Norwegian University of Science and Technology

DT8122 **Project Assignment — Uncertainty Quantification**

PROBABILISTIC AI

dt8122@idi.ntnu.no

Summer 2021

1 Introduction

Uncertainty modeling and quantification is relevant in many real world applications. One possible approach to this task is to use Bayesian methods in order to obtain the probability density of the quantities of interest and use the obtained full densities to compute other quantities of interest such as prediction intervals (PI). Traditionally, we would resort to linear models in order to have a computationally feasible solution to the inference problem. However, with the recent advances in Probabilistic AI, we are able to extend the space of possible models with non-linear models. The objective of this project assignment is to explore some of those techniques in a set of prediction tasks.

2 Task

The task concerns implementation and evaluation of probabilistic deep learning methods for uncertainty quantification. First, you are asked to implement two deep generative modeling techniques (Section 2.1). We then ask you to identify shortcomings or limitations of your chosen techniques. Finally, we ask you to implement a non-trivial possible improvement for one of your chosen methods (Section 2.2).

Each part should be accompanied by empirical evaluation (Section 2.4) using the provided datasets (Section 2.3) and discussion of the results. All this should be compiled into a final report (max. 10 pages) accompanied with the code to reproduce your results.

2.1 Models

You are expected to implement two deep probabilistic regression models for uncertainty estimation and use these model to generate prediction intervals and point predictions. For that you have to choose two of the following techniques:

1. Structured and Efficient Variational Deep Learning with Matrix Gaussian Posteriors [1]
2. Multiplicative Normalizing Flows [2]
3. A Simple Baseline for Bayesian Uncertainty in Deep Learning [3]
4. Noisy Natural Gradient as Variational Inference [5] (Either Adam or K-FAC.)

2.2 Research of Shortcomings, Limitations and Possible Improvements

Discuss the shortcomings or limitations of the two approaches you have chosen to implement. Suggest and implement at least one possible improvement. This improvement should be non-trivial, so hyperparameter optimization does not count as a possible improvement. Justify your modification, compare with previous results, and report your findings.

This part of your work is what we will emphasize the most when assessing your submission. We do not necessarily expect you to come up with a new publishable method, but we do expect your implemented change to be non-trivial.

2.3 Datasets

The provided datasets are the same as in Zhang et al. [5], and Louizos, Welling [1]. All datasets have target values (y) placed in the last column. Ensure that your implementation can run on all datasets (see also Section 3 for required command line interface)

2.4 Evaluation

For evaluation you will use the last 10% of rows in the dataset as the test set. For both of your implemented methods, as well as your modified method, measure and report the following metrics on the test set:

- Root mean square error (RMSE),
- Prediction interval coverage probability (PICP) for 95% PI,
- Mean prediction interval width (MPIW) for 95% PI.

Following the notation from Pearce et al. [4], let $\boldsymbol{x}_i \in \mathbb{R}^D$ be the i-th D dimensional input features corresponding to target observation y_i, where $1 \leq i \leq n$ for n data points. The predicted lower and upper bounds of PI are denoted by $\hat{y}_{Li}$ and $\hat{y}_{Ui}$. The PI is then defined as

$$P\left(\hat{y}_{Li} \leq y_i \leq \hat{y}_{Ui}\right) \geq \gamma,$$

where γ is 0.95 for 95% prediction interval.
PICP is calculated as following:

$$k_i = \begin{cases} 1 & \text{if } \hat{y}_{Li} \leq y_i \leq \hat{y}_{Ui}; \\ 0 & \text{otherwise,} \end{cases}$$

$$PICP = \frac{1}{n}\sum_{i=1}^{n} k_i.$$

MPIW is calculated as following:

$$MPIW = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_{Ui} - \hat{y}_{Li}.$$

3 Submission Requirements

We expect you to submit the following:

- **Code** with your implementation in Python using Pyro/PyTorch or TensorFlow (Probability) of two of the four techniques of uncertainty modeling with neural networks.
  - Your deliverable should contain a *main.py* file that can be executed with the following command and arguments:
    python -m main

    --dataset accepts any string matching one of the supplied datasets

    --method accepts integers 0–4. 0 is your modified method and 1–4 follows the same order as listed in Section 2.1. For the methods not implemented it should return a *NotImplementedError*.

    For instance, use

    python -m main --dataset ./wine.txt --method 3

    to run Maddox et al. [3]'s method on data stored in *./wine.txt*.

    You can choose to accept more arguments, however, all other parameters should have appropriate default values such that the script can be executed with only the arguments specified with default values per dataset the script should reproduce results in the paper. The script should train a model on the training-portion of the specified dataset, and return the RMSE, PICP, and MPIW values for the test portion. If required, your script can have load pre-trained models that you also supply.

  - You should make sure the code is not dependent on a system specific configuration. To make sure that we can execute your code, provide a PIPFILE or REQUIREMENTS.TXT file.

- **Report** that should include your model choices, inference methods, results from empirical evaluation, findings and discussion. Max. 10 pages.

All assignment artifacts are to be sent as a single ZIP file to the email address dt8122@idi.ntnu.no. The deadline is August 18th 2021 AoE (Anywhere on Earth).

References

[1] *Louizos Christos, Welling Max.* Structured and Efficient Variational Deep Learning with Matrix Gaussian Posteriors // Proceedings of The 33rd International Conference on Machine Learning. 48. New York, New York, USA: PMLR, 20–22 Jun 2016. 1708–1716. (Proceedings of Machine Learning Research).

[2] *Louizos Christos, Welling Max.* Multiplicative Normalizing Flows for Variational Bayesian Neural Networks // Proceedings of the 34th International Conference on Machine Learning. 70. International Convention Centre, Sydney, Australia: PMLR, 06–11 Aug 2017. 2218–2227. (Proceedings of Machine Learning Research).

[3] *Maddox Wesley J, Izmailov Pavel, Garipov Timur, Vetrov Dmitry P, Wilson Andrew Gordon.* A Simple Baseline for Bayesian Uncertainty in Deep Learning // Advances in Neural Information Processing Systems. 32. 2019.

[4] *Pearce Tim, Brintrup Alexandra, Zaki Mohamed, Neely Andy.* High-Quality Prediction Intervals for Deep Learning: A Distribution-Free, Ensembled Approach // Proceedings of the 35th International Conference on Machine Learning. 80. Stockholmsmässan, Stockholm Sweden: PMLR, 10–15 Jul 2018. 4075–4084. (Proceedings of Machine Learning Research).

[5] *Zhang Guodong, Sun Shengyang, Duvenaud David, Grosse Roger.* Noisy Natural Gradient as Variational Inference // Proceedings of the 35th International Conference on Machine Learning. 80. 10–15 Jul 2018. 5852–5861. (Proceedings of Machine Learning Research).