Multidimensional Scaling for Longitudinal Data Embeddings in Pharmacometrics

Mohamed Tarek, Lucas Pereira

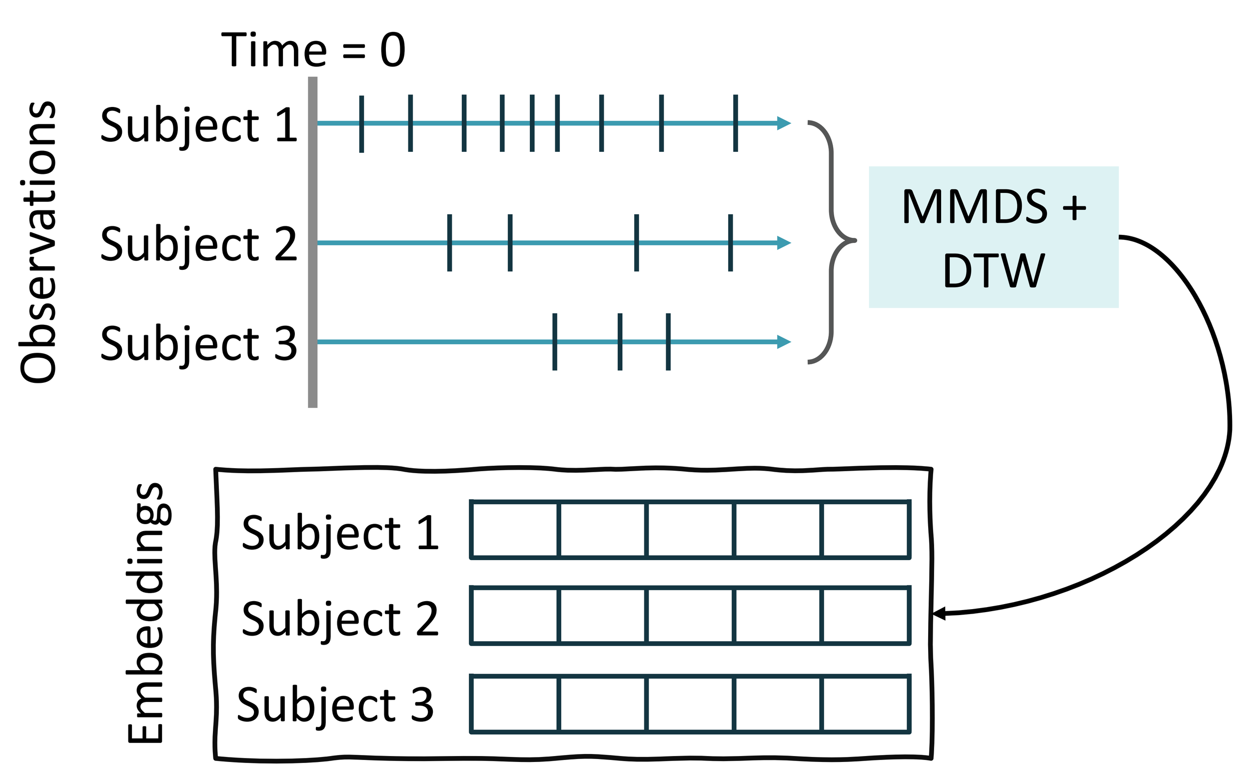
PumasAI, USA

Introduction

Nonlinear mixed effects (NLME) models are the standard approach for modeling, but they can be computationally intensive for **many input covariates and output biomarkers and endpoints**. Some machine learning (ML) methods can quickly eliminate useless covariates and biomarkers. Many ML models require fixed-size data as inputs and outputs. Such a tabular representation of a more complex data structure is commonly known as an embedding.

Objectives

In this work, we generate **dissimilarity-preserving embeddings for longitudinal** data commonly used in pharmacometrics. We use metric multi-dimensional scaling (MMDS) along with dynamic time warping (DTW) to generate fixed-size embeddings for each time-varying variable of each subject in a population. An experiment on a synthetic dataset shows that the procedure can generate useful embeddings that preserve neighborhood structures. This has potential applications in covariate and biomarker elimination, and in model evaluation.



Methods

The MMDS formulation has the following definition, d and δ are the dissimilarity functions in the input and embedding spaces, respectively.

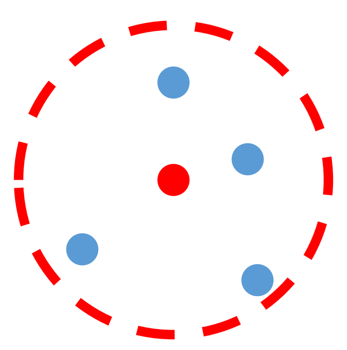
$$\min_{Y} \left(\frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(d(x_i, x_j) - \delta(y_i, y_j) \right)^2}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d(x_i, x_j)^2} \right)^{\frac{1}{2}}$$

For temporal variables, DTW can be used to compute pairwise dissimilarities between subjects based on their time-varying observations, **even if the observation sequences do not have the same length**.

