Understanding Wasserstein t-SNE

University of Tübingen

Fynn Bachmann

24.09.2021

Contents

Chapter 1

Introduction: Visualizing Structure in Datasets

In the modern world data is collected on a daily basis
GB per day
leading to lots of interesting datasets.
Getting overview of the dataset first thing are visualization techniques.
However when trying to visualize structure

1.1 Examples and Motivation

Recent improvement in data visualization have led to methods like tSNE ...
Thus a complex multidimensional dataset can be embedded onto the 2D plane and give away structure as in Figure
BAWU

1.1.1 Embedding Information

We can also encode other external information into an embedding, for example the average income leading tio Figure 2
BAWU

1.2 Hierarchical Data

In this thesis efforts are done to anaylse hierarchical data, that is . ...

We want to cluster or visualize the higher level using samples of the probability distribution as lower level

1.2.1 Loosing Information by taking the mean

In a later used dataset we find that correlation plays an interesting role as well, thus we lose information by collapsing the whole distribution into the mean.

Chapter 2

Theory: Methods and Distances

We have seen in t he introduction that it can be very valueable to visualize structure of datasets in as low dimensional space. However there exist a mutlitude of algorithms that put emphasis on different aspects. I will briefly give an overvierw to the

2.1 Dimension Reduction

The most basic algorithm has been known since ...

2.1.1 PCA

Principal Component Analzysis finds the axes of largest variation and projects the whole dataset onto these vectors.

2.1.2 t-SNE

A non-linear alternative in Dimension Reduction is called t Stochastic Neighbor Embedding. We define a distance measure such as Euclidean distance in the high dimensional space and t
RThe points are then moved along the gradient until it results in a local maximum where no point can be slighlty moved without resulting in a worse embeding than before
When interpreting t-SNE embedding is is important to keep in mind that the choice of distance measure puts emphasis on close points. Points the

are embedded far from each other don't need to be far away in the high dimensional dtaset.
Initilazation

2.2 Wasserstein Distance

We have learned that for a meaningful embedding it is necessary to have a distance measure in the high diomensional space. When dealing with hierarchical data this becomes non trivial as there is no default distacne measure for probability distribution. A standard way is to collapse the dsitrbition into the mean and then use the Euclidian distance of the means. But one can easily imagine that this technique can loose arbitryily much of the information. A standard metric in computer Science is called the Wasserstein metric and has been widely used to compare distibutions. Its downside is the complex compuation durance. However, for small datasets we have been able to counter this problem and computed Wasserstein distances. I will thus give a brief overview of the theory and then explain the computation.

2.2.1 Exact Formulation

The Wasserstein metrix is formally edfined by

2.2.2 Special Case: Gaussians

For multivariate normals there exists a closed form solution of the 2-Wasserstein metric. it reads
formalu
and the derivation can be found in Frechet distance blabla

2.2.3 Convex Interpolation Method

We first nore, that the first part of the sum is just the Euclidiean distance of the mean of the distribution. The Wasserstein distance can therfore be seen as an extension of the Euclidiean distance. As Landau etc showed, the second summand in euqation 1 is a proper meric on covarainces. We can therefore combine these two distances in any way other than

This leads to the convex generalization of the Wasserstein distance which ziels both other distances for $\lambda = 0$ or $\lambda = 1$ respectively.

2.3 Linear Programming

We have earlier stated that the Wasserstein distance is hard to compute for continuous distributions. However for discrete distributions it boils down to the linear program described in Equaton
LINEAR PROGRAMM

2.3.1 Scalability in Participants

2.3.2 Scalability in Features

Chapter 3

Analysis: Is there structure in Covariance?

3.1 Synthetic Data

3.1.1 Hierarchical Gaussian Mixture

3.1.2 Proof of Concept

3.2 German Election 2017/2021

3.2.1 Finer Structure

3.2.2 Covariance Embeddings

3.2.3 Results

3.3 European Value Study 2017-2020

3.3.1 Embeddings

3.3.2 Discrete Data

3.3.3 Comparison to Exact Wasserstein Embedding

3.4 Big5 Personality Traits Survey

3.4.1 Results

Chapter 4

Outlook and Discussion

4.1 Complexity Analysis

4.1.1 Improving the runtime of Exact Wasserstein

4.2 Finding more Use-Cases

4.2.1 Medical Data