Analyzing the Dynamics of Gated Auto-encoders

Daniel Jiwoong Im & Graham W. Taylor
School of Engineering, University of Guelph
imj@uoguelph.ca

UNIVERSITY of GUELPH
CHANGING LIVES IMPROVING LIFE

Introduction

- Kamyshanska and Memisevic [1] have recently shown how scores can be computed from an auto-encoder by interpreting it as a dynamical system.
- Gated auto-encoders (GAEs) are an interesting and flexible extension of auto-encoders which can learn transformations among different images or pixel covariances within images.
- We cast the GAE as a dynamical system driven by a vector field in order to analyze the model.
- As a consequence, we derived some relationships between Guassian Bernoulli Factored RBMs and Gated Auto-encoders, and mean-covariance RBMs and mean-covariance Auto-encoders.

Gated Auto-encoders

- Consist of an encoder $h(\cdot)$ and decoder $r(\cdot)$ that processes pairs of datapoints $(\mathbf{x}, \mathbf{y})$:

$$h(\mathbf{x}, \mathbf{y}) = \sigma(W^M((W^F\mathbf{x}) \odot (W^F\mathbf{y}))) \quad (1)$$

where $\odot$ is element-wise multiplication and $\sigma(\cdot)$ is an activation function.

- Reconstruct $\mathbf{y}$ given $\mathbf{x}$ or $\mathbf{x}$ given $\mathbf{y}$ or both:

$$r(\mathbf{x}|\mathbf{y}, h) = (W^F)^T((W^F\mathbf{y}) \odot (W^M)^T h(\mathbf{x}, \mathbf{y})), \quad (2)$$
$$r(\mathbf{y}|\mathbf{x}, h) = (W^F)^T((W^F\mathbf{x}) \odot (W^M)^T h(\mathbf{x}, \mathbf{y})). \quad (3)$$

- Intuitively, the GAE learns *relations* between the inputs, rather than representations of the inputs themselves.
- Minimize a reconstructive objective, e.g. for real-valued data:

$$J = \frac{1}{2}\|r(\mathbf{x}|\mathbf{y}) - \mathbf{x}\|^2. \quad (4)$$

Mean Covariance Auto-encoders

- Covariance Auto-encoder : In the case that $\mathbf{x} = \mathbf{y}$ (i.e. the input is copied), the mapping units learn pixel covariances.
- Mean Covariance Auto-encoder : Auto-encoder + Gated Auto-encoder with $\mathbf{x} = \mathbf{y}$

