Implicit Reparameterization Gradients

Michael Figurnov, Shakir Mohamed, Andriy Mnih

Poster: Room 210 #33

DeepMind

Reparameterization gradients

Core part of variational autoencoders, automatic variational inference, etc.

Backpropagation in graphs with continuous random variables

Reparameterization gradients

Core part of variational autoencoders, automatic variational inference, etc.

Backpropagation in graphs with continuous random variables

$$\frac{\partial}{\partial \phi} \mathbb{E}_{\underset{\substack{\uparrow \\ \text{continuous} \\ \text{(Normal, ...)}}}{q_\phi(z)}} \underset{\substack{\uparrow \\ \text{differentiable} \\ \text{(ELBO, ...)}}}{[f(z)]} = \mathbb{E}_{q_\phi(z)} \underbrace{\left[\frac{\partial f(z)}{\partial z} \frac{\partial z}{\partial \phi} \right]}_{\text{backpropagation}}$$

Reparameterization gradients

Core part of variational autoencoders, automatic variational inference, etc.

Backpropagation in graphs with continuous random variables

$$\frac{\partial}{\partial \phi} \mathbb{E}_{\underset{\substack{\uparrow \\ \text{continuous} \\ \text{(Normal, ...)}}}{q_\phi(z)}} \big[\underset{\substack{\uparrow \\ \text{differentiable} \\ \text{(ELBO, ...)}}}{f(z)}\big] = \mathbb{E}_{q_\phi(z)} \underbrace{\left[\frac{\partial f(z)}{\partial z} \boxed{\frac{\partial z}{\partial \phi}} \right]}_{\text{backpropagation}}$$

requires a *tractable* inverse transformation!
Normal, Logistic, …

Reparameterization gradients

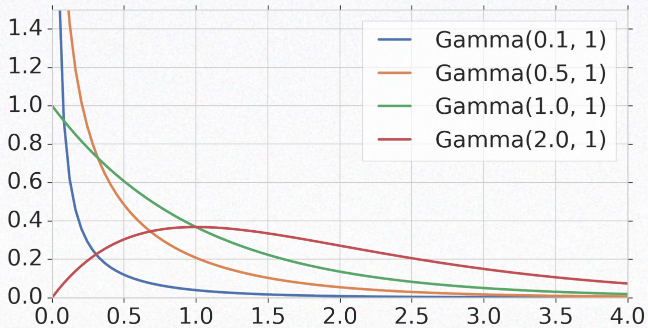
Core part of variational autoencoders, automatic variational inference, etc.

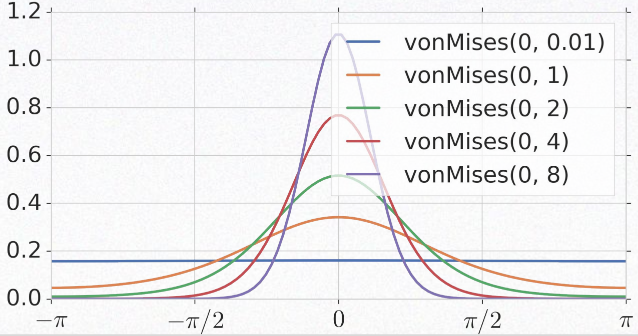
Backpropagation in graphs with continuous random variables

$$\frac{\partial}{\partial \phi} \mathbb{E}_{q_\phi(z)}[f(z)] = \mathbb{E}_{q_\phi(z)} \left[\frac{\partial f(z)}{\partial z} \boxed{\frac{\partial z}{\partial \phi}} \right]$$

requires a *tractable* inverse transformation!
Normal, Logistic, …

We show how to use *implicit differentiation* for reparameterization of other continuous random variables, such as Gamma and von Mises





Explicit and implicit reparameterization

Cumulative density function $F(z|\phi) \equiv \int_{-\infty}^{z} q_\phi(t)dt = u \sim \mathrm{Uniform}(0, 1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit		
Implicit		

Explicit and implicit reparameterization

Cumulative density function $F(z|\phi) \equiv \int_{-\infty}^{z} q_\phi(t)dt = u \sim \text{Uniform}(0,1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit	$u \sim \text{Uniform}(0,1)$ $z = F^{-1}(u\vert\phi) \sim q_\phi(z)$	
Implicit		

Explicit and implicit reparameterization

Cumulative density function $F(z|\phi) \equiv \int_{-\infty}^{z} q_\phi(t)dt = u \sim \text{Uniform}(0,1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit	$u \sim \text{Uniform}(0,1)$ $z = F^{-1}(u\|\phi) \sim q_\phi(z)$	$\frac{\partial z}{\partial \phi} = \frac{\partial F^{-1}(u\|\phi)}{\partial \phi}$
Implicit		

