Computational Statistics

Hyperspherical VAE

Victor Deng Inès Vati

École Normale Supérieure Paris-Saclay, Master MVA

10^{th} JAN 2024

Table of contents

Table of contents

Introduction

2018 paper from Tim R. Davidson *et al.* [DFDC$^+$]

- Replacing the Gaussian prior and approximate posterior distributions with a von Mises-Fisher distribution
- Goal: better model data with a hyperspherical latent structure
- Various experiments, where the $\mathcal{S}$-VAE (von Mises-Fisher distributions) often outperforms the $\mathcal{N}$-VAE (Gaussian distributions) in low dimensions

Table of contents

Algorithm 1 Overview of the sampling method from $\mathcal{S}(\mu, \kappa)$

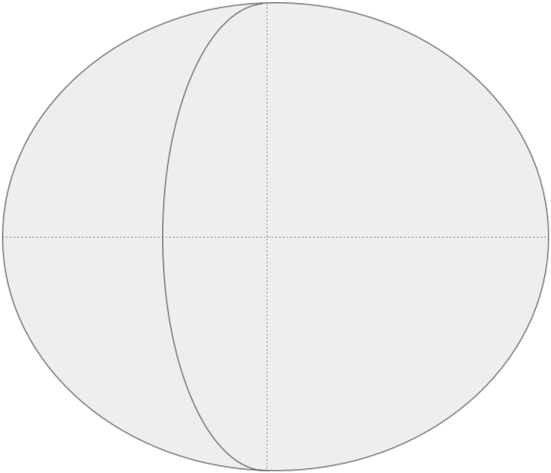
1: Sample $z \sim q(z|e_1, \kappa)$ where $e_1 = (1, 0, \dots, 0)$
2: Compute Householder reflection $U(\mu)$ so that $U(\mu)e_1 = \mu$
3: **return** $z' = U(\mu)z$

Sampling z from vMF

Algorithm 2 Overview of the sampling method from $\mathcal{S}(\mu, \kappa)$

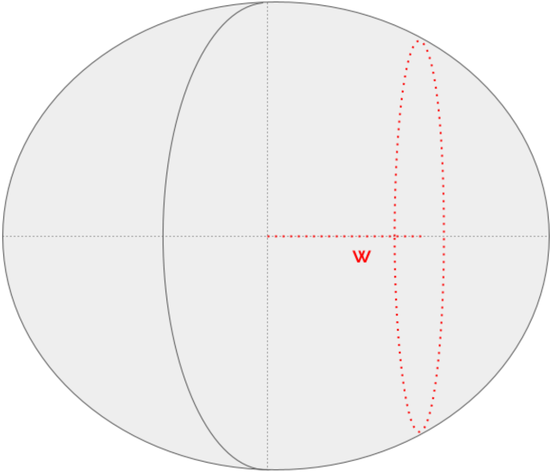
Sample $z \sim q(z|e_1, \kappa)$ where $e_1 = (1, 0, \ldots, 0)$
Sample $w \in \mathbb{R} \sim g(w|\kappa)$ by acceptance rejection sampling
Sample $v \in \mathbb{R}^{d-1} \sim \mathcal{U}(S^{d-2})$ (uniform on the hypersphere S^{d-2})
$z \leftarrow (w, \sqrt{1-w^2}v^T)^T$
Compute Householder reflection $U(\mu)$ so that $U(\mu)e_1 = \mu$
return $z' = U(\mu)z \sim q(z'|\mu, \kappa)$

