37th Conference on Uncertainty in Artificial Intelligence (UAI, July 27-30, 2021)

Correlated Weights in Infinite Limits of Deep Convolutional Neural Networks

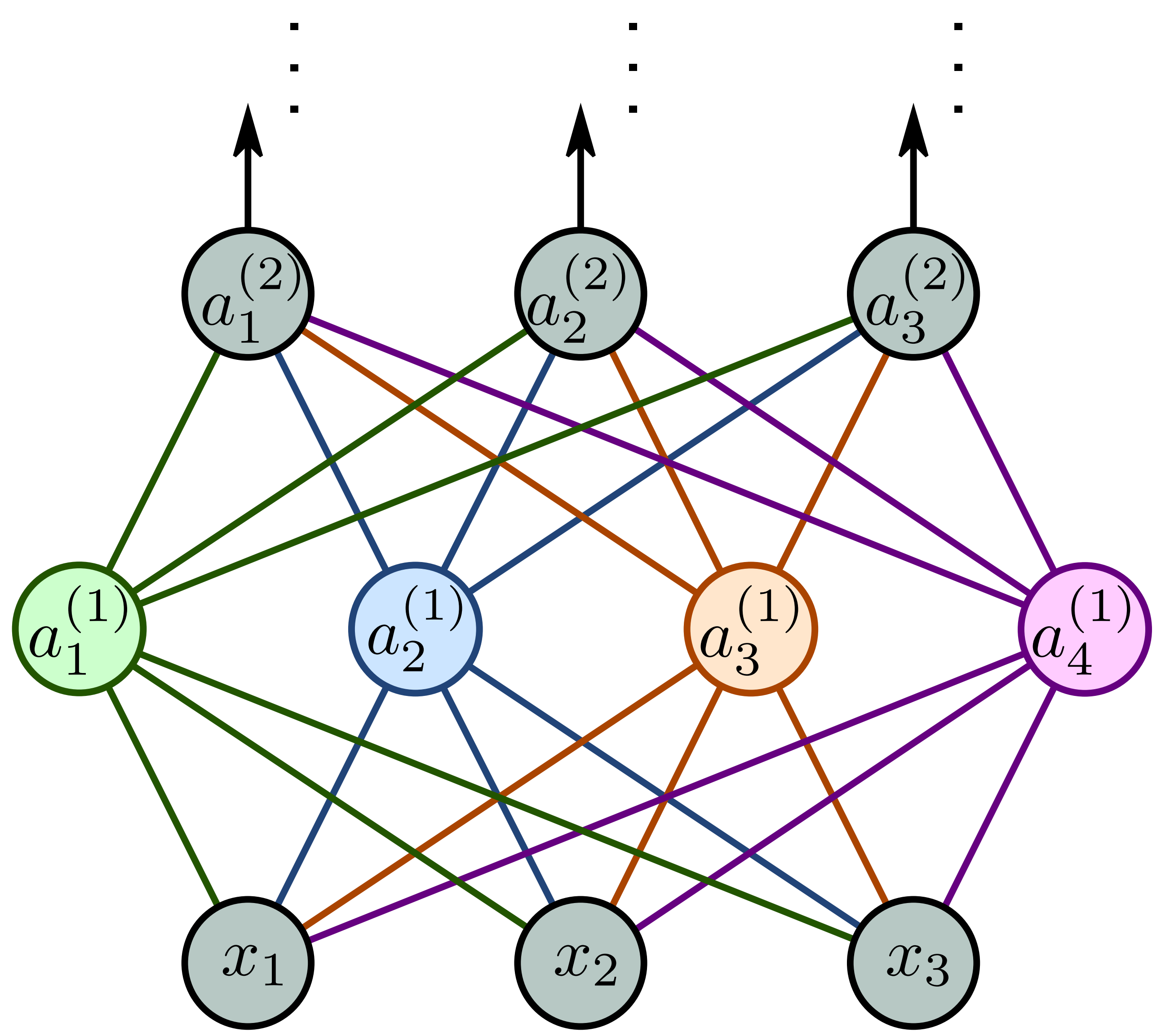
Adrià Garriga-Alonso
Mark van der Wilk

Bayesian neural network

Gaussian process (GP)

