Effects of noise injection in artificial neural networks

Ben Poole, Jascha Sohl-Dickstein, Surya Ganguli

Intrinsic variability, or noise, is an extremely salient feature of biophysical systems. Many studies have shown that neural systems have evolved to efficiently process information in the face of this noise. Yet the question still persists as to whether this noise could be utilized to solve hard computational problems. Our work builds on a growing body of evidence in machine learning that show the benefits of noise injection for learning representations in neural networks.

We first consider the case of injecting noise into the hidden layer of an autoencoder, a simple one hidden layer neural network used for representation learning. By analytically marginalizing out the noise, we derive a set of penalties that can be related to existing explicit regularization strategies for neural networks and autoencoders. In particular, we find that injecting Poisson-like noise into the hidden layer yields a sparsity-inducing penalty. Thus Poisson variability in neural systems may provide a computationally simple mechanism for learning sparse representations.

We empirically evaluate the impact and utility of noise in learning by training neural networks for image denoising and digit classification. In both cases, we find that noise injection yields hidden representations that are sparser, more decorrelated, and more robust to random removal of neurons than their noiseless counterparts. Furthermore, the networks trained with noise injection yield improved performance on denoising natural image patches, and achieve state-of-the-art performance on a specific digit recognition task. These results indicate that noise can be beneficial in learning sparse distributed representations of sensory inputs.