Using Learning Dynamics to Understand Neural Network Generalisation

Arnu Pretorius*

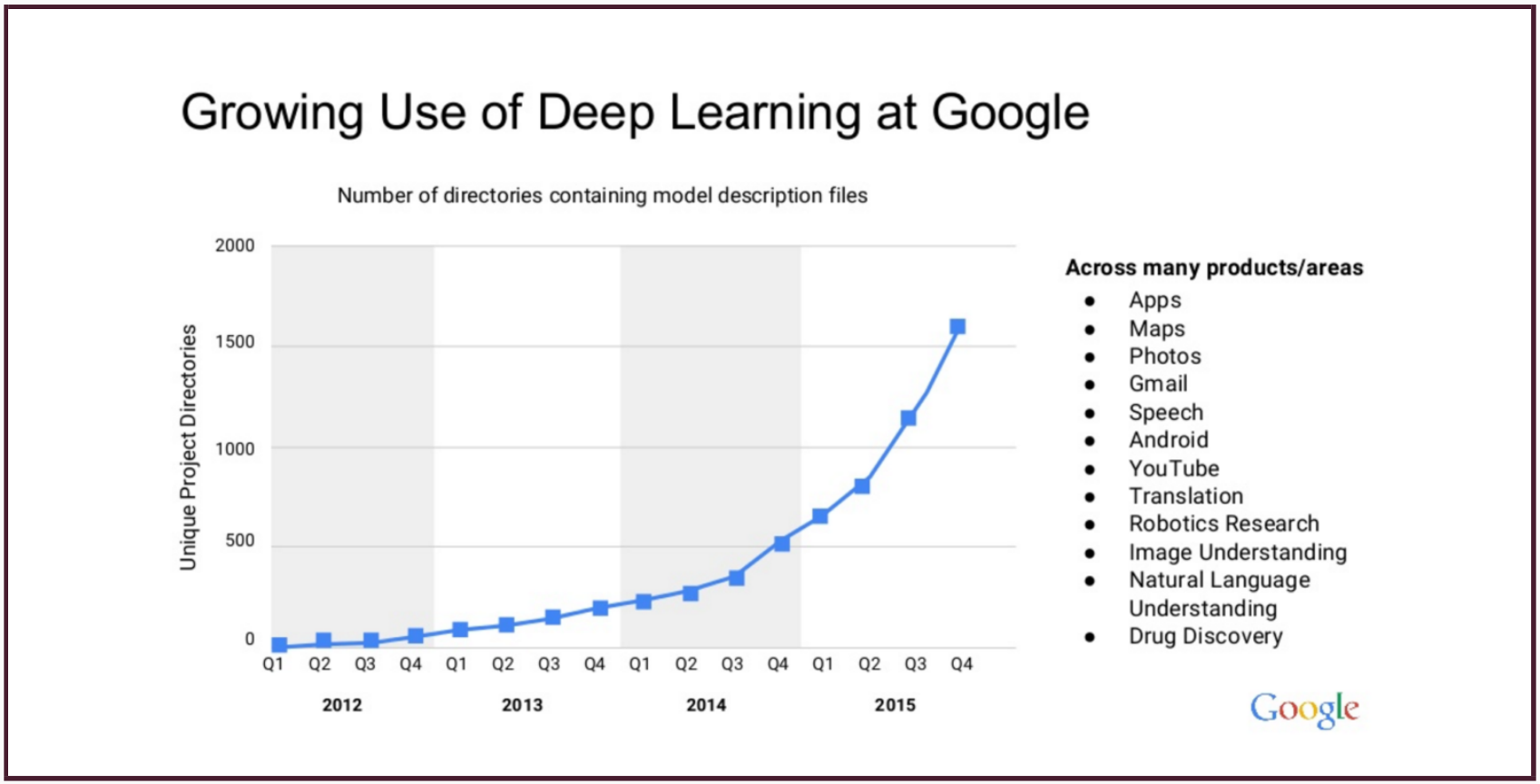
Supervisor: Dr. Steve Kroon (Computer Science*)
Co-supervisor: Dr. Herman Kamper (EE Engineering+)

