Revising the progress on BANN project

Alex DongHyeon Seo

Our goal:

Our goal for this research is that we want to finish our paper about our algorithm BANN. This paper is about understanding the flow of information in deep probabilistic models. The initial paper was submitted but rejected due to lack of significant experiments and we want to make our paper accepted and officially published with additional experiments using this algorithm.

Understanding BANN algorithm:

Bayesian Neural Networks generally refer to a class of algorithms that treat neural network models in a ‘Bayesian manner’, which is well suited for uncertainty estimates. In this paper our goal is to design a deep neural network with structure and interpretability in mind, with easily amenable to uncertainty estimation, which we call ‘Bayesian Additive Neural Network’. From this algorithm we can produce posterior distribution approximations that provide uncertainty estimates, and moreover, achieve interpretability in deep probabilistic models to understand the flow of information.

The main structure that we will be using is the ‘Computational Skeleton’ that is used to study the relationship between neural networks and kernels. We will use this to design Bayesian Neural Network block by block, such that it can help efficiently approximate the posterior of Deep Gaussian Processes. Addition to this structure we will be using two kinds of blocks to design the model structure. Function block, which will allow every node in the computational skeleton to replicate a few times that will help define Bayesian priors and posteriors. Random feature block, which is used to construct random feature approximations for kernels to leverage the expressive power of Deep Gaussian Process. From this framework we apply additive structure to detect statistical interaction and define this through ANOVA decomposition for interpretability.

What we have accomplished so far (Experiments done so far):

The initial experiment we had on the paper was to evaluate the performance on prediction accuracy and interaction detections by comparing with different types of AddNNs that provide uncertainty estimates. We compared AddNN with BNN, BART and NID and showed that our model has competitive prediction performance but with much more compact design and outperforms on interaction detection. Additionally, we produced an average interaction function and the uncertainty function as a heatmap which can be very useful for interpretability.

For more comparison, Similar to our work, to show flow of the information in deep neural network models, there are few other attempts to achieve this. Soft Decision Tree developed by Frosst and Hinton proposes a novel way of resolving the tension between generalization and interpretability. However, Soft Decision Tree and BANN deliver different types of interpretability. While Soft Decision Tree uses model-based interpretability using decision trees risking its interpretability-accuracy tradeoff, BANN uses model-agnostic interpretability using feature interaction. We compared these two models and provide different points of view of interpretability and flow of information using their respective visualization, interaction heatmap and tree-based decision boundary visualization. For this experiment, hard label and soft label using knowledge distillation were both used to compare prediction accuracy for two models. While the Soft Decision Tree showed a 10% increase of prediction accuracy from 80% to 90%, our model BANN was consistent with 88% of prediction accuracy for both hard and soft labels. This shows that our model has competitive prediction performance compared to other models, even with soft labels, that provide different interpretation perspectives than ours.

Next steps:

So far, our experiments were mainly focused on comparison with different models regarding prediction accuracy, interaction detection and perspective on interpretation. However, to show why our model can be unique and critical, just comparing performances with other models is not enough. Our model’s potential lies on the flexibility of the model structure and providing interpretability at the same time.

Flexibility of models has been well studied across the field and could be critical to solve the underlying problems on Neural Networks. For example, on [Lee et al. (2018)](https://openreview.net/pdf?id=B1EA-M-0Z), it provides an extended perspective on connection between infinitely wide neural networks and kernel methods. In more details, the exact equivalence between infinitely wide deep networks and GP were derived. From this, we can use the resulting GPs to perform Bayesian inference for wide deep networks and see that GP uncertainty is strongly correlated with network’s prediction error. This milestone can be significant since by using wide deep networks, we can overcome the tradeoff between training error and generalization gap: According to [Belkin et al (2019)](https://arxiv.org/abs/1812.11118), they suggest that as the model complexity increases with over parameterized model the test error actually drops, which is against the common sense of U-shaped curve for the test error.

Taking this one step further, on [Jacot et al. (2020)](https://arxiv.org/pdf/1806.07572.pdf), they propose a kernel called Neural Tangent Kernel that takes inner product between the gradients of the network outputs with respect to the network parameter. This leads to a conclusion of, a properly randomly initialized sufficiently wide deep neural network trained by gradient descent with infinitesimal step size is equivalent to a kernel regression predictor with a deterministic kernel called NTK. On top of this, on [Arora et al. (2019)](https://arxiv.org/pdf/1904.11955.pdf), they improve this result to the non-asymptotic setting where the width of every layer only needs to be greater than a certain finite threshold. They also showed the results of the extension of NTK to convolutional neural nets (CNTK) which yield a significant performance for kernel-based methods on CIFAR-10.

Like these experiments above, we could use our model’s flexibility to translate to other model architecture. More specifically, we can derive the connection between our model and Deep Gaussian processes, and combine the connection between Deep Gaussian processes and wide deep neural networks to tie them all together. Similarly, on [Matthews et al. (2018)](https://arxiv.org/pdf/1804.11271.pdf), they showed how a wide deep neural network converges in distribution to a Gaussian Process and compared finite Bayesian deep networks to that Gaussian processes showing their connection with MCMC inference. Most recently, on [He et al. (2020)](https://arxiv.org/pdf/2007.05864.pdf), they tried to combine Bayesian Deep Ensembles that enables posterior interpretation with connecting GPs and infinite wide deep neural networks through the lens of NTK, both before and after training. Therefore, with all these discussed papers in account, our new experiment can be focused on connecting DGP that could represent infinite wide deep neural networks using NTK to our flexible BANN model using computational skeleton structure by benchmarking the modification that was used on [He et al. (2020)](https://arxiv.org/pdf/2007.05864.pdf). This could be a significant milestone since through our additive structure, we can also show that we could achieve interpretability using feature interaction. Ultimately, we could present a connection between an infinite wide deep neural network and our finite BANN model that could provide interpretability using interaction with help of both being related to GPs.