

Wasserstein Coreset via Sinkhorn Loss

JSM 2024

Haoyun Yin

Department of Statistics, Purdue University

July 25, 2024



- ① Introduction
- ② Backgrounds
- ③ Methodology

1 Introduction

2 Backgrounds

3 Methodology

Introduction

This presentation discusses the challenges in data science, including missing, noisy, and imbalanced data. We will explore the concept of Wasserstein Coreset via Sinkhorn Loss.

1 Introduction

2 Backgrounds

Wasserstein Distance & Sinkhorn Loss
Coresets

3 Methodology

Related Work

- The Wasserstein distance is discussed in [Villani, 2009].
- The Sinkhorn loss and algorithm is introduced in [Cuturi, 2013].
- The relationship between Wasserstein distance, MMD and Sinkhorn loss is discussed in [Feydy et al., 2019].

Related Work

- The Wasserstein distance is discussed in [Villani, 2009].
- The Sinkhorn loss and algorithm is introduced in [Cuturi, 2013].
- The relationship between Wasserstein distance, MMD and Sinkhorn loss is discussed in [Feydy et al., 2019].

Related Work

- The Wasserstein distance is discussed in [Villani, 2009].
- The Sinkhorn loss and algorithm is introduced in [Cuturi, 2013].
- The relationship between Wasserstein distance, MMD and Sinkhorn loss is discussed in [Feydy et al., 2019].

Coresets

- Coreset is a small weighted subset of the data that approximates the original data.
- The concept of Wasserstein Coreset is discussed in [Claici et al., 2018].

Coresets

- Coreset is a small weighted subset of the data that approximates the original data.
- The concept of Wasserstein Coreset is discussed in [Claici et al., 2018].

- 1 Introduction
- 2 Backgrounds
- 3 Methodology**

Methodology

- We propose a new method to construct Wasserstein Coreset via Sinkhorn Loss.
- The method is suitable for high-dimensional data.
- The method is efficient and stable.

Methodology

- We propose a new method to construct Wasserstein Coreset via Sinkhorn Loss.
- The method is suitable for high-dimensional data.
- The method is efficient and stable.

Methodology

- We propose a new method to construct Wasserstein Coreset via Sinkhorn Loss.
- The method is suitable for high-dimensional data.
- The method is efficient and stable.

References I

Sebastian Claiici, Aude Genevay, and Justin Solomon. Wasserstein measure coresets. *arXiv (Cornell University)*, 2018. ISSN 2331-8422.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. volume 26, 2013.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and MMD using Sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2681–2690, 2019.

Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

Thanks!