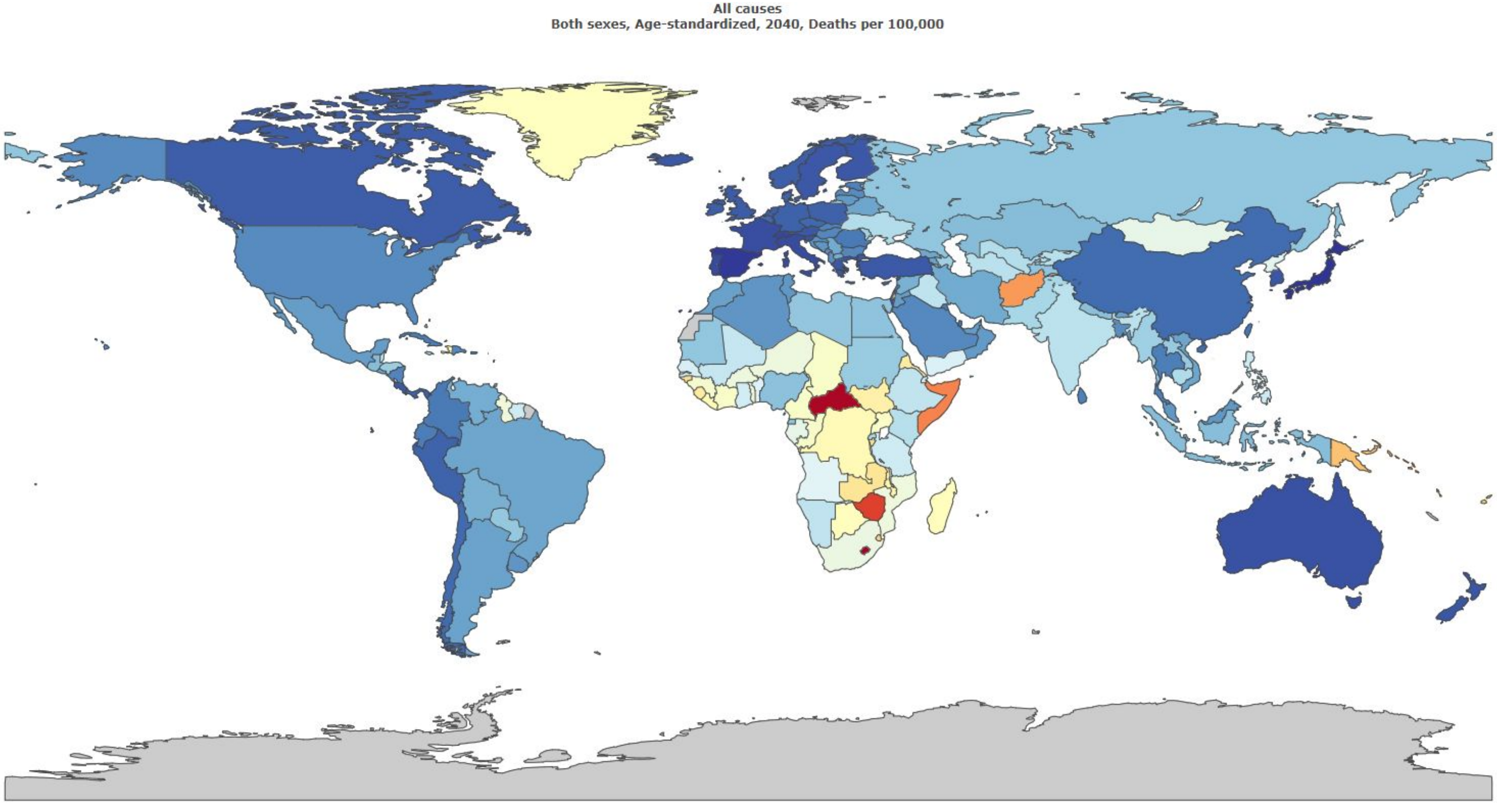
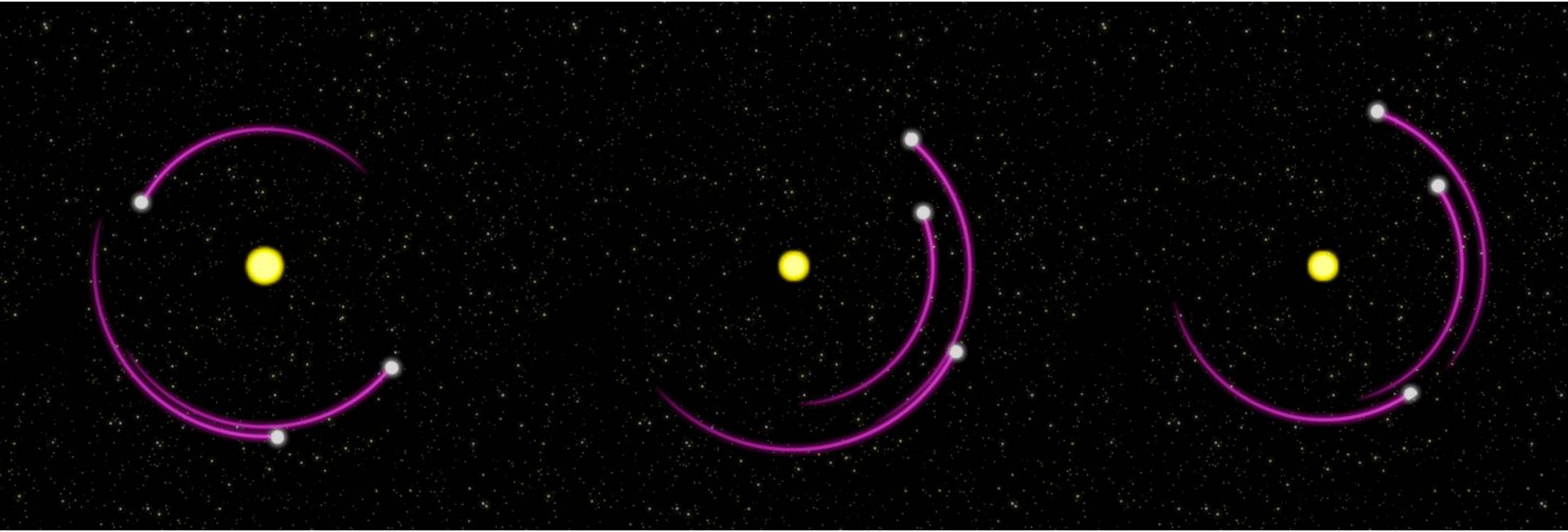
A look at Bayesian Deep Learning

