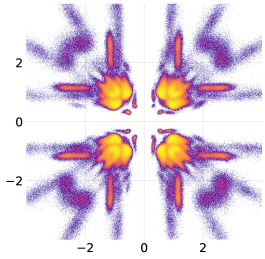
On the Interplay of Priors and Overparametrization in Bayesian Neural Network Posteriors

Julius Kobialka*, Emanuel Sommer*, Chris Kolb, Juntae Kwon, Daniel Dold, David Rügamer

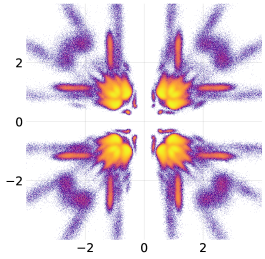
AISTATS 2026

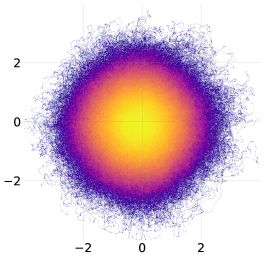
MUNIQ.AI | MCML | LMU Munich

BNN Posteriors are Hard!



BNN Posteriors are Hard!

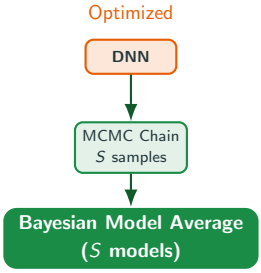




Setup

Setup:

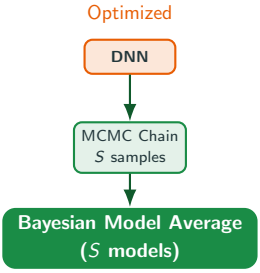
- $p(\boldsymbol{w} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{w}) p(\boldsymbol{w})$.
- Gaussian prior on weights $p(\boldsymbol{w}) \propto \exp(-\lambda R(||\boldsymbol{w}||_2^2))$.



Setup

Setup:

- $p(\boldsymbol{w} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{w}) p(\boldsymbol{w})$.
- Gaussian prior on weights $p(\boldsymbol{w}) \propto \exp(-\lambda R(||\boldsymbol{w}||_2^2))$.



Scope:

- Whenever likelihood has symmetries/flat directions, the prior determines how posterior mass is distributed.
- Different priors therefore induce different forms of **symmetry breaking**.

$\Rightarrow$ Our paper: Analyze interplay specifically for the **Gaussian isotropic prior**.

Idea of Analysis

- Full posterior characterization is intractable
  $\Rightarrow$ study local posterior geometry
- Restrict attention to high-likelihood/minimum-norm regions

Within a hidden layer

Analyze
additive overparametrization.

One neuron contribution can be reallocated across many copies.

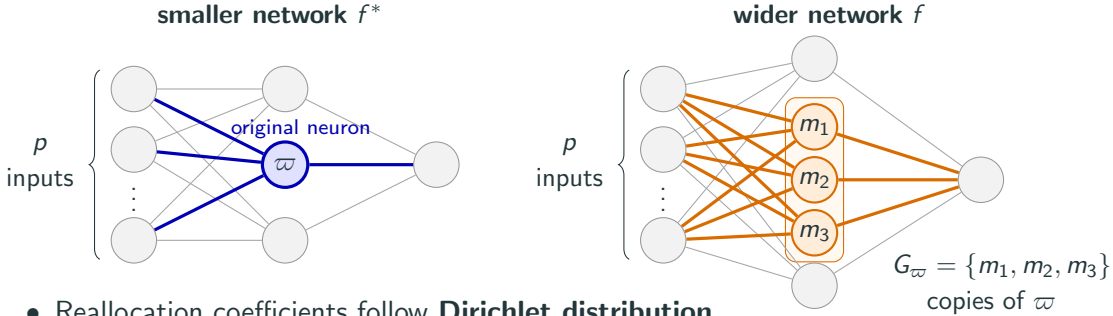
Across layers

Analyze
multiplicative overparametrization.

Rescaling adjacent layers can preserve the represented function.

Additive Overparametrization: Assignment Map

If one essential neuron's "work" is shared among $|G_\varpi|$ redundant neurons, how does the prior distribute that work in the Bayesian posterior?



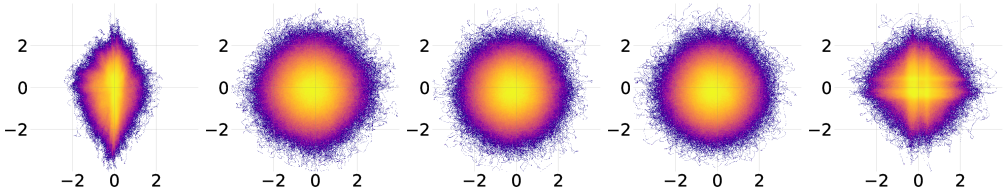
- Reallocation coefficients follow **Dirichlet distribution**.
- Dependence on incoming weights p of a given neuron: Larger p means more concentration around equal-norm-split.

Multiplicative Overparametrization: Layerwise Balancedness

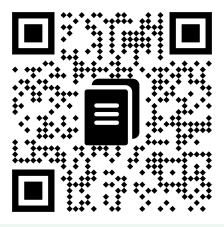
How does the prior distribute mass multiplicatively across different layers?

$\Rightarrow$ punishes unequal norms across adjacent factors.

$$\tau_l^{-2}\,\mathbb{E}_\pi\big[\|\boldsymbol{W}_l\|_F^2\big] \;-\; \tau_{l+1}^{-2}\,\mathbb{E}_\pi\big[\|\boldsymbol{W}_{l+1}\|_F^2\big] \;=\; d_l - d_{l+1}.$$



Thank you!



Let's continue the discussion!

Come by our poster today at 3pm, poster #182!

Contact: J.Kobialka@lmu.de

Thank you for your attention, see you this afternoon!