

Training Normalizing Flows

Maximum likelihood training to find parameters θ 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: ℓ_2 -norm minimization of transformed samples
– volume change term.

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

How does likelihood maximization work?

Many models are trained on some form of likelihood maximization ...

but why do we expect this to learn distributions?