of *any* architecture

with independent, zero-mean
weight prior,
Gaussian bias prior

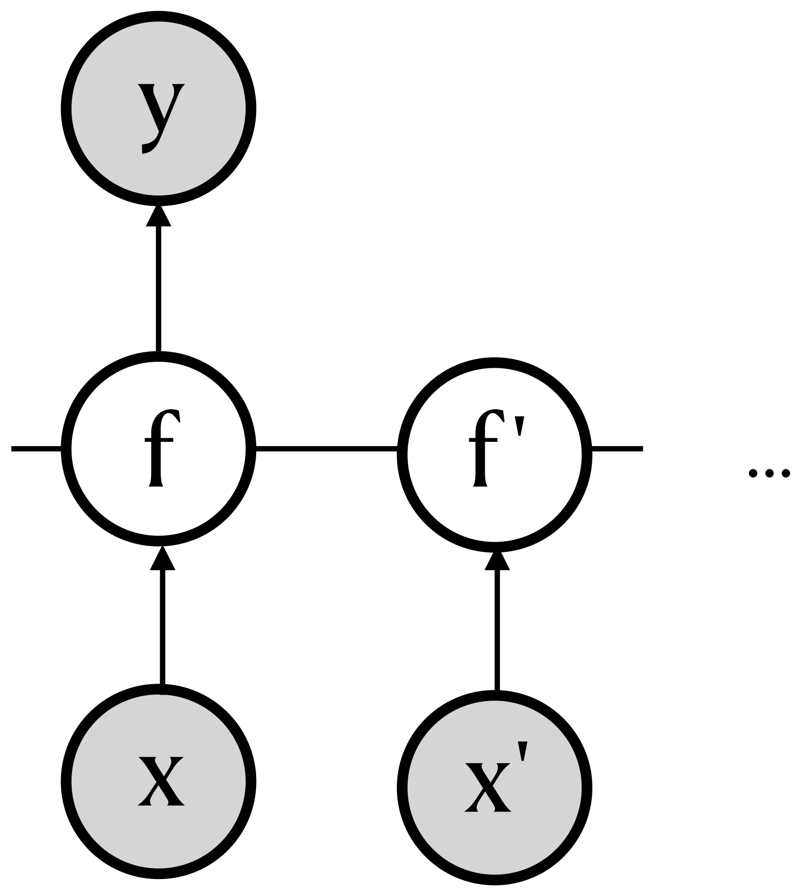


$\lim \text{width} \to \infty$

see e.g. Yang, 2019

First noted by Neal (1996)

Williams & Rasmussen (1995)



Bayesian neural network	Gaussian process
- Hard to infer posterior	+ Easy to infer posterior
+ Successful at modelling functions	+ Successful at modelling functions
+ Learns feature functions from data	- Feature functions fixed (by the kernel function of the GP)

Can GPs really replace NNs?

"Have we thrown the baby out with the bathwater?"

David J. C. MacKay, 1998

This paper: convolutions

Bayesian convolutional NN

\+ Applies the same (random) function to each image patch

↓

Spatially **correlated** activations

Corresponding GP

\- Applies a different random function to each image patch

(Locally connected network, LeCun, 1989. Noted by Novak et al. (2019)

↓

Spatially **uncorrelated** activations

Can we avoid this?

This paper: convolutions

Bayesian convolutional NN		Corresponding GP
Spatial correlation in weight prior	→	Spatial correlation between activations in the ∞-width limit
D-dimensional weight convolution	→	2D-dimensional covariance tensor convolution

This paper: generalize

Independent weights

$$\mathrm{Cov} = \begin{array}{|c|c|c|c|}\hline 1 & & & \\ \hline & 1 & & \\ \hline & & 1 & \\ \hline & & & 1 \\ \hline \end{array}$$

$\longleftrightarrow$

Interpolate

(stationary kernel
on positions)