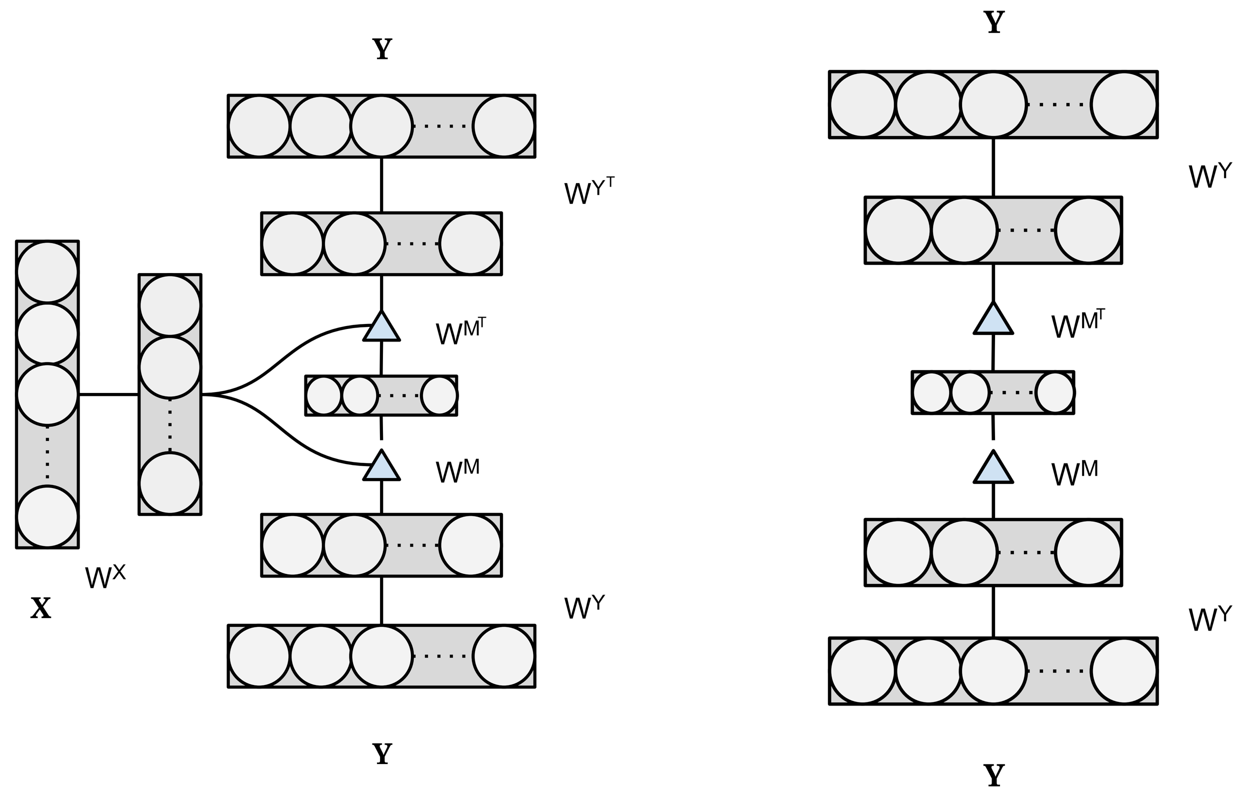

Figure 1: (a) Gated Auto-encoder, which takes as input $(\mathbf{x}, \mathbf{y})$ and learns the relationship between $\mathbf{x}$ and $\mathbf{y}$. (b) Covariance Auto-encoder, which takes only input $\mathbf{y}$ and learns correlations among input dimensions. The triangle notation refers to three-way (multiplicative) connections.

Gated Auto-encoder Scoring

- Measures how much a GAE "likes" a given pair of inputs $(\mathbf{x}, \mathbf{y})$ up to a normalizing constant.

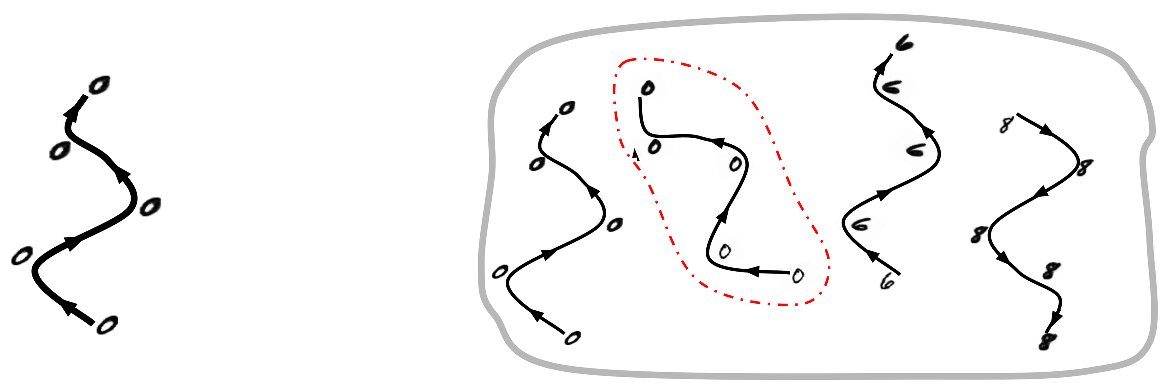

Figure 2: (a) Illustrates the path generated by starting from a data point $\mathbf{y}|\mathbf{y}$ and recursively generating the next point $r(r \cdots r(\mathbf{y}|\mathbf{y})|\mathbf{y})$ by applying the GAE. For visualization, each image plotted is the result of applying the recursion three times. (b) Illustrates the path given on a 3-dimensional manifold. Let's assume the manifold is a rectangular box. Conditioning on $\mathbf{y}$ along the horizontal axis, we get a path that lies on a 2D plane (slice of 3D manifold).

Vector field representation

$$F(\mathbf{y}|\mathbf{x}) = r(\mathbf{y}|\mathbf{x}) - \mathbf{y}. \quad (5)$$

- The vector field represents the linear transformation that $\mathbf{y}|\mathbf{x}$ undergoes as a result of applying the reconstruction function $r(\mathbf{y}|\mathbf{x})$. Repeatedly applying the reconstruction function to an input $\mathbf{y}|\mathbf{x} \rightarrow r(\mathbf{y}|\mathbf{x}) \rightarrow r(r(\mathbf{y}|\mathbf{x})) \rightarrow \cdots r(\cdots r(\mathbf{y}|\mathbf{x}))$ yields a trajectory whose dynamics, from a physics perspective, can be viewed as a force field.
- Satisfies Poincaré's integrability criterion: For some open, simple connected set $\mathcal{U}$, a continuously differentiable function $F : \mathcal{U} \rightarrow \Re^m$ defines a gradient field if and only if

$$\frac{\partial F_i}{\partial y_j} = \frac{\partial r_i(\mathbf{y}|\mathbf{x})}{\partial y_j} - \delta_{ij} = \frac{\partial r_j(\mathbf{y}|\mathbf{x})}{\partial y_i} - \delta_{ij} = \frac{\partial F_j}{\partial y_i}$$
$$\frac{\partial F_i}{\partial x_j} = \frac{\partial r_i(\mathbf{x}|\mathbf{y})}{\partial x_j} = \frac{\partial r_j(\mathbf{x}|\mathbf{y})}{\partial x_i} = \frac{\partial F_j}{\partial x_i} \; \forall i, j = 1 \cdots n.$$

GAE Scoring

- The vector field in Eq. 5 is the derivative of a scalar field $\Rightarrow$ this vector field is a conservative field.

$$r(\mathbf{y}|\mathbf{x}) - \mathbf{y} = \nabla E$$

- The line integral of a conservative vector field is path independent, which allows us to take the anti-derivative of the scalar function:

$$E(\mathbf{y}|\mathbf{x}) = \int (r(\mathbf{y}|\mathbf{x}) - \mathbf{y}) d\mathbf{y}$$
$$= \int h(\mathbf{u}) d\mathbf{u} - \frac{1}{2}\mathbf{y}^2 + \text{const.}$$

where $\mathbf{u}$ is an auxiliary variable such that $\mathbf{u} = W^M((W^F\mathbf{y}) \odot (W^F\mathbf{x}))$ and $\frac{d\mathbf{u}}{d\mathbf{y}} = W^M(W^F \odot (W^F\mathbf{x} \otimes \mathbf{1}_D))$, and $\otimes$ is the Kronecker product.

