

Supplemental Material: Overdispersed Variational Autoencoders

Harshil Shah and David Barber

Department of Computer Science
University College London

Neural network architectures

Below, we describe the architecture of all of the neural networks used for the parameters of the generative and variational distributions. Note that all hidden units used the tanh nonlinearity.

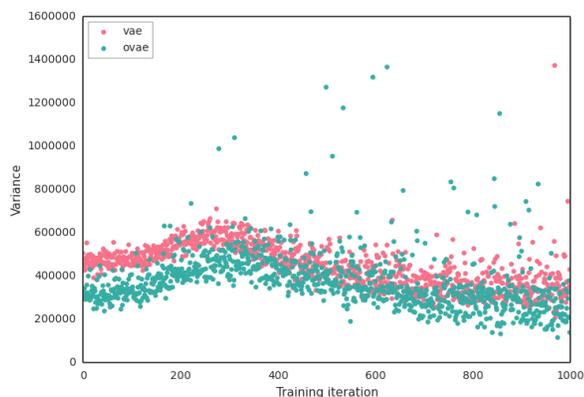
		Hidden layers	Output nonlinearity
MNIST	$\pi_{\theta}(\mathbf{h})$	2×200 units	sigmoid
	$\eta_{\phi}^{MNIST}(\mathbf{x})$	2×200 units	linear
	$\Omega_{\phi}^{MNIST}(\mathbf{x})$	2×200 units	exp
FF	$\mu_{\theta}(\mathbf{h})$	2×100 units	sigmoid
	$\Sigma_{\theta}(\mathbf{h})$	2×100 units	exp
	$\eta_{\phi}^{FF}(\mathbf{x})$	2×100 units	linear
	$\Omega_{\phi}^{FF}(\mathbf{x})$	2×100 units	exp

Variance of gradient estimates

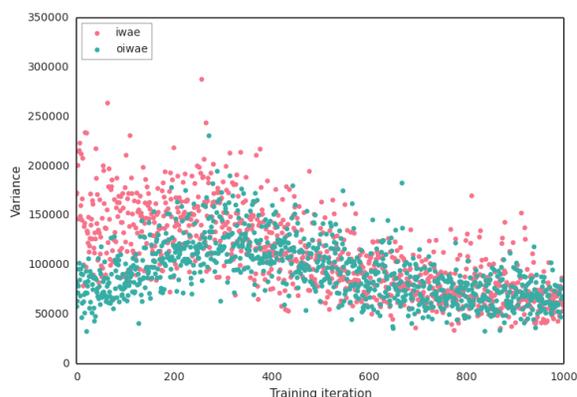
Below, we show the sample variances of the gradient estimates during the first 1,000 iterations of training, for both datasets.

MNIST

VAE vs. OVAE



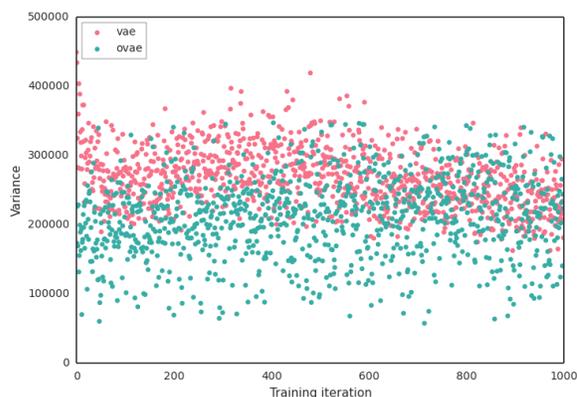
IWAE vs. OIWAE



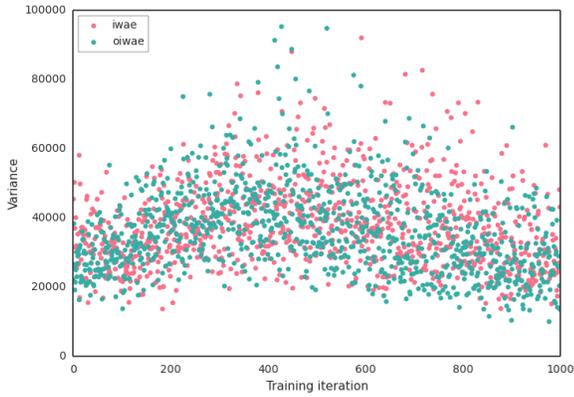
For the MNIST dataset, The OVAE gradient estimates have noticeably lower variance than do the VAE updates. Comparing the OIWAE against the IWAE, this is less clear, except in the first 300 training iterations.

Frey Faces

VAE vs. OVAE



IWAE vs. OIWAE



OVAE: Posterior



As with MNIST, for the Frey Faces dataset, the OVAE gradient estimates have noticeably lower variance than do the VAE updates. However, this is not the case for the OIWAE when compared against the IWAE, which have very similar variances.

OIWAE: Prior



Generated output

Below we show sample outputs generated from the prior and posterior of the learned models, for both datasets.

MNIST

OVAE: Prior



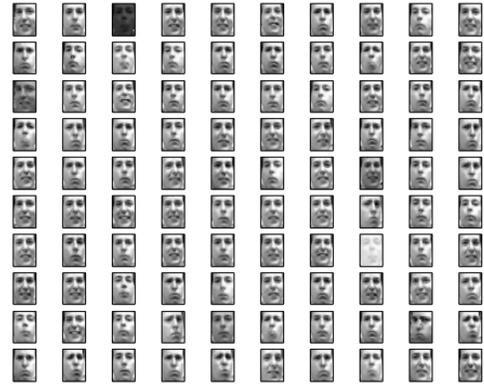
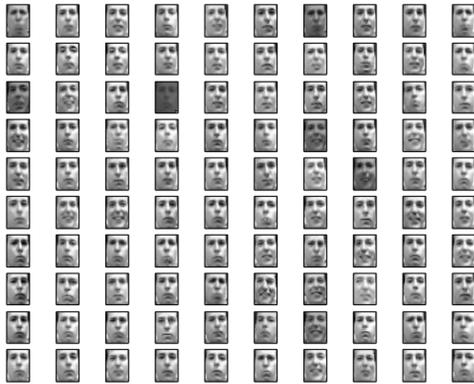
OIWAE: Posterior



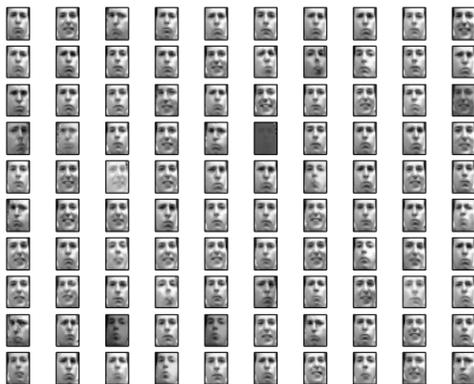
Frey Faces

OIWAE: Posterior

OVAE: Prior



OVAE: Posterior



OIWAE: Prior