Sampling w from $g(w|\kappa, \theta)$



S^2 : unit sphere in $\mathbb{R}^3$

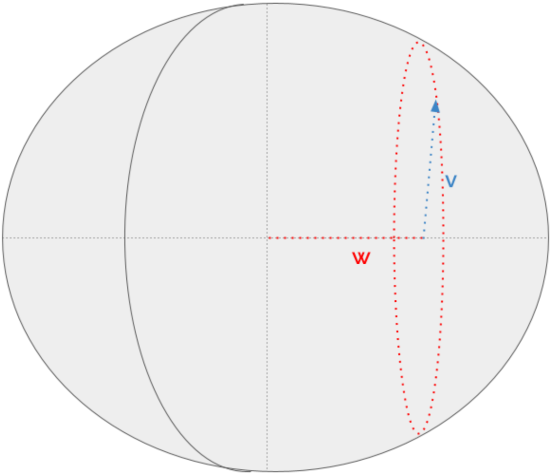
Sampling w from $g(w|\kappa, \theta)$



Sample $w \in \mathbb{R} \sim g(w|\kappa, d)$ by acceptance rejection sampling

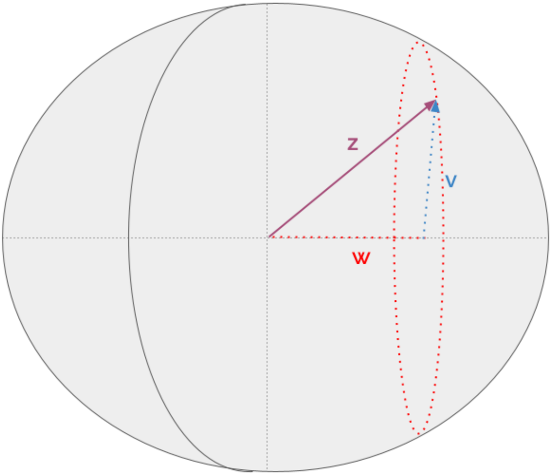
Sampling w from $g(w|\kappa)$

- Case $d = 3$
- Case $d > 3$



Sample $v \in \mathbb{R}^{d-1} \sim \mathcal{U}(S^{d-2})$

Sampling z from $q(z|e_1, \kappa)$



$$z = (w, \sqrt{1 - w^2} v^T)^T$$

Transform z

Algorithm 3 Overview of the sampling method from $\mathcal{S}(\mu, \kappa)$

1: Sample $z \sim q(z|e_1, \kappa)$ where $e_1 = (1, 0, \ldots, 0)$
2: Compute Householder reflection $U(\mu)$ so that $U(\mu)e_1 = \mu$
3: $u \leftarrow Normalize(e_1 - \mu)$
4: $U \leftarrow I - 2uu^T$
5: **return** $z' = U(\mu)z$

Table of contents

Reparameterization Trick

The authors use a reparameterization trick that has been extended to distributions that can be sampled using rejection sampling [NRLB20].

Algorithm 4 Reparameterized Rejection Sampling (from [NRLB20])

$i \leftarrow 0$
repeat
    $i \leftarrow i + 1$
    Propose $\varepsilon_i \sim s(\varepsilon)$
    Simulate $u_i \sim \mathcal{U}[0, 1]$
until $u_i < \frac{g(h(\varepsilon_i,\theta);\theta)}{r(h(\varepsilon_i,\theta);\theta)}$
return ε_i

Monte Carlo estimation

By noting $\pi(\varepsilon|\theta)$ the distribution of the resulting ε, we have

$$\nabla_\theta \mathbb{E}_{g(\varepsilon|\theta)}[\ldots] = \mathbb{E}_{\pi(\varepsilon|\theta)}[\ldots] \text{ ''} = \mathbb{E}_{(\varepsilon_i, U_i)_i}[\ldots]\text{''}$$

Problem: $(\varepsilon_i, U_i)_{i\in\mathbb{N}}$ is not a random variable (it is a stochastic process)
No reference to a convergence proof in [DFDC$^+$, NRLB20, PBJ12, MG14]

Table of contents

Experiments on link prediction

- Link prediction on a graph dataset: given a graph with some edges removed, predict the likelihood for each pair of nodes to be connected by an edge
- Cora dataset [MNRS00]: 2708 publications, 5429 links, 1433-dimensional feature vectors
- Using a Variational Graph Auto-Encoder [KW]: a variational encoder which uses a graph neural network (GNN) as encoder
- Reconstruction loss:

$$\mathbb{E}_{q(\mathbf{Z}|\mathbf{X},\mathbf{A})}(\log p(\mathbf{A}|\mathbf{Z})) \quad \text{where} p(\mathbf{A}|\mathbf{Z}) = \prod_{i=1}^{N}\prod_{j=1}^{N} p(A_{i,j}|\mathbf{z}_i, \mathbf{z}_j)$$

- Negative sampling: in the sum $\sum_{i,j} \log p(A_{i,j}|\mathbf{z}_i, \mathbf{z}_j)$, keep all positive edges and one randomly sampled negative edge per positive edge

TODO

reproduire l'experience
- ~~data~~ (Ines)
- ~~implementer les modèles~~ (Victor VGAE)
- gradients pour le hyperspherical VAE (Inès)
- courbes d'entraînement dans le cas normal (Victor)
- entrainement et evaluation

Table of contents

Conclusion and Discussion

- Quite meaningful contribution in low dimensions
- Algorithm not really useful in high dimensions, due to vanishing surface problem and soap bubble effect of the $\mathcal{N}$-VAE
- Much less variance parameters (1 vs. d for $\mathcal{N}$-VAE), so possibly less expressivity
- vérifier différentes dimensions de l'espace latent
- et algo vraiment utile en petite ou moyenne dimension ?

References

Tim R. Davidson, Luca Falorsi, Nicola De Cao, Thomas Kipf, and Jakub M. Tomczak.
Hyperspherical variational auto-encoders.

Thomas N. Kipf and Max Welling.
Variational graph auto-encoders.

Andriy Mnih and Karol Gregor.
Neural variational inference and learning in belief networks, 2014.

Andrew Kachites McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore.
Automating the construction of internet portals with machine learning.
Information Retrieval, 3(2):127–163, 2000.

Christian A. Naesseth, Francisco J. R. Ruiz, Scott W. Linderman, and David M. Blei.
Reparameterization gradients through acceptance-rejection sampling algorithms, 2020.