Relationship to Restricted Boltzmann Machines

- The energy is identical (up to a constant) to the free energy of a Factored Gated Conditional Restricted Boltzmann Machine (FCRBM) with Gaussian visible units and Bernoulli hidden units:

$$E_\sigma(\mathbf{y}|\mathbf{x}) = \int (1 + \exp -(\mathbf{u}))^{-1} d\mathbf{u} - \frac{1}{2}(\mathbf{y}^2) + \text{const},$$
$$= \sum_k \log\left(1 + \exp\left(W^M(W^F_{k\cdot}\mathbf{y} \odot W^F_{k\cdot}\mathbf{x})\right)\right) - \frac{\mathbf{y}^2}{2} + \text{const.}$$

- Similarly, considering $\mathbf{y} = \mathbf{x}$, the energy function of the covariance auto-encoder with dynamics $r(\mathbf{y}|\mathbf{x}) - \mathbf{y}$ is equivalent to the free energy of a Covariance RBM:

$$E(\mathbf{y}, \mathbf{y}) = \sum_k \log\left(1 + \exp\left(W^M(W^F\mathbf{y})^2 + \mathbf{b}\right)\right) - \frac{\mathbf{x}^2}{2} + \text{const}$$

- The energy function of a Mean-covariance auto-encoder and the free energy of a Mean-covariance RBM (mcRBM) with Gaussian-distributed visibles and Bernoulli-distributed hiddens are the same:

$$E = \sum_k \log\left(1 + \exp\left(-W^M(W^F\mathbf{x})^2 - \mathbf{b}\right)\right) + \sum_k \log(1 + \exp(W\mathbf{x} + \mathbf{c}) - \mathbf{x}^2 + \text{const} \quad (6)$$

Experiments

We considered classification using the Gated Softmax Classifier with biases [2]. Output probabilities using the GAE and mcAE are respectively:

$$P_{GAE}(y_i|\mathbf{x}) = \frac{\exp(E^C_i(\mathbf{x}) + B_i)}{\sum_j \exp(E^C_j(\mathbf{x}) + B_j)}, \quad (7)$$
$$P_{mcAE}(y_i|\mathbf{x}) = \frac{\exp(E^M_i(\mathbf{x}) + E^C_i(\mathbf{x}) + B_j)}{\sum_j \exp(E^M_j(\mathbf{x}) + E^C_j(\mathbf{x}) + B_j)}. \quad (8)$$

The training procedure is as follows:

1. Train a (denosing/contractive) mean covariance (gated) autoencoder for each class with tied input weights and tied inputs on gated version.
2. Train the mean covariance (gated) autoencoder scoring coefficients based on Equation 8.

DATA	SVM RBF	RBM	DEEP SAA_3	GSM	AES	GAES	mcAES
RECT	2.15	4.71	2.14	0.56	0.84	0.61	**0.54**
RECTIMG	24.04	23.69	24.05	22.51	21.45	22.85	**21.41**
CONVEX	19.13	19.92	18.41	**17.08**	21.52	21.6	20.63
MNISTROT	11.11	14.69	**10.30**	11.75	11.25	16.5	13.42
MNISTRAND	14.58	9.80	11.28	10.48	**9.70**	18.65	16.73
MNISTROTIM	55.18	52.21	51.93	55.16	47.14	39.98	**35.52**

Table 1: Classification error rates on the Deep Learning Benchmark dataset. SAA_3 stands for three-layer Stacked Auto-encoder. SVM and RBM results are from [3], DEEP and GSM are results from [2], and AES is from [1].

Conclusion

We applied a dynamical systems view to GAEs, deriving a means of GAE scoring, and drawing connections to RBMs and score matching. Specifically, we

- Derived some theoretical results for the GAE that enable us to gain more insight and understanding of its operation.
- Showed that the GAE could be scored according to an energy function.
- Demonstrated the equivalency of the GAE energy to the free energy of conditional RBMs.

References

[1] Hanna Kamyshanska. On autoencoder scoring. In *Proceedings of the 28th International Conference on Machine Learning (ICML)*, pages 720–728, 2013.

[2] Roland Memisevic, Christopher Zach, Geoffrey Hinton, and Marc Pollefeys. Gated softmax classification. In *Neural Information Processing Systems (NIPS).*, 2010.

[3] Pascal Vincent. A connection between score matching and denoising auto-encoders. *Neural Computation*, 23(7):1661–1674, 2010.