Explicit and implicit reparameterization

Cumulative density function $F(z|\phi) \equiv \int_{-\infty}^{z} q_\phi(t)dt = u \sim \text{Uniform}(0, 1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit	$u \sim \text{Uniform}(0, 1)$ $z = F^{-1}(u\|\phi) \sim q_\phi(z)$	$\frac{\partial z}{\partial \phi} = \frac{\partial F^{-1}(u\|\phi)}{\partial \phi}$
Implicit	$z \sim q_\phi(z)$ using any sampler (e.g., rejection sampling)	

Explicit and implicit reparameterization

Cumulative density function $F(z|\phi) = \int_{-\infty}^{z} q_\phi(t)dt = u \sim \text{Uniform}(0,1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit	$u \sim \text{Uniform}(0,1)$ $z = F^{-1}(u|\phi) \sim q_\phi(z)$	$\frac{\partial z}{\partial \phi} = \frac{\partial F^{-1}(u|\phi)}{\partial \phi}$
Implicit	$z \sim q_\phi(z)$ using any sampler (e.g., rejection sampling)	$\boxed{\frac{\partial z}{\partial \phi} = -\frac{\frac{\partial F(z|\phi)}{\partial \phi}}{q_\phi(z)}}$

Explicit and implicit reparameterization

Cumulative density function $F(z|\phi) = \int_{-\infty}^{z} q_\phi(t)dt = u \sim \text{Uniform}(0,1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit	$u \sim \text{Uniform}(0,1)$ $z = F^{-1}(u\|\phi) \sim q_\phi(z)$	$\frac{\partial z}{\partial \phi} = \frac{\partial F^{-1}(u\|\phi)}{\partial \phi}$
Implicit	$z \sim q_\phi(z)$ using any sampler (e.g., rejection sampling)	$\boxed{\frac{\partial z}{\partial \phi} = -\frac{\frac{\partial F(z\|\phi)}{\partial \phi}}{q_\phi(z)}}$

Derivation: implicit differentiation $\frac{dF(z|\phi)}{d\phi} = \frac{du}{d\phi}$

$$\underbrace{\frac{\partial F(z|\phi)}{\partial z}}_{q_\phi(z)} \frac{\partial z}{\partial \phi} + \frac{\partial F(z|\phi)}{\partial \phi} = 0$$

Explicit and implicit reparameterization

Cumulative density function $F(z|\phi) = \int_{-\infty}^{z} q_\phi(t)dt = u \sim \mathrm{Uniform}(0,1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit	$u \sim \mathrm{Uniform}(0,1)$ $z = F^{-1}(u\|\phi) \sim q_\phi(z)$	$\frac{\partial z}{\partial \phi} = \frac{\partial F^{-1}(u\|\phi)}{\partial \phi}$
Implicit	$z \sim q_\phi(z)$ using any sampler (e.g., rejection sampling)	$\boxed{\frac{\partial z}{\partial \phi} = -\frac{\frac{\partial F(z\|\phi)}{\partial \phi}}{q_\phi(z)}}$

often not implemented in numerical libraries

Derivation: implicit differentiation $\frac{dF(z|\phi)}{d\phi} = \frac{du}{d\phi}$

$$\frac{\partial F(z|\phi)}{\partial z}\frac{\partial z}{\partial \phi} + \frac{\partial F(z|\phi)}{\partial \phi} = 0$$

$q_\phi(z)$

How to compute $\frac{\partial F(z|\phi)}{\partial \phi}$?

Relative metrics (lower is better)

	Gamma		Von Mises	
Method	Error	Time	Error	Time
Automatic differentiation of the CDF code	**1x**	**1x**	**1x**	**1x**
Finite difference	832x	2x	514x	1.2x
Jankowiak & Obermeyer (2018) concurrent work; closed-form approximation	18x	5x	-	-

Jankowiak, Obermeyer "Pathwise Derivatives Beyond the Reparameterization Trick." ICML, 2018

How to compute $\frac{\partial F(z|\phi)}{\partial \phi}$?