Kyle Heuton
6/11/2021

Do you care about uncertainty?

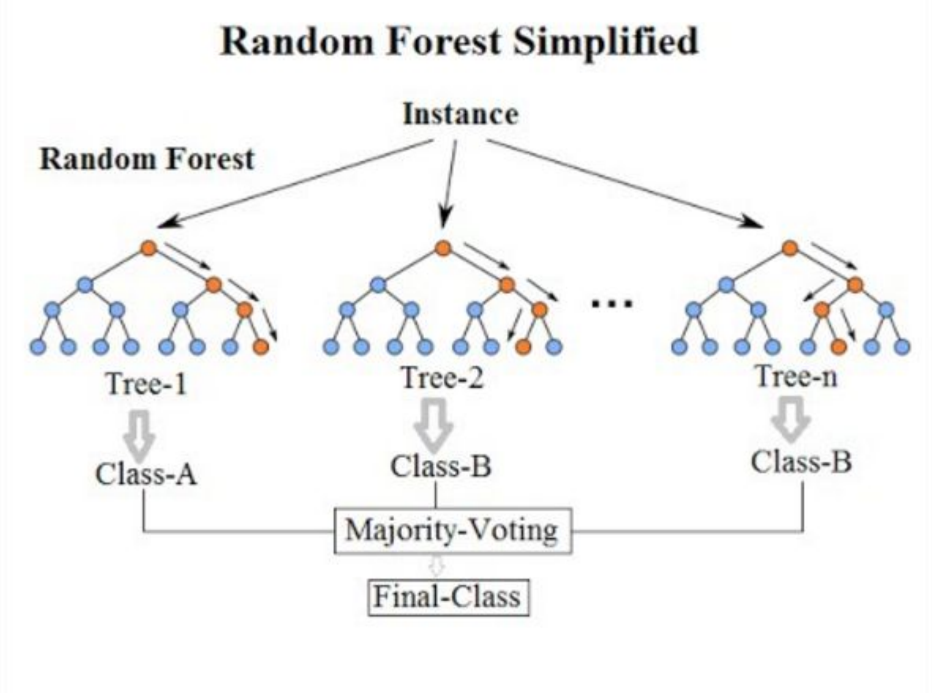


The problem



What Previous approaches have been tried?