20 October 2017

Maties Machine Learning, Stellenbosch University

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

The success of Deep Learning

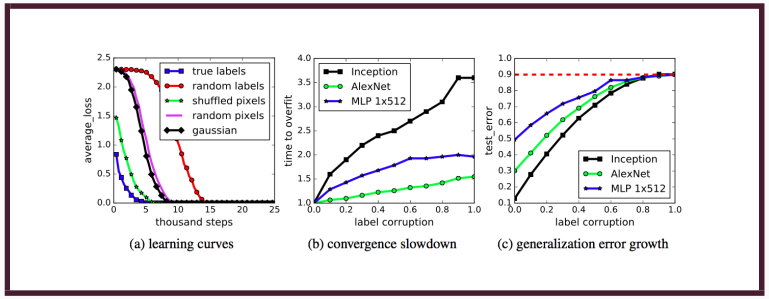


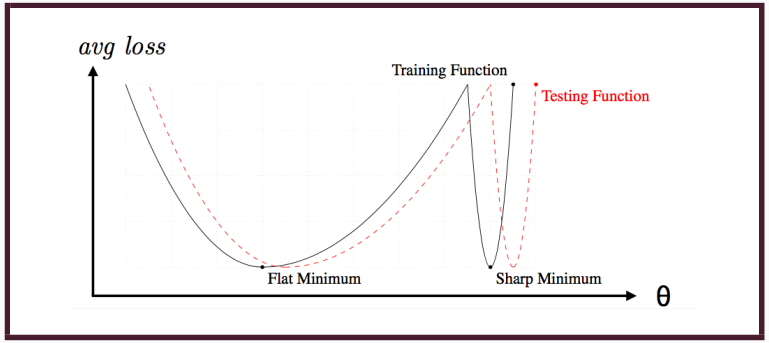
Scale of deep learning deployment at Google over time.

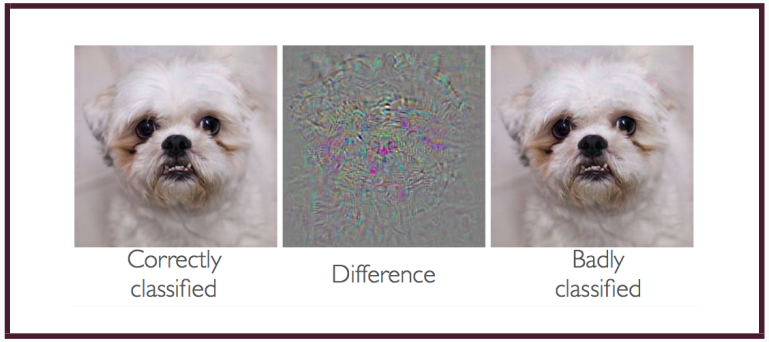
Large-scale deep learning for intelligent computer systems, Dean. BayLearn keynote speech, 2015.

The mysteries of Deep Learning

- Generalisation
- The role of explicit versus implicit regularisation
- Non-convex optimisation
- Adversarial examples
- and more ...





