# Training Normalizing Flows

Maximum likelihood training to find parameters $\theta$ that make our training samples likely under our parameterized model.

$$\max_{\theta} \mathbb{E}_{\mathbf{x}\sim p_x} p_\theta(\mathbf{x})$$

Change of variables makes this: $\ell_2$-norm minimization of transformed samples – volume change term.

$$\min_{\theta} \frac{1}{N} \sum_{\mathbf{x}\in X_{train}} [\frac{1}{2}\|T_\theta(\mathbf{x})\|_2^2 - \log \left| \det \frac{\partial T_\theta(\mathbf{x})}{\partial \mathbf{x}} \right| ]$$

# How does likelihood maximization work?

Many models are trained on likelihood maximization …

but why do we expect this to learn distributions?

$$\max_{\theta} \mathbb{E}_{\mathbf{x}\sim p_x} p_\theta(\mathbf{x}) = \min_{\theta} \; \mathbb{E}_{\mathbf{x}\sim p_x} - \log p_\theta(\mathbf{x})$$