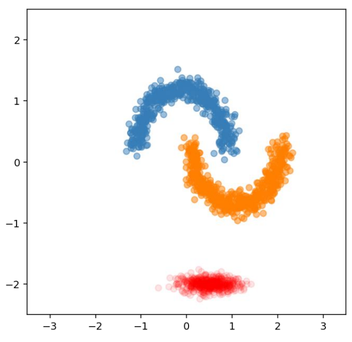
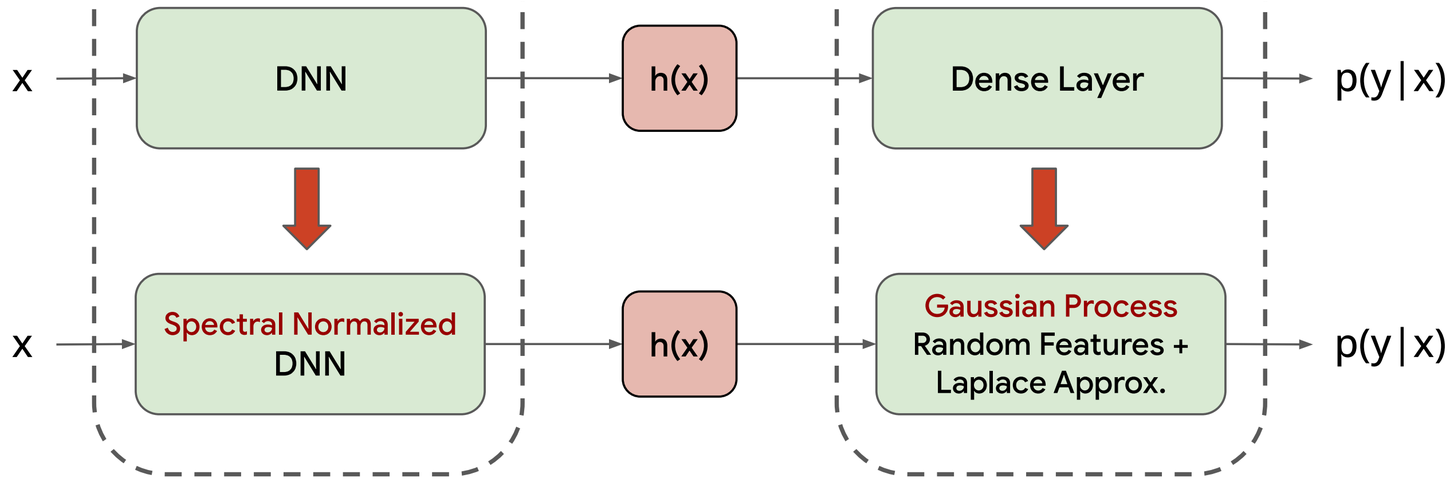
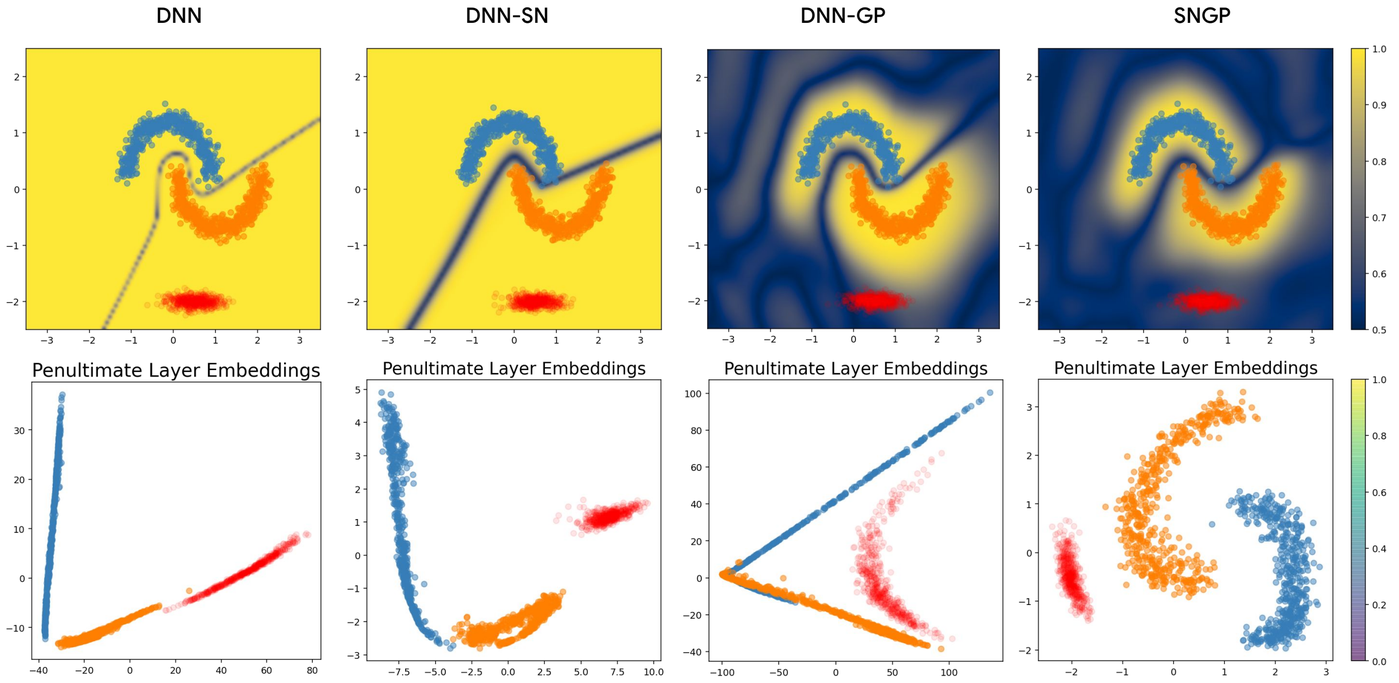
A Simple Approach to Improve Single-Model Deep Uncertainty via Distance-Awareness

