Variational Auto-encoders: Representations for image generation and semi-supervised learning

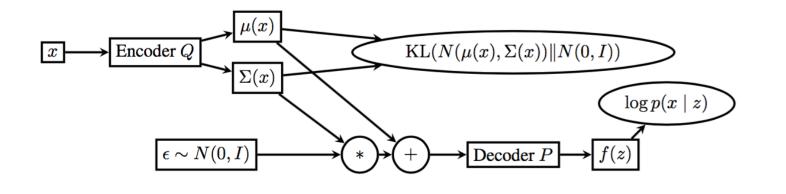
JIANBO CHEN, BILLY FANG, CHENG JU Departments of Statistics and Biostatistics, UC Berkeley CS 294-129, FALL 2016

BACKGROUND, DATA, TOOLS

- Variational auto-encoders are useful for generating new examples from observed data
- Learns latent encoding of data
- Can be used for semi-supervised learning
- Data: MNIST digits dataset and SVHN dataset
- Tools: TensorFlow, GeForce GTX 770 GPU

VARIATIONAL AUTO-ENCODER (VAE)

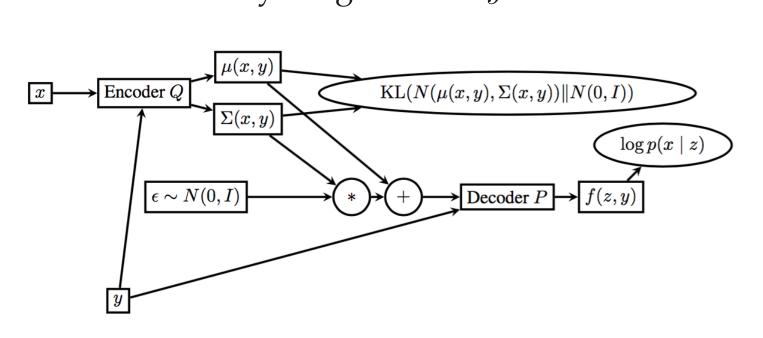
- latent variable model: $z \sim \mathcal{N}(0, I)$, $x \mid z \sim f(x; z, \theta)$ (e.g. Bernoulli)
- variational inference: maximize lower bound on log likelihood



 $\log p(x) \ge \mathbb{E}_{z \sim Q(\cdot \mid x)}[\log p(x \mid z)] - \mathrm{KL}(Q(z \mid x) || p(z)).$

CONDITIONAL VAE (CVAE)

• condition everything on label y



 $\log p(x \mid y) \ge \mathbb{E}_{z \sim Q(\cdot \mid y, x)}[\log p(x \mid y, z)] - \mathrm{KL}(Q(z \mid x, y) || p(z \mid y)).$

CVAE FOR COMPLETION

Test image	Left half	Generated image	Test image	Left half	Generated image
4	4	4	200	10	(6)
Test image	Left half	Generated image	Test image	Left half	Generated image
7	7	7	1:	1	1
Test image	Left half	Generated image	Test image	Left half	Generated image
ł	1	l	3	Æ	1
Test image	Left half	Generated image	Test image	Left half	Generated image
2	2	3	3	2	3
Test image	Left half	Generated image	Test image	Left half	Generated image
4	*	8)4	32	3-5
Test image	Left half	Generated image	Test image	Left half	Generated image
\circ	C	C	7		1
Test image	Left half	Generated image	Test image	Left half	Generated image
2	2	2	108)(101
Test image	Left half	Generated image	Test image	Left half	Generated image
7	7	7	16	26	26
Test image	Left half	Generated image	Test image	Left half	Generated image
Y	4	Y	180	1 8	181
Test image	Left half	Generated image	Test image	Left half	Generated image
3	<u> </u>	3	26		2