$$F_g = \frac{G m_1 m_2}{r^2}$$



Try Bayesian Deep Learning

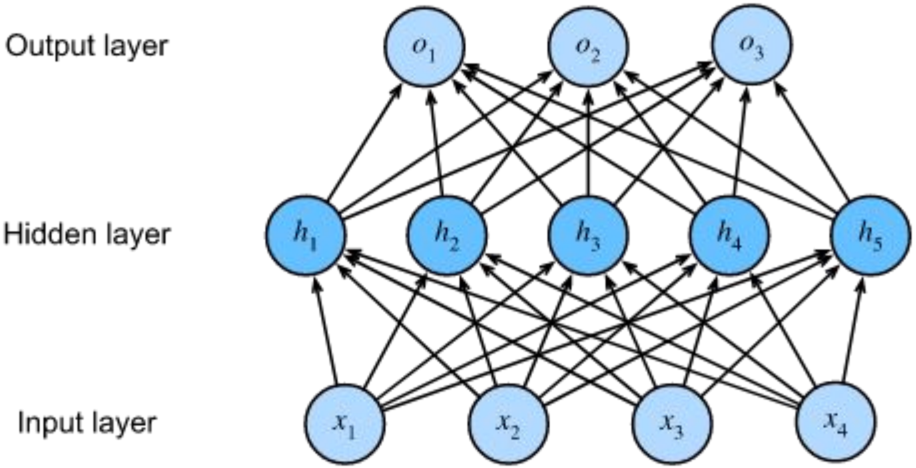
Bayesian Deep Learning Marginalizes over parameters to give:

- Accurate uncertainty quantification
- Out-of-distribution generalization

A Bayesian neural network predicts the dissolution of compact planetary systems

Miles Cranmer[a,†], Daniel Tamayo[a], Hanno Rein[b,c], Peter Battaglia[d], Samuel Hadden[e], Philip J. Armitage[f,g], Shirley Ho[g,a,h], and David N. Spergel[g,a]

