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 nonlinear-
			ity
ST	${m \pi}_{ heta}({f h})$	2×200 units	sigmoid
Ĕ	$\boldsymbol{\eta}_{\phi}^{MNIST}(\mathbf{x})$	2×200 units	linear
Σ	$\mathbf{\Omega}_{\phi}^{MNIST}(\mathbf{x})$	2×200 units	exp
FF	$\boldsymbol{\mu}_{ heta}(\mathbf{h})$	2×100 units	sigmoid
	$\Sigma_{\theta}(\mathbf{h})$	2×100 units	exp
	$\boldsymbol{\eta}_{\phi}^{FF}(\mathbf{x})$	2×100 units	linear
	$\mathbf{\Omega}_{\phi}^{FF}(\mathbf{x})$	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.

VAE vs. OVAE

MNIST



Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

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



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 OI-WAE when compared against the IWAE, which have very similar variances.

Generated output

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

OVAE: Prior

MNIST



OVAE: Posterior



OIWAE: Prior

\$	9	4	ŗ	7	9	0	Ζ	٤	G
6	ĩ	3	6	Y	2	3	2	5	3
ą	Z	B	3	6	1	3	0	3	Ø
9	1	2	9	Ч	1	0	7	1	Ŋ
5	9	5	0	¥	2	3	B	7	3
5	1	8	4	3	5	0	4	3	0
1	Ð	ß	С	3	Š	Э	1	4	9
4	ų	0	B	3	₽	9	4	Ĝ	Y
3	9	97	1	1	IJ	6	2	1	3
7	4	11	4	\$	Э	\mathcal{I}	9	0	1

OIWAE: Posterior



Frey Faces

10

2-0 문화 문화 문화 문·0 문·0 문·0 문·0

OIWAE: Posterior

10 m	114	26	100	150	1998	1) C	13.0	0.0	8-50
100	O a li	0.4	0.0	1)1	100	0 × (1)	1.0	100	10.0
8-15	8 × ((950	8 × (()	() C	8 × (i	a.a	150	9 × (f)	())(
20	1)< 0	1)1	8 × (()	13.8	15.0	の	100	8 × (i	() C 4
13.0	1)6	8 × (f)	1	13.8	0 . (L	ujs e		B = (()	1) e 4
13.8	13 B	8 (A)	100	8 × (i	a sta	8-5(T)	1 × (8)	1) + 6	0.6.
a sta	8 × (I)	1)48	100	8 » ((13.8	10 m	0.4	8 × (i)	調
1		8.A.	10	8 × (i	150	1)15	10 m	*	1) 5
0.0	13.8	0.6	1×10	10 × (1	-	1)(B	100	10	100
10	1)15	0 t ((1)0	1	の	1)	8 × (6	13 B	13.0

13+8 (1) +8 948 844 948 848 848 848 848 848 848 2+0 2+0 2+0 2+0 2+0 2+0 2+0 2+0 13.0 문사 문사 문사 문사 문사 문사 문사 문·태 문·대 문·대 문·대 문·대 문·대 문·대 11년 18년 18년 18년 18년 18년 18년 18년 100 000 000 000 000 000 000 000 000 1114 Ball 1940 3·등 3·등 3·등 3·등 3·등 3·등 3·등 1)< # 100 1) (1 + 1 1)*** [1]** 1) C 1/2 () C 8×8 1 × (t)

OVAE: Prior

OVAE: Posterior

00	100	1. Co	100	1)15	8×(i	100	100	19×10	13.0
()(c	10	a.c.	() e .		43(A	1	0.0	0 × (i)	10
50	() C .	8 a (i)	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	104	10.0	8 × (()	13.0	1 x (()	-
100	P.B.	B. C.	8.50 B	* x (l)	-	0 × (c)	1) v 8	100	0.50
8×(()		時間	13.8	1) C	1)0	a.u	1)(t)	13.8	1) e 8
\$ × (1)	13.0	13.8	1)4	0 × (()	8 × (()	13.0	8 × (()	100	1.00
	() C .	0.4.50	13.0	13.0	314	19. E	0.0	100	() C
a sta	-	1) + ¹	0.0	8 × (i)	2×(0)	101 C	8-5m	1)+ 10	1) 1 8
1.10	2.00	040	8 × (()	0.0	950	\$ x (t)	1)14	13.8	8 × (()
19.6	0)6	8 × (8	84.0	8 b (i	()) (-	130	9.6	19.40	11.4

OIWAE: Prior

199.0	8 a (t)	1) 1 B	194	100	1948	8- ft	a sta	848	10.	
194	1 AL	1) C	33.0	1) t .	0 × (i)	() C	15 a	We B	9 a (i	
20.00	9.14C	0 × (i)	19×6	194	na m	0.101	9+6	0 × (i)	3.(1)	
14 B	19.6	13.0	1) C	19 e	104	10	19.40	0 × (i	10	
B	100	8×(()	1)ca	8 × (i)	0.00	8+(s)	9.40	1 × (1)	100	
日の	948	a s a	200	0 . (1)	19.4 B	1000	1 a. (b)	1) C	150	
10 B	10 a 6	100	19.60	940	前の後	1)4 W	10	18 B	0 r.(i)	
	1 a (t)	1) + 6	194	0.54	aste.	33(6)	1)4 ⁸	2 × (R)	13 °	
1	1) v (1) 4 B	1) 4 ¹ / ₁	14 M	8 a. (i		3	10.0	0 . ft	
	arte	9×6	1120	Sec.	910	0.6	10.0	1.(1)	0.6	