Jeremiah Liu*, Shreyas Padhy*, Jie Ren*, Zi Lin, Yeming Wen, Ghassen Jerfel, Zack Nado, Jasper Snoek, Dustin Tran, Balaji Lakshminarayanan

UNIVERSITY OF CAMBRIDGE Google Research







Distance-Preservation w/ Spectral Normalisation

$$L_1 \left\| \mathbf{x}_1 - \mathbf{x}_2 \right\|_X \leq \left\| h(\mathbf{x}_1) - h(\mathbf{x}_2) \right\|_H \leq L_2 \left\| \mathbf{x}_1 - \mathbf{x}_2 \right\|_X$$

⇩

- Bi-Lipschitz Condition on residual layers is easily satisfied by Spectral Norm $\|\mathbf{W}\|_2 \leq c$
  - Perform $\mathbf{W}_l = c\dfrac{\mathbf{W}_l}{\|\mathbf{W}_l\|}$

Distance-Awareness w/ Gaussian Processes

$$g_{N\times 1} \sim N\left(\mathbf{0}_{N\times 1}, \sigma^2 * \mathbf{K}_{N\times N}\right), \mathbf{K}_{i,j} = \exp\left(-\left\| h_i - h_j \right\|_2^2 / 2\right)$$

⇩

- Approximations for scalability -
  - Random Fourier Feature expansions
    - $\mathbf{K} \approx \Phi\Phi^T \quad \rightarrow \quad g(h_i) \approx \Phi(h_i)^T \mathbf{W}_L$
  - Laplace Approximation for Posterior
    - $p(\mathbf{W}_L \,|\, \mathcal{D}) \approx \mathcal{N}(\mathbf{W}_{L,\text{MAP}}, \hat{\boldsymbol{\Sigma}})$

Competitive on many benchmarks

- Vision: CIFAR10/100, ImageNet (ResNet-50)
- NLP: Intent Detection (BERT)
- Genomics: Sequence Prediction

Method	Accuracy (↑) Clean	Accuracy (↑) Corrupted	ECE (↓) Clean	ECE (↓) Corrupted	NLL (↓) Clean	NLL (↓) Corrupted
Single Model						
DNN	76.2 ± 0.01	40.5 ± 0.01	0.032 ± 0.002	0.103 ± 0.011	0.939 ± 0.01	3.21 ± 0.02
DNN-SN	**76.4 ± 0.01**	40.6 ± 0.01	0.079 ± 0.001	0.074 ± 0.001	0.96 ± 0.01	3.14 ± 0.02
DNN-GP	76.0 ± 0.01	**41.3 ± 0.01**	0.017 ± 0.001	0.049 ± 0.001	0.93 ± 0.01	3.06 ± 0.02
SNGP (Ours)	76.1 ± 0.01	41.1 ± 0.01	**0.013 ± 0.001**	**0.045 ± 0.012**	**0.93 ± 0.01**	**3.03 ± 0.01**
Ensemble Model						
MC Dropout	76.6 ± 0.01	42.4 ± 0.02	0.026 ± 0.002	**0.046 ± 0.009**	0.919 ± 0.01	2.96 ± 0.01
Deep Ensemble	77.9 ± 0.01	**44.9 ± 0.01**	**0.017 ± 0.001**	0.047 ± 0.004	0.857 ± 0.01	2.82 ± 0.01
SNGP Ensemble (Ours)	**78.1 ± 0.01**	**44.9 ± 0.01**	0.039 ± 0.001	0.050 ± 0.002	**0.851 ± 0.01**	**2.77 ± 0.01**

SNGP is an uncertainty building block!

- Orthogonal Performance to other methods
  - Ensembling: Deep Ensembles, MC Dropout
  - Augmentations: AugMix