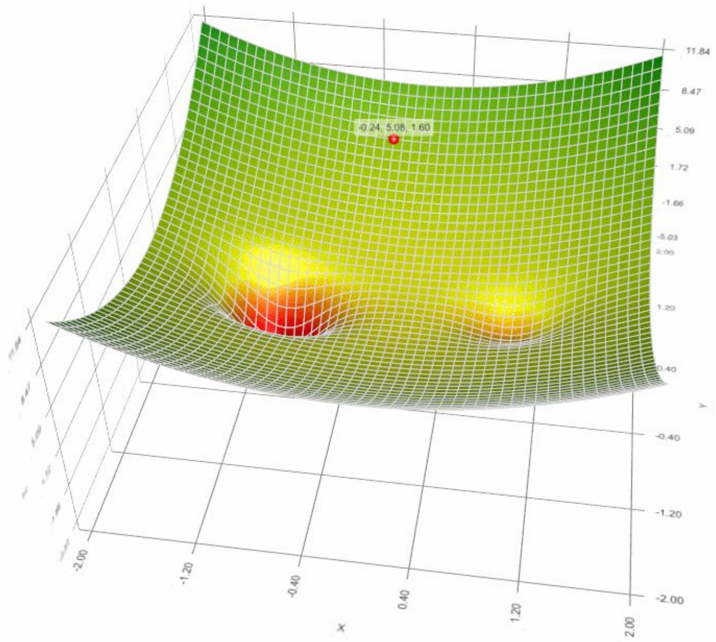
What is deep learning



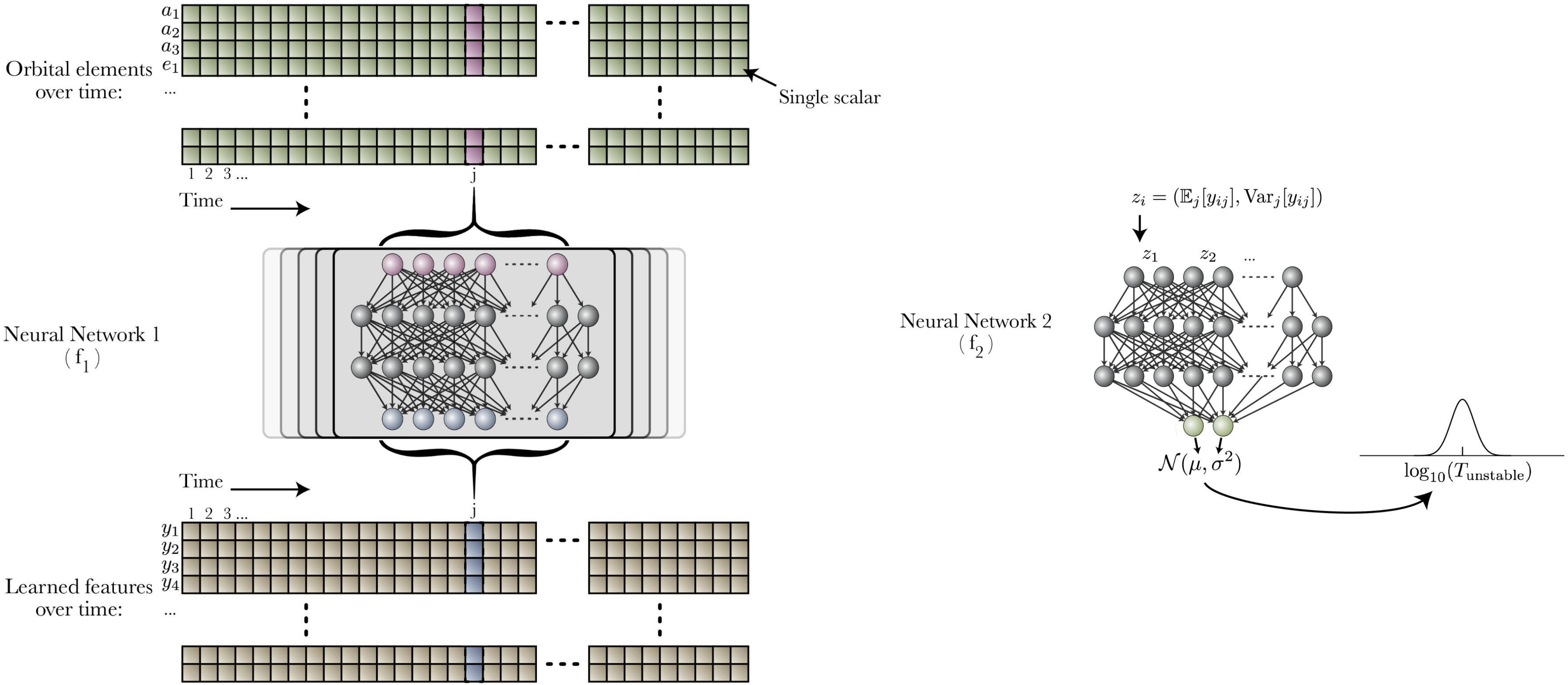
$$\mathbf{H} = \sigma(\mathbf{X}\mathbf{W}^{(1)} + \mathbf{b}^{(1)}),$$
$$\mathbf{O} = \mathbf{H}\mathbf{W}^{(2)} + \mathbf{b}^{(2)}.$$

http://d2l.ai/chapter_multilayer-perceptrons/mlp.html

Can train these parameters with Stochastic Gradient Descent



Miles's Model



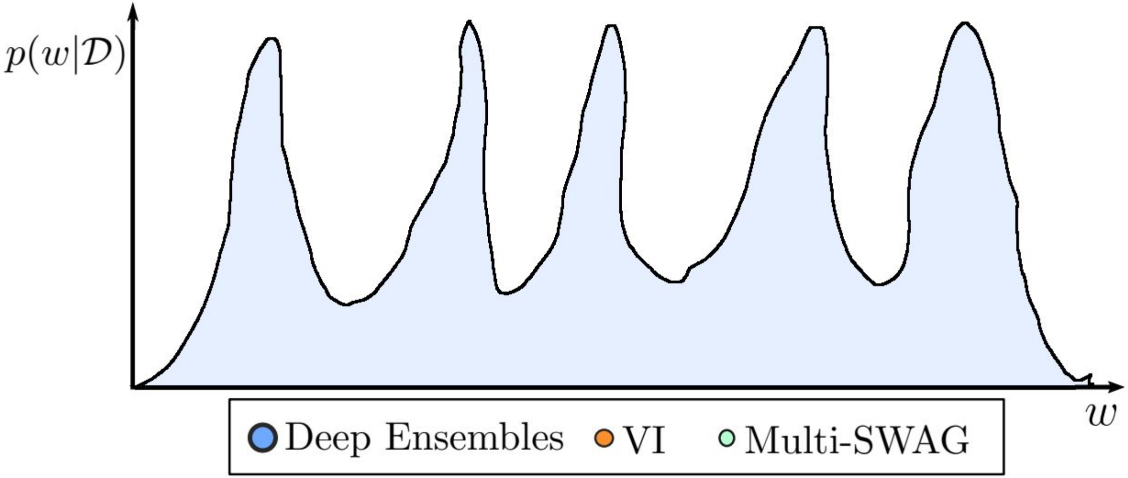
How do you do *Bayesian* deep learning?