CVAE FOR STYLE TRANSFER

											Input	Gen. 0	Gen. 1	Gen. 2	Gen. 3	Gen. 4	Gen. 5	Gen. 6	Gen. 7	Gen. 8	Gen. 9
Input	Gen. 0	Gen. 1	Gen. 2	Gen. 3	Gen. 4	Gen. 5	Gen. 6	Gen. 7	Gen. 8	Gen. 9	:01	01				4	51	6	7	ŏ	91
7	D	1	2	3	4	5	6	7	8	9	1;	01	11	21	31	4	15 I	6	7	8	91
0	0	ı	2	3	4	5	6	7	8	9	3	0	1	2	3	4	5	6	2	18	9
1	0	/	2	3	4	5	6	7	8	9	3	0	3	2	3	4	5	6	2	8	9
5	0	3/	2	\mathcal{Z}	4	5	6	4	8	9	4	01	51	21	31	41	51	61	71	81	91
9	0	1	2	3	4	5	6	7	8	9	7	0		2	3	4	5	6	2	8	9
0	0	1	2	3	4	5	6	7	8	9	6	63	-	21	3	4	5	6	79	8	9
9	0	1	2	3	4	5	6	7	8	9	110	[8]	Ħ	121	13	6	5	61	17	6	
0	0	1	3	3	4	5	6	7	8	9		0			3	4	5	6	177	8	9
j	٥	1	2	3	4	5	6	7	8	9	108	101	KI	121	3	41	5	15	2	181	191
7	D	7	2	Э	4	5	6	7	8	9	16	10	J# 1	[2]	[6]	333	15	16	17	18	19
4	0	l	2	3	4	5	6	7	8	9	32	10	13	12	13	14	15	16	17	18	19
9	0	/	3	3	4	5	6	7	8	9	2.	8				1		6	7		Q
,	•	•		•	,		•	,	V	,											

SEMI-SUPERVISED LEARNING (SSL) VAE

- Handle datasets with missing labels
- Models label distribution
- Labeled and unlabeled examples enter loss differently

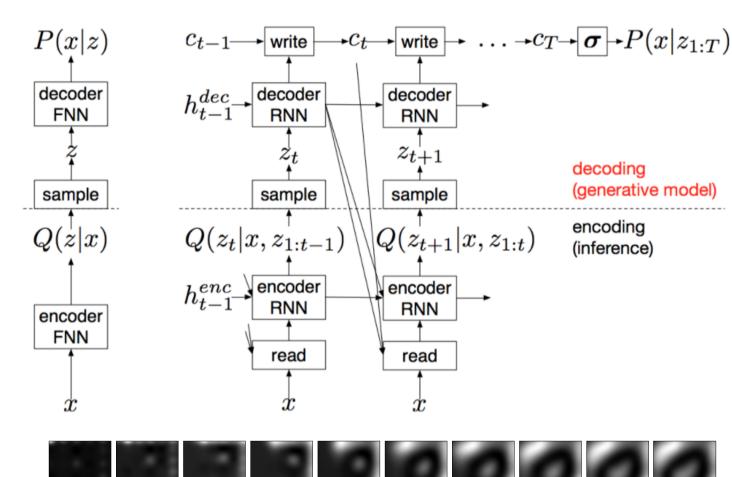
 $\log p(x,y) \ge \mathbb{E}_{z \sim Q(z|x,y)} [\log p(x \mid y, z) + \log p(y)]$ $- \text{KL}(Q(z \mid x, y) || p(z)) =: -\mathcal{L}(x, y)$ $\log p(x) \ge \sum q(y \mid x) (-\mathcal{L}(x, y)) + H(q(y \mid x))$