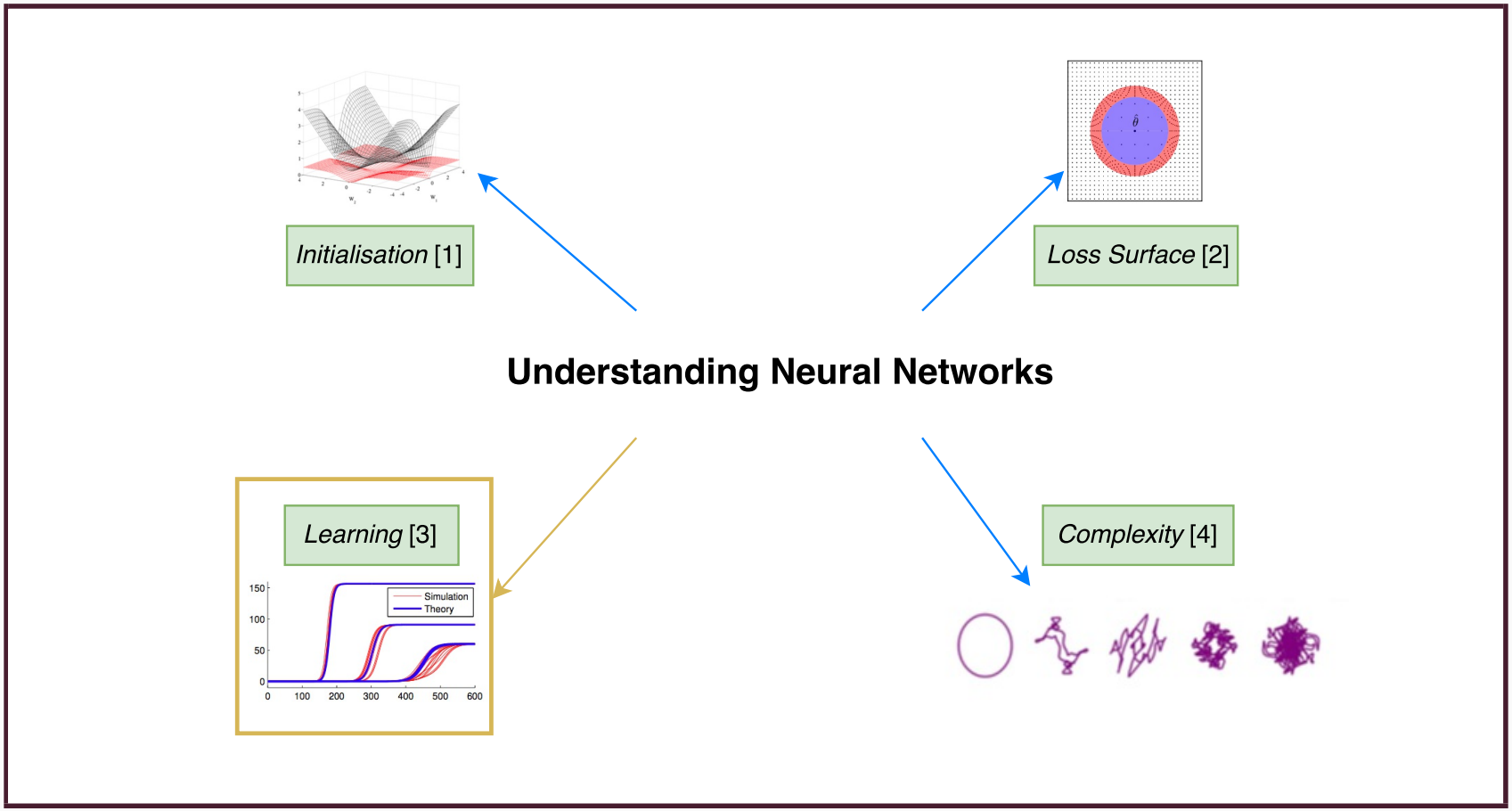


Understanding deep learning requires rethinking generalization, Zhang, Bengio, Hardt, Recht, Vinyals. ICLR, 2017.

On large-batch training for deep learning: generalization gap and sharp minima, Keskar, Mudigere, Smelyanskiy, Tang. ICLR, 2017.

Intriguing Properties of Neural Networks, Szegedy, Zaremba, Sutskever, Bruna, Erhan, Goodfellow, Fergus. ICLR, 2014.

Approaches Towards Understanding Neural Networks



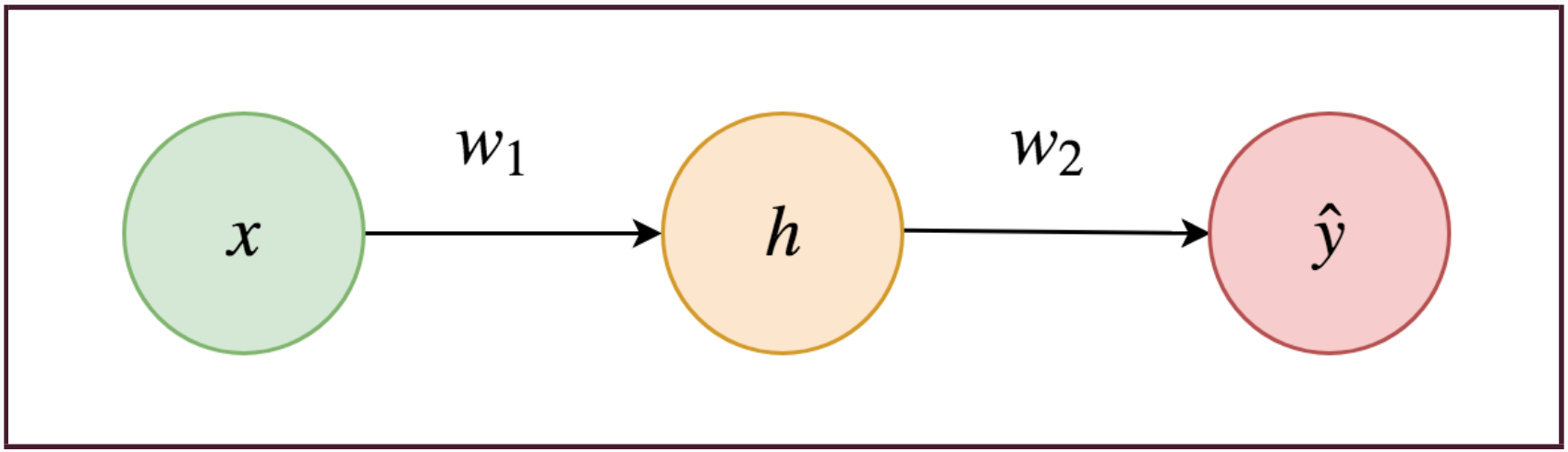
Approaches towards understanding neural networks.