$$p(y|x, \mathcal{D}) = \int p(y|x, w)p(w|\mathcal{D})dw$$

$$p(w|\mathcal{D}) = \frac{p(\mathcal{D}|w)p(w)}{p(\mathcal{D})}$$

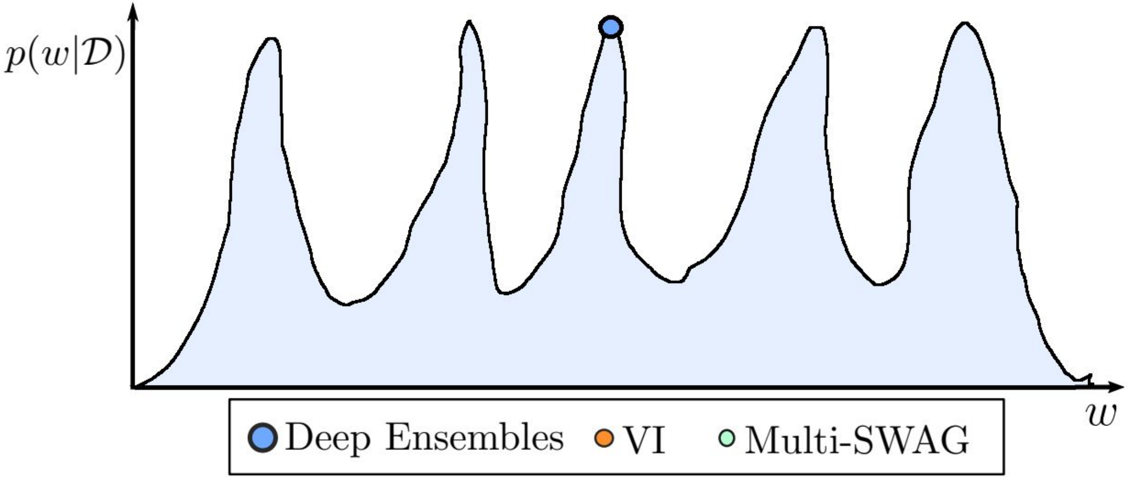
How do you get samples from the posterior over parameters?

$$p(w|\mathcal{D}) = \frac{p(\mathcal{D}|w)p(w)}{p(\mathcal{D})}$$



How do you get samples from the posterior over parameters?

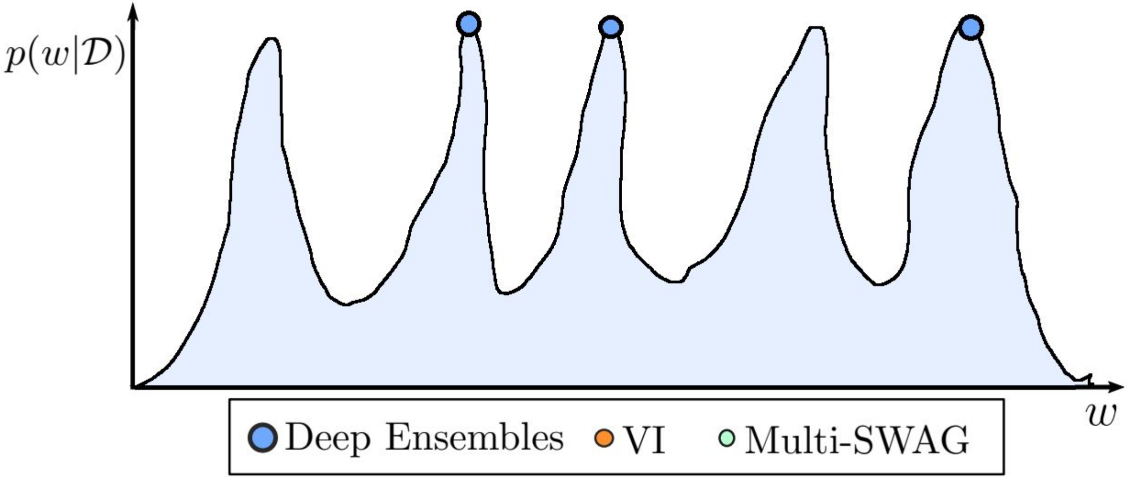
$$p(w|\mathcal{D}) = \frac{p(\mathcal{D}|w)p(w)}{p(\mathcal{D})}$$



$$\hat{w} = \arg\max_w \log p(w|\mathcal{D}) = \arg\max_w (\log p(\mathcal{D}|w) + \log p(w) + \text{constant}).$$

How do you get samples from the posterior over parameters?

