AM207 Final Project: Checkpoint 3

Avriel Epps-Darling, Sangyoon Park, Shih-Yi Tseng, Paulina Toro Isaza, Arthur Young

The Learned Uncertainty-Aware (LUNA) based regression method is a training procedure for Neural Linear Models (NLMs) by Thakur, Lorsung, Yacoby, Doshi-Velez, and Pan. The development of LUNA was motivated by the recognition that NLMs, a popular Bayesian deep learning architecture with easily interpreted inference properties, generate problematic uncertainty estimates in sparse regions of support using traditional learning methods. The authors address this shortcoming by proposing LUNA as an alternative training procedure, augmenting the objective function with an additional term designed to “diversify” priors at a computational cost equivalent to training one neural-network and one Bayesian linear regression. In short, an NLM can be considered in two parts: the first is a neural network that projects the covariates $\mathbf{x} \in \mathbb{R}^D$ into an embedding called $\phi_\theta(\mathbf{x}) \in \mathbb{R}^L$, parameterized by θ. Then, the embedded manifold ϕ_θ is passed to a layer of Bayesian linear regression parameterised by $\mathbf{w} \sim \mathcal{N}(0, \alpha\mathbf{I})$ to yield predictions y. The model is trained by seeking point estimate values for θ and then performs inference using the last layer parameterised by $\mathbf{w}$. Next, we discuss how the LUNA method modifies the training regime.

LUNA based regression encourages functional diversity of the feature map ϕ_θ learned by the neural network via training M auxiliary linear regressors. The core of LUNA is a two-part objective function:

$$\mathcal{L}_{LUNA}(\Psi) = \mathcal{L}_{Fit}(\Psi) - \lambda \cdot \mathcal{L}_{Diverse}(\Psi)$$

where Ψ includes shared parameters θ from the neural network and M sets of $\mathbf{w}$ for the auxiliary regressors. The first part, $\mathcal{L}_{Fit}(\Psi)$, represents the log likelihood of M regressors on training data with l_2 regularization, whereas the second part, $\mathcal{L}_{Diverse}(\Psi)$, encodes the cosine similarity of the gradients of the M regressors and thus penalizes non-orthogonal gradients. After training the NLM with $\mathcal{L}_{LUNA}(\Psi)$ which generates diversified feature basis ϕ_θ, we toss the auxiliary regressions and perform standard linear Bayesian regression over $\phi_\theta(\mathbf{x})$ and analytically compute the posterior of the weights $\mathbf{w}$.

The authors test LUNA against various models including traditional NLM on a selection of toy and UCI "gap" datasets as well as for Bayesian optimization. On the toy data, LUNA achieves the ideal aleatoric and epistemic uncertainty estimations of Bayesian Neural Networks with HMC sampling and Gaussian Processes, far outperforming NLM, MC Dropout (MCD), MAP, and Ensemble Bootstraps. In the UCI Gap Datasets, LUNA performs better than traditional NLM and MCD by properly discriminating between data-poor and data-rich regions and demonstrating higher epistemic uncertainty in the data poor regions without substantially decreasing the test log-likelihood. Other experiments show, when considering a small number of features, LUNA outperforms traditional NLM in generalizing well to test data and in transfer learning scenarios. In the Bayesian optimization experiment, LUNA exhibits much faster convergence than GP and slightly faster convergence to NLM while producing lower variance than NLM but more variance than GP. **For our project, we hope to reproduce these results while also finding failure modes where these results do not hold.**

In conclusion, the authors posit that the reason that traditional NLM posterior predictives cannot distinguish between data-poor and data-rich regions is because training methods for these models do not encourage diversity in functions that span the learning feature basis. Using a LUNA approach to learning feature bases mitigates this problem by maximizing uncertainty awareness alongside likelihood. The authors prove that this approach outperforms traditional NLMs on uncertainty estimation, generalization and transfer learning. In short, the authors assert that uncertainty-aware frameworks for hyperparameter selection in Bayesian Linear Regression are a promising path forward.