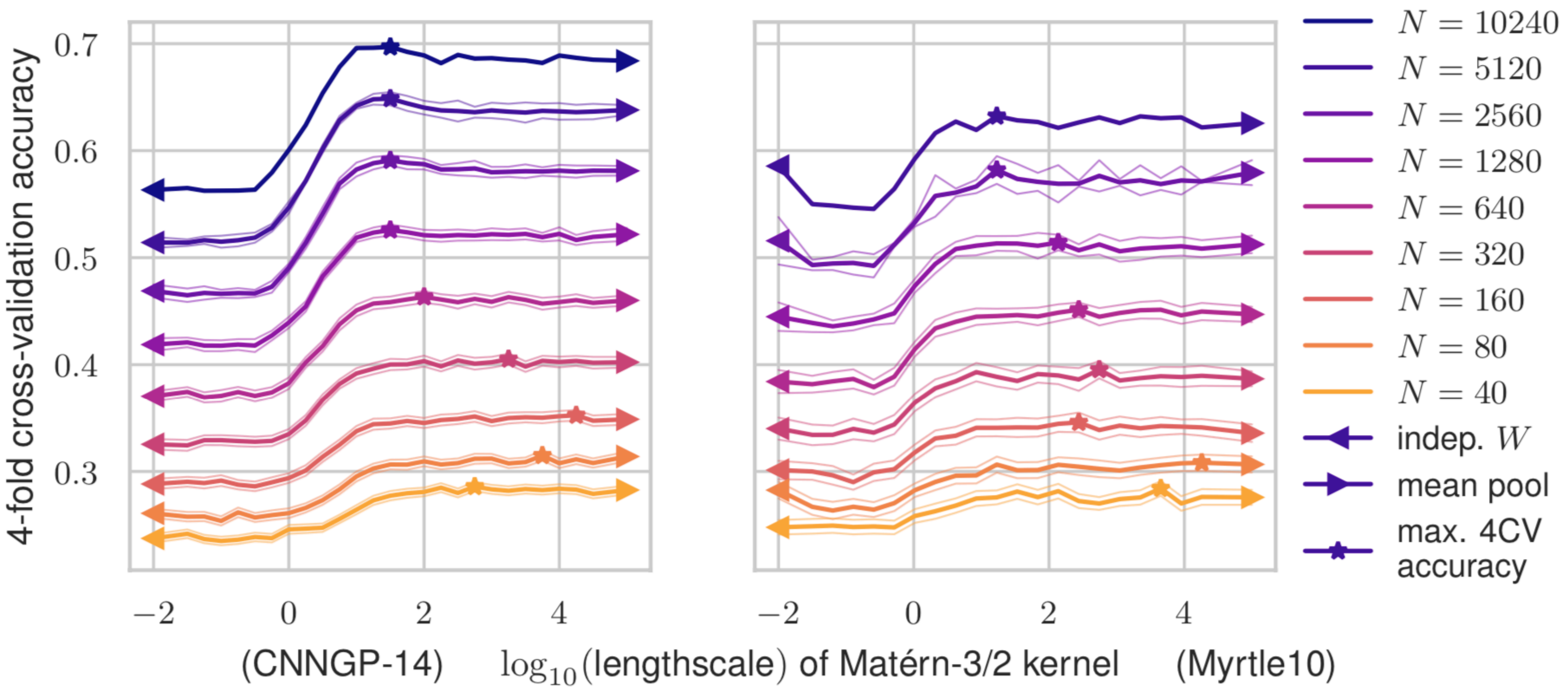
Mean pooling

$$\mathrm{Cov} = \frac{1}{16} \begin{array}{|c|c|c|c|}\hline 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 \\ \hline \end{array}$$

Convolutional Gaussian processes
(van der Wilk et al., 2017)