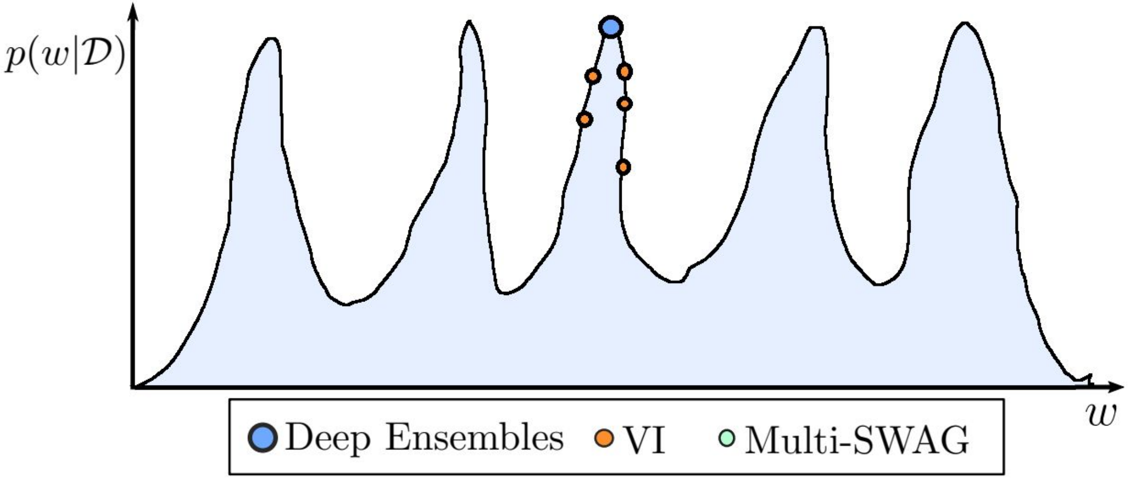
$$p(w|\mathcal{D}) = \frac{p(\mathcal{D}|w)p(w)}{p(\mathcal{D})}$$



$$\hat{w} = \arg\max_w \log p(w|\mathcal{D}) = \arg\max_w (\log p(\mathcal{D}|w) + \log p(w) + \text{constant}).$$

How do you get samples from the posterior over parameters?

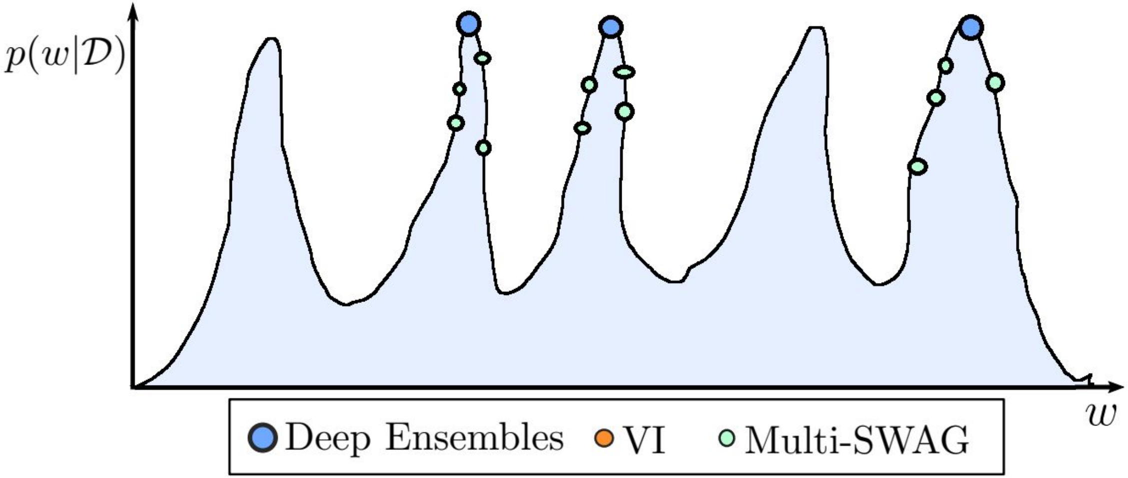
$$p(w|\mathcal{D}) = \frac{p(\mathcal{D}|w)p(w)}{p(\mathcal{D})}$$



$$\mathcal{F}(\boldsymbol{\theta}) = \mathop{\mathbb{E}}_{W \sim q_{\boldsymbol{\theta}}}[\log p(\mathcal{D} \mid W)] - \mathrm{D_{KL}}(q_{\boldsymbol{\theta}} \,\|\, p).$$

