Inverting VAEs for Improved Generative Accuracy

Ian Gemp*[1], Mario Parente[2], Sridhar Mahadevan[1]

[1] College of Information and Computer Sciences, University of Massachusetts at Amherst

[2] Electrical and Computer Engineering, University of Massachusetts at Amherst

***corresponding author:** imgemp@cs.umass.edu

Problem: When labeled data (**x**,**y**) are scarce, semi-supervised learning can improve model performance by leveraging large amounts of unlabeled data (**x**). Under this setting, the deep semi-supervised VAE (M2) learns a generative latent variable model of the data. However, there are cases when this model does not recognize data that *it* generated! How do we encourage the learned model to be internally consistent with desired semantics?

MNIST Inconsistency

Desired semantics (**y**): $[0, 1, 2, 3, 4, 8, 9, 5, 7, 6]$

Generate image of 8 with random style:

$y \leftarrow [0, 0, 0, 0, 0, 1, 0, 0, 0, 0]$

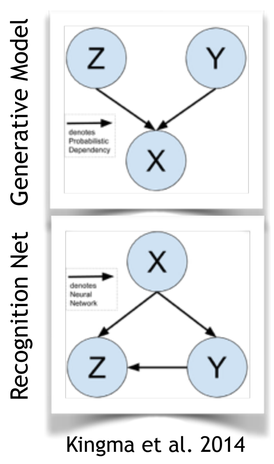
$z \sim p(\mathbf{z})$

$x \sim p(\mathbf{x}|y, z)$

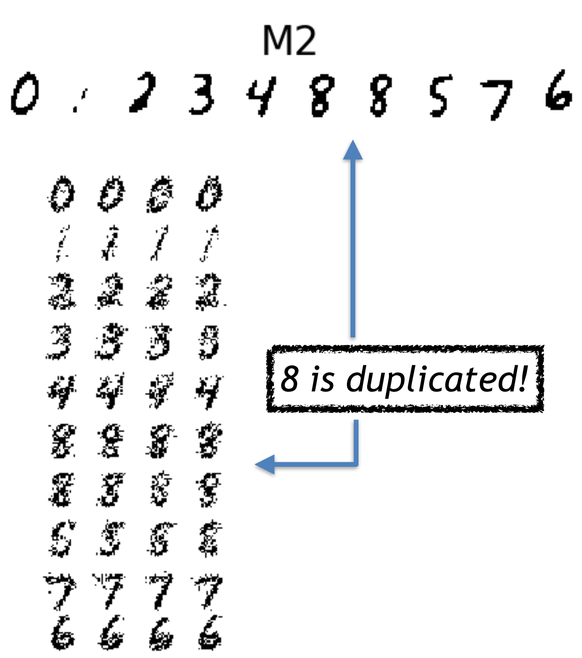
mismatch!

$y' \sim q(\mathbf{y}|x)$

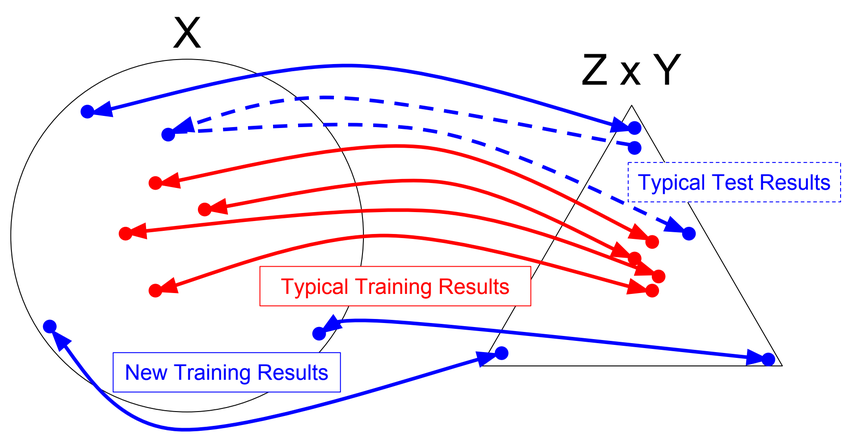
$[0, 0, 0, 0, 0, 0, 1, 0, 0, 0]$



Kingma et al. 2014



General Problem



Model: To encourage internal consistency, we "invert" the semi-supervised VAE and train under the adversarial scenario exhibited above. We derive the lower bound for maximizing the marginal likelihood of un*featured* labels (**y**) and add this to the standard M2 objective.

$$\log q_\phi(\mathbf{y}) \geq \mathbb{E}_{p(\mathbf{z})p_\theta(\mathbf{x}|\mathbf{y},\mathbf{z})}\Big[\log q_\phi(\mathbf{y},\mathbf{z}|\mathbf{x}) - \log p_\theta(\mathbf{x}|\mathbf{y},\mathbf{z}) + \log q(\mathbf{x}) - \log p(\mathbf{z})\Big] = ELBO_r^u$$

$$ELBO_r^u = \log q_\phi(y) - KL\Big(p_\theta(x,z|y)\Big|q_\phi(x,z|y)\Big)$$

Bound becomes tight as standard and "inverted" conditionals become similar

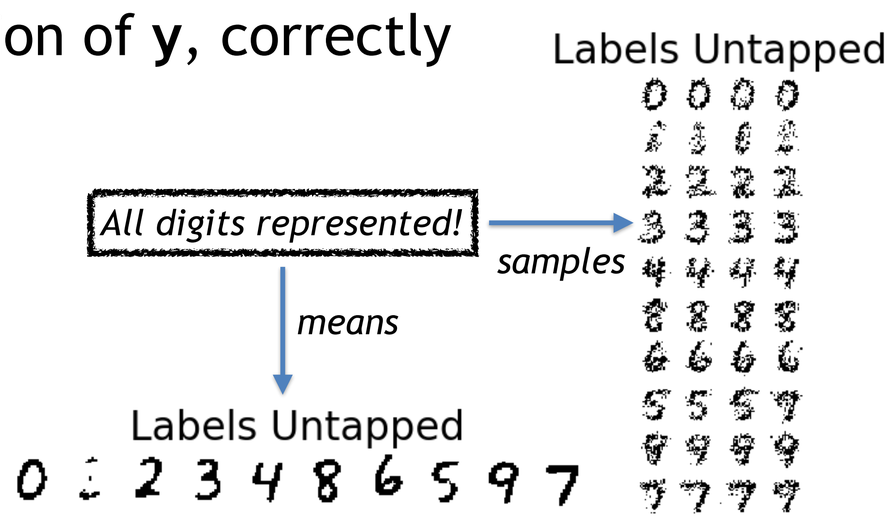
Experiment: We compare the proposed model, *Untapped*, with the original semi-supervised VAE model, M2, by Kingma et al. on the following MNIST task. For both models, labeled data (**x**,**y**) is provided for digits 0-4 and unlabeled data (**x**) for digits 0-9. In addition to these, *Untapped*, leverages un*featured* data (**y**) for digits 0-9 by maximizing the above **e**vidence lower **bo**und. Our hypothesis is that *Untapped* is better at disentangling the semantics of **y**.

Results: *Untapped* assigns a unique digit to each dimension of **y**, correctly disentangling semantics.

Improving the generative model results in improved discriminative performance as well.

Closest semantic permutation for digits learned with pretrained logistic regression classifier. Models tested on classifying held-out digits 5-9.

	M2	Untapped
Cross-Entropy	2.30	2.07



Conclusion:

- Improved disentangling of semantics using *Untapped* over M2 VAE model.
- Future work: improving prior q(**x**) tightens lower bound - improves model performance.