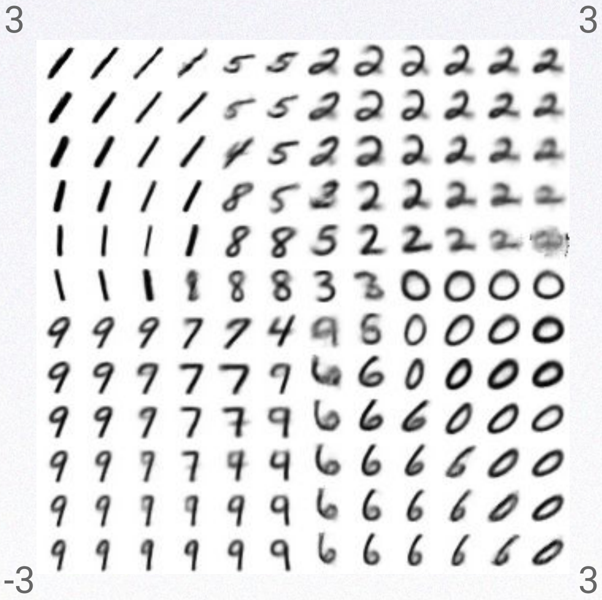
Relative metrics (lower is better)

	Gamma		Von Mises	
Method	Error	Time	Error	Time
Automatic differentiation of the CDF code	**1x**	**1x**	**1x**	**1x**
Finite difference	832x	2x	514x	1.2x
Jankowiak & Obermeyer (2018) concurrent work; closed-form approximation	18x	5x	-	-
Knowles (2015) approximate *explicit* reparameterization	2840x	63x	-	-

Knowles, "Stochastic gradient variational Bayes for Gamma approximating distributions." arXiv, 2015
Jankowiak, Obermeyer "Pathwise Derivatives Beyond the Reparameterization Trick." ICML, 2018