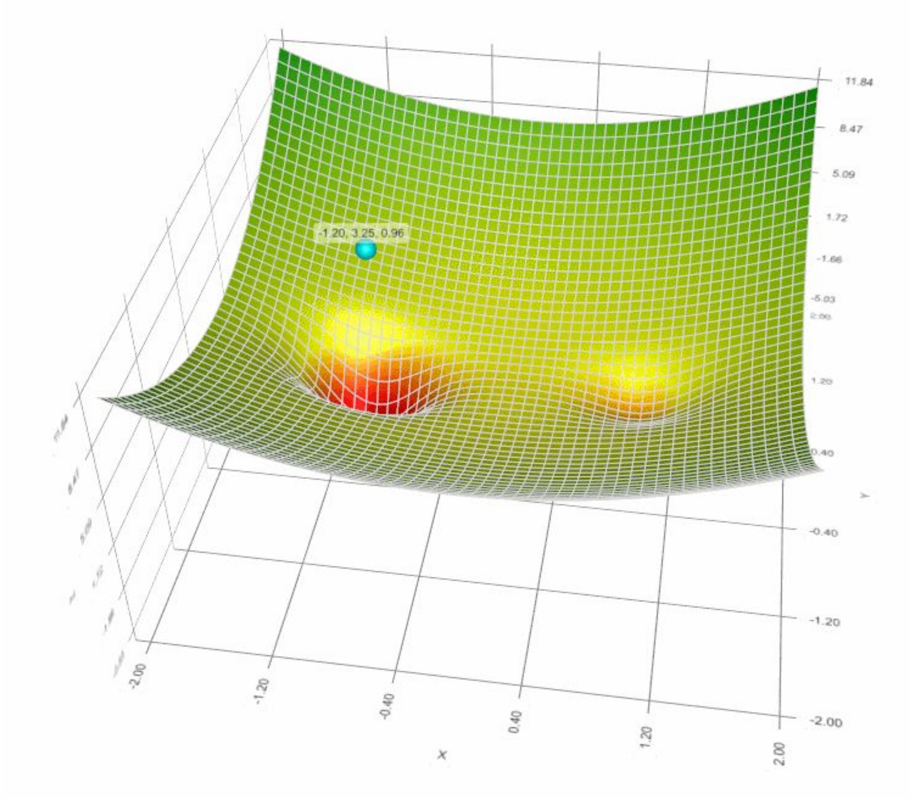
How do you get samples from the posterior over parameters?

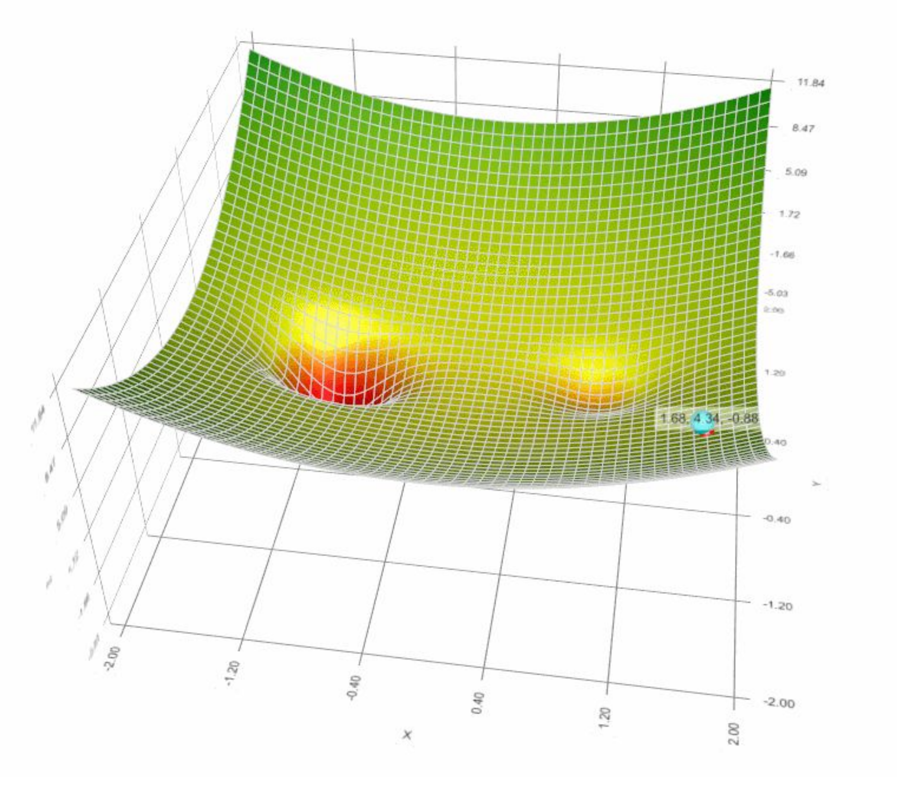
$$p(w|\mathcal{D}) = \frac{p(\mathcal{D}|w)p(w)}{p(\mathcal{D})}$$



