

# Noisy Recurrent Neural Networks

Soon Hoe Lim

KTH Royal Institute of Technology, Stockholm, Sweden

In this talk, we discuss a general framework for studying recurrent neural networks (RNNs) trained by injecting noise into hidden states. Specifically, we consider RNNs that can be viewed as discretizations of stochastic differential equations driven by input data. This framework allows us to study the implicit regularization effect of general noise injection schemes by deriving an approximate explicit regularizer in the small noise regime. We find that, under reasonable assumptions, this implicit regularization promotes flatter minima; it biases towards models with more stable dynamics; and, in classification tasks, it favors models with larger classification margin. Sufficient conditions for global stability are obtained, highlighting the phenomenon of stochastic stabilization, where noise injection can improve stability during training. Our theory is supported by empirical results which demonstrate that the RNNs have improved robustness with respect to various input perturbations. This is joint work with N. Benjamin Erichson, Liam Hodgkinson and Michael Mahoney.