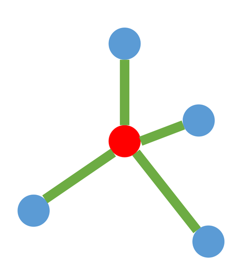
To evaluate how well structure is preserved with the resulting embedding, two experiments were performed:

1. Determine the neighbors of each point (subject)
2. Calculate the average distance to neighbors

Closest neighbors



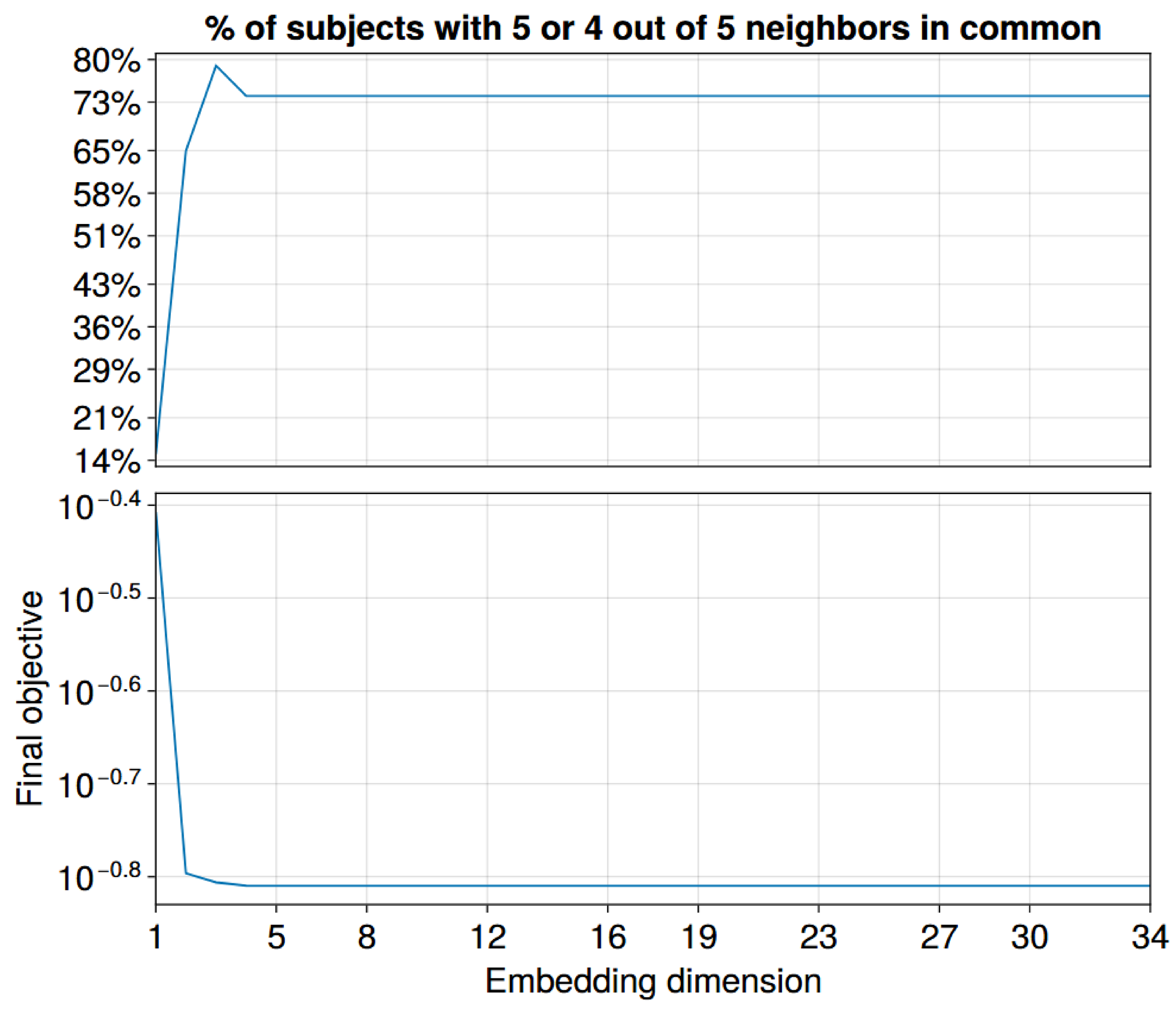
Distance to neighbors



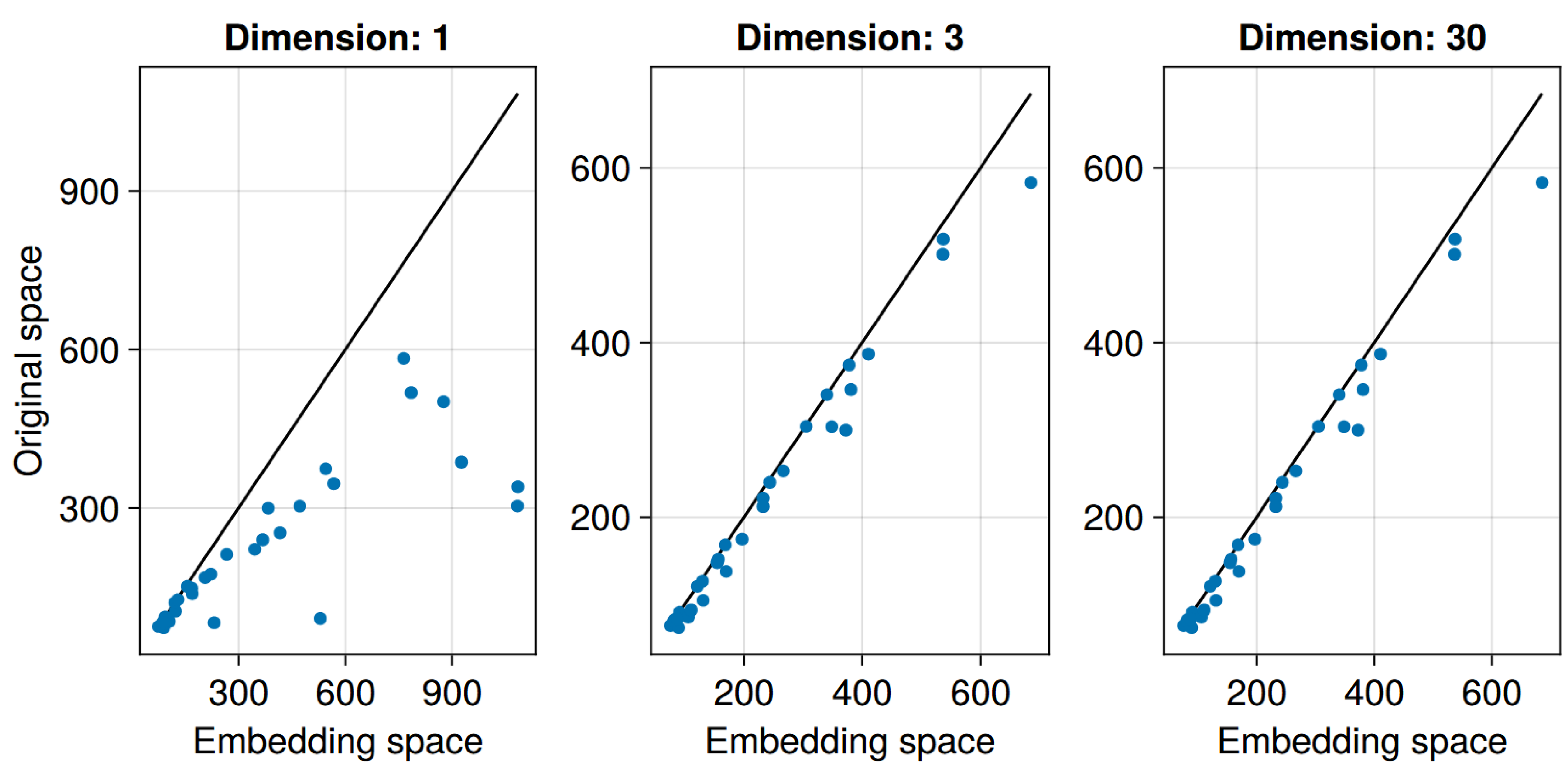
CONTACT
www.pumas.ai/resources
lucas@pumas.ai

Results

Below, we illustrate the percentage of subjects that have 4 or 5 overlapping neighbors (out of k = 5) in their 2 sets of neighbors as a function of the embedding dimension. As can be seen in the top plot, a considerable overlap between the neighborhoods in both spaces was achieved. This indicates **good structure preservation in the embedding procedure**.



Below is a scatter plot of the average dissimilarity from each subject to its neighbors defined in each space. Each subject is a point with the average dissimilarities representing the coordinates.



Conclusions

Key Takeaways

- Applied MMDS and DTW to longitudinal PK data for tabular embeddings
- Neighborhood structure preserved in the embedding space
- Procedure generates dissimilarity-preserving embeddings for longitudinal data
- Future: Covariate/biomarker filtering via tabular ML to drop uninformative variables before NLME fitting
- Future: Embedding-based model evaluation inspired by Fréchet inception distance

References

1. Review of the Development of Multidimensional Scaling Methods. A, Mead. Journal of the Royal Statistical Society. Vol. 41. No. 1 (27-39). Royal Statistical Society, Wiley. 1992.
2. Using Dynamic Time Warping to Find Patterns in Time Series. Berndt, Donald; C., J. Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining. Pg 359-370. 1994.