Posterior is a mixture of Gaussians

MultiSWAG is easy

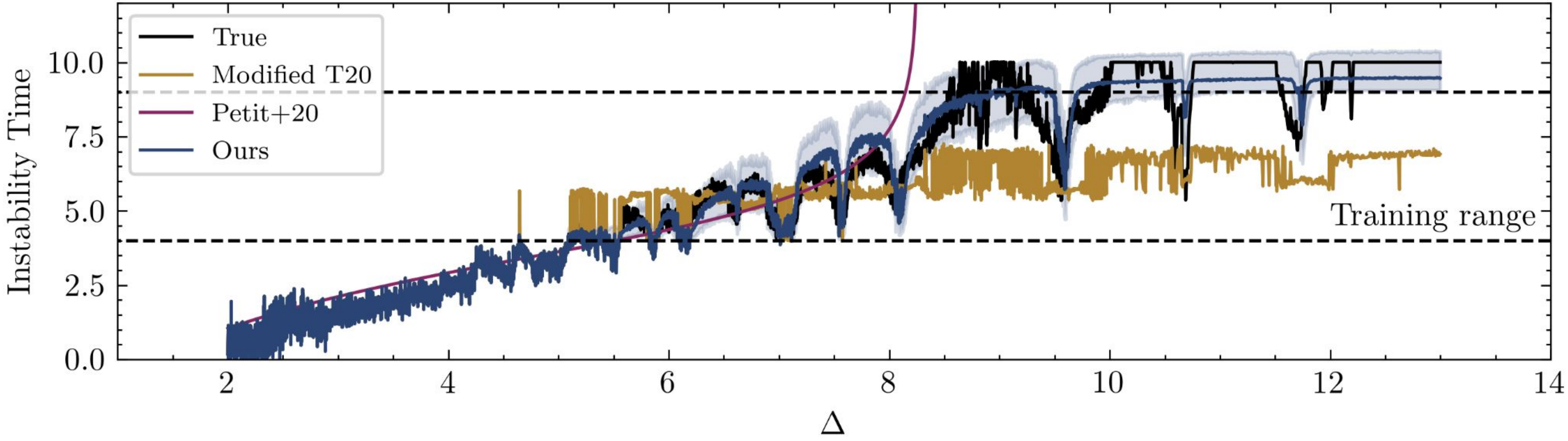




What does Bayesian Deep Learning get you?

Model	RMSE	Classif. accur.	Bias† (4, 5)	Bias (8, 9)
Obertas et al. (2017)	2.12	0.628	1.04	-1.71
Petit et al. (2020)	3.22	0.530	3.99	0.54

Bayesian Deep Learning can be robust to out-of-distribution data



Resources

- Notebook building a Bayesian Neural Net with Bayes by backprop (warning, old tensorflow): https://nbviewer.jupyter.org/github/krasserm/bayesian-machine-learning/blob/dev/bayesian-neural-networks/bayesian_neural_networks.ipynb
- Complicated Notebook for BNNs and uncertainty estimates: https://nbviewer.jupyter.org/github/krasserm/bayesian-machine-learning/blob/dev/noise-contrastive-priors/ncp.ipynb
- Notebook building a Bayesian Neural Net with PyMC3 and VI:
- https://docs.pymc.io/notebooks/bayesian_neural_network_advi.html
- The Case for Bayesian Deep Learning: https://cims.nyu.edu/~andrewgw/caseforbdl/
- Bayesian Deep Learning and a Probabilisitic Perspective of Generalization: https://arxiv.org/abs/2002.08791
- NeurIPS Bayesian Deep Learning workshop: http://bayesiandeeplearning.org/
- Approximate Bayesian Inference Competition: https://izmailovpavel.github.io/neurips_bdl_competition/
- "Hands-on" Bayesian Deep Learning Tutorial https://arxiv.org/pdf/2007.06823.pdf
- Dropout as Bayesian Deep Learning- http://mlg.eng.cam.ac.uk/yarin/blog_3d801aa532c1ce.html
- Good, interactive deep learning book: http://d2l.ai/index.html
- Good, conceptual deep learning book: https://www.deeplearningbook.org/