[1] *Understanding the difficulty of training deep feedforward neural networks*, Glorot, Bengio. AISTATS, 2010.
[2] *Sharp minima can generalize for deep nets*, Dinh, Pascanu, Bengio, Bengio. arXiv 2017.
[3] *Exact solutions to the nonlinear dynamics of learning in deep linear neural networks*, Saxe, McClelland, Ganguli. ICLR, 2014.
[4] *On the Expressive Power of Deep Neural Networks*, Raghu, Poole, Kleinberg, Ganguli, Dickstein. ICML, 2017.

Scalar linear neural network

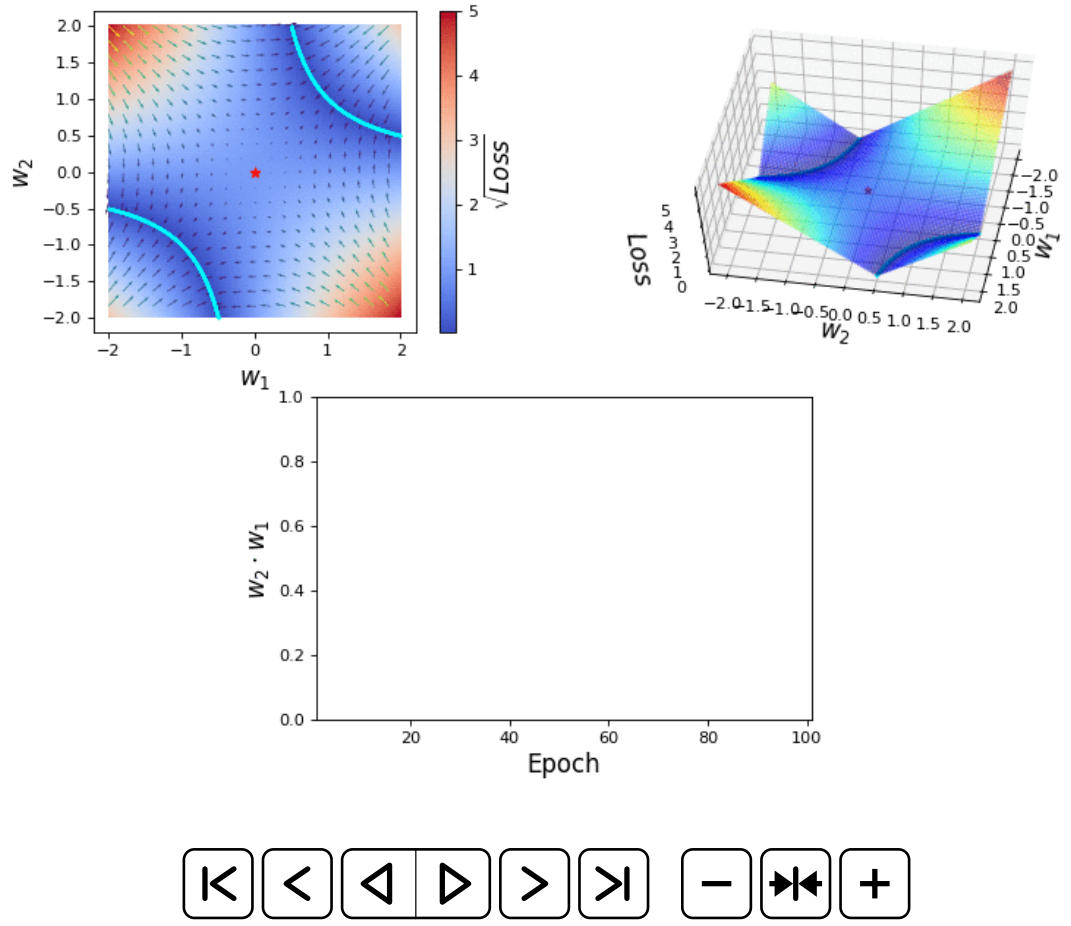
Let $w_2, w_1, x \in \mathbb{R}$,

$$\hat{y} = w_2 w_1 x \tag{1}$$



Scalar linear neural network.