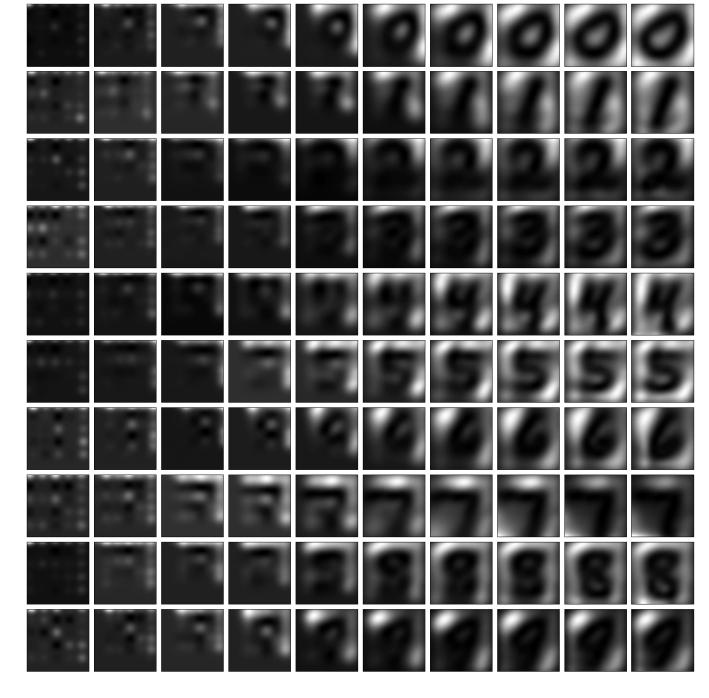
Validation/test error on MNIST (55000 training examples)

(88888)	or children or contribution	100)
	1000 labeled	600 labeled
Fully connected	4.7%/ 5.1%	11.5%/12.0%
Convolutional	4.2%/4.8%	6.0%/6.2%
Kingma et al. [3]	2.4%	2.6%

DRAW

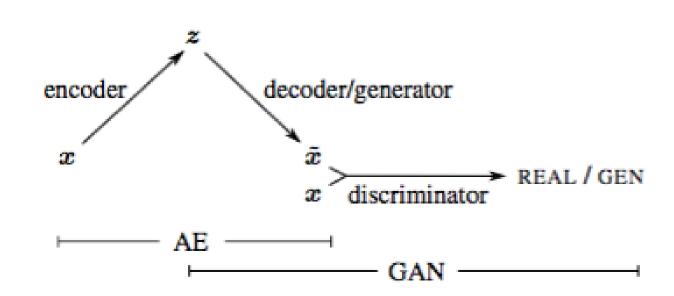
- attention-based sequential generation
- RNN structure



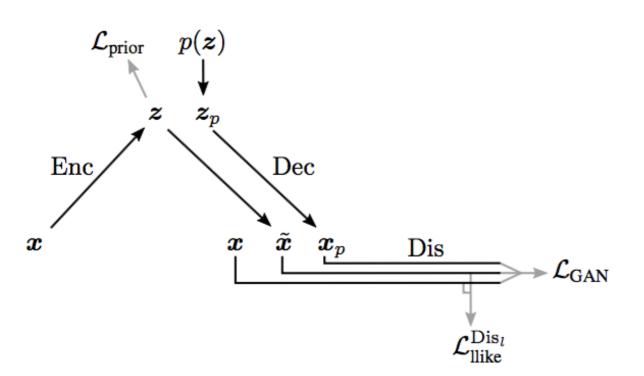


ADDING GANS

- VAE output is often blurry
- Add discriminator to encourage sharpness



Replace decoder loss with comparison of discriminator layers



CVAE WITH GAN

CVAEGAN results here

SSL WITH GANS

CVAEGAN results here

DRAW WITH GANS

CVAEGAN results here

FUTURE DIRECTIONS

а

REFERENCES

- [1] Carl Doersch. Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908, 2016.
- [2] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra. DRAW: A recurrent neural network for image generation. arXiv preprint arXiv:1502.04623, 2015.
- [3] Diederik P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems*, pages 3581–3589, 2014.
- [4] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [5] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, and Ole Winther. Autoencoding beyond pixels using a learned similarity metric. *arXiv preprint arXiv:1512.09300*, 2015.