Linearised Laplace Inference in Networks with Normalisation Layers and the Neural g-Prior

UNIVERSITY OF CAMBRIDGE
UNIVERSITY OF AMSTERDAM
Computational and Biological Learning, University of Cambridge
Max Planck Institute for Intelligent Systems

Javier Antorán (ja666@cam.ac.uk), James Allingham, Dave Janz, Erik Daxberger, Eric Nalisnick, José Miguel Hernández-Lobato

- **Normalisation layers break the Laplace model evidence for model selection.**
- **The Jacobian of a NN has both very large and very small entries, complicating prior choice.**

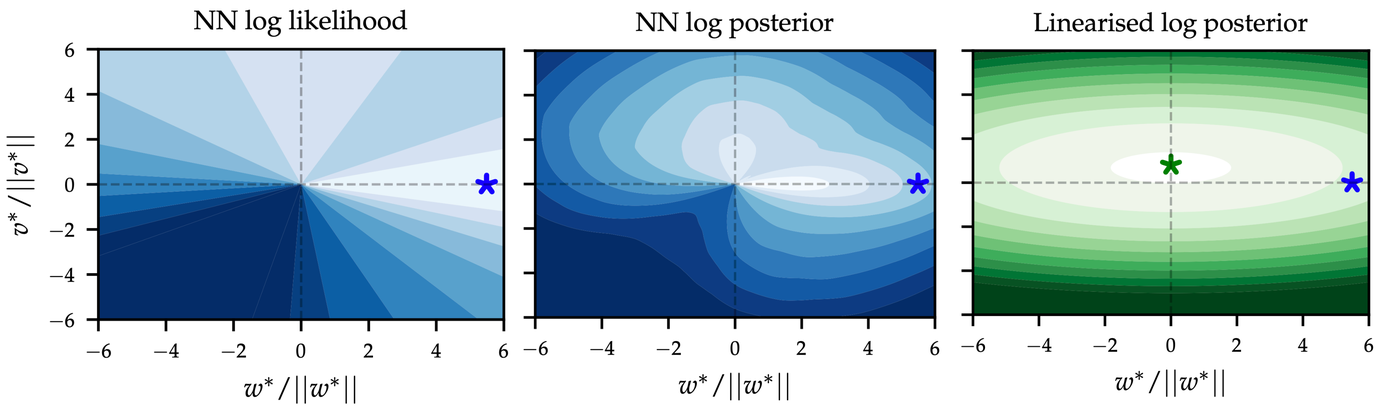
Normalisation Layers

- **SGD can not find a mode of NN posterior or linear model posterior biasing model evidence estimates**

➔ ***Use the mode of the linearised model: $v^\star$***

$$-\frac{1}{2}[\ ||w^\star||_\Lambda + \log\det\left(\frac{H+\Lambda}{\Lambda}\right)] + C$$

$$||v^\star||_\Lambda$$



★ **Linearisation point found with SGD $w^\star$ — not a mode of any posterior**

★ **Mode of linearised model posterior $v^\star$ is well defined**

g-Prior for Heterogeneous Jacobian Features

- **Isotropic priors leave some weights underspecified while overly constraining others**

➔ ***The 'scale invariant' g-prior induces the same posterior as scale-normalising our Jacobian features.***

$$p(w) = \mathcal{N}(w;\ 0, g \cdot \mathrm{diag}(\mathcal{F})^{-1})$$