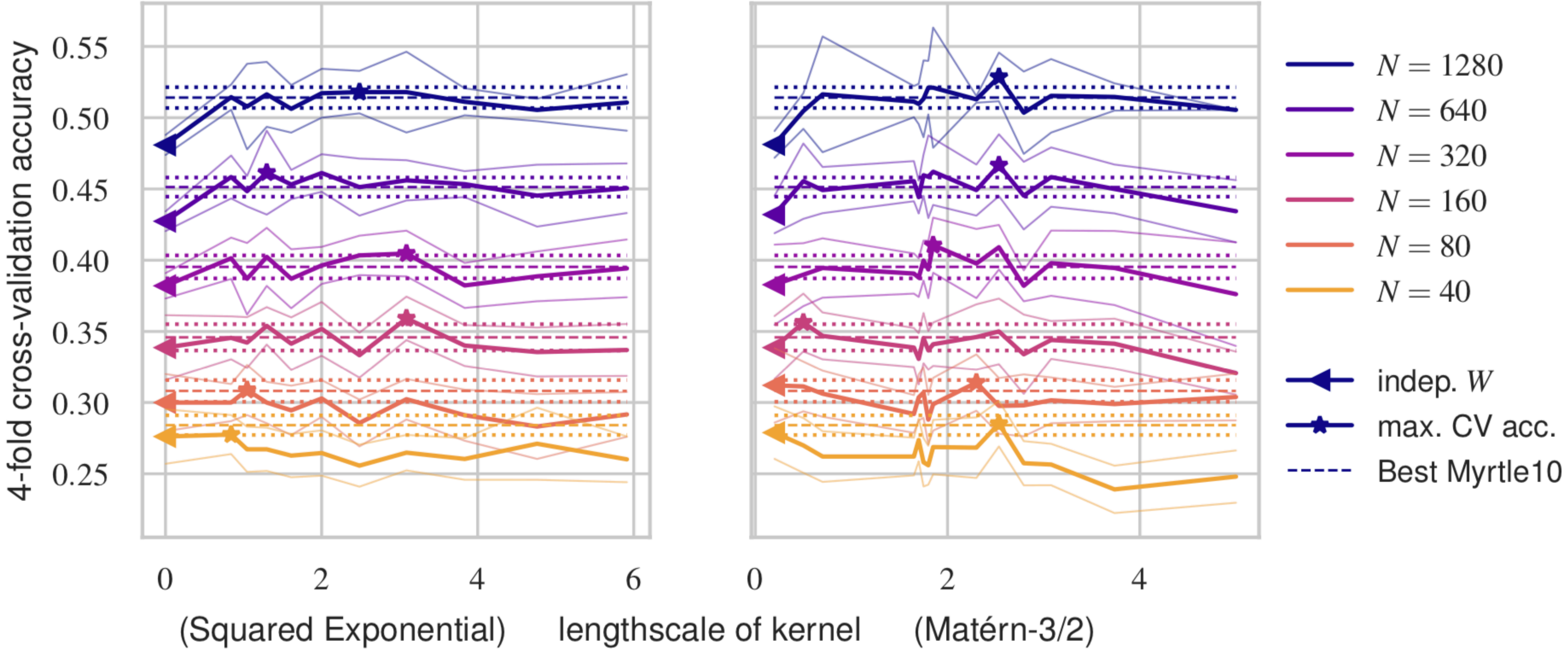
Experiments

Last layer correlation



Dataset: CIFAR-10

Intermediate layers correlation



Dataset: CIFAR-10. Network: Myrtle10 (Shankar et al. 2020).
Replace mean-pooling and convolution by just convolution.

Take-home message

- Infinite limit of independent-weight CNN has no spatial correlations.
  - Recover them with spatial correlations in the weights.
  - Successful prior for finite Bayesian NNs (Fortuin et al., 2021)

- Prior and kernel generalize existing full-independence and mean-pooling

- Competitive performance by tuning continuous
  "spatial correlation" parameters

References

- Greg Yang. Wide feedforward or recurrent neural networks of any architecture are Gaussian processes. NeurIPS 2019. http://papers.nips.cc/paper/9186-wide-feedforward-or-recurrent-neural-networks-of-any-architecture-are-gaussian-processes.pdf
- Radford M. Neal. Bayesian learning for neural networks, volume 118. Springer, 1996.
- David J. C. MacKay. Introduction to Gaussian processes. In C. M. Bishop, editor, Neural Networks and Machine Learning, NATO ASI Series, pages 133–166. Kluwer Academic Press, 1998.
- Yann LeCun. Generalization and network design strategies. Connectionism in perspective, 19:143–155, 1989.
- Roman Novak, Lechao Xiao, Yasaman Bahri, Jaehoon Lee, Greg Yang, Jiri Hron, Daniel A. Abolafia, Jeffrey Pennington, and Jascha Sohl-Dickstein. Bayesian deep convolutional networks with many channels are Gaussian processes. ICLR 2019. https://openreview.net/forum?id=B1g30j0qF7
- Mark van der Wilk, Carl Edward Rasmussen, and James Hensman. Convolutional Gaussian processes. NeurIPS 2017. http://papers.nips.cc/paper/6877-convolutional-gaussian-processes.pdf
- Shankar, V., Fang, A., Guo, W., Fridovich-Keil, S., Ragan-Kelley, J., Schmidt, L., & Recht, B. (2020, November). Neural kernels without tangents. In International Conference on Machine Learning (pp. 8614-8623). PMLR.
- Fortuin, V., Garriga-Alonso, A., Wenzel, F., Rätsch, G., Turner, R., van der Wilk, M., & Aitchison, L. (2021). Bayesian neural network priors revisited. https://arxiv.org/abs/2102.06571