Variational Autoencoder

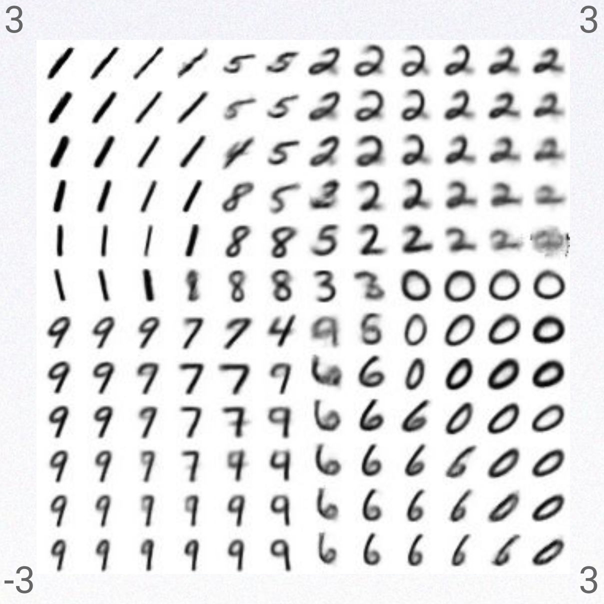
2D latent spaces for MNIST



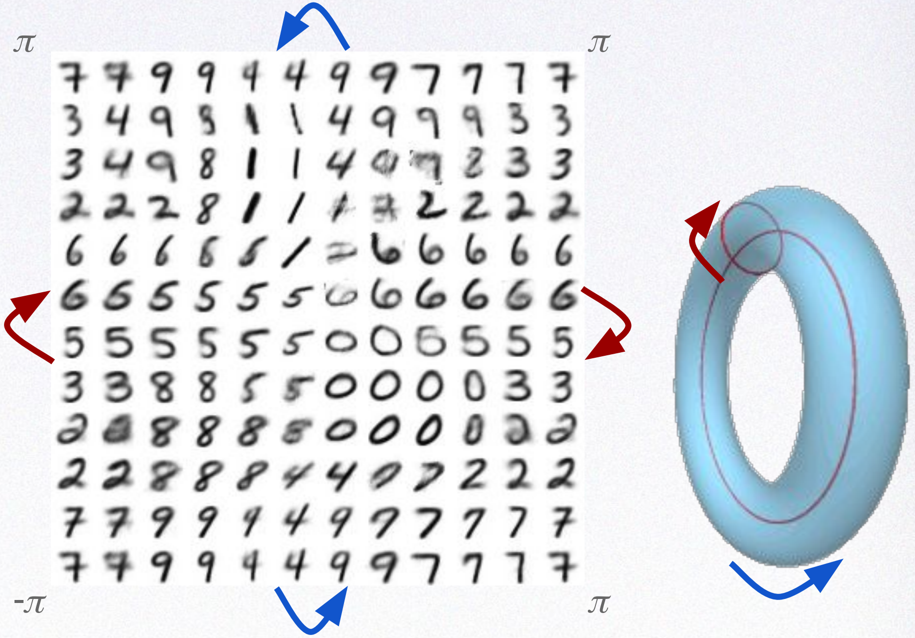
Normal prior and posterior

Variational Autoencoder

2D latent spaces for MNIST



Normal prior and posterior

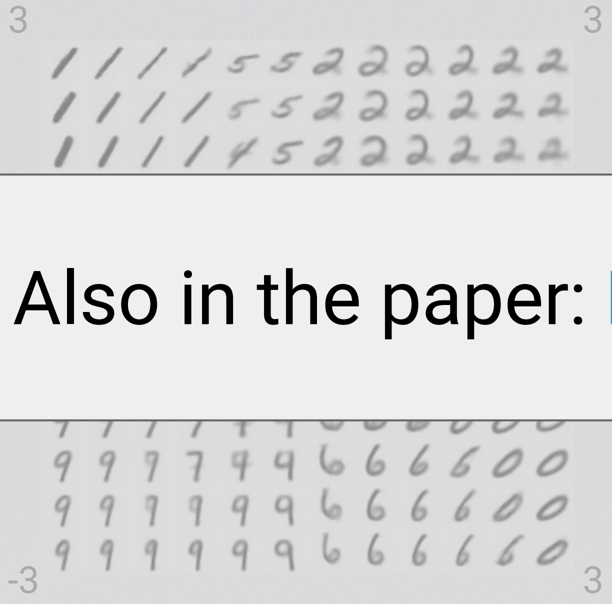


Uniform prior, von Mises posterior

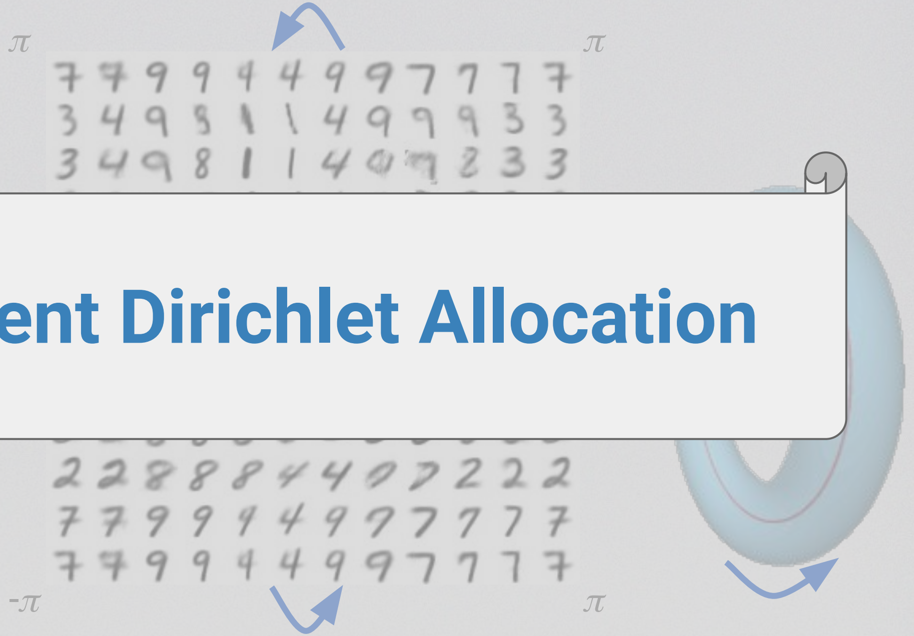
Torus adapted from https://en.wikipedia.org/wiki/Torus#/media/File:Sphere-like_degenerate_torus.gif

Variational Autoencoder

2D latent spaces for MNIST



Normal prior and posterior



Uniform prior, von Mises posterior

Also in the paper: Latent Dirichlet Allocation

Torus adapted from https://en.wikipedia.org/wiki/Torus#/media/File:Sphere-like_degenerate_torus.gif

Implicit Reparameterization Gradients

Michael Figurnov, Shakir Mohamed, Andriy Mnih

- A more general view of the reparameterization gradients
  - Decouple sampling from gradient estimation
- Reparameterization gradients for Gamma, von Mises, Beta, Dirichlet, ...
  - Faster and more accurate than the alternatives
  - Implemented in TensorFlow Probability: `tfp.distributions.{Gamma,VonMises,Beta,Dirichlet,...}`
- Move away from making modelling choices for computational convenience

Poster: Room 210 #33