Scalar learning dynamics



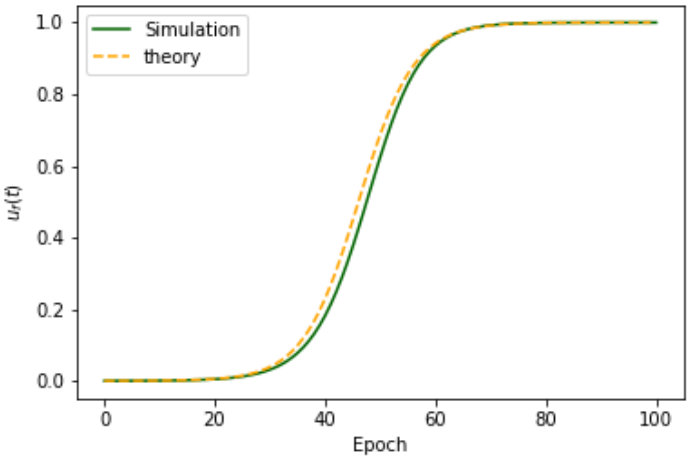
Loss surface $(1 - w_2 w_1)^2$, GD path (orange), learning dynamics (green).

Exact solutions for scalar learning dynamics

The learning dynamics for a scalar linear neural network starting from small initial values $u_0 \equiv w_{2(0)} w_{1(0)}$ is given by

$$u_f(t) = \frac{yE}{x(E-1) + y/u_0}, \tag{2}$$

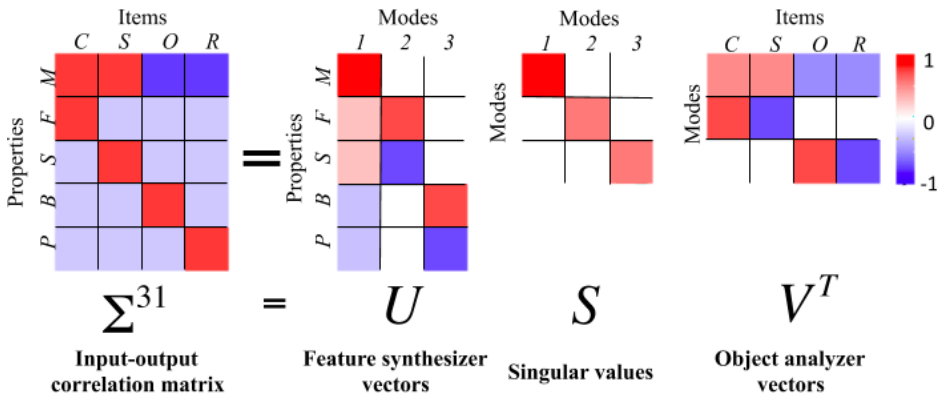
where $E = e^{2yxt\alpha}$, α is the learning rate and t is measured in epochs.

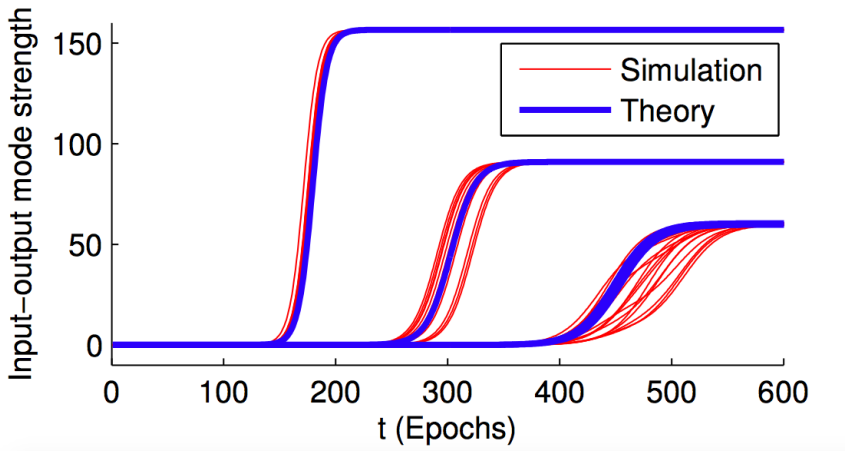


Simulated versus theoretical learning dynamics.

