on Toy data
    - 2d models
    - Heterogeneous models
- Train models for FFJORD data (make table of results)
- Make learning curve on data from FFJORD paper (likelihood vs free-parameters)
- Train heterogeneous models on Adult dataset
    - Learning curves
    - Sample, histograms
- Add a small value to sigma softplus to ensure numerical stability.

Week 9: 29.04.21

Presentation of results since last meeting

- Plots for heterogeneous toy data
- Results on FFJORD data
- Heterogen fit on Adult data

Action points for next week

- Finish last runs of to get test losses for comparison with
- Finish report