Exact solutions to the nonlinear dynamics of learning in deep linear neural networks, Saxe, McClelland, Ganguli. ICLR, 2014.

Learning dynamics for deep linear networks





Linear neural network learning dynamics.

Learning hierarchical category structure in deep neural networks, Saxe, McClelland, Ganguli. CogSci, 2013.

Our current focus

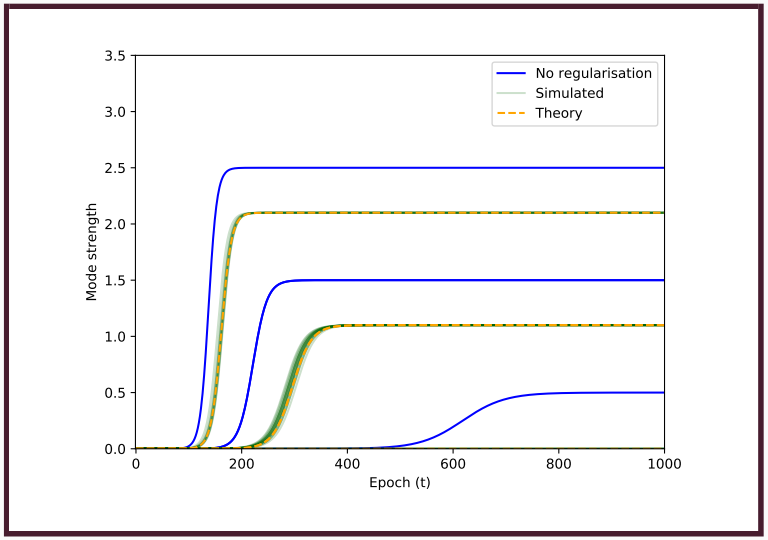
- The role of regularisation
  - *Weight decay*

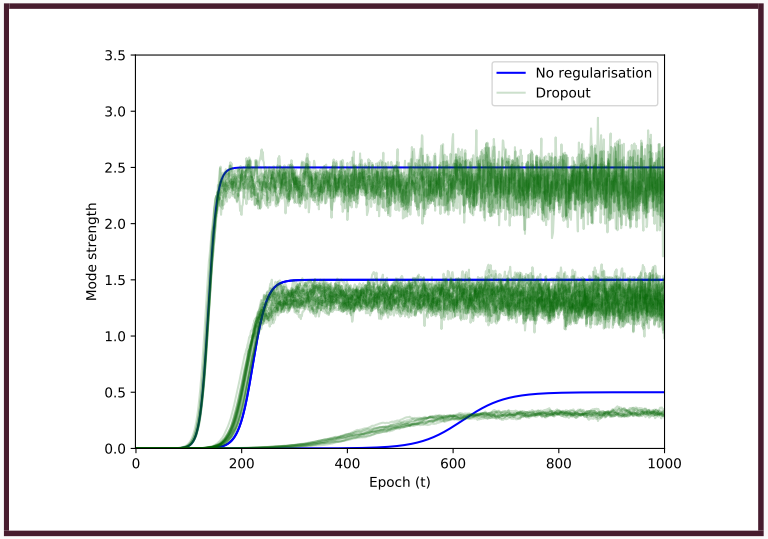
$$u(t, s, u_0, \lambda^*) = \frac{(s - \lambda^*)e^{2(s-\lambda^*)t/\tau}}{e^{2(s-\lambda^*)t/\tau} - 1 + (s - \lambda^*)/u_0}$$

  - *Dropout*

$$E_{d\sim Bern(p)}[t] \leq \frac{\tau}{sp} \ln\left(\frac{s}{\epsilon}\right) = \mathcal{O}\left(\frac{\tau}{sp}\right)$$

- Autoencoder networks
- Generalisation





Summary

- Deep learning has been hugely successful in solving large and complex machine learning task, however many mysteries remain.
- Better understanding deep neural networks might be achieved via several different routes.
- Studying the learning dynamics of neural networks may help us understand how neural networks learn.
- We hope to use this learning dynamics approach to study generalisation in deep